UNIR LA UNIVERSIDAD EN INTERNET

"If we do it right, we might be able to evolve a form of work that taps into our uniquely human capabilities and restores our humanity. The ultimate paradox is that this technology may become the powerful catalyst that we need to reclaim our humanity."

John Hagel

Editor's Note

THE International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI (ISSN 1989 - 1660) is a quarterly journal which provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on artificial intelligence tools, methodologies, systems, architectures integrating multiple technologies, problems including demonstrations of effectiveness, or tools that use AI with interactive multimedia techniques. Indexed now in the Journal Citation Reports by Clarivate Analytics, within the categories "Computer Science, Artificial Intelligence" and "Computer Science, Interdisciplinary Applications" [1], the IJIMAI Editorial Board reiterates its appreciation for their support to authors, reviewers and readers.

The last numbers where very exciting; we had a special issue where some of the hottest topics in soft computing were presented [2]. Then, a regular issue included, for the first time and almost necessarily, works related to the sadly famous COVID-19 [3]. Finally, the last special issue integrated together two very interesting fields like artificial intelligence and blockchain [4].

The present regular issue consists of 10 works with a great variety of topics. It begins with two interesting health-related works: the first one, reviews different approaches that can help fight COVID-19 [5] and the second one, tries to solve an optimal control problem of cancer treatment using an artificial neural network [6]. It continues with algorithms that can help in industrial processes such as reducing the temperature of electronic circuits [7] or carrying out chromium layer thickness forecast [8]. Radically changing the subject, the next paper proposes a method to identify the most influential nodes in social networks that can be the source of rumor spreading [9]. Another different work focuses on how to interact with virtual 3D environments in a cheap way by using multiple embedded markers in a specialized manner [10]. In this issue, a work has also been carried out to improve the identification of polysemy in natural language processing tasks [11]. The following two works are related to assess students satisfaction: on the perception of students when they are evaluated by an artificial intelligence and not by a human [12], and on the quality of the learning content and the methodology at unit level for any course and at any time [13]. Finally, the issue closes with a paper in which a tool can classify the emotions of users based only on non-invasive techniques like their keyboard typing and mouse usage pattern [14].

The first article, written by Lezcano et al. [5], reviews the most recent research efforts and approaches related to new data driven techniques, tools and datasets intended to manage the coronavirus disease. These approaches can assist medical staff in efficient decision making with complex and very heterogeneous data, even in stressful scenarios. The paper also describes how these new approaches can assist in the process of easing both the care of infected patients in Intensive Care Units and in the discovery of new treatments.

The next article, written by Heydarpour et al. [6], focuses on machine learning approaches for cancer treatment. The paper proposes a system of ordinary differential equations with state variables for immune cells, tumour cells, healthy cells, and drug concentration. The underlying idea is to anticipate the tumour growth showing the interactions of cells in the body. To achieve the goals, an artificial neural network is applied with results that show both a good performance and efficiency, showing the optimal dose of drugs needed for treatments.

Dash and Pradhan [7] propose a multi-objective heuristic approach to select an efficient input variable polarity for simultaneous optimization of area, power and temperature in chips. The idea of the work is to achieve temperature minimization at the logic level instead of at the physical level, reducing the cooling cost of circuits. The proposed algorithm saves around 76.20% silicon area, 29.09% power dissipation and reduces 17.06% peak temperature in comparison with the reported values in the literature.

Nieto et al. [8] build a novel nonparametric method for the hard chromium layer thickness forecast, which is very important for many different industrial applications (e.g., avoiding corrosion). The study is based on machine learning, employing a hybrid support vector machines model. The outcomes indicate that the model, together with radial basis function kernel, has permitted to satisfactorily foretell the thickness of the chromium layer created in the industrial process with a coefficient of determination of 0.9952.

Jain et al. [9] use a game-theoretic approach called Shapley Value to find the top 10 most influential nodes on three social network datasets. They have compared the results to the ones obtained by using primitive centrality measures showing that the new approach is better at understanding the marginal contribution, and therefore the actual influence of each node to the entire network. That is very important to identify nodes and edges that play a crucial role in a social network to find the most influential sources of rumour spreading.

Rehman et al. [10] propose an interaction technique that provides different ways of interaction on complex objects in a simple and computationally cheap way. The interaction is based on the use of multiple embedded markers in a specialized manner. The proposal can perform any type of interaction in a 3D virtual environment for working with Augmented Reality and Mixed Reality scenarios. A biological virtual learning application is developed for evaluation and experimentation, showing less task execution time than the alternatives.

Mudigonda and Sharma [11] present a novel idea for handling polysemy in natural language processing tasks by generating Multi-Sense Embeddings using synonym sets and hypernyms information of words. The work derives embeddings of a word by understanding the information of a word at different levels, starting from sense to context and definitions. The proposed sense embeddings of words obtain prominent results when tested on word similarity tasks, outperforming several state-of-the-art systems.

Sánchez-Prieto et al. [12] show a technology acceptance model for the artificial intelligence-based assessment on education. The different aspects are perceived: 1) usefulness; 2) ease of use; 3) attitude towards use; 4) behavioural intention; and 5) actual use. The outcomes can be relevant to the research community, since there is a lack of this kind of proposal in the literature to deal with the students perceptions of being assessed by a non-conscious software entity like a machine learning model or any other artificial intelligence application.

The penultimate article, written by Cantabella et al. [13], proposes a new tool to assess student satisfaction using emoticons. The idea is to evaluate the quality of the learning content and the methodology at unit level for any course and at any time. The results indicate that the assessment of student satisfaction is sensitive to the period when the survey is performed and to the level of study. Moreover, the results of this proposal are compared to the satisfaction results using traditional surveys, showing more accuracy data.

The last article, written by Magdin et al. [14], focuses on the

classification of emotions based on behavioral characteristics of users such as keyboard typing and mouse usage pattern, which is an effective and non-invasive way of gathering data. The work obtains the data using an application developed for the study (Emotnizer), that can classify emotional states with an average success rate of 82.31% among 4 emotional categories from Russel's complex circular model: happiness, anger, sadness and the state of relaxation.

As a novelty, we are pleased to announce that this is the first regular issue in IJIMAI that contains a special section with a collection of works on a specific topic. In this case, the section is presented under the title "Artificial Intelligence and Sensor Informatics: Exploring Smart City and Construction Business Implications" and is edited by Prof. Shaofei Wu from the Wuhan Institute of Technology (China). The section includes 7 exciting works which topics range from industrial robots to energy estimation based on machine learning algorithms. Enjoy!

Editor:
Dr. Vicente García-Díaz,
University of Oviedo, Spain

From the Guest Editor

With the continuous development of the world and the rapid growth of the urban population, the process of human urbanization has entered a new era. At the same time, with the rapid development of urbanization, more and more urban problems have been exposed, such as traffic congestion, population explosion, uneven distribution of resources, and threats to public security, which require an adaptive scheme to solve these urban governance problems. On the other hand, the development of internet technology and information and communication technology provides a technical basis for solving these problems. The concept of smart city is put forward under the action of practical demand and technical basis, which aims to manage the production and living conditions of human beings in a refined and dynamic way through information technology. By embedding or equipping sensors in the power supply system, water supply system, transportation system, buildings, and other aspects of production and life, the connection between the Internet of Things and the Internet is formed to integrate the human society and physical system, and then the rich data in the Internet of Things are processed through cloud computing to form a new way of urban governance.

Generally speaking, artificial intelligence refers to the technology of using ordinary computers to present human intelligence. The core of the technology is to construct the ability of reasoning, learning, perception, planning, etc., which are similar to or even exceed those of humans. Artificial intelligence has now surpassed human levels in chess games, image recognition, and other aspects. At present, there are a large number of tools using artificial intelligence, including mathematics, search optimization, logic deduction.

The extensive connection of digital entities is driving the construction of smart cities, which are all based on sensor networks in cities. The sensor network combines sensor technology, embedded computing technology, distributed information processing technology, and wireless communication technology. Real-time monitoring, data collection of various objects, and cloud computing data transmission processing are achieved through mutual coordination, finally the results are provided to users.

Ruhlandt proposed in a research that smart cities should connect advanced integrated materials, sensors, electronic devices, and networks to computer networks composed of databases, trackers, etc.,

and use these materials, sensors, electronic devices, and networks to design, build, and maintain physical or virtual structures [15]. Grimaldi and Fernandez mentioned in the research that through the application of information and communication technology, smart cities have improved the living standards of residents, achieved sustainable development, and made urban resource management more intelligent [16].

Traditional urban planning and architectural planning rely on experienced architects and planners to judge the problems between buildings and cities. Nowadays, urban and architectural planning not only relies on traditional statistical data, but also on various big data including satellites, sensors, maps, and remote sensing data. The changes in the breadth and speed of construction and planning projects brought about by artificial intelligence are obvious.

A smart city is a complex system, and the management of the city must have a systematic perspective. Especially under the various urban problems brought about by the rapid development of the current era, when the Internet of Things and artificial intelligence are applied to the construction of smart cities, the efficiency and quality of urban management can be better improved and the sustainable development of cities can be achieved. Besides, the construction of smart cities based on artificial intelligence can meet the high standard requirements of residents for modern public services. However, most of the construction of smart cities is practice-first, the theory lags behind, which lacks the development strategy of integration and adaptation to the history of the city. Therefore, it is necessary to strengthen research in this area and find a smart city construction plan that conforms to the historical development of the city.

We are grateful to the authors who submitted papers to this special section. We would also like to thank the reviewers for their hard work and their valuable feedback to the authors. Finally, we would like to express our sincere gratitude to Dr. Elena Verdú Pérez, the Managing Editor, for providing the opportunity and assistance to edit this special section in the International Journal of Interactive Multimedia and Artificial Intelligence.

Guess Editor:
Prof. Shaofei Wu,
Wuhan Institute of Technology, China

References

[1] R. González Crespo and E. Verdú, "IJIMAI Editor's Note-Vol. 5 Issue 7," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 7, pp. 4–5, 2019.

[2] J. A. Morente-Molinera, "IJIMAI Editor's Note-Vol. 6 Issue 1," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 1, pp. 4–5, 2020.

[3] E. Verdú, "IJIMAI Editor's Note-Vol. 6 Issue 2," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 2, pp. 4–5, 2020.

[4] A. Baldominos and F. Mochón, "IJIMAI Editor's Note-Vol. 6 Issue 3," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 3, pp. 4–5, 2020.

[5] S. M. Lezcano, F. C. L. Hernández, and A. C. Bellot, "Data Science Techniques for COVID-19 in Intensive Care Units," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 8–17, 2020.

[6] F. Heydarpour, E. Abbasi, M. J. Ebadi, and S. M. Karbassi, "Solving an Optimal Control Problem of Cancer Treatment by Artificial Neural Networks," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 18–25, 2020.

[7]   A. Das and S. N. Pradhan, "An Elitist Non-Dominated Multi-Objective Genetic Algorithm Based Temperature Aware Circuit Synthesis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 26–38, 2020.

[8]   [P. J. G. Nieto, E. García-Gonzalo, F. S. Lasheras, and A. B. Sánchez, "Chrome Layer Thickness Modelling in a Hard Chromium Plating Process Using a Hybrid PSO/ RBF–SVM–Based Model," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 39–48, 2020.

[9]   M. Jain, A. Jaswani, A. Mehra, and L. Mudgal, "Rumour Source Detection Using Game Theory," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 49–56, 2020.

[10]   I. Rehman, S. Ullah, and D. Khan, "Multi Layered Multi Task Marker Based Interaction in Information Rich Virtual Environments," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 57–67, 2020.

[11]   K. S. P. Mudigonda and P. Sharma, "Multi-sense Embeddings Using Synonym Sets and Hypernym Information from Wordnet," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 68–79, 2020.

[12]   J. C. Sánchez-Prieto, J. Cruz-Benito, R. Therón, and F. J. García-Peñalvo, "Assessed by machines: Development of a TAM-based tool to measure AI-based assessment acceptance among students," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 80–86, 2020.

[13]   M. Cantabella, R. Martínez-España, B. López, and A. Muñoz, "A Fine-Grained Model to Assess Learner-Content and Methodology Satisfaction in Distance Education," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 87–96, 2020.

[14]   M. Magdin, D. Držík, J. Reichel, and Š. Koprda, "The possibilities of classification of emotional states based on user behavioral characteristics," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 97–104, 2020.

[15]   R. W. S. Ruhlandt, "The governance of smart cities: A systematic literature review," *Cities*, vol. 81, pp. 1–23, 2018.

[16]   D. Grimaldi and V. Fernandez, "The alignment of University curricula with the building of a Smart City: A case study from Barcelona," *Technol. Forecast. Soc. Change*, vol. 123, pp. 298–306, 2017.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

The International Journal of Interactive Multimedia and Artificial Intelligence is covered in Clarivate Analytics services and products. Specifically, this publication is indexed and abstracted in: *Science Citation Index Expanded, Journal Citation Reports/ Science Edition, Current Contents®/Engineering Computing and Technology.*

Data Science Techniques for COVID-19 in Intensive Care Units

Sergio Muñoz Lezcano*, Fernando Carlos López Hernández, Alberto Corbi Bellot

Universidad Internacional de La Rioja, Logroño (Spain)

Abstract

Data scientists aim to provide techniques and tools to the clinicians to manage the new coronavirus disease. Nowadays, new emerging tools based on Artificial Intelligence (AI), Image Processing (IP) and Machine Learning (ML) are contributing to the improvement of healthcare and treatments of different diseases. This paper reviews the most recent research efforts and approaches related to these new data driven techniques and tools in combination with the exploitation of the already available COVID-19 datasets. The tools can assist clinicians and nurses in efficient decision making with complex and heavily heterogeneous data, even in hectic and overburdened Intensive Care Units (ICU) scenarios. The datasets and techniques underlying these tools can help finding a more correct diagnosis. The paper also describes how these innovative AI+IP+ML-based methods (e.g., conventional X-ray imaging, clinical laboratory data, respiratory monitoring and automatic adjustments, etc.) can assist in the process of easing both the care of infected patients in ICUs and Emergency Rooms and the discovery of new treatments (drugs).

Keywords

I. Introduction

BY the end of 2019, doctors in the city of Wuhan (China, province of Hubei) began to detect evidence of the expansion of a new type of severe acute respiratory syndrome (SARS) among the population of the aforementioned city. This event, although different from, had strong similarities to the SARS coronavirus (CoV) outbreak that took place there during 2002-2004, hence the name SARS-CoV-2 or CoVID-19, for short. This new coronavirus is considered to be a strain from the 2002 one and has, by the time of this writing, brought about a novel global pandemic with potentially lethal consequences for infected people. The new disease, COVID-19, was considered as a public health emergency by the World Health Organization (WHO) on January 30, 2020. In March 11, its condition was raised to the status of global pandemic [1].

As the original SARS-CoV coronavirus, SARS-CoV-2 is transmitted through respiratory droplets from person to person and the condition's symptoms appear after an incubation period that lasts from 2 to 14 days [2]. The great expansion rate of the disease has led to a shortage in intensive care equipment and to a greatly increased demand for technical resources (e.g. mechanical ventilation, Personal Protective Equipment (PPE), drug treatments, etc.). In many situations, the health care systems have risked or even surpassed the point of collapse.

This pandemic has also had a huge impact on the generation of clinical and research data. Of course, part of this dataset has been generated in the Intensive Care Units (ICUs) and Emergency Rooms (ERs). However, the exploitation of this information is complex due

* Corresponding author.

E-mail address: smunozle@gmail.com

to the stressful situations that clinicians experience in ICUs. Some pieces of this data may turn out to be too complex and too varied for obtaining practical and correct diagnoses. In this context, Artificial Intelligence (AI), Image Processing (IP) and Machine Learning (ML) techniques can contribute to manage this problem of information overload. The information flow is represented in Fig. 1, where AI has a pivotal role in the fight against COVID-19 in intensive care units [3].
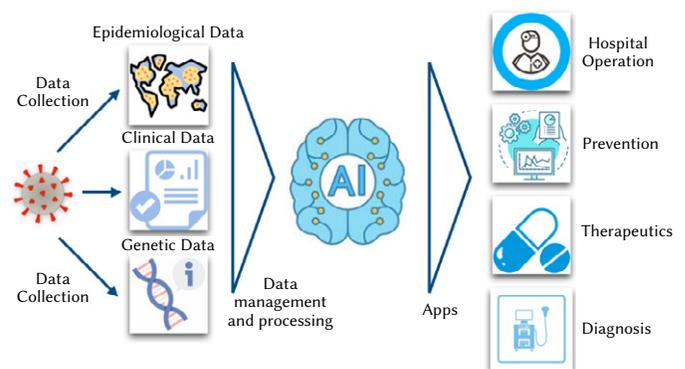
Fig. 1. Areas of interest where artificial intelligence and machine learning could play a key role in the fight against COVID-19 (e.g., biochemistry and clinical data of COVID-19 patients can be studied by ML techniques to focus not only in risk prediction, but also for disease treatment). Model adapted from the figure published [3].

The rest of this paper is organized as follows. First, we review the COVID-19 outbreak, its complexity, symptoms and how healthcare professionals are dealing with ICU patients. Then, we provide a broad prospective on the utility of machine learning in data science. Next we explain how ML is applied in medicine with some examples and

applications. The next aim is to contextualize the applicability of AI+IM+ML in COVID-19 focusing on its critical areas of interest. Specifically, this review covers mechanical ventilation, medical imaging and clinical data. Finally, we outline the most important findings and prospective future research lines.

II. The COVID-19 Outbreak

Dr. Tedros Adhanom Ghebreyesus[1], the World Health Organization (WHO) Director-General, reported on February 11, 2020 the new Coronavirus Disease SARS-CoV-2 or COVID-19 intensive epidemic in Wuhan. This is the third identified type of dangerous acute respiratory syndrome from coronavirus (SARS-CoV-2) in the last twenty years, but the one with the lowest level mortality rate [4], as evinced in Table I. Until July 13, there were a total of 12,859,834 confirmed cases, and more than 567.957 deaths in 188 countries and territories around the world[2]. The countries in which the outbreak has had the most severe impact in terms of cases are in Table II: United Stated of America, Brazil, India, Russian Federation, Peru, Chile, Mexico, The United Kingdom, South Africa, Iran, Spain and Pakistan.

TABLE I. CoV Mortality Rate

CoV	Year	Origin	Mortality Rate
SARS	2002	Guandong province, China	10%
MERS	2013	Saudi Arabia	34%
COVID-19	2019	Wuhan, China	3.4%

Mortality rate comparing different coronavirus outbreaks from 2002 to present

TABLE II. Situation By Country

Country	Cases	Deaths	Death Rate (%) of cases
United Stated of America	3,163,581	133,486	4.2
Brazil	1,800,827	70,398	3.9
India	849,553	8,196	0.9
Russian Federation	727,162	11,335	1.5
Peru	319,646	11,500	3,6
Chile	312,029	6,881	2.2
Mexico	289,174	34,191	11.8
The United Kingdom	288,957	44,798	15.5
South Africa	264,184	3,971	1.5
Iran	255,117	12,635	4.9
Spain	253,908	28,403	11,2
Pakistan	248,872	5,197	2

Globally, as of 4:49 pm CEST, 12 July 2020, there have been 12.552.765 confirmed cases of COVID-19, including 561.617 deaths, reported by the WHO in https://covid19.who.int.

Scientists think that COVID-19 made a leap from animals to humans as a new zoonotic [5] human disease at the Huanan Seafood Market in Wuhan. In this place, bats, snakes and other wild animals are sold as culinary delicacies. As the Fig. 2 shows, even though the zoonotic root of SARS-CoV-2 has not been corroborated, the series-based analysis indicates that the virus jumped from bats (as this species is an important coronavirus reservoir) to humans perhaps through an intermediate host, the pangolin [6]. Once the virus had the capability of infecting humans, it rapidly spread up to 109 countries in November and December 2019.
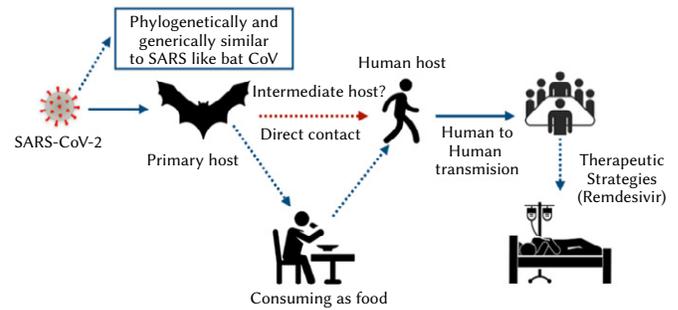
Fig. 2. Representation of the transmission of SARS-CoV-2. Model adapted from the figure published [6].

Direct transmission [7] requires direct physical contact between the infected and predisposed individuals, via bites and handling of infected animal tissue. Direct contact and bites involve the formation of skin rash and ulcerations, whereas the ingestion of infected meat/water involves digestive tract problems and diseases transmitted through the intake of inhalation of contaminated particles could cause pneumonia.

After a massive outbreak, now driven by human-to-human transmission, the infection continued its expansion thanks to national and migratory movements and global tourism, starting with the Chinese New Year holidays. SARS-CoV-2 is a highly diffusible virus, carried by airborne droplets, direct contact or contact with contaminated objects. Major risk factors are:

- Age greater than 60 (increasing with age).
- Male gender.
- Subjacent non-communicable diseases (NCDs): diabetes, hypertension, cardiac disease, chronic lung disease, cerebrovascular disease, chronic kidney disease, immunosuppression and cancer.

In many cases, the most common symptoms at inception of COVID-19 disease could include [8] –[9]:

- Fever experienced by 83–99% of the patients [10].
- Neurological manifestations [11] such as anosmia (loss of sense of smell) and dysgeusia (perception of altered taste).
- Dyspnea or shortness of breath (31–40%).
- Dry cough (59–82%).
- Fatigue (44–70%).
- Myalgia (11–35%).
- Diarrhea.

Moreover the "Massachusetts General Hospital treatment guidance for critically ill patients with COVID-19" [12] reported that it would be expected that a percentage of these patients (5%) would need ICU admissions due to bilateral pneumonia, acute respiratory distress syndrome (ARDS), hypoxemia, respiratory failure, shock and multiple organ dysfunction syndrome (MODS). As Fig. 3 discloses, critically events lead to further complications [13]. In the most severe cases, a cytokine storm syndrome is triggered that involves an overall hyper-inflammation that leads to sepsis and septic shock, thromboembolism, and/or multiorgan failure, that includes in many cases acute kidney injury and cardiac injury. When these events take place, the mortality rate is in the range of 20% to 60% after ICU admission. In Wuhan [14], between January 10 and February 29, 2020 at least a median(interquartile range) of 429(25-1143) patients with COVID-19 were reported and admitted in the ICU while 1521 (111-7202) patients were diagnosed with severe sickness each day. During the epidemic peak, 19,425 patients (24.5 per 10,000 adults) were hospitalized, 9,689 (12.2 per 10 000 adults) were considered in severe illness, and 2,087 (2.6 per 10,000 adults) needed critical care per day. After the first 7-8

days in the ICU, patients required massive intensive care management (lung protective ventilation, avoidance of fluid overload and support of organ function) as well as the application and intervention by a wide spectrum of medical technologies and therapies such as:

- ventilators in most cases,
- RT-PCR tests,
- aerosol generating therapies,
- laboratory equipment,
- X-ray and CT Imaging, and
- other healing therapies previously used for SARS and MERS.
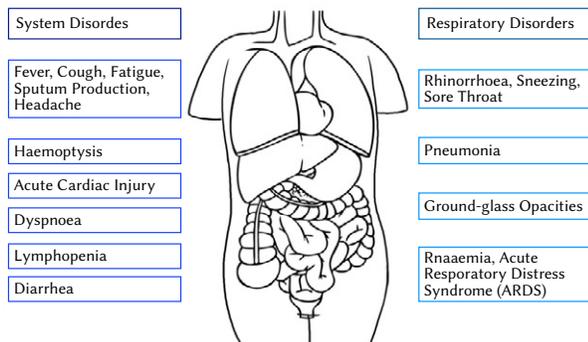
Fig. 3. Severe and abnormal symptoms caused by critical disease complications. e.g. respiratory failure. Model adapted from the figure published [13].

Regarding the use of therapies, even when different drugs have in vitro activity against different coronaviruses, until today there has been no clinical evidence that argues the efficacy and safety of these treatments for coronavirus infection, including SARS-CoV-2. What is more, compassionate treatments that have not been previously approved for clinical purposes could involve serious harmful consequences when taking into account that many patients suffer from previous disorders such as cardiac arrhythmias. In these circumstances, taking into consideration that nonproven specific therapies (NPTs) are available, a large number of patients have received compassionate use therapies such as Remdesivir [15], which had been previously targeted at fighting against the Ebola virus, Lopinavir–Ritonavir (an HIV type 1 aspartate protease inhibitor [16]), chloroquine (generally-used anti-malarial and autoimmune disease drug [17]) and hydroxychloroquine (an analogue of chloroquine [15]) among other therapies.

Bearing in mind all these concerns, a significant number of doubts still need to be addressed. In summary, until a vaccine or effective treatment is available, the most efficient measure for reducing person-to-person transmission of COVID-19 needs to focus on the use of masks, social distancing and hand washing.

III. Machine Learning

In 1959, Arthur Samuel defined machine learning as a "field of study that gives computers the ability to learn without being explicitly programmed" [18]. We might say, very broadly, that machine learning refers to the changes in systems regarding all those tasks associated with AI. These tasks have a wide range of functions, from recognition, to diagnosis and prevention. Many techniques in machine learning arrive from the efforts of psychologist to find similarities to biological learning. But psychologists are not the only ones who have contributed to the study and evolution of ML. Other disciplines have also made relevant contributions, such as:

- **Brain modeling**, which tries to interpret learning phenomena in terms of artificial neural networks.

- **Statistical machine learning**, which can be considered another instance of artificial learning.
- **Adaptive control**, which manages processing in real time by the use of mathematical estimations in real time.
- **Artificial intelligence**, which has been associated with ML from the very beginning, since past actions have a direct effect on the future.
- **Evolutionary algorithms**, which have proposed certain aspects of biological evolution as learning methods. Genetic algorithms are the most representative of these algorithms.

Traditionally, the training of a ML algorithm has been classified into the three broad categories shown in Fig. 4, which depend on the available data [19].

- **Supervised learning**, which teaches the computer how to do something and then lets it use its new knowledge to do the task.
- **Unsupervised learning**, by which the computer learns without instruction and uses this knowledge to structure and find patterns in the dataset.
- **Reinforcement learning**, by which the computer interacts with a changing ecosystem where it needs to perform a specific action or objective, such as driving an autonomous vehicle.
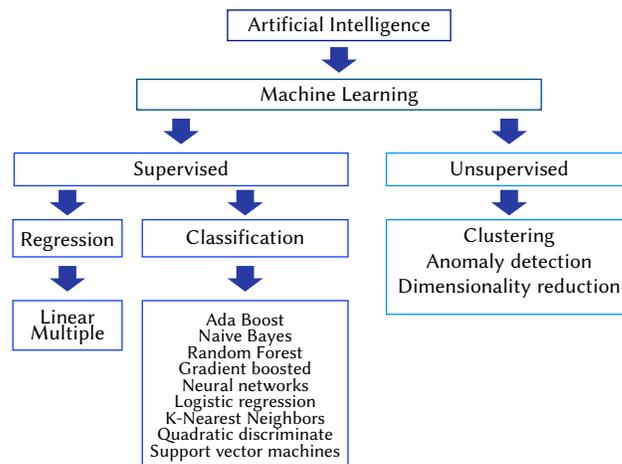
Fig. 4. The major learning approaches in ML. Examples of algorithms of the corresponding type are shown under each branch. Model adapted from the figure published [19].

ML has had an impact on different applied disciplines such as physics, chemistry, engineering, as well as medicine. These techniques can be applied to problems of prediction, classification or control in a wide spectrum of medical fields from the assessment of a patient's state to the prognosis of disease outcome from genetic information.

IV. Machine Learning in Medicine

The next sections survey machine learning applied to medicine and then focus on its prospective applications for addressing COVID-19, medical imaging, clinical data and the automatic monitoring of ventilators. Until today, experience and heuristics have been the gold standard in medicine, whereby knowledge is acquired through daily practice in self-learning, considering the high variable healthcare environments. Computational resources development as well as data storage capacity has become the big enabler in automatic or semi-automatic data-driven and evidence-based medical decisions. The use of ML in this context relies on the collection and analysis of huge and complex datasets to recognize and learn patterns, limit false

classifications and make the following decisions related to:

- **Diagnosis**. Errors in diagnosis happen on a daily basis and this problem is not limited to complex situations. In this scenario, ML could reduce the uncertainty by helping clinicians make better decisions and at the same time reduce errors. In this context, computer-based medical IP (e.g. electrocardiogram, topographies, etc.) enables the non-intrusive identification and medical assessment of the level of damage in different regions of the body.

- **Prognosis**. ML applications can learn patterns from a vast number of patients. Using data belonging to laboratories for classification of biomarkers, clinicians can anticipate future events going beyond the individual physician's clinical experience and reducing bias.

- **Treatment**. In a large health care system, there could exist tens of thousands of physicians, and a huge number of patients with related conditions. ML could help clinicians automate, program and at the same time customize treatment in order to apply better treatment pathways. A clear field of application is the assessment of genomic attributes to find personalized treatments.

V. Artificial Intelligence Applied to Covid-19

COVID-19 has rapidly become a global concern that needs faster, more sensitive, and at the same time customizable support. Currently, artificial intelligence, in particular ML, can help dealing with the key challenges associated with the SARS-CoV-2 outbreak, which are:

- **Diagnostic Capacity**. The huge number of cases overwhelms diagnostic potential, accenting the need for an agile pipeline for sample processing and diagnosis.

- **Diagnosis Specificity**. SARS-CoV-2 is nearly connected to other significant coronavirus subspecies and species, so diagnostic assays can produce false positives if they are not finely specific to SARS-CoV-2.

- **Pathogen Characterization**. Presumed SARS-CoV-2 patients sometimes suffer from a dissimilar respiratory viral infection and in worst cases have co-infections with SARS-CoV-2 and other respiratory viruses. Consequently, it is important to identify these other pathogens, for both patient diagnostics and pandemic response. For these reasons, the authors in [20] identify (in Table III) an inherent COVID-19 virus genomic signature and compare the performance of their propose algorithm with six machine learning algorithms (Linear Discriminant, Linear SVM, Quadratic SVM, Fine KNN, Subspace Discriminant, and Subspace KNN) for an ultra-rapid adaptable, and extremely accurate classification of whole COVID-19 virus genomes. Table III shows the effectiveness of these approaches.

Taking into account the previously mentioned difficulties, clinicians and nurses need to deal with a vast amount of data, work with multidimensional problems or identify patterns based on a huge dataset. To face this problem, the combination of AI techniques and high-performance computing networks will be the new clinician's tools for health care provision. This new data-driven approach enables decision-taking processes that involve high complexity, are more accurate and with a quicker response, to cure more people when dealing with a lack of resources.

Considering COVID-19 unique circumstances, AI efforts have been mainly focused on ICU activity. Fundamentally, the spotlight has been put on diagnosis, monitoring, prognosis and drug discovery. At the same time, the fast spread of COVID-19 and the lack of resources in all countries have made possible the development of a myriad of applications based on AI to deal with the following clinical problems:

- **Patient diagnoses and severity status**. Clinical asset restrictions

TABLE III. Classification Accuracy Scores of Betacoronavirus

Dataset	Clusters	Nº Seq	Classification Model	Acc.
Test-4:	Embecovirus	49	LinearDiscriminant	97.6
BetaCV;	Merbecovirus	18	LinearSVM	98.4
124 seq.;	Nobecovirus	10	QuadraticSVM	98.4
MaxL:31526	Sarbecovirus	47	FineKNN	97.6
MinL:29107			SubspaceDiscriminant	98.4
MdL:30155			SubspaceKNN	97.2
MnL:30300			AverageAccuracy	97.6
Tes-5:	Embecovirus	49	LinearDiscriminant	98.6
Beta&CV19	Merbecovirus	18	LinearSVM	97.4
152 seq.	Nobecovirus	10	QuadraticSVM	97.4
MaxL:31526	Sarbecovirus	47	FineKNN	97.4
MinL:29107	COVID-19	29	SubspaceDiscriminant	98.7
MdL:29891			SubspaceKNN	96.1
MnL:30217			AverageAccuracy	97.5
Test-6:	Sarbecovirus	47	LinearDiscriminant	100
Sarb&CV19	COVID-19	29	LinearSVM	100
76 seq.;			QuadraticSVM	100
MaxL:30309			FineKNN	100
MinL:29452			SubspaceDiscriminant	100
MdL:29478			SubspaceKNN	100
MnL:29772			AverageAccuracy	100

Relationship between viral sequences, taking into account different levels of taxonomies and the score of Classification's accuracy.

(ICU equipment, etc.) imply that the treatments depend on the seriousness of the patient's status. Since asymptomatic patients could become seriously or even critically ill very fast, it is critical to identify them as soon as possible and administer the right treatment in order to reduce mortality. AI techniques can help to support the analysis, prediction and treatment of COVID-19 infections. Until now, AI applications based on COVID-19 clinical needs have been focused on diagnosis relying on medical imaging to analyze computer tomography (CT) and scans [21]. Authors in [22] and [23] have demonstrated that ML+CT or ML+X-ray combinations give extra time to radiologists for COVID-19 diagnoses, as shown in Fig. 5. Results from this research demonstrate, that ML+CT combinations are quicker and less expensive than the Reverse Transcription Polymerase Chain Reaction (RT-PCR or) test; the current gold standard for COVID-19 diagnosis. This approach makes more sense due to test global shortage and expense of the PCRs. However, not all ICUs have X-ray equipment (or even CT scanners). With the assistance of deep learning techniques, authors in [23] propose classifying the patient status as either negative or positive for COVID-19. This research has reported a 94.92% sensibility and 92.00% specificity in COVID-19 detection. These results demonstrate the utility of deep learning models to classify COVID-19 medical imaging.

- **Patient's health progress**. Patients' clinical data and blood tests contain many analyzable features that help doctors in recognizing high-risk patients [24], as explained in the flowchart in Fig. 6. The sooner the patient is tested, the sooner the prognosis improves. Taking into account this premise, a patient's health status can improve, and the mortality can decline for people seriously sick. So far, this methodology facilitates the interpretability of the most remarkable elements regarding the physiological progress of COVID-19, specifically in cell injury, cell immunity, and tissue inflammation.

Input

Model

Output
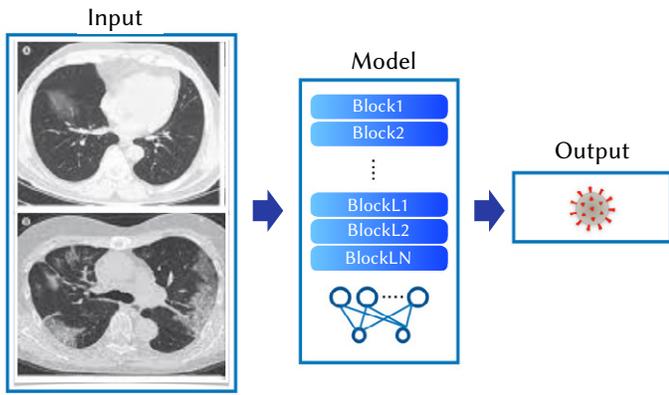
Fig. 5. Deep learning algorithm framework. For each patient's computed tomography scan, ROI (regions of interest) have been randomly selected and used in the inception network to extract features and then make a prediction. Model adapted from the figure published [21].
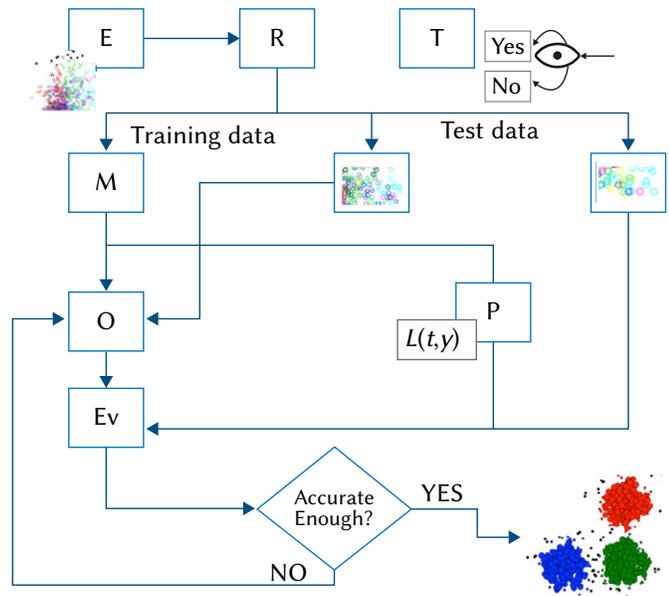
Fig. 6. The flow chart represents different concepts involved in a typical supervised ML process. "E" means experience (data); "R", is related to the task; "T", the classifying task; "M", represents the ML Model (the ML algorithm or the classifier); "O", is the optimization phase; "P", is related to metrics; "Ev", is in charge of the evaluation, here represented as a "loss function" $L(t, y)$. The final feedback involves the "learning" by the machine. Model adapted from the figure published [24].

- **Computational biology and drug discovery**. This approach is based on disease dynamic modeling [25]. It includes the use of data analytics, mathematical modeling and computational simulation procedures to study the COVID-19 biology. Authors in [26] propose a new pathway that contributes to the understanding of the effect of parameters that govern the proliferation of the disease. The results, as Fig. 7 shows, put the focus on a new AI algorithm that has been designed for identifying drugs that might block the viral infection process. In this way, Baricitinib [26] could decrease the potential of the virus to cause infection in lung cells.

Taking into account this enormous number of combinations, the scenario of COVID-19 gets more complex especially in ICUs with COVID-19 patients. Electronic health records from patients admitted at ICUs has been growing exponentially during the past months.

Clear examples of this new scenario are common ICU equipment such as pressure and flow transducers, infusion pumps, pulse
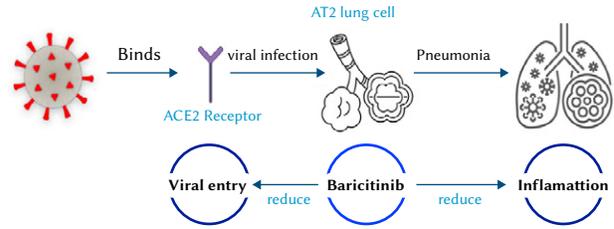
Fig. 7. The receptor ACE2 might be used by 2019-nCoV to infect lung AT2 alveolar epithelial cells. Baricitinib could be appropriate for patients with 2019-nCoV acute respiratory disease in order to reduce both the viral entry and the inflammation in patients caused by cytokines. Model adapted from the figure published [26].

oximeters, cardiac output monitors, and mechanical ventilators. This medical equipment stores data records and includes interoperable interfaces that allow for their interconnection using standards such as the Health Level Seven[3] (HL7). Through HL7, ICU medical devices communicate with a hub of electronic systems through busses and interchangeable, hot pluggable interfaces.

A. Adaptive Ventilation Settings

In this section, we focus on ventilation, medical imaging and clinical data topics that seem to be the most relevant areas where the combination of AI and COVID-19 promises better results.

Mechanical ventilation has saved many lives since it was implemented in the 1950s to tackle poliomyelitis. From that time, the use of ventilation support has been intensified with the past of the years. This increase has been directly associated with the evolution of critical care units [27]. The most important goal of mechanical ventilation (Table IV) is to enhance alveolar gas exchange. Taking into consideration this purpose, the ventilation system is composed of three elements:

- the ventilator together with its inspiratory and expiratory circuits,
- the patient,
- the interface between these two that controls the gas exchange.

TABLE IV. Objectives of Mechanical Ventilation

Make better pulmonary gas Exchange	Reverse hypoxemia
	Improve acute respiratory acidosis
Alleviate respiratory distress	Reduce oxygen cost of breathing
	Revoke respiratory-muscle fatigue
Modify pressure-volume relations	Prevent and reverse atelectasis
	Improve Compliance
	Prevent further injury
Allow lung and airway healing	
Avoid complications	

Main and secondary targets

The ventilation process is simple. In fact, it is the product of the respiratory frequency and the lung volume. This process is defined as the units of ventilation with gas inhaled or exhaled from a person's lungs per minute. Even when it is an apparently simple relationship, ventilation is related with a biological process and for this reason there are always complexities. For instance, the changes in the inspiratory and expiratory dynamics change the mean inspiratory output in

---

[3] https://www.hl7.org

volume-preset ventilators, the time for filling the alveolus with gas, the presence of airflow difficulties and finally expiratory volume. Moreover, mechanical ventilation without spontaneous breathing leads to respiratory muscle atrophy. For this reason, clinicians prefer assisted modes initiated by patients' respiratory activity. The most usual procedures are:

- **Assist-control ventilation:** the ventilator provides a breath independently when triggered by a patient's respiratory activity or not.
- **Intermittent mandatory ventilation:** the patient gets regular positive pressure breaths from the ventilator at a preset volume and rate. Spontaneous breath is also allowed.
- **Pressure-support ventilation:** the clinician sets an initial level of pressure (instead of volume) to increase the spontaneous breathing activity by the patient. Airway pressure is preserved at a preset level until the patient's inspiratory flow descends below a certain level.

As indicated in chapter II of this review, COVID-19 disease involves, in many cases, ICU admissions due to bilateral pneumonia, ARDS, hypoxemia, respiratory failure, shock and MODS. These situations require in most cases mechanical ventilation and intubation for helping the patient. Recent studies based on public data from COVID-19 patients from England, Wales, and Northern Ireland, reported that until March 24, 2020, two thirds of patients (i.e., 132 people) admitted at ICU required mechanical ventilation [28]. In most cases, with the shortage of advanced ventilators, the ventilation process needs a manual setup process before they can operate. In this situation, the patient's healing depends on the clinician's ability to modify the patient's respiration. Therefore, it is important to take into consideration the following questions for optimally managing the ventilation support:

- What are the clinical aims?
- Is there a complete comprehension of the foundations and functional limitations of the medical devices?
- What is the medical intention in changing the patient's prescription?

In many cases, there aren't any homogenous treatments, and ventilation involves a huge degree of personalization, so clinical bias takes place. Even when ventilators work in a proper way in delivering air to the lungs, the design of this equipment is based (in many cases) on an *open loop flow* system, which implies patient having to adapt to the machine. This means that the input ventilation does not depend on the outflow so the patient's response to the flow and the delivered breath can't be assessed. The Fig. 8 represents this open loop flow process.
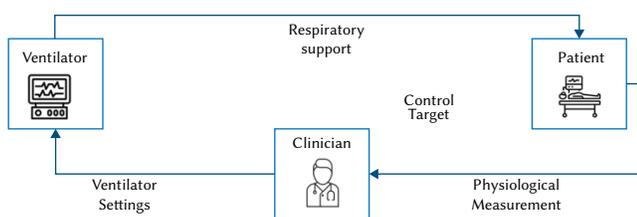
Fig. 8. Open Loop Ventilators Systems.

In order to avoid this issue, *automated closed-loop* systems (with feedback control) allow doctors to keep a patient at a specified clinical target without permanent clinical interaction. This means that a centralized module, *the controller*, takes the control over the task of adapting ventilator settings. Fig. 9 represents the closed loop flow.
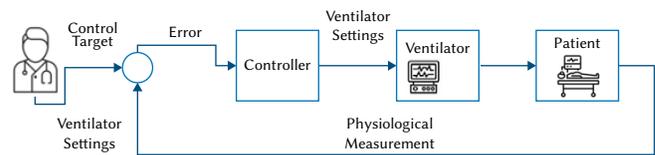
Fig. 9. Closed Loop Ventilators Systems.

From this point of view, machine learning plays a key role in automatic adaptation to customized patient ventilation. Gutierrez [19] mentions various applications regarding the development of AI artifacts that could help with the design of autonomous ventilators, making feasible the continuous monitoring of the ventilation process feasible. These artifacts could adjust some parameters automatically, such as asynchrony in breathing, which is one of the main problems with patient intubation. These techniques could improve advanced modes of closed loop ventilation systems, which imply sharing of information between the ventilator and the patient. For this purpose, the authors of [28] have described one of the most advanced algorithms. This research has compared ML predictions by working with a 1024 breaths' selection based on random criteria that take into account 16 patients. The results were compared to those made by five experts. The assessment reported very good results with sensitivity of 91.5%, specificity of 91.7%, positive predictive value (PPV) of 80.3% and negative predictive value (NPV) of 96.7%. Additionally, 1,378 breath cycles from 11 mechanically ventilated critical care patients took part on studies related to patient ventilator asynchrony [29]. In this paper, the use of random forest algorithms permitted the detection of the presence or absence of cycling asynchrony with valuable results. In this way, three different situations such as delayed termination, premature termination and no cycling asynchrony were analyzed. The sensitivity (specificity) was respectively of 89% (99%), 94% (98%), and 97% (93%) . The authors of this research work also include a variable *kappa*, a statistical metric that compares the assessment of the alignment between ML and clinicians' results. The kappa coefficient ranges from 0.90 to 0.91, which means an almost perfect model.

Other research related to ventilation carried out by authors in [30] have pointed out the need for developing promising methods based on machine learning techniques for prolonged mechanical ventilation situations and potential complications. Research was focused on reaching improvements in areas such as:

- prediction of arterial blood gases,
- successful extubation,
- spontaneous breathing trails, and
- development of autonomous ventilator setting adjustments.

A common difficulty for all this research work is the lack of data sources necessary for reducing bias and getting better results supported by more clinical cases. This fact has become more acute when considering the COVID-19 crisis.

B. Image Processing For the Diagnosis of COVID-19

ML has been used for the detection, segmentation and classification of medical imaging. Some examples of it are:

- The analysis of histological sections of tumors for detecting the presence or absence of metastases,
- the application on radiological images in order to segment different areas, and
- the categorization of diagnostics based on the classification of images.

In regard to COVID-19 and IP, one of the main problems is the similarity of symptoms related to bacterial and viral pneumonia in chest radiographies. In this way, deep learning can be applied to detect and differentiate clinical nuances. According to the Treatment Guidance

for Critically Ill Patients with COVID-19 from Massachusetts General Hospital [12], one of the clinical futures that characterizes this condition is chest abnormalities, including bilateral opacities. These abnormalities are due in many cases to bacteria that were acquired during pneumonia and viral pneumonia outbreaks. In fact, both conditions can generate confusion, particularly in those with severe illness requiring admission to ICU and mechanical ventilation. In order to differentiate between COVID-19 induced pneumonia and other types of pneumonias, there are some distinctive trends and distribution of CTs (Table V) evidences that could be taken into account [31] in order to apply ML techniques for isolating COVID-19 diagnosis:

- Ground glass opacification (GGO) (88.0%),
- Bilateral involvement (87.5%),
- peripheral distribution (76.0%),
- and multilobar (more than one lobe) involvement (78.8%).

TABLE V. Patterns and Distribution of CT Studies

Imaging Insights	Nº Studies	Nº(%) of registered cases/ total nº of patients
Bilateral involvement	12	435/407(87.5)
Peripheral distribution	12	92/121(76.0)
Posterior involvement	1	41/51(80.4)
Multilobar involvement	5	108/137(78.8)
Ground-glass opacification	22	346/393(88.0)
Consolidation	10	65/204(31.8)

Most usual patterns and distribution on initial CT Studies of 919 Patients with Coronavirus Disease (COVID-19).

Nevertheless, authors in [32] suggest that daily chest X-rays have no positive effect over the outcome in the ICUs once COVID-19 has been diagnosed. These scientific study results are aligned with the IP and AI diagnostic techniques results. Medical images reduce the over exposure to radiation and are more predictive during the first stages of the condition.
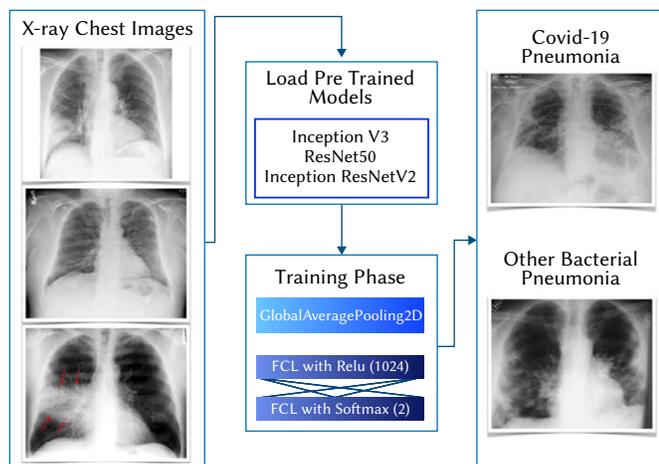
Fig. 10. Schematic representation of the pre-trained models to predict COVID-19 by comparing patients with and without the disease. Model adapted from the figure published [22].

In this way authors in [22] have recently detected COVID-19 by applying IP and convolutional neural networks (CNNs) techniques to the patients' X-ray images. Moreover, this approach is more relevant if we consider that results are more accurate than the "gold standard" for COVID-19 diagnosis, and that the RT-PCR does not have enough sensitivity for diagnosis purposes. This means that many patients

could not receive adequate treatment. Authors in [22] propose using a pre-trained neural network (ResNet50, InceptionV3 and Inception-ResNetV2) processing applied to X-ray images as these are faster and get accurate diagnoses with better outcomes than PCR. Fig. 10 shows a schematic representation of this network comparing X-rays of patients with COVID-19 and those of patients without the disease. The images dataset available is small (100 images in total) so the results should be interpreted with caution. The results of the study show that with this dataset the ResNet50 network has the best performance, with an accuracy of 98%.

Fig. 11 shows the accuracy results coming from the implementation of ResNet50, InceptionV3 and Inception-ResNetV2, which try to make predictions about patients with COVID-19 based on X-ray images. The best accuracy curve model has been obtained by ResNet50.
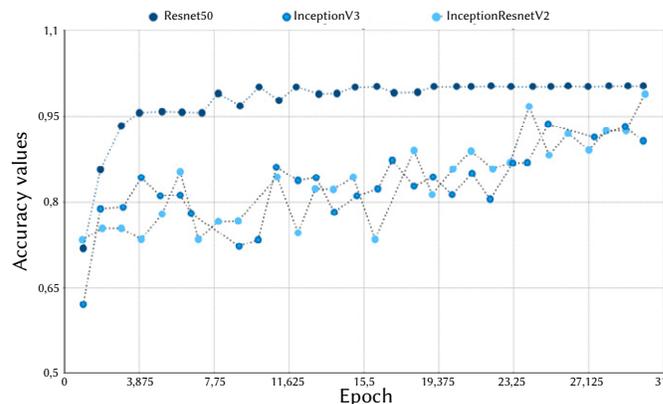
Fig. 11. Authors have used three pre-trained models (ResNet50, InceptionV3 and Inception resNetV2) to make predictions of coronavirus patients based on chest X-ray images. This figure shows the accuracy curve for each model. In this case, the highest accuracy has been reached with ResNet50. Model adapted from the figure published in [22].

Authors in [33] applied a deep learning system to diagnose pneumonia in COVID-19 patients. However, in this case, the research was conducted with CT images used for diagnostic purposes. The research work follows the same argument that authors in [32] made, which is that there is a low positive rate in early stages of COVID-19 obtained from PCR. In this case, the dataset was bigger than the one used in [22], which sum a total of 618 CTs from 110 patients with COVID-19 and 224 patients with Influenza-A pneumonia. In this case, the accuracy was 86.7%. Even with these promising results, this technique is relatively questionable due to use of CTs for routine diagnostic purposes. In terms of health care system efficiency, CTs are not a kind of medical equipment that could be used in situation of ICU and ER collapse. Moreover, CT rooms could lead to a high percentage of viral transmission if these rooms are crowded.

Authors in [4] published in March 2020 a COVID-19 study applying a combination of deep learning architectures (AlexNet, VGG16, VGG19, GoogleNet, ResNet18, ResNet50, ResNet101, InceptionV3, InceptionResNetV2, DenseNet201 and XceptionNet) for extracting relevant features that were used for feeding an SVM classifier. In this study, authors have considered not only patients suffering from COVID-19 but also SARS and MERS. This approach involves similar diseases but with some specific characteristics that present differences among groups. The comparative study (see Table VI) shows that ResNet50 model had the best result, compared to the others. However, the dataset with X-ray images is relatively small to be able to conclude that this system can work properly in real time.

A different approach used the COVID-Net convolutional neural network model [34]. This research project has been proposed as a

chest X-Ray image screening method. This ML tool can be considered as a faster alternative to PCR or at least, as a complement to this in vitro diagnosis method. Moreover, this network has been pre-trained with the ImageNet dataset [35] in order to get better results. After this pre-trained process, the network has been trained with a new dataset COVIDx that comprises 16,756 chest images from 13,645 COVID-19 patients. The COVIDx dataset groups:

- the COVID-19 image dataset [36]. This dataset contains frontal view X-rays images and is the greatest public database for COVID-19 image and prognostic data, and

- the Radiological Society of North America (RSNA) Pneumonia Detection Challenge Dataset [37] is a subset of 30,000 chest radiograph (CXR) exams taken from the NIH CXR14 dataset [38]. In accordance with the whole group of exams, 15,000 exams had positive results for pneumonia or similar pathological conditions such as consolidation and infiltrate.

TABLE VI. The Same Letter(s) Within A column Means That Results Are Not Statically Significant (*P=0.05*) According to Duncan's Multiple Range Test (SPSS Version 26)

Classification Model	Accuracy	Sensitivity	Specificity
AlexNet	.933235d	.934117b,c	.932352c
DenseNet201	.938823d	.943529c,d	.934117c
GoogleNet	.914411b,c	.898235a	.930588b,c
Inceptionv3	.910882b	.911176a,b	.910588b,c
ResNet18	.910882b	.911176a,b	.910588b,c
ResNet50	.953823c	.972941d,c	.934705c
ResNet101	.892647a	.912352a,b	.872941a
VGG16	.927647c,d	.970745c	.880588a
VGG19	.929117d	.951176c,d,c	.907058b
XceptionNet	.939117d	.947647c,d,c	.930588b,c
Inceptionresnetv2	.933235d	.852941c,d,c	.913529b,c

Statistical analysis for SVM models based on features extracted from different CNN models

COVID-Net's results offer three possibilities: non-infection, COVID-19 viral infection and other non-COVID-19 infection. This methodology helps clinicians to stratify patients in an easy way and decide which is the most adequate treatment for each patient. COVID-Net has achieved a 92.6% of accuracy. This percentage is very relevant for taking into consideration the complexity of the ICUs. At the same time, results regarding its sensitivity (Table VII) are quite good with respect to epidemiological requirements such as the limitation over missed COVID-19 cases.

TABLE VII. COVID-Net Sensitivity (%)

Normal	Non-COVID19	COVID-19
97.0	90.0	87.1

Sensitivity effectiveness related to COVID-Net results

C. ML Biomarker Processing For COVID-19

The COVID-19 pandemic has been a one in a century event, whose symptoms are fever, cough, dyspnea, and viral pneumonia. For this reason, diagnostic tests have become very useful for confirming suspected cases, screening patients, and conducting virus surveillance. Currently, the gold standard technique that is used to detect COVID-19 is the PCR test, which is based on the diagnostic protocol provided by the WHO on 13 January 2020 [39]. Fig. 12 shows the steps that are part of the process [40] for identifying active COVID-19 infection in a patient. The whole process consists in three steps:

1. The first one consists in taking a sample from the patient's nose or mouth.

2. After that, the second step continues with a RNA isolation from a patient's sample.

3. The third and final step is the conversion of the RNA into complementary DNA (cDNA) using an enzyme called reverse transcriptase. This component is then re-loaded into a plate and place in a PCR machine. After that the results from the PCR machine are evaluated in a computer. If the level of COVID-19 RNA in the sample is over the threshold, the diagnosis is confirmed, and the patient is positive for COVID-19.

However, COVID-19 patients frequently present additional findings in laboratory testing, which are carried out before ICU admission. ML has been commonly used for the interpretation of data in different fields of biomedicine such as genomics, transcriptomics and proteomics pathways [16]-[17]. At the same time, it has also been used in biological materials in clinical laboratory medicine for identifying biomarkers. Typically, the ML approach is based on rules that are learned by the machine. In general terms, a discussion of rules is not appropriate because predictions are based on non-linear parameters that hide decision-making around the data.
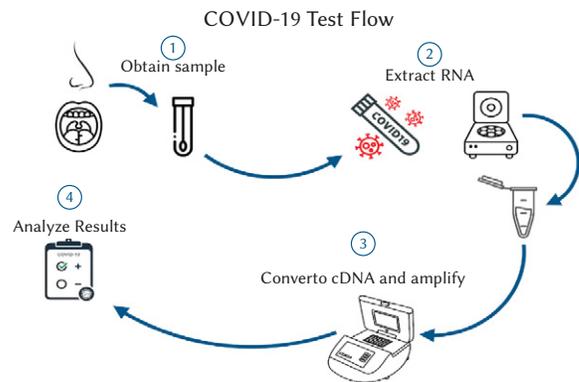
Fig. 12. Workflow for identifying patients with active COVID-19 infection. Model adapted from the figure published [40].

In reference to COVID-19, authors in [41] have found some clinical biomarkers that present abnormal values in the white blood cell and absolute lymphocyte count. In order to isolate which of them should be considered as survival predictors, AI algorithms analyze all these biomarkers, in combination with other clinical findings that can be analyzed. In this way, authors in [42], have applied ML models (XGBoost) by selecting three biomarkers from the *Electronic Health Record System* EHRS: Lactic Dehydrogenase (LDH), Lymphocyte and High-sensitivity C-Reactive Protein (hs-CRP).

The training and test datasets are composed of 1,523 blood samples with complete measures coming from 375 patients. Additionally, the project adds 228 more blood samples for the validation set of 29 patients. Survival prediction has been about 90%. From the point of view of clinical decisions, the biomarkers selected by the algorithm described in Fig. 13 make sense because:

- The increase in Serum LDH is associated with a prognostic marker of lung injury.

- hsCRP could identify risk of death since its increase is associated with ARDS prognosis and inflammation.

- Lymphocyte associated Lymphopenia is common in COVID-19 patients with severity and mortality.
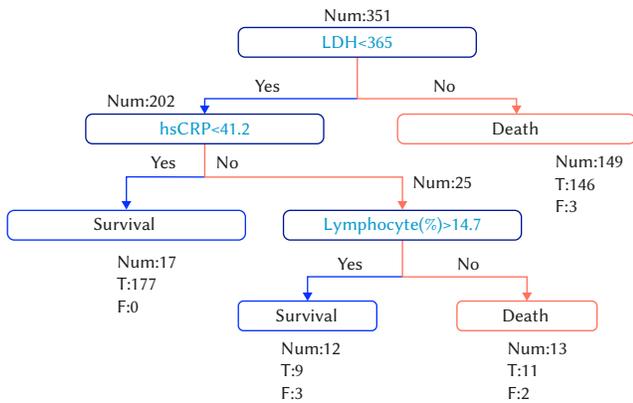
Fig. 13. Based on Yan et al. results, the best performing tree based discover three (LDH, hs-CRP and Lymphocyte) key features and their thresholds values over a population of 351 patients with the aim of being able to identify high risk patients before an irreversible lesion occur. Model adapted from the figure published [42].
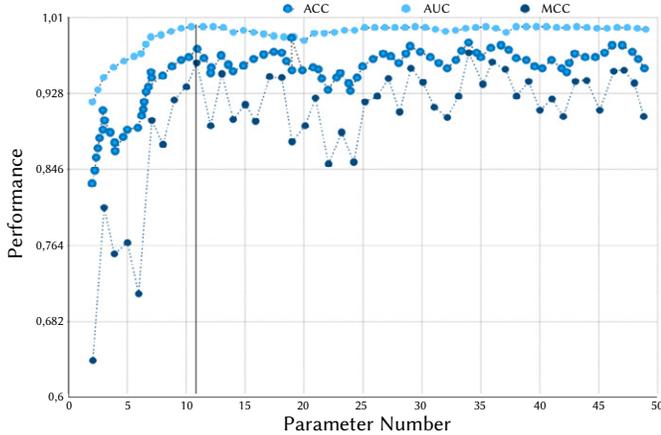
Fig. 14. Model performance: ACC (0.98), MCC (0.95) and AUC (1.0). Model adapted from the figure published [43].

Authors in [43] have developed a tool to triage suspected COVID-19 patients. The accuracy has been 97.95% and 96,97% for the cross-validation set and test set, respectively. However, the model has been also assessed with an external validation obtaining sensitivity, specificity, and overall accuracy of 95.12%, 96,97% and 95.95%, respectively. The predictive model has been conducted based on Random Forest algorithm (RF), which has been able to identify 11 laboratory parameters: Total Proteins, Glucose, Calcium, Creatine Kinase Isoenzyme, Magnesium, Basophil, Total bilirubin, Creatinine, Lactate Dehydrogenase, Kalium and Platelet Distribution. These results (see Fig. 14) could simplify the laboratory blood processes and at the same time, reduce the time scope for applying treatments in COVID-19 patients and improve the disease monitoring.

VI. Conclusions

COVID-19 has shown the need for new clinical pathways focused on prevention, as well as a need for improvements in methods and processes for diagnosis, prognosis and treatment. This pandemic has also had a huge impact on the generation of associated clinical and research data. An important part of these datasets has been generated in ICU and ER rooms. The new illness in many countries has led to or surpassed the collapse in health care systems, due to a shortage of intensive care professionals and a high demand for ICU

equipment. Similar situations have been related to compassionate use of drugs that had not been previously approved for specific COVID-19 clinical use. This unprecedented scenario requires a cross clinical and technical disciplinary approach to tackle the pandemic. Moreover, when it is necessary to take into account patient data, its automatic and centralized management is beneficial. For this reason, COVID-19 patient care will require even more automatic integration and analysis of multiple parameters and data sources.

In this situation computational clinical tools based on AI systems have a huge impact on improvements for dealing with COVID-19 patients. The next goals will focus on respiratory status monitoring in real time, the design of new therapeutic treatments for SARS-CoV-2, hemodynamic management, imaging and laboratory research, as well as reducing possible delays in patient admission and optimizing the clinician's time. Notwithstanding, we must not forget and remember that ML systems are in place to support and not replace the clinician.

References

[1] World Health Organization. (2020). WHO statement regarding cluster of pneumonia cases in Wuhan, China. Beijing: WHO, 9.

[2] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PloS one, vol. 15, no. 3, pp. e0230405, 2020.

[3] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, X. Cheng. Artificial Intelligence & Machine Learning to Fight COVID19. 2020.

[4] P. K. Sethy, S. K. Behera. Detection of coronavirus Disease (COVID-19) based on Deep Features, 2020.

[5] F.A. Murphy. Emerging zoonoses. Emerg Infect Dis. Vol. 4, no. 3, pp. 429-435, 1998.

[6] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir R. Siddique. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. Journal of Advanced Research, 2020.

[7] G. Venkatesan, V. Balamurugan, P. N. Gandhale, R. K. Singh, V. Bhanuprakash. Viral zoonosis: a comprehensive review. Asian Journal of Animal and Veterinary Advances, vol. 5, no. 2, pp. 77-92, 2010.

[8] S. Wan, Y. Xiang, W. Fang, Y. Zheng, B. Li, Y. Hu, X. Huang. Clinical features and treatment of COVID-19 patients in northeast Chongqing. Journal of medical virology. 2020.

[9] T. Klopfenstein, L. Toko, P. Royer, Q. Lepiller, V. Gendrin, S. Zayet Features of anosmia in COVID-19. Médecine et Maladies infectieuses, 2020.

[10] World Health Organization. Clinical management of COVID-19: interim guidance, 27 May 2020 (No. WHO/2019-nCoV/clinical/2020.5). World Health Organization.

[11] L. A. Vaira, G. Salzano, G. Deiana, G. De Riu. Anosmia and ageusia: common findings in COVID-19 patients. The Laryngoscope, 2020.

[12] Massachusetts General Hospital Treatment Guidance for Critically Ill Patients w/ COVID-19. https://bit.ly/mghcovid. Last accessed July 2020.

[13] H. A. Rothan, S. N Byrareddy. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of autoimmunity, 102433, 2020.

[14] R. Li, C. Rivers, Q. Tan, M. B. Murray, E. Toner, M. Lipsitch. Estimated Demand for US Hospital Inpatient and Intensive Care Unit Beds for Patients With COVID-19 Based on Comparisons With Wuhan and Guangzhou, China. JAMA network open, vol. 3, no. 5, pp. e208297-e208297, 2020.

[15] P. Gautret, J. C. Lagier, P. Parola, L. Meddeb, M. Mailhe, B. Doudier, S. Honoré. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. International journal of antimicrobial agents, 105949, 2020.

[16] B. Cao, Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, X. Li. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. New England Journal of Medicine, 2020

[17] M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, G. Xiao. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell research, vol. 30, no. 3, pp. 269-271, 2020.

[18] P. Simon. Too Big to Ignore: The Business Case for Big Data. (Vol. 72).

John Wiley & Sons, 2013.

[19] G. Gutierrez. Artificial Intelligence in the Intensive Care Unit. Critical Care, vol. 24, article number 101, pp. 1-9, 2020.

[20] G. S. Randhawa, M. P. Soltysiak, H. El Roz, C. P. de Souza, K. A. Hill, L. Kari. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. Plos one, vol. 15, no. 4, pp. e023239, 2020.

[21] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, B Xu. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). MedRxiv, 2020.

[22] A. Narin, C. Kaya, Z. Pamuk. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849, 2020.

[23] F.A. Saiz, I. Barandiaran. COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach. International Journal of Interactive Multimedia and Artificial Intelligence, vol 6. n. 12, pp. 11-15, 2020.

[24] F. Cabitza, G. Banfi. Machine learning in laboratory medicine: waiting for the flood? Clinical Chemistry and Laboratory Medicine (CCLM), vol. 56, no. 4, pp. 516-524, 2018.

[25] N. Ferguson, D. Laydon, G. Nedjati, N. Imai, K. Ainslie, M. Baguelin, A. Dighe. Impact of non-pharmaceutical interventions to reduce COVID19 mortality and healthcare demand, 2020.

[26] P. Richardson, I. Griffin, C. Tucker, D. Smith, O. Oechsle, A. Phelan, J. Stebbing. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet (London, England), vol. 395. No. 10223, pp. e30-e31, 2020.

[27] M. J. Tobin. Mechanical Ventilation. New England Journal of Medicine, vol. 330, no. 15, pp. 1056–1061, 1994.

[28] L. Blanch, B. Sales, J. Montanya, U. Lucangelo, O. Garcia-Esquirol, A. Villagra, G Murias. Validation of the Better Care system to detect ineffective efforts during expiration in mechanically ventilated patients: a pilot study. Intensive Care Medicine, vol. 38, no. 5, pp. 772–780, 2012.

[29] B. Gholami, T.S. Phan, W.M. Haddad, A. Cason, J. Mullis, L. Price, J.M. Bailey. Replicating human expertise of mechanical ventilation waveform analysis in detecting patient ventilator cycling asynchrony using machine learning, Computers in Biology and Medicine, 2018.

[30] M. T. Kwong, G. W. Colopy, A. M. Weber, A. Ercole, J. H. Bergmann. The efficacy and effectiveness of machine learning for weaning in mechanically ventilated patients at the intensive care unit: a systematic review. Bio-Design and Manufacturing, vol. 2, no. 1, pp. 31-40, 2019

[31] S. Salehi, A. Abedi, S. Balakrishnan, A. Gholamrezanezhad. Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. American Journal of Roentgenology, pp. 1-7, 2020.

[32] Y. Oba, T. Zaza. Abandoning daily routine chest radiography in the intensive care unit: meta-analysis. Radiology, vol. 255, no. 2, pp. 386-395, 2010.

[33] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, Y. Li. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv preprint, 2020.

[34] L. Wang, A. Wong. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv preprint arXiv:2003.09871, 2020.

[35] J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255, 2009.

[36] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Duong, M. Ghassemi. COVID-19 Image Data Collection: Prospective Predictions Are the Future. arXiv preprint, 2020.

[37] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. Summers. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2097-2106), 2020.

[38] Radiological Society of North America. RSNA pneumonia detection challenge. https://bit.ly/kagglersna. Last accessed July 3, 2020.

[39] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, D. G. Mulders. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance, vol. 25, no. 3, pp. 2000045, 2020.

[40] Explained: how to identify active covid-19 infection in people. https://www.researchamerica.org/blog/explained-how-identify-active-covid-19-infection-people by 06/07/2020. Last accessed July 3, 2020.

[41] M. Arentz, E. Yim, L. Klaff. Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. JAMA. Published online March 19, 2020.

[42] L. Yan, H. T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, X. Huan. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. medRxiv, 2020.

[43] J. Wu, P. Zhang, L. Zhang, W. Meng, J. Li, C. Tong, M. Zhao. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv, 2020.

Sergio Muñoz Lezcano

Sergio Muñoz Lezcano is a full-time innovation and digital health manager at the Spanish Federation of Healthcare Technology Companies. Actually, he is a PhD Student at the Universidad Internacional de La Rioja (UNIR) involved in artificial intelligence research applied in healthcare focus on intensive care units (with emphasis on sepsis condition in adults and neonates). He obtained a MSc in Artificial Intelligence in 2019 at UNIR and MBA in 2000 at ICADE - Universidad Pontificia Comillas. He is also a recognized business mentor working with EU entrepreneurship and startups programs in health care ecosystems and coordinates Spanish innovation Platform in Health Technology.

Fernando Carlos López Hernández

Fernando Carlos López Hernández is a full-time associate professor at the Universidad Internacional de La Rioja (UNIR). His current research interests lie in image processing, computer vision, statistical machine learning and data-driven science. He is the education manager of the Doctorate Program in Computer Science, and a of Course of Robotics for Education. In addition, he works as a lecturer in the Computer Science degree (Statistics, Algebra, Discrete Mathematics, Algorithms Complexity, Image Processing, Signal Processing, Computer Graphics, Compilers). Until 2015, he was the director, education manager as well as a lecturer in the Master in Mobile Apps degree (iOS technologies) at UNIR. Before joining UNIR, he worked for the Universidad Autónoma de Madrid (UAM) as a post-doctoral researcher in European research projects. He obtained his EU-wide recognized PhD degree in Computer Science and Telecommunications in 2010 at the Video Processing and Understanding Lab (VPU Lab) of the UAM, a period during which he worked in 2 European research projects related to his thesis.

Alberto Corbi Bellot

Alberto Corbi Bellot is a senior researcher at the Research Institute for Innovation & Technology in Education (iTED) and as an assistant professor at the Engineering School, which are both part of the Universidad Internacional de La Rioja (UNIR). He is currently involved in a variety of research fields: e-learning standards, systems interoperability, medical physics, radiological protection, STEM education, the monitoring of physical activities, the social implications of technology (with emphasis on social networks), e-health advancement (with emphasis on Alzheimer's disease characterization and clinical information standards) and environmentalism. He has published over 20 research papers on all these subjects, and he is a speaker and knowledge disseminator at radios, podcasts, scientific workshops, magazines, academic settings, and outreach events.

Solving an Optimal Control Problem of Cancer Treatment by Artificial Neural Networks

F. Heydarpour[1], E. Abbasi[2], M. J. Ebadi[3], S. M. Karbassi[1] *

[1] Department of Applied Mathematics, Yazd University, Yazd (Iran)
[2] Department of Computer Science, Yazd University, Yazd (Iran)
[3] Department of Mathematics, Chabahar Maritime University, Chabahar (Iran)

Abstract

Cancer is an uncontrollable growth of abnormal cells in any tissue of the body. Many researchers have focused on machine learning and artificial intelligence (AI) based on approaches for cancer treatment. Dissimilar to traditional methods, these approaches are efficient and are able to find the optimal solutions of cancer chemotherapy problems. In this paper, a system of ordinary differential equations (ODEs) with the state variables of immune cells, tumor cells, healthy cells and drug concentration is proposed to anticipate the tumor growth and to show their interactions in the body. Then, an artificial neural network (ANN) is applied to solve the ODEs system through minimizing the error function and modifying the parameters consisting of weights and biases. The mean square errors (MSEs) between the analytical and ANN results corresponding to four state variables are 1.54e-06, 6.43e-07, 6.61e-06, and 3.99e-07, respectively. These results show the good performance and efficiency of the proposed method. Moreover, the optimal dose of chemotherapy drug and the amount of drug needed to continue the treatment process are achieved.

Keywords

I. Introduction

CANCER is a class of diseases caused by disorders in the natural function of the body cells and it is one of the main reasons of death in the world. Development of cancer happens when the normal control mechanism of the body does not work. In this situation, the former cells do not die but they will grow without control and new abnormal cells are developed. This making additional cells may cause a mass of tissue which is called a tumor. There are many types of cancer, such as lung cancer, prostate cancer, breast cancer, and colorectal cancer. The main and usual treatments for cancer therapy are surgical procedures, hormone therapy, radiation therapy, organic therapy and chemotherapy.

The history of anticancer chemotherapy mathematical modeling comes back to more than forty years ago. Recently, plenty of attempts have been made to mathematically model the dynamics of the population of tumor cell and optimal control theory in order to provide optimal strategies for cancer treatment. New progress in computer and biological sciences have drawn attention to the role of computational methods in the wide scope of cancer research.

We are witnessing the improvement and development of many mathematical models for anticipating the growth of tumor after execution of chemotherapy and for inhibiting the disease progression

within treatment time. This can be done by minimizing the number of cancer cells, minimizing the doses of the drug as well as lessening side effects by the use of optimization methods in which finding optimal drug administration is considered as an optimal control problem (OCP) of a dynamical system [1]-[12]. The drug effect as well as interplays between normal, immune, and tumor cells in these models are seen as a result of the chemotherapy treatment drug for cancer. Based on the optimal measured dosage of the drug and the population number of tumor cells, Calzada et al. [9] and De Pillis and Radunskaya [6] proposed a performance index to decrease the tumor density. El-Gohary [7] investigated the optimal control of tumor, the disorders behavior and the stability of the system in the equilibrium states of tumor's model prior and posterior to the injection. Shuo Wang and Heinz Schattler [11] offered an optimal cancer therapy to decrease the tumor density and its destructive impacts during a specific time period. Urszula Ledzewicz et al. [12] argued the influences of tumor growth disorders and drug resistance on the process of treatment and optimal control analysis of a mathematical model.

The interplay between cancerous, immune, and healthy cells can be modeled as an ordinary differential equations (ODE) system [13]. To explain natural phenomena by using the differential equation, as an application of differential equations, one can convert the natural phenomena and related problems into the differential equations and then solve those differential equations. Since achieving the analytical solution of ODE's systems may not be easy, the researchers have attempted to find the approximate solutions for these systems. Many algorithms, for instance, finite element [14], predictor–corrector [15], finite difference, Runge–Kutta [16], and other methods have been studied and developed to solve these kinds of systems. In

---

* Corresponding author.

E-mail addresses: smkarbassi@yazd.ac.ir (S. M. Karbassi), dr.f.heidarpoor@gmail.com (F. Heydarpour), e.abbasi@yazd.ac.ir (E. Abbasi), ebadi@cmu.ac.ir (M. J. Ebadi).

these methods, the domain needs to be discretized into some finite domains or some points where we have the local approximation of the functions. Therefore, several AI methods especially artificial neural network (ANN) methods have been extensively applied to solve ODEs. Some of the related methods [17]-[30] are given in Table I.

TABLE I. Some ANN Methods for Solving ODEs

Reference	Method
[17]	Hopfield NN
[18][19]	Feedforward NN
[20]	Multi-layer perceptron (MLP)
[21] [22]	Hybrid of NN and optimization techniques
[23]	An unsupervised version of kernel least mean square algorithm
[24]	New algorithms based on NN
[27]	Multilayer perceptron and RBFNN
[29]	Multi-Quadric RBFNN
[30]	Regression-based NN
[41]	Adaptive resonance theory and boosted fuzzy classifier
[42]	ART1 network
[43] [44]	Machine Learning
[45]	Single layer Legendre Neural Network (LeNN)
[26] [46]	ANN
[47]	ANN time-series

In the field of computer sciences, the ANN has been known as a technology which has broad applications in many areas of engineering and science such as robot control, medicine, optimal control, manufacturing system design, pattern recognition, signal processing, business, classification and education (see [31]-[39] and the references therein). The real-time operation, Adaptive learning, Fault-tolerance through surplus information coding and self-organization are some benefits of ANNs.

An ANN is a configuration of artificial intelligence on the base of algorithms which imitate the human brain operation. It uses the existing patterns in the data and makes a generalization from these patterns to classify or predict. Neural networks (NNs) are effective particularly for interpreting the nonlinear data that usually occur in the domain of medicine and biology research. Fundamentally, it is a mathematical model inspired by the structure of systems of biological neurons, similar to the brain processes of the human. A neuron in the ANN is a computational unit whose inputs and outputs are numbers. Fig. 1 shows a simple neuron. Neuron inputs (X_1, X_2, X_3) have their own weights (W_1, W_2, W_3). The output of each neuron is calculated as follows:

$$output = f(W1X1 + W2X2 + W3X3)$$

The function f can be a binary step, linear or non-linear function such as sigmoid.
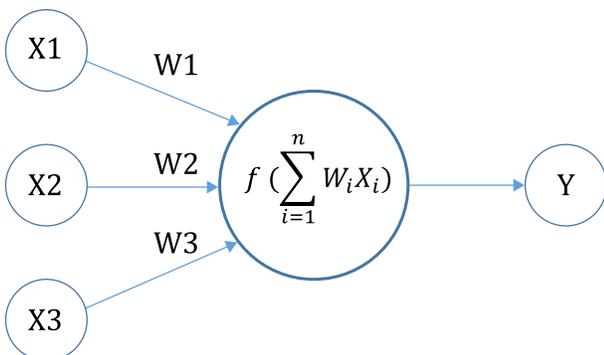
Fig. 1. A simple neuron (neuron is a computational unit, which its inputs and outputs are numbers).
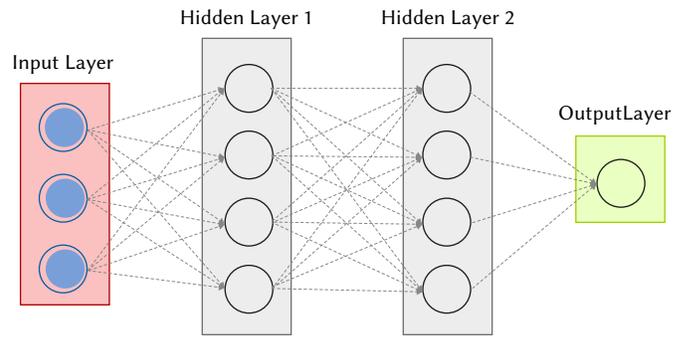
Fig. 2. An ANN Model (a feedforward network).

Fig. 2 is a sample fully connected feedforward neural network consisting of an input layer, two hidden layers and an output layer. The connections among these neurons similar to biological neurons specify the network's behavior which can be realized through the process of backpropagation [40]. In backpropagation, a given data set forms the network input, then network output is constantly compared with ANN output in order to minimize the MSE by making adjustments to the weights of the network. The network achieves higher accuracy after several repeated adjustments and becomes ready for solving complex problems [40]. Nowadays, it is found that biomedical systems should be demonstrated through nonlinear systems that represent ANNs as rich computing sources of biological investigation. In the last three decades, ANNs have been implemented in different domains of cancer medicine (several samples are given in Table I. [41]-[44]). To find the solutions of a nonlinear differential equation numerically, radial basis function (RBF) and multilayer perceptron neural networks with a novel unsupervised training model are utilized by Shirvany et al. [27]. Yazdi et al. [23] proposed and developed a new method based on a version of unsupervised kernel least mean square algorithm (KLMS) to solve ordinary differential equations of first and second order. A neural network on the base of regression was proposed by Mall and Chakraverty [30] to solve ordinary differential equations of higher and lower orders. Mall and Chakraverty [45] proposed a novel model for solving initial as well as boundary value problems on the base of Legendre Neural Network (LeNN) method with a single layer. A neural network model is proposed by Naushad et al. [46] for predicting the possibility of a woman growing breast cancer by analyzing the interplay of nutrients, genes, and demographic indicators. Another study having applications in quality warranty research and having modeled the treatment data resulted from the use of ANNs for cancer patients' radiotherapy demonstrated computation superiority for time-series ANN modeling over traditional techniques [47].

Although all the presented ordinary differential equations (ODE) systems used for cancer therapy have attempted to find the optimal doses of chemotherapy drugs, none of them has focused on the prediction of the drug dose needed to continue the treatment process. To fill the existing gap, the present work seeks new solutions to the system of ordinary differential equations (ODEs) for achieving the optimal dose of chemotherapy drugs and anticipating the amount of drug needed to continue the treatment process. To this end, a new method including Feedforward Neural Network (FFNN) is presented to find solutions to ODE at the same time predict cancer issues. Showing cancerous, immune and healthy cells interaction in the body and anticipating the tumor growth, this method predicts the optimum drug dose that should be injected with the most beneficial effect on tumor, immune, and normal cells. The dynamic behaviour of the system is analyzed by investigating the different equilibrium points in which the population number of tumor cells is zero or few. On the base of the ANN, the trial solution involves two components:

the first component with no adjustable parameters satisfies the initial/boundary conditions and the second component comprises the output of ANN. There are some advantages to the approximate solution of ANN comparing to the traditional numerical methods. First, we have the differentiability and closeness to the analytical solution for the approximate solution. Second, when the training points number increases, we do not encounter with more computation complexity. Third, the speed with which the parameters of the ANN (biases and weights) are modified increases to reduce error function. The proposed method is compared with two machine-learning methods, Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) as well as two well-known numerical methods, Finite Element and Runge–Kutta.

The organization of this paper is as follows. In Section II, the modeling of the tumor and its equilibrium points are presented. Section III gives the proposed method including the formulation of the first order ODE and the structure of ANN. The numerical results are discussed in Section IV. Finally, some conclusions are given in Section V.

II. The Tumor Model

The tumor expansion has attracted the attention of technicians, mathematicians and medical specialists, which can be modeled as systems of partial differential equations (PDEs) or ordinary differential equations (ODEs). The treatment of malignancy tumor is crucial. Hence, our model represents an optimal medical therapy for malignancy tumor derived from [13], [48]. The emphasis of this model is not on a specific kind of tumors, but it is a dynamic model to deal with the tumor expansion problem in its society [13], [48]. The explanation of our proposed model, its reduction and stability are brought in the following.

A. Tumor With Drug Mathematical Model and Stability of the Equilibrium States

The tissue close to the tumor site is targeted by the model consisting of three anomalous cell populations. $I(t)$, $T(t)$, and $N(t)$ illustrate immune cells, tumor cells, and normal cells at the time respectively. To show the impact of drug therapy, a state D is added representing the concentration of the drug in the bloodstream. The nonlinear differential system of equations expresses each component t dynamics for the treatment of chemotherapy which is shown below [13], [48].

$$\dot{T}(t) = T(t)[r_1(1 - b_1T(t)) - c_2I(t) - c_3N(t) - a_1(1 - e^{-D(t)})];$$

$$\dot{N}(t) = N(t)[r_2(1 - b_2N(t)) - c_4T(t) - a_2(1 - e^{-D(t)})];$$

$$\dot{I}(t) = s + I(t)\left[\frac{r_3T(t)}{(\alpha + T(t))} - c_1T(t) - d_1 - a_3(1 - e^{-D(t)})\right];$$

$$\dot{D}(t) = -d_2D + \upsilon(t), \qquad (1)$$

where $\upsilon(t)$ is used for the amount of drug consumed orally or by injections before impacting on the bloodstream and d_2 is the number of death per capita for the drug. Here, because of physiological reasons, $I(t)$, $T(t)$ and $N(t)$ values and their parameters should be positive. A situation that a system does not alter is called the equilibrium point. When a dynamical system is represented by the differential equation to find the equilibrium point, one can take the first derivative of the system equal to zero.

Definition 2.1. A point $x^* \in R^n$ is called an equilibrium point of the ODEs system $\dot{x} = F(x)$ if it satisfies $F(x^*) = 0$, where

$$F(x) = \begin{bmatrix} F_1(x_1, x_2, \dots, x_n) \\ F_2(x_1, x_2, \dots, x_n) \\ \vdots \\ F_n(x_1, x_2, \dots, x_n) \end{bmatrix} \quad [49].$$

In this case, a patient is called the treated patient if the dynamic system of tumor model without drug is reached to the equilibrium points. In the equilibrium points, the population number of tumor cells is zero or few. Therefore, it is necessary to find these points. After finding the equilibrium points of the system, it is essential to investigate the stability of the model [13], [49]. According to the parameters values of the system, there are three kinds of equilibria for a system without drug input as follows [13].

B. Tumor-free Equilibrium Point

In this case, the population of tumor cells is zero and the normal cells remain alive. If

$$r_1 < c_3 + {c_2 s}/{d_1}$$

then, we have the asymptotical stability of this equilibrium point whose form is as follows:

$$T_f{}^* = ({1}/{b_2}, 0, {s}/{d_1})$$

1. Dead Equilibrium Points

If the population number of normal cells is zero, this equilibrium point is said to be "dead". We have two types of these equilibrium points. In the first type, both population number of the tumor and normal cells are zero and the dead equilibrium point is as follows

$$D_1^* = (0, 0, {s}/{d_1})$$

in the second type, the population number of normal cells is zero but the tumor cells remain alive and the dead equilibrium point is as follows

$$D_2^* = (0, x, f(x))$$

Where x is a solution of (2) which is nonnegative value and $f(x)$ is introduced in (3):

$$x + \left({c_2}/{r_1 b_1}\right)f(x) - {1}/{b_1} = 0 \qquad (2)$$

$$f(x) = \frac{s(x+a)}{c_1 x(x+a) + d_1(x+a) - \rho x} \qquad (3)$$

D_1^* is always an unstable equilibrium point. Depending on the system parameters values, we may have D_2^* as stable or unstable equilibrium point.

2. Coexisting Equilibrium Point

In this kind, both tumor and normal cells remain alive while they have coexistence and their population number is not zero. The form of this kind can be seen below:

$$C^* = (g(y), y, f(y))$$

where y is a solution of the following equation with nonnegative value:

$$y + \left({c_2}/{r_1 b_1}\right)f(y) + \left({c_3}/{r_1 b_1}\right)g(y) - {1}/{b_1} = 0 \qquad (4)$$

And $g(y)$ is as follows:

$$g(y) = {1}/{b_2} - \left({c_4}/{r_2}\right)y$$

Relying on the parameters of the system, this equilibrium point is fairly stable or unstable. For the chemotherapy remedy, the system must ideally come close to two equilibrium points including T_f^* and any coexisting equilibrium point when the value of y is small and

$g(y)$ is near to 1. The objective of chemotherapy treatment is that the system is brought into an attraction area in which equilibrium point can be tumor free or can be in the presence of only small population number of tumors.

C. Reducing Parameters of the System

In this section, we will decrease the parameters of the system by determining the new parameters of the system and defining new variables. In order to decrease the parameters of the system, we use some new variables as

$$\psi_1(t) = b_1 T(t), \psi_2(t) = b_2 N(t), \psi_3(t) = \frac{d_2 I(t)}{s}, \psi_4(t) = D(t), \ v(t) = d_2 u_4(t), \tau = d_2 t \tag{5}$$

We have the following system by substitution (5) into (1)

$$\dot{\psi}_1 = \psi_1 \left[k_1 (1 - \psi_1) - n_2 \psi_3 - n_3 \psi_2 - m_1 (1 - e^{-\psi_4}) \right]$$
$$\dot{\psi}_2 = \psi_2 \left[k_2 (1 - \psi_2) - n_4 \psi_1 - m_2 (1 - e^{-\psi_4}) \right]$$
$$\dot{\psi}_3 = 1 + \psi_3 \left[\frac{k_3 \psi_1}{v_1 + \psi_1} - n_1 \psi_1 - v_2 - m_3 (1 - e^{-\psi_4}) \right]$$
$$\dot{\psi}_4 = u - \psi_4 \tag{6}$$

Where the relations between the new parameters of the system and old parameters of the system are as follows:

$$k_i = \frac{r_i}{d_2}, \qquad m_i = \frac{a_i}{d_2}, \qquad (i = 1,2,3),$$
$$n_1 = \frac{c_1}{b_1 d_2}, \quad n_2 = \frac{c_2 s}{d_2^2}, \ n_3 = \frac{c_3}{b_2 d_2},$$
$$n_4 = \frac{c_4}{b_1 d_2}, \quad v_1 = \alpha b_1, \quad v_2 = \frac{d_1}{d_2} \tag{7}$$

In comparison with the system (1), solving the system (6) is simpler because of the reduction in the numbers of the system parameters.

III. Proposed Method

A. First Order ODEs System Formulation

Now, a system of first order initial value differential equations is considered as follows:

$$\frac{d\psi_k}{dt} = f_k(t, \psi_1, \psi_2, \psi_3, \psi_4), k = 1,2,3,4 \text{ and } t \in [a, b] \tag{8}$$

Where $\psi^k(a) = A_k, k = 1,2,3,4$ and ψ_{T_k} indicates the kth trial solution based on ANN which can be shown here as:

$$\psi_{T_k}(t, p_k) = A_k + (t-a)N_k(t, p_k), k = 1,2,3,4 \tag{9}$$

Where $N_k(t, p_k)$ is the ANN's output corresponding to input data t and parameters p_k. The initial condition is satisfied by the trial solution $\psi_{T_k}(t, p_k)$. The derivative of the trial solution $\psi_{T_k}(t, p_k)$ is as follows:

$$\frac{d\psi_{T_k}(t, p_k)}{dt} = N_k(t, p_k) + (t - a)\frac{dN_k(t, p_k)}{dt}, k = 1,2,3,4 \tag{10}$$

B. The Error Function and Elements of the ANN

In this case, the error function is given with

$$E(p) = \sum_{i=1}^{n} \sum_{k=1}^{4} \left[\frac{d\psi_{\tau_k}(t_i, \ p_k)}{dt} \right.$$
$$\left. - f_k(t_i, \psi_{\tau_1}(t_i, p_1), \psi_{\tau_2}(t_i, p_2), \psi_{\tau_3}(t_i, p_3), \psi_{\tau_4}(t_i, p_4)) \right]^2 \tag{11}$$

where n is the number of samples. The error function is composed of two terms. The first term described in (10) consists of the network output and its derivative. The second term involves the desired output. The error function (11) is minimized by applying *fminunc*

according to (12) which is a toolbox function of Matlab for solving the unconstrained optimization problem.

$$X = \text{fminunc}(\text{fun}, X_0) \tag{12}$$

In equation (12), the output of the *fminunc* is the local minimum of the function described in *fun* with respect to the parameters initialized in X_0. In this case, *fun* is the error function that is presented in (11) and the parameters are weights that are initialized randomly. With regard to a multilayer perceptron neural network with three layers including one input neuron and one output neuron in the input and output layers and a hidden layer with m neurons, the neural network output is computed for the input t as below

$$N(t, p) = \sum_{j=1}^{m} v_j \sigma(z_j) \tag{13}$$

where m is the number of neurons in the hidden layer, v_j's are the weights between hidden and output layers corresponding to j^{th} neuron in the hidden layer, and the activation function of each neuron in the hidden layer is considered as sigmoid given by:

$$\sigma(z_j) = \frac{1}{1 + e^{-z_j}} \tag{14}$$

where z_j is:

$$z_j = w_j t + u_j \tag{15}$$

where w_j's are weights between input and hidden layers and the biases indicated by u_j. The network output derivative with respect to input t is as follows:

$$\frac{dN(t, p)}{dt} = \sum_{j=1}^{m} v_j w_j \sigma'(z_j) \tag{16}$$

In the next section, the numerical results will be shown while this strategy is applied for therapy and control of cancer modeled as an ODEs system.

IV. Numerical Results

In this section, the results of ANN are compared with the analytical solution using MATLAB R2015b, on a 2.4 GHz Intel(R) Core(TM) i7-4510U laptop running Windows 10 Ultimate with 12.00 GB main memory. The values of the ODEs system parameters presented in Section II are determined according to the following [7]:

$$k_1 = 30, \ k_2 = 48, k_3 = 29, n_1 = 2, n_2 = 1.3, n_3 = 0.47,$$
$$n_4 = 8, m_1 = 9, m_2 = 15, m_3 = 4, v_1 = 0.25, v_2 = 5,$$
$$\gamma_1 = 10, \gamma_2 = 20, \gamma_3 = 8, \gamma_4 = 5, \gamma = 15$$

By considering these parameter values, a four-coupled first order ODEs system is obtained as follows:

$$\dot{\psi}_1 = \psi_1 \left[30(1 - \psi_1) - 1.3\psi_3 - 0.47\psi_2 - 9(1 - e^{-\psi_4}) \right]$$
$$\dot{\psi}_2 = \psi_2 [48(1 - \psi_2) - 8\psi_1 - 15(1 - e^{-\psi_4})]$$
$$\dot{\psi}_3 = 1 + \psi_3 \left[\frac{29\psi_1}{0.25 + \psi_1} - 2\psi_1 - 5 - 4(1 - e^{-\psi_4}) \right]$$
$$\dot{\psi}_4 = u - \psi_4, \tag{17}$$

where the initial conditions are $\psi_1(0) = 2.5$, $\psi_2(0) = 2.5$, $\psi_3(0) = 1.55$ and $\psi_1(0) = 1.35$. According to (9), the trial solutions are given by

$$\psi_{T1}(t, p) = 2.5 + tN_1(t, p_1),$$
$$\psi_{T2}(t, p) = 0.25 + tN_2(t, p_2),$$
$$\psi_{T3}(t, p) = 1.55 + tN_3(t, p_3),$$
$$\psi_{T4}(t, p) = 1.35 + tN_4(t, p_4).$$

According to the proposed method in Section III, four NNs

associated with ψ_1 to ψ_4 equations are trained with 156 samples by considering the step size of 0.0008 within the interval [0 0.125]. Fig. 3 illustrates the ANN and analytical results. The analytical results are obtained by Pontryagin minimum principle.

As it can be seen in Fig. 3, ANN results are a good approximation of analytical solutions because the MSEs between the analytical and ANN results corresponding to equations ψ_1 to ψ_4 are 1.54e-06, 6.43e-07, 6.61e-06 and 3.99e-07, which are denoted by E_1, E_2, E_3, and E_4, respectively.

In Table II, a comparison is provided between analytical solutions and ANN results for the different number of neurons in the interval [2, 8] with the step size 2. In Table II, for each equation except for ψ_4, the best results are obtained for six neurons in the hidden layer. As it can be seen from Table II, when the number of neurons in the hidden layer increases from 2 to 4, the MSEs decrease to the amount of 1.5206e-03, 2.8931e-03, and 4.709e-06, for ψ_1, ψ_3, and ψ_4 while the MSE does not change for equation ψ_2. In addition, when the number of neurons increases from 4 to 6, the MSEs of the equations ψ_1, ψ_2, and ψ_3, except for ψ_4, increase. By increasing the number of neurons from 6 to 8, the MSEs increase for ψ_1, ψ_2, and ψ_3, whereas the MSE decreases for ψ_4. The running time of ANN corresponding to each equation ψ_1 to ψ_4 is given in Table III.

TABLE II. Mean Square Error Between Analytical and ANN Results for 2, 4, 6 and 8 Neurons in Hidden Layer of ANN

Number of Neurons MSEs	2	4	6	8
ψ_1	0.0016	7.94e-05	1.54e-06	3.44e-06
ψ_2	0.00013	0.00013	6.43e-07	0.00013
ψ_3	0.0029	6.84e-06	6.61e-06	0.00013
ψ_4	5.07e-06	3.61e-07	3.99e-07	3.84e-07

TABLE III. Running Time of ANN in Seconds for Six Neurons in Hidden Layer

Optimal control equation	ψ_1	ψ_2	ψ_3	ψ_4
Running time	11.88	9.46	9.44	8.84

Ideally, the modeling should improve treatment strategies and lead to the development of new approaches. One of the goals of mathematical models is dosimetry, accurate prediction of dose response or simulation of different dose programs. Fig. 3 (d) shows the amount of drug used for different time intervals during treatment, and Figs. 3(a), 3(b) and 3(c) show the effect of the drug on tumor, normal and immune cells, respectively.

Fig. 3(d) shows that we increase the amount of drug used to counteract tumor cells till t = 0.08. However, by decreasing tumor cells and increasing immunity after t=0.08, we gradually decrease the amount of drug. From now on, tumor cells will decline due to the competition between them and immune cells. In other words, as the immune system increases, the amount of used drug and the tumor cells decrease. Where the amount of normal cells is increasing to prevent additional toxicity of chemotherapy injection, we try to keep them at 0.25 during treatment. To prevent self-control of the immune system on cancer cells, we assume the number of primary immune cells so small and the number of tumor cells so large that tumor growth is unavoidable unless chemotherapy is used. Support Vector Regression (SVR) is a version of SVM for solving regression problem which was proposed by Drucker, H. et al. [50].
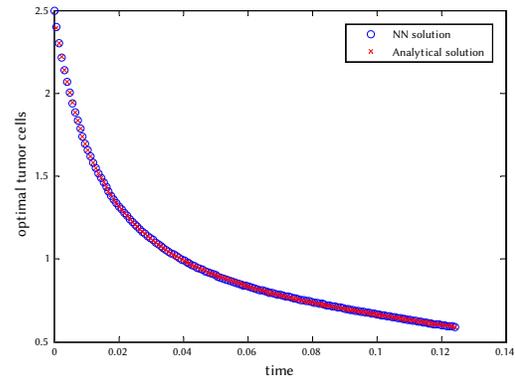
(a)
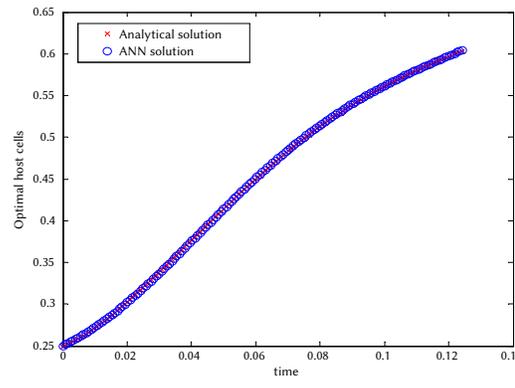
(b)
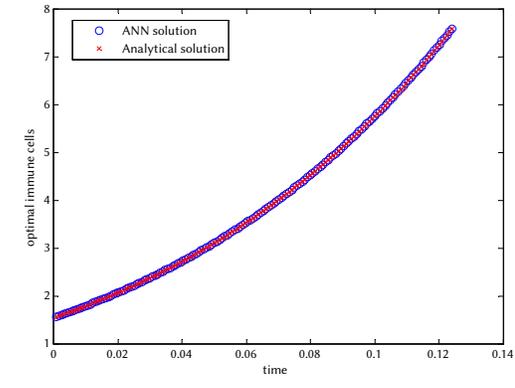
(c)
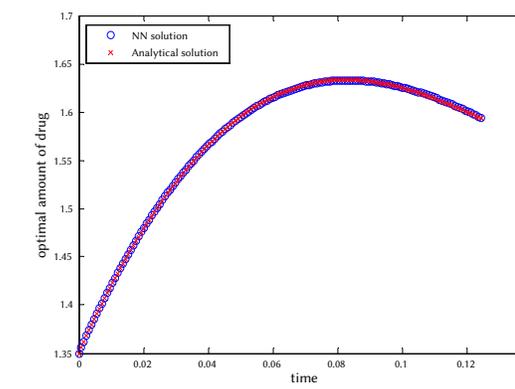
(d)

Fig. 3. Analytical and ANN solution for (a) Optimal tumor cells, (b) Optimal host cells, (c) Optimal immune cells and (d) Optimal amount of drug.

Currently, therapies are expensive and time-consuming. They often result in variable treatment outcomes among patients and their harmful effects are often unpredictable [51]. We compare the results of ANN with the SVR model. The Optimal amount of drug by using SVR and ANN is shown in Fig. 4. The MSEs between the analytical and SVR results corresponding to equations are 1.48, 0.007, 3.5816 and 0.0583 respectively. By comparing the results of ANN and SVR, it can be clearly seen that ANN performs better than SVR. In addition, Table IV shows the MSEs comparisons between the proposed method, SVR, MLP, Finite Element, and Runge–Kutta methods. We can clearly see that the proposed method provides more accurate results compared to the SVR and MLP machine learning methods as well as Finite Element and Runge–Kutta numerical methods.
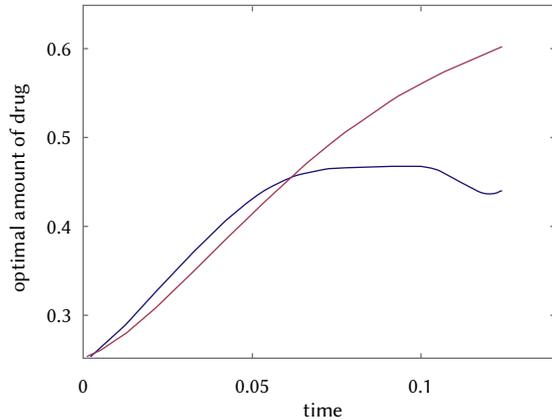
Fig. 4. Optimal amount of drug by using SVR(Red) and ANN(Blue).

Remark 4.1: In this paper, the goal is not just to find solutions to the ODEs system because the given ODE system can be solved by different numerical methods. Solving the system, the NN allows us to anticipate the process of cancer treatment here providing a great advantage over other methods for solving ODE systems. Using the proposed NN, the optimal dose of the chemotherapy drug injected is shown in Fig. 3 (d).

TABLE IV. MSEs Comparisons Between SVR, MLP, Finite Element, Runge–Kutta, and Proposed Method

MSEs / Methods	SVR	Finite Element	Here	MLP	Runge-Kutta
ψ_1	1.48	3.24e-04	**1.54e-06**	7.29e-05	1.003
ψ_2	0.007	2.12e-05	**6.43e-07**	8.78e-06	0.002
ψ_3	3.5816	1.37e-04	**6.61e-06**	5.61e-05	1.247
ψ_4	0.0583	7.51e-05	**3.99e-07**	7.82e-07	0.0018

V. Conclusion

In this paper, a system of ODEs was used to anticipate the tumor growth and to show cancerous, immune and healthy cells interaction in the body. Also, the dynamic behaviour of the system was analyzed by investigating the different equilibrium points in which the population number of tumor cells is zero or few. Using the proposed ANN, not only the optimal dose of chemotherapy drug that should be injected was shown, but also the amount of the drug needed to continue the treatment process was predicted. This can lead to decreasing the number of tumor cells and increasing the number of immune and healthy cells to their utmost. Simulation results had shown the good performance and efficiency of the proposed method. Compared to existing methods in cancer treatment and diagnostic centers that require advanced equipment with expensive costs, this method can be used without trial and error to predict the process of cancer treatment. In the future works, we will design and implement effective closed loop control systems to control the dose of a chemotherapy drug to reduce the side effects of receiving an additional drug.

References

[1] F. Michor and K. Beal, "Improving cancer treatment via mathematical modeling: surmounting the challenges is worth the effort," *Cell 163*, vol. 3, no. 5, pp. 1059-1063, 2015, doi: 10.1016/j.cell.2015.11.002.

[2] P. Khalili, R. Vatankhah, and S. Taghvaei, "Optimal sliding mode control of drug delivery in cancerous tumor chemotherapy considering the obesity effects", *IET Systems Biology*, vol. 12, no. 4, pp. 185-189, 2018, doi: 10.1049/iet-syb.2017.0094.

[3] X. Yang, H. He, D. Liu, and Y. Zhu, "Adaptive dynamic programming for robust neural control of unknown continuous-time non-linear systems," *IET Control Theory & Applications*, vol. 11, no. 14, pp. 2307-2316, 2017, doi: 10.1049/iet-cta.2017.0154.

[4] B. Ribba, B. You, M. Tod, P. Girard, B. Tranchand, V. Trillet-Lenoir, and G. Freyer, "Chemotherapy may be delivered based on an integrated view of tumor dynamics," *IET Systems Biology*, vol. 3, pp.180-190, 2009, doi: 10.1049/iet-syb.2008.0104.

[5] E. Shakeri, G. Latif-Shabgahi, and A. E. Abharian, "Adaptive non-linear control for cancer therapy through a Fokker–Planck observer," *IET Systems Biology*, vol. 12, no. 2, pp. 73-82, 2018, doi: 10.1049/iet-syb.2017.0032.

[6] L. G. De Pillis and A. Radunskaya, "A mathematical tumor model with immune resistance and drug therapy: an optimal control approach," *Computational and Mathematical Methods in Medicine*, vol. 3, no. 5, pp. 79-100, 2001, doi: 10.1080/10273660108833067.

[7] A. El-Gohary, "Chaos and optimal control of equilibrium states of tumor system with drug," *Chaos Solitons & Fractals*, vol. 41, no. 2, pp. 425-435, 2009, doi: 10.1016/j.chaos.2008.02.003.

[8] M. Engelhart, S. Lebiedz, and S. Sager, "Optimal control for selected cancer chemotherapy ODE models: a view on the potential of optimal schedules and choice of objective function," *Mathematical Biosciences*, vol. 229, no. 1, pp. 123-134, 2011, doi: 10.1016/j.mbs.2010.11.007.

[9] M. C. Calzada, E. Fernández-Cara, and M. Marín, "Optimal control oriented to therapy for a free-boundary tumor growth model," *Journal of Theoretical Biology*, vol. 325, no. 7, pp. 1-11, 2013, doi: 10.1016/j.jtbi.2013.02.004.

[10] T. Agrawal, M. Saleem, and S. K. Sahu, "Optimal control of the dynamics of a tumor growth model with Hollings' type-II functional response," *Computational and Applied Mathematics*, vol. 33, no. 6, pp. 591–606, 2014, doi: 10.1007/s40314-013-0083-x.

[11] S. Wang and H. Schattler, "Optimal control of a mathematical model for cancer chemotherapy under tumor heterogeneity," *Mathematical Biosciences and Engineering*, vol. 13, pp. 1223-1240, 2016, doi: 10.3934/mbe.2016040.

[12] U. Ledzewicz, S. Wang, H. Schattler, N. André, M. A. Heng, and E. Pasquier, "On drug resistance and metronomic chemotherapy: A mathematical modeling and optimal control approach," *Mathematical Biosciences and Engineering*, vol. 14, pp. 217-235, 2017, doi: 10.3934/mbe.2017014.

[13] L. G. Pillis and A. Radunskaya, "The dynamics of an optimally controlled tumor model: A case study," *Mathematical and Computer Modelling*, vol. 37, no. 11, pp. 1221-1244, 2003, doi: 10.1016/S0895-7177(03)00133-X.

[14] J. N. Reddy, *An introduction to the finite element method*, New York, USA: McGraw—Hill, 2006.

[15] J. Douglas and B. F. Jones, "On predictor–corrector methods for nonlinear parabolic differential equations," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 1, pp. 195–204, 1963, doi: 10.1137/0111015.

[16] A. Wambecq, "Rational Runge–Kutta methods for solving systems of ordinary differential equations," *Computing*, vol. 20, pp. 333-342, 1978, doi: 10.1007/BF02252381.

[17] H. Lee and I. S. Kang, "Neural algorithms for solving differential equations," *Journal of Computational Physics*, vol. 91, no. 1, pp. 110–117, 1990, doi: 10.1016/0021-9991(90)90007-N.

[18] A. J. Meade Jr and A. A. Fernandez, "The numerical solution of linear

ordinary differential equations by feed forward neural networks," *Mathematical and Computer Modelling,* vol. 19, no. 12, pp. 1-25, 1994, doi: 10.1016/0895-7177(94)90095-7.

[19] A. J. Meade Jr and A. A. Fernandez, "Solution of nonlinear ordinary differential equations by feed forward neural networks," *Mathematical and Computer Modelling,* vol. 20, no. 9, pp. 19-44, 1994, doi: 10.1016/0895-7177(94)00160-X.

[20] I. E. Lagaris, A. Likas, and D. I. Fotiadis, "Artificial neural networks for solving ordinary and partial differential equations," *IEEE Transactions on Neural Networks,* vol. 9, no. 5, pp. 987-1000, 1998, doi: 10.1109/72.712178.

[21] B. Liu and B. Jammes, "Solving ordinary differential equations by neural networks," in *13th European Simulation Multi-Conference Modelling and Simulation: A Tool for the Next Millennium,* Warsaw, Poland, June 14, 1999.

[22] A. Malek and R. S. Beidokhti, "Numerical solution for high order deferential equations using a hybrid neural network-optimization method," *Applied Mathematics and Computation,* vol. 183, no. 1, pp. 260-271, 2006, doi: 10.1016/j.amc.2006.05.068.

[23] H. S. Yazdi, M. Pakdaman, and H. Modaghegh, "Unsupervised kernel least mean square algorithm for solving ordinary differential equations," *Neurocomputing,* vol. 74, no. 12-13, pp. 2062-2071, 2011, doi: 10.1016/j.neucom.2010.12.026.

[24] J. Abdul Samath and N. Selvaraju, "Solution of matrix Riccati differential equation for nonlinear singular system using neural networks," *International Journal of Computer Applications*, vol. 1, no. 29, pp. 48–54, 2010, doi: 10.1.1.206.3895.

[25] L. P. Aarts and P. Van der Veer, "Neural network method for solving partial differential equations," *Neural Processing Letters*, vol. 14, pp. 261–271, 2001, doi: 10.1023/A:1012784129883.

[26] S. A. Hoda and H. A. Nagla, "Neural network methods for mixed boundary value problems," *International Journal of Nonlinear Science*, vol. 11, no. 3, pp. 312–316, 2011.

[27] Y. Shirvany, M. Hayati, and R. Moradian, "Multilayer perceptron neural networks with novel unsupervised training method for numerical solution of the partial differential equations," *Applied Soft Computing,* vol. 9, no. 1, pp. 20-29, 2009, doi: 10.1016/j.asoc.2008.02.003.

[28] K. McFall and R. Mahan, "Artificial neural network for solution of boundary value problems with exact satisfaction of arbitrary boundary conditions," *IEEE Transactions on Neural Networks*, vol. 20, no. 8, pp. 1221–1233, 2009, doi: 10.1109/TNN.2009.2020735.

[29] N. Mai-Duy and T. Tran-Cong, "Numerical solution of differential equations using multi quadric radial basis function networks," *Neural Networks,* vol. 14, no. 2, pp. 185-199, 2001, doi: 10.1016/S0893-6080(00)00095-2.

[30] S. Mall and S. Chakraverty, "Regression based neural network training for the solution of ordinary differential equations," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 4, no. 2, pp. 136-149, 2013, doi: 10.1504/IJMMNO.2013.055203.

[31] R. T. Rockafellar, "Linear-quadratic programming and optimal control," *SIAM Journal on Control and Optimization,* vol. 25, no. 3, pp. 781-814, 1987, doi: 10.1137/0325045.

[32] Y. S. Xia, G. Feng, and J. Wang, "A primal-dual neural network for online resolving constrained kinematic redundancy in robot motion control," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 35, no. 1, pp. 54-64, 2005, doi: 10.1109/TSMCB.2004.839913.

[33] D. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, New York, USA: Academic press, 2014.

[34] J. Wang, "Primal and dual neural networks for shortest-path routing," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* vol. 28, no. 6, pp. 864-869, 1998, doi: 10.1109/3468.725357.

[35] Q. Tao, X. Liu, and X. Cui, "A linear optimization neural network for associative memory," *Applied Mathematics and Computation,* vol. 171, no. 2, pp. 1119-1128, 2005, doi: 10.1016/j.amc.2005.01.123.

[36] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear programming: theory and algorithms*, New York, USA: John Wiley & Sons, 2013.

[37] M. J. Ebadi, A. Hosseini, and M. M. Hosseini, "A projection type steepest descent neural network for solving a class of nonsmooth optimization problems," *Neurocomputing*, vol. 235, pp. 164-181, 2017, doi: 10.1016/j.neucom.2017.01.010.

[38] M. J. Ebadi, M. M. Hosseini, and S. M. Karbassi, "An efficient one-layer

recurrent neural network for solving a class of nonsmooth pseudoconvex optimization problems," *Journal of Theoretical and Applied Information Technology,* vol. 96, no. 7, pp. 1999-2015, 2018.

[39] F. C. Morabito, M. Campolo, C. Ieracitano, J. M. Ebadi, L. Bonanno, A. Bramanti, S. Desalvo, N. Mammone, and P. Bramanti, "Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings," in *IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI),* Bologna, Italy, pp. 1-6, 2016, doi: 10.1109/RTSI.2016.7740576.

[40] P. J. Drew and J. R. T. Monson, "Artificial neural networks," *Surgery*, vol. 127, no. 1, pp. 3-11, 2000, doi: 10.1067/msy.2000.102173.

[41] H. Takahashi, Y. Murase, T. Kobayashi, and H. Honda, "New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index," *Biochemical Engineering Journal,* vol. 33, no. 2, pp. 100-109, 2007, doi: 10.1016/j.bej.2006.08.004.

[42] S. Swathi, G. Babu, R. Sendhilkumar, and S. N. Bhukya, Performance of ART 1 Network in the Detection of Breast Cancer, in *International conference on Computer design and Engineering (ICCDE)*, Singapore 2015, pp. 100-105.

[43] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol 13, pp. 8-17, 2015, doi: 10.1016/j.csbj.2014.11.005.

[44] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, pp. 59-77, 2006, doi: 10.1177/117693510600200030.

[45] S. Mall and S. Chakraverty, "Application of Legendre Neural Network for solving ordinary differential equations," *Applied Soft Computing,* vol. 43, pp. 347-356, 2016, doi: 10.1016/j.asoc.2015.10.069.

[46] S. M. Naushad, M. J. Ramaiah, M. Pavithrakumari, J. Jayapriya, T. Hussain, S. A. Alrokayan, S. R. Gottumukkala, R. Digumarti, and V. K. Kutala, "Artificial neural network-based exploration of gene-nutrient interactions in folate and xenobiotic metabolic pathways that modulate susceptibility to breast cancer," *Gene,* vol. 580, no. 2, pp. 159-168, 2016, doi: 10.1016/j.gene.2016.01.023.

[47] Q. Li and M. F. Chan, "Predictive time-series modeling using artificial neural networks for Linac beam symmetry: an empirical study," *Annals of the New York Academy of Sciences,* vol. 1, pp. 84-94, 2016, doi: 10.1111/nyas.13215.

[48] M. Itik, M. U. Salamci, and S. P. Banks, "Optimal control of drug therapy in cancer treatment," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 71, no. 12, pp. 1473-1486, 2009, doi: 10.1016/j.na.2009.01.214.

[49] S. Bunimovich-Mendrazitsky, E. Shochat, and L. Stone, "Mathematical model of BCG immunotherapy in superficial bladder cancer," *Bulletin of Mathematical Biology,* vol. 69, no. 6, pp.1847-1870, 2007, doi: 10.1007/s11538-007-9195-z.

[50] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems,* vol. 9, pp. 155-161, 1997, doi: 10.1.1.644.3302.

[51] WHA70.12: Cancer prevention and control in the context of an integrated approach, Accessed: Nov. 18, 2020. [Online]. Available: http://apps.who.int/gb/ebwha/pdf_files/WHA70-REC1/A70_2017_REC1-en.pdf
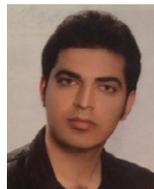
F. Heydarpour

F. Heydarpour received his BSc degree in Mathematics education from Shahid Rajaee University of Kerman'shah, Iran in 2010 and his MSc degree in Numerical Analysis from Shahid Rajaee Teacher Training University of Tehran, Iran in 2012. In 2019, he received his Ph.D. in Applied Mathematics from Yazd University, Iran. He has accepted several international papers in high rank journals. His research interests include numerical optimization, operations research, neural networks, numerical analysis, and optimal control.

E. Abbasi

E. Abbasi received the M.Sc. in computer science from Department of Computer Science of Amirkabir University of Technology, Tehran, Iran, in 2011. In 2016, she earned her PhD from Amirkabir University of Technology. Now, she is an assistant professor with the Department of Computer Science of Yazd University. Her major is data mining, soft computing, intelligent systems and bioinformatics.

M. J. Ebadi

M. J. Ebadi received his BSc degree in Applied Mathematics from Shahid Bahonar University of Kerman, Iran in 2003 and his MSc degree in Applied Mathematics from S&B University, Iran in 2006. In the same year, he joined the Department of Mathematics at Chabahar Maritime University, Iran, as a faculty member. In 2018, he received his Ph.D. in Applied Mathematics from Yazd University, Iran. He is a member of editorial board of two international reputed journals and a reviewer of more than 20 international reputed journals. He has published several international papers in high rank journals. Currently, he is assistant professor in Applied Mathematics at the Department of Mathematics, Chabahar Maritime University, Iran. His research interests include numerical optimization, deep learning, neural networks, numerical analysis, optimal control, fuzzy optimization, EEG signals classification, and image processing.

S. M. Karbassi

Professor S. M. Karbassi obtained his BSc degree in Electrical Engineering (1969) and MSc degree in Engineering Mathematics (1970) from University of Newcastle, England. He began his academic career at Isfahan University as an instructor at the Physics Department. Later, he obtained his PhD degree in Control Engineering from University of Salford, UK (1978). He is now a retired Professor of Applied Mathematics and Control of Yazd University. He has continued his research since his retirement (2007) by supervising research of postgraduate students mostly in different areas of Engineering Mathematics and Control.

An Elitist Non-Dominated Multi-Objective Genetic Algorithm Based Temperature Aware Circuit Synthesis

Apangshu Das*, Sambhu Nath Pradhan

Department of Electronics & Communication Engineering, National Institute of Technology Agartala, Tripura 799046 (India)

Abstract

At sub-nanometre technology, temperature is one of the important design parameters to be taken care of during the target implementation for the circuit for its long term and reliable operation. High device package density leads to high power density that generates high temperatures. The temperature of a chip is directly proportional to the power density of the chip. So, the power density of a chip can be minimized to reduce the possibility of the high temperature generation. Temperature minimization approaches are generally addressed at the physical design level but it incurs high cooling cost. To reduce the cooling cost, the temperature minimization approaches can be addressed at the logic level. In this work, the Non-Dominated Sorting Genetic Algorithm-II (NSGA-II) based multi-objective heuristic approach is proposed to select the efficient input variable polarity of Mixed Polarity Reed-Muller (MPRM) expansion for simultaneous optimization of area, power, and temperature. A Pareto optimal solution set is obtained from the vast solution set of 3n ('n' is the number of input variables) different polarities of MPRM. Tabular technique is used for input polarity conversion from Sum-of-Product (SOP) form to MPRM form. Finally, using CADENCE and HotSpot tool absolute temperature, silicon area and power consumption of the synthesized circuits are calculated and are reported. The proposed algorithm saves around 76.20% silicon area, 29.09% power dissipation and reduces 17.06% peak temperature in comparison with the reported values in the literature.

Keywords

I. Introduction

PARAMETERS such as area, power, and delay of highly complex logic circuits used in the field of digital systems can be appropriately optimized by utilizing AND-XOR based circuits rather than AND-OR based Boolean function [1]-[2]. Besides, XOR-based circuits are well suited for testability [3]-[4] and easily mapped into the Field Programmable Gate Arrays (FPGAs). Minimization of AND-XOR nodes (area count) is possible by sharing the sub-functions/product terms within the multi-output functions. With the reduction of node count, the signal transitions among the sub-function get reduced. 'Espresso,' the two-level AND-OR minimizer was developed to eliminate redundant literals from canonical form using unite function decomposition [5]. Low power approaches are established by searching a suitable input variable polarity for maximum sharing of internal nodes to reduce the switching activity of a Boolean function [6]-[8]. There are several sub-classes of AND-XOR circuit synthesis which are of interest. The most general 2-level AND-XOR form is EXOR Sum-Of-Product (ESOP). Due to its non-canonical nature, ESOPs are very hard for optimization. It is observed that any Boolean function can be represented in modulo-2 AND-XOR based algebraic expressions. These expressions are elaborated with the help of Davio

functions [9], and are termed as Reed-Muller (RM) expansions. A Boolean function represented in RM form is unique and canonical in nature, consumes less area, reduces power dissipation, and is in readily testable form [4]-[6], [8]. Depending on application-specific advantages, RM circuits are represented in the Positive Polarity Reed-Muller (PPRM) expansion, Fixed Polarity Reed-Muller (FPRM) expansion and Mixed Polarity Reed-Muller (MPRM) expansion. A thorough search on literature review has shown that optimization of MPRM is superior over the FPRM expansions on circuit performance regarding the area, switching activity (dynamic power) and/or delay [6]-[8], [10]-[11]. A fast minimization algorithm (FMA) using the binary differential evolution (BDE) method to minimize the FPRM product term is proposed in [12]. A comparative study of proposed FMA with the genetic algorithm (GA) and simulated annealing genetic algorithm (SAGA) is also reported. An incompletely specified FPRM (ISFPRM) acquisition algorithm is proposed by He et al. in [13]. The authors proposed a chromosome conversion technique to convert zero polarity ISFPRM to the FPRM for power reduction. A hybrid simulated annealing (SA) and discrete particle swarm optimization (DPSO) based area optimization approach is proposed in [14]. Authors in [8], considered the NSGA-II algorithm to find an optimal polarity for power and area optimization of MPRM network. Authors in [8] proposed a chromosome encoding method based on ternary input polarity and binary don't care allocation. By exploiting the don't care condition, authors in [15] proposed a delay optimization approach for MPRM based logic. In [15], the authors minimized the weighted

* Corresponding author.

E-mail address: apangshuextc@gmail.com

path length using the Huffman tree construction algorithm. A shared mixed polarity RM (SMPRM) network is proposed using weighted search GA (WSGA) in [16] to find the optimal polarity based on area, power, and temperature. The trade-off analysis is also reported among the area, power, and temperature. But it is very difficult to find the optimum polarity in WSGA as one parameter may dominate the other. Other than [16], none of the articles have considered temperature as one of the cost metric for RM network synthesis. But consideration of temperature as a cost metric is very much essential because, ICs need to operate within a stipulated temperature zone prescribed by the manufacturers. The commercial devices and industrial devices are operating within the temperature zone of (0 °C to 70 °C) and (-40 °C to 85 °C) respectively, which are much lower than the aerospace and military devices operating zone (-55 °C to 125 °C) [17]-[18]. Due to aggressive device scaling and package density, most of the integrated circuits (ICs) are burnt just because of over-heating. Overheating is build-up due to excessive power density generation within the chip for the inclusion of a vast number of complex functionality within a small silicon area. So far most of the researchers paid attention to physical design domain for temperature minimization [19]-[20], but the cooling solutions are rising at $ 1-3 or more per watt of power dissipation [21]. The cooling cost of high-performance processor increases exponentially with the growth of power density. So, design-time thermal-aware techniques can be used to improve the power and thermal characteristics of integrated circuits. Few works report temperature minimization by reducing the power density at logic synthesis level [22]-[26]. The power density finds a direct relation with temperature generation within a chip by the following expression [27].

$$T_{chip} = T_{amb} + R_{th} \cdot \left[\frac{P_T}{A_T} \right]$$

(1)

In equation (1), T_{chip} and T_{amb} are the average chip temperature and ambient temperature respectively. Where, R_{th} is the summative equivalent thermal resistance of the substrate (Si) layer, package, and heat sink (m² °C/W). Total power dissipation is represented by P_T (in W). And A_T (in m²) referred as total silicon core area of the chip. This area does not include the package area, but it consists of all cell area and routing area of Silicon chip. Earlier researchers ignore the thermal issues in higher levels (logic synthesis and circuit design) of very-large-scale integration (VLSI) design synthesis. In [28], Shang and Dick reported that rise in chip temperature set back reliability, performance, cost and power consumption. It is reported in [28] that 30% cost of IC packaging is contributed by cooling arrangement. The temperature determining parameter inclusion in logic synthesis level may reduce the cooling cost.

An exact or exhaustive search method can be used for small-sized circuits, but this strategy is not feasible for middle or large-sized circuits. The problem of determination of exact input variable polarity for getting minimum cost is a non-deterministic polynomial-time hard (NP-hard) problem. No known algorithm can solve this problem in polynomial time. Non-exhaustive or heuristic search approaches have been introduced to solve such NP-hard problems. Detail of NP-hard problems can be found in [29]. The proposed work presents a fast converging heuristic technique called Non-Dominated Sorting based Genetic Algorithm-II (NSGA-II) for the thermal-aware problem. Compared to existing optimization approaches, the contributions of the proposed approach are as follows:

- The thermal-aware AND-XOR logic synthesis is done suitably using MPRM expansion methodology.
- NSGA-II is used to get the optimum solution in terms of area, power, and power density for MPRM circuits. Parameters of the NSGA-II algorithm are tuned suitably to get the optimum solution.

- The simulation result of the proposed approach is reported by calculating the absolute temperature, total power consumption, and silicon area. 'HotSpot' tool [30] is used to report the absolute temperature. Cadence 'Innovus' tool [31] is used to report the total power consumption (dynamic and leakage) and silicon area at 45nm technology.

In the proposed approach, we considered the ternary input variable polarity for chromosome encoding and then modified the NSGA-II approach at crossover and mutation level to find the better offspring. Parent chromosomes for crossover and mutation are chosen from the elite group or entire population based on threshold value. Two-point crossover methodologies are used to generate the offspring chromosomes. Random bit positions are chosen to increase the mutation diversity within the offspring. In the proposed work, power density is considered as a cost metric to reduce the thermal effect in MPRM network. Finally, Electronic Design Automation tools (Cadence and HotSpot) are used for actual area, power and temperature calculation.

The rest of the paper is organized as follows: Section II demonstrates the motivation and basic terminologies used in RM expansion. Section III presents the Thermal-aware mixed polarity problem formulation using tabular technique approach. NSGA-II based thermal-aware realization is described in section IV. Section V details the results, and finally, section VI draws the conclusion.

II. Reed-Muller Preliminaries and Motivation

A. Reed-Muller Expansion

Any n-input m-output Boolean function can be represented canonically as AND-OR based Sum-Of-Product (SOP) form. The SOPs are expanded with 2^n different product terms as shown below:

$$f\left(x_n, \ldots, x_1\right) = \sum_{i=1}^{2^n} p_i . m_i$$

(2)

Where 'm_i' represents the minterms and $p_i \in \{0, 1\}$ represents the absence or presence of minterms. Suffix 'i' accounts for the number of terms which varies from 1 to 2^n. If all the input variables are used to represent the minterms of an expression, then it is said to be Canonical Sum-Of-Product (CSOP). In CSOP logic function all OR gates can be replaced with XOR gates and provides ExOR Sum-Of-Product (ESOP) function. The ESOP form can be represented as:

$$f\left(x_n, \ldots, x_1\right) = \oplus \sum_{i=1}^{2^n} p_i . m_i$$

(3)

Here, $\oplus$ represent the ExOR operation. The expanded ESOP form can be written as:

$$f\left(x_n, \ldots, x_1\right) = p_1 \overline{x_n} \ldots \overline{x_2} \, \overline{x_1} \oplus p_2 \overline{x_n} \ldots \overline{x_2} x_1 \oplus \ldots \oplus p_{2^n} x_n \ldots x_2 x_1$$

(4)

Eq. (4) is also being represented as Reed-Muller (RM) expansions based on each variable appearance. Variables can be appeared as true form (x_i) or complemented form ($\overline{x_i}$) or mixed form (x_i and $\overline{x_i}$).

B. Fixed Polarity Reed-Muller Expansions

When each variable appears in true or complemented form but not both at the same time as shown in eq.(4) is known as Fixed Polarity Reed-Muller (FPRM) expansion. FPRM expansion provides 2^n different polarities or expansions for a given problem. Example 1 demonstrates the formation of an FPRM expansion.

Example 1: Consider a Boolean expression with the function given by:

$$f_1(x_3, x_2, x_1) = \sum m(3,5,6) = x_1 x_2 \overline{x_3} + x_1 \overline{x_2} x_3 + \overline{x_1} x_2 x_3 \quad (5)$$

FPRM expansion polarities are defined with binary numbers as

$$< p_j > = < p_n, p_{n-1}, ... p_1 >$$

$$p_j = \begin{cases} 0 & \text{if} \quad x_k \quad \text{appears in complemented form} \\ 1 & \text{if} \quad x_k \quad \text{appears in} \quad \text{true} \quad \text{form} \end{cases} \quad (6)$$

If polarity $(101)_2$ is assigned to a function $f_1(x_3, x_2, x_1)$, then the variables x_1 and x_3 are expressed in true form, and variable x_2 is in complemented form by utilizing $x_i = (1 \oplus \overline{x_i})$ and $\overline{x_i} = (1 \oplus x_i)$ respectively.

For the given polarity the FPRM expansion for function f_1 is given by:

$$f_1(x_3, x_2, x_1) = x_1 x_2 \overline{x_3} + x_1 \overline{x_2} x_3 + \overline{x_1} x_2 x_3$$
$$= \left[x_1(\overline{x_2} \oplus 1)(x_3 \oplus 1) \right] \oplus \left[x_1 \overline{x_2} x_3 \right] + \left[(x_1 \oplus 1)(\overline{x_2} \oplus 1) x_3 \right]$$
$$= x_1 \oplus x_3 \oplus x_1 \overline{x_2} \oplus \overline{x_2} x_3 \oplus x_1 \overline{x_2} x_3 \quad (7)$$

C. Mixed Polarity Reed-Muller Expansions

If each variable in eq. (4) is represented by true or complemented form at the same time, then this form of representation is known as Mixed Polarity Reed-Muller (MPRM) Expansion. 3^n different polarities or expansions are possible in MPRM expansion. The 3^n polarities of MPRM expansion include 2^n polarities of FPRM expansion. Hence, probability of getting a better solution in MPRM is more than that of FPRM. Example 2 illustrates the formation of MPRM expansion.

Example 2: Example considered for FPRM expansion (in example 1) is taken to illustrate the MPRM expansion.

$$f_1(x_3, x_2, x_1) = \sum m(3,5,6) = x_1 x_2 \overline{x_3} + x_1 \overline{x_2} x_3 + \overline{x_1} x_2 x_3 \quad (8)$$

Ternary variable is used to represent the polarities of MPRM expansion.

$$< p_j > = < p_n, p_{n-1}, ... p_1 >$$

$$p_j = \begin{cases} 0 & \text{if} \quad x_k \quad \text{appears in complemented form} \\ 1 & \text{if} \quad x_k \quad \text{appears in} \quad \text{true} \quad \text{form} \\ 2 & \text{if} \quad x_k \quad \text{appears in} \quad \text{mixed} \quad \text{form} \end{cases} \quad (9)$$

If function $f_1(x_3, x_2, x_1)$ is encoded as $(201)_3$, where x_1 is expressed in true polarity, x_2 is in complementary form, and x_3 is represented in mixed form, the MPRM expansion for function f_1 by given polarity is expressed as:

$$f_1(x_3, x_2, x_1) = x_1 x_2 \overline{x_3} \oplus x_1 \overline{x_2} x_3 \oplus \overline{x_1} x_2 x_3$$
$$= [x_1(\overline{x_2} \oplus 1)\overline{x_3}] \oplus [x_1 \overline{x_2} x_3] \oplus [(x_1 \oplus 1)(\overline{x_2} \oplus 1) x_3]$$
$$= x_1 \overline{x_2} \overline{x_3} \oplus x_1 \overline{x_3} \oplus x_1 \overline{x_2} x_3 \oplus x_1 \overline{x_2} x_3 \oplus x_1 x_3 \oplus \overline{x_2} x_3 \oplus x_3$$
$$= x_1 \overline{x_2} \overline{x_3} \oplus x_1(x_3 \oplus 1) \oplus x_1 x_3 \oplus \overline{x_2} x_3 \oplus x_3$$
$$= x_1 \overline{x_2} \overline{x_3} \oplus x_1 \oplus \overline{x_2} x_3 \oplus x_3 \quad (10)$$

It is inferred from eq. (7) and (10) that judicious choice of input variable polarity in MPRM expansion can provide a better solution than the FPRM expansion. Nine (9) literals are required to represent the function given in example 1 using FPRM. Whereas, only seven (7) literals are sufficient to represent the same function using MPRM expansion. It is expected that the number of switching activity is also get reduced with the literal minimization. The next section describes the tabular technique implementation for MPRM thermal-aware problem realization.

III. Proposed Thermal Aware Mixed Polarity Reed-Muller Approach Using Tabular Technique

A. Area Computation

The thoughtful conversion of Boolean function into MPRM for maximum sharing of product terms considering the optimization parameters by efficient input variable encoding is carried out in this work. A multi-input multi-output Boolean function in the form of pla file is considered as input for the proposed synthesis process. The following steps illustrate the tabular technique implementation for MPRM thermal-aware problem realization. A brief description for polarity conversion procedure is given below.

All the terms present in the Boolean function are listed in Binary form. Don't care conditions are realized in true as well as complementary form to generate canonical representation. Input variables are encoded in mixed polarity, as shown by eq. 9. Then, the input functions are decomposed based on encoding.

Inter polarity conversion takes place according to the chromosome encoding. There can be any of the following three cases –

- 2 to 2 conversion: When a variable is initially in mixed form, '2' and after conversion also the polarity of that variable is '2', then the variable is in mixed form. For such case, the bits of the corresponding variable remain unchanged.

- 2 to 1 conversion: When the variable is initially in mixed form, '2' and the final polarity of the variable is '1', i.e., the variable exists in true form in the final expression, for that all the '0's of the variable are to be replaced by '1' and thus a new term with don't care is generated in the table.

- 2 to 0 conversion: When the variable is in mixed form, '2' initially, and the final polarity of the variable is '0', i.e., the variable exists in complementary form, for that all the '1's of the variable are to be replaced by '0', and thus a new term with don't care is generated in the table.

After generating all the possible new terms for a single variable, they are to be compared with the existing terms to cancel out the similar terms and the table is updated. If two input cubes having same output, but the input is varied by only one literal then that literal is replaced by don't care symbol (–). In this way, steps are repeated for all the input variables in the function to get the reduced MPRM expression.

.i	4				
.o	2				
.p	8				
x	y	z	w	f_1	f_2
1	0	0	0	0	1
1	0	0	1	0	1
0	0	1	0	1	0
0	0	1	1	1	0
0	1	0	0	1	0
1	1	1	0	1	1
1	1	0	1	1	1
0	1	1	1	1	0
.e			(a)		

x	y	z	w
2	2	1	0

(b)

Fig. 1. (a) '.pla' file representation of Boolean function ('i', 'o' and 'p' represent number of inputs, number of outputs and product terms respectively); (b) Input variable encoding ('x' and 'y' are in mixed form; 'w' is in complementary form and 'z' is in true form).

An arbitrary Boolean function is considered as an example case and it is shown in Fig. 1(a). The chromosome encoding for the example case is shown in Fig. 1(b). The translation of input Boolean function and area computation using tabular technique is shown below.

The two output functions are:

$$f_1 = \overline{x}\,\overline{y}z\overline{w} + \overline{x}\,\overline{y}zw + \overline{x}\,y\overline{z}\,\overline{w} + xyz\overline{w} + xy\overline{z}\,w + \overline{x}yzw \tag{11}$$

$$f_2 = x\overline{y}\,\overline{z}\,\overline{w} + x\overline{y}\,\overline{z}\,w + xyz\overline{w} + xy\overline{z}\,w \tag{12}$$

Generally, Boolean functions are expressed in terms of AND- OR function. As f_1 and f_2 are represented in disjoint cube form, so it can be represented as:

$$f_1 = \overline{x}\,\overline{y}z\overline{w} \oplus \overline{x}\,\overline{y}zw \oplus \overline{x}\,y\overline{z}\,\overline{w} \oplus xyz\overline{w} \oplus xy\overline{z}\,w \oplus \overline{x}yzw \tag{13}$$

$$\text{and, } f_2 = x\overline{y}\,\overline{z}\,\overline{w} \oplus x\overline{y}\,\overline{z}\,w \oplus xyz\overline{w} \oplus xy\overline{z}\,w \tag{14}$$

Table I shows the input polarity conversion based on encoding, as shown in Fig. 1(b). Variable x and y are represented in mixed polarity form so, no new term is generated. But new term will be generated for z and w, where the variables are expressed in true and complementary form respectively. The redundant terms noted with *a*, *b*, *c*, *d* and *e* get eliminated, and the terms noted with *f* forms a new term by replacing one literal with don't care.

After polarity conversion, the final MPRM output for function f_1 and f_2 are represented as:

$$f_1 = \overline{x}y\overline{w} \oplus xyz \oplus xy\overline{w} \oplus xy \oplus \overline{x}z \tag{15}$$

$$f_2 = x\overline{y}z \oplus x\overline{y} \oplus xyz \oplus xy\overline{w} \oplus xy \oplus \overline{x}z \tag{16}$$

Shared terms are: $xyz, xy\overline{w}, xy$

It is observed that primary function requires 8 product terms with 32 literals whereas, final function requires 7 product terms with 18 literals (where 3 product terms are shared among function f_1 and f_2).

B. Power Estimation Using Switching Activity

In CMOS circuits, the dynamic dissipation is the main contributor to power consumption, which is caused by charging and discharging the load capacitances. It can be modeled as:

$$P_{dyn} \approx P_{swt} = \alpha_L C_L V_{DD}^2 f + \sum_i \alpha_i C_i V_{DD} (V_{DD} - V_T) f \tag{17}$$

Where, P_{dyn} and P_{swt} represent the dynamic and switching power respectively. α_L and α_i are the switching activity at the load and internal node respectively. The capacitance at the load and internal gates are represented by C_L and C_i respectively. Supply voltage, threshold voltage, and frequency of operation are given by V_{DD}, V_T, and f respectively.

Eq. (17) illustrates that, except for those of switching activity, all other parameters are user/manufacturer defined at a particular technology. Switching activity is the only parameter that needs to be estimated for technology-independent power optimization. Expected number of signal transitions at the outputs of the gates of a combinational logic circuit is defined as switching activity. This work follows the same procedure used in the reference [32] to estimate switching activity. Let us consider that initial inputs are uncorrelated and statically independent of each other, represented as:

$$\text{Prob(input} = 1) = \text{Prob(input} = 0) = 0.5 \tag{18}$$

The probability of the output of a gate when its inputs are changed from the previous state is estimated by:

$$\text{Prob(present} = 0) \cdot \text{Prob(past} = 1)$$
$$+ \text{Prob(present} = 1) \cdot \text{Prob(past} = 0) \tag{19}$$

The switching probability follows the stationary random process, and probabilistic description does not change over a given period. Then, switching activity of logic gate (α_g) is given by:

$$\alpha_g = 2 \cdot \text{Prob}_{op=0} \cdot \text{Prob}_{op=1} \tag{20}$$

TABLE I. Input Variable Polarity Transformation Using Tabular Technique

Input		Term generated by z		Term generated by w		Terms get canceled/modified			Remaining terms	
x y z w	$f_1 f_2$	x y z w	$f_1 f_2$	x y z w	$f_1 f_2$	x y	z w	$f_1 f_2$	x y z w	$f_1 f_2$
1 0 0 0	0 1	1 0 1 0	0 1	1 0 1 0	0 1	1̶0̶	0̶1̶	a		
		1 0 – 0	0 1	1 0 – 0	0 1	1̶0̶	0̶1̶	b		
1 0 0 1	0 1	1 0 1 1	0 1	1 0 1 0	0 1	1̶0̶	0̶1̶	a	1 0 1 –	0 1
				1 0 1 –	0 1					
		1 0 – 1	0 1	1 0 – 0	0 1	1̶0̶	0̶1̶	b	1 0 – –	0 1
				1 0 – –	0 1					
0 0 1 0	1 0	0 0 1 0	1 0	0 0 1 0	1 0	0̶0̶	1̶0̶	c		
0 0 1 1	1 0	0 0 1 1	1 0	0 0 1 0	1 0	0̶0̶	1̶0̶	c		
				0 0 1 –	1 0	0 0	1 0	f		
0 1 0 0	1 0	0 1 1 0	1 0	0 1 1 0	1 0	0̶1̶	1̶0̶	d	0 1 – 0	1 0
		0 1 – 0	1 0	0 1 – 0	1 0					
1 1 1 0	1 1	1 1 1 0	1 1	1 1 1 0	1 1	1̶1̶	1̶1̶	e		
1 1 0 1	1 1	1 1 1 1	1 1	1 1 1 0	1 1	1̶1̶	1̶1̶	e	1 1 1 –	1 1
				1 1 1 –	1 1					
		1 1 – 1	1 1	1 1 – 0	1 1				1 1 – 0	1 1
				1 1 – –	1 1				1 1 – –	1 1
0 1 1 1	1 0	0 1 1 1	1 0	0 1 1 –	1 0	0 1	1 0	f	0 – 1 –	1 0
				0 1 1 0	1 0	0̶1̶	1̶0̶	d		

The generalized expression for switching activity for an '*i*' input AND gate (α_{AND}) with input switching probability '0.5' is given by:

$$\alpha_{AND} = 2 \cdot (0.5)^i \cdot [1 - (0.5)^i] \tag{21}$$

Second level ON-probability of XOR gates may be computed by '$P \cdot 0.5^i$'. Where 'i' is the inputs realization of a function with '*P*' ON-terms. The probable switching activity of the node is given by:

$$\alpha_{XOR} = 2 \cdot \left[P \cdot 0.5^i \right] \cdot \left[1 - (P \cdot 0.5^i) \right] \tag{22}$$

The power consumption of a MPRM circuit is the sum of power of AND gates and XOR gates. Assuming that 'n' is the set of nodes in MPRM circuits, then the total switching activity is given by:

$$\alpha_{total} = \sum_n \alpha_{AND} + \sum_n \alpha_{XOR} \tag{23}$$

C. Power Density

The amount of power drawn per unit area defines the power density of a circuit. It can be calculated by taking the ratio of total switching activity and area of the circuit.

$$Pd_{MPRM} = \alpha_{total} / A_{MPRM} \tag{24}$$

Where, Pd_{MPRM}, α_{total} and A_{MPRM} represent the power density, overall switching activity and total area of an MPRM realized network. The power density is estimated for a particular offspring chromosome to determine the thermal effect. Lower the power density better is the distribution of temperature among the different modules within a chip. This has also been verified by finding the absolute temperature (in °C) using Cadence and HotSpot tool.

IV. Non-Dominated Sorting Based Genetic Algorithm-II For Proposed Thermal-aware Realization

Classical search techniques like genetic algorithm (GA) disperse the optimum solution throughout the search space and can find one optimal solution for a given weight combination in a single run when multiple objectives are there. All possible weight combinations are mandatory to go through to obtain the optimum solution. For which the execution time consumes much delay to find the optimum solution. An elitist non-dominated sorting based multi-criteria decision-making algorithm called non-dominated genetic algorithm-II (NSGA-II) is employed to overcome the above inconsistency. NSGA-II is a fast and improved multi-objective evolutionary algorithm (MOEA) with computational complexity $O(XY^2)$, where 'X' is the number of objective parameters, and 'Y' is the population size. Fitness estimation or sharing parameters are replaced with the rank assignment and front selection using non-dominated sorting and crowding distance calculation in NSGA-II for better elitism and fast convergence toward an optimum solution. The detailed procedure of NSGA-II is discussed in [33]. Configurable parameters, optimization objectives and constraints used for proposed algorithm are discussed elaborately in this section.

A. Chromosome Structure

Efficient chromosome structure can be encoded for an 'm' input combinational logic circuit by ternary bit string of length 'm'. The 'm' input variables (l_1, l_2, l_3, ..., l_m) represents complementary, true and mixed polarity based on ternary operator bits {0, 1, 2}. If the pth bit is '1', it denotes that the pth input variable is implemented in true polarity whereas, if the qth and r^{th} bits are '0' and '2' respectively, it symbolizes that the qth and r^{th} inputs are realized in complementary and mixed polarity respectively.

B. Front Selection and Rank Assignment Based on Non-domination

Chromosomes in each front are assigned fitness based on their rank values or the front in which they exist. Chromosomes in the first front are designated with the highest rank value as 'one'(1) and individuals in the second are assigned the rank value as two (2) and so on.

1. **Crowding Distance Calculation**:Crowding distance (i_{dist}) is another fitness parameter which depicts the density of a solution in a population. i_{dist} can be calculated for each objective function by evaluating the Euclidean distance between individual chromosomes in a front by considering '*n*' objective functions in the '*n*' dimensional hyperspace.

2. **Parent selection**:A chromosome is selected as a parent if its rank is lesser than the other. If the ranks of chromosomes are same then the individual having higher crowding distance is selected. The selected parent chromosomes generate next-generation chromosomes using crossover and mutation operators.

C. Genetic Operators

Crossover and mutation are the two inherent mechanisms of the NSGA-II algorithm. They introduce the variation within the generated offspring and converge the output solution towards the optimal solution. It is observed from the literature that better offspring is generated by considering 90% crossover and 10% mutation in NSGA-II based multi-objective evolutionary algorithms [8]. For the proposed approach, the same method is followed. However, three other experiments (70% crossover and 30% mutation, 80% crossover and 20% mutation, 100% crossover) were carried out by varying crossover and mutation percentage but it has been observed that more diversity in population is there if 90% crossover and 10% mutation is considered and good result is obtained.

1. **Crossover**: During crossover operation, two-parent chromosomes '*x*' and '*y*' from the initial population mates to produce two new offspring 'co$_1$' and 'co$_2$' at randomly selected crossover points. Two-point crossover methods converge the solution faster towards the optimum solution than that of single-point crossover. Parent chromosomes selection is biased towards the chromosomes with better fitness value ('elite group'). Chromosomes with rank one (1) are considered as elite group. The selection of parent chromosomes from elite groups or from entire population to participate in crossover operation depends on the generation of a uniform random number between '0' and '1'. If the number is greater than or equal to '0.5' then the parents for crossover is chosen from the elite group; otherwise, parents are selected from the entire population. The threshold for elite group is considered as '0.5' for selecting best-fit chromosome to participate in the crossover operation to generate better offspring. Let, the size of population is 'p' and the cardinality of elite group is 'q'. Then the probability of selecting a chromosome from elite group is 0.5/q + 0.5/p.Whereas, probability of chromosome selecting from entire population is 0.5/p. The probability of selecting chromosome from elite group is more than that of entire population, because 'q' is much smaller than 'p'. This method selects best-fit chromosomes to participate in the crossover operation and generates better offspring as compared to truly random one [34]. Two crossover positions (cp$_1$ and cp$_2$) are randomly selected within the chromosome string length, and the alleles are exchanged between the two selected individuals as shown in Fig. 2(a) and 2(b). Fig. 2 (a) and (b) show the different outcomes of the same parent chromosomes using crossover operation method 1 and method 2 respectively. A check is made after each generation with the already generated chromosomes, and duplicate chromosomes are eliminated.
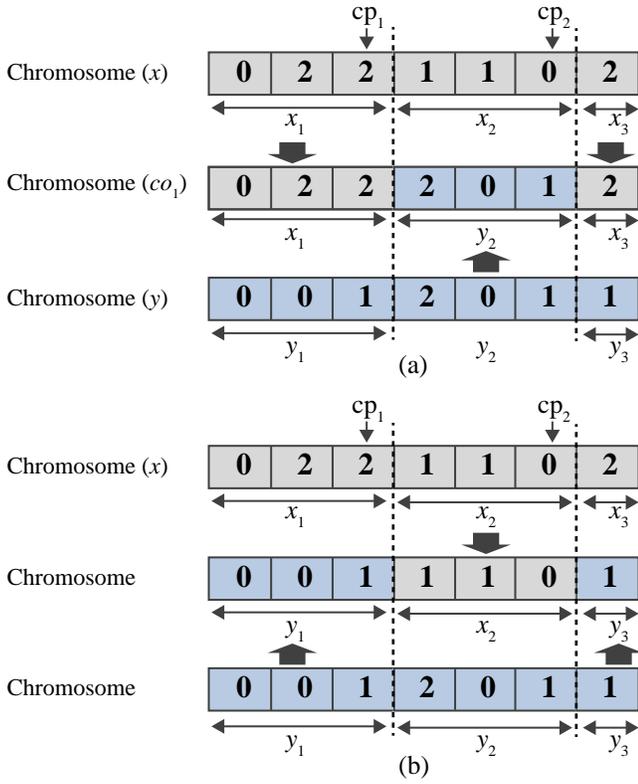
Fig. 2. (a) Crossover operation method 1; (b) Crossover operation method 2.

Fig. 3 (a) Mutation operation; (b) Roulette wheel criteria for bit manipulation in mutation operation.

2. **Mutation**: Mutation enables the genetic diversity from generation to generation. Mutation prohibits falling off all solutions in the population into a local optimum. 10% of the 'N' offspring population is contributed by mutant chromosomes using mutation. Mutation operation is performed by selecting few random bit positions called mutation points (mp) and the polarity of that selected position is altered by the roulette wheel selection methodology as shown in Fig. 3(a). To increase the randomness, the mutation points are chosen randomly within a range of 1 to 'n' (where 'n' is the length of the chromosome). For an example case, a chromosome (m) is participating in mutation operation from the present generation; randomly three positions are selected as a mutation point (mp_1, mp_2, and mp_3). Inter-conversion of polarity that is, the positive, negative and mixed polarity is done using roulette wheel criterion and remaining bits get unaltered. The newly generated offspring becomes the chromosome of the next generation.

Fig. 3(b) illustrates the operation of the roulette wheel criterion. A random number (R_n) between '0' and '1' is generated for each mutation point, and if the generated random number (R_n) is greater than or equal to '0.5', the wheel position moves clockwise otherwise anti-clockwise. Depending on the elevated position, the polarity of the mutation point will change. Choosing a random number based on some prior information (like range, mean, variance, etc.) is a convex optimization problem (which is determined by entropy of objective function). With finite range, maximum entropy is given by uniform probability distribution function. Other distributions will have less entropy than the uniform probability distribution in the same range.

The proposed NSGA-II algorithm contributes 'N' chromosomes using the selected parents by crossover and mutation methods. Generated 'N' offspring and 'N' parents contribute as '2N' numbers of next-generation population.
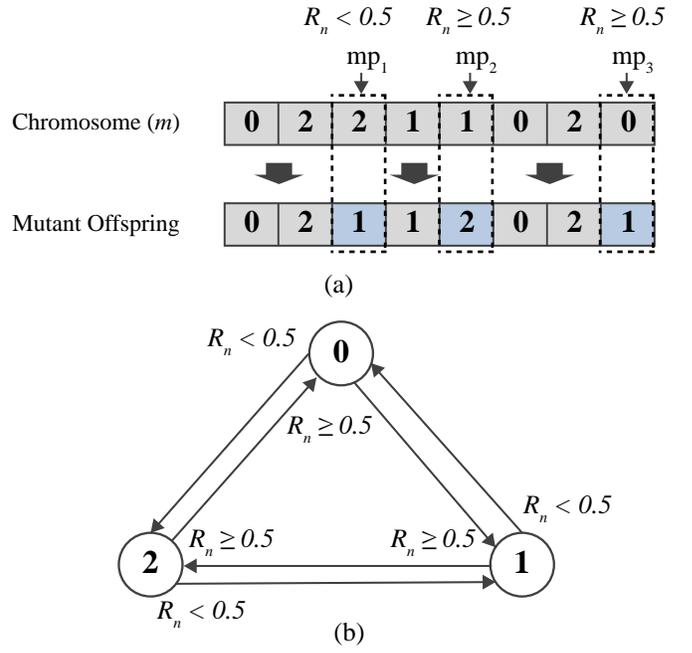
V. Results and Related Discussions

Proposed thermal-aware mixed polarity AND-XOR realization of logic circuits have been implemented using NSGA-II in LINUX based C++ platform on a Pentium IV machine with 3-GHz clock frequency and 4-GB RAM memory. The algorithm is applied to MCNC and LGSynth93 benchmark suite [35] for experimentation. In the proposed work, NSGA-II based optimization approach is proposed for simultaneous reduction of area, power, and temperature. We have targeted logic level for optimization to reduce the cooling cost through heat-sink. At logic level, absolute values of area, power, and temperature are unknown. Therefore, for area reduction, reduction of product term is considered. To optimized power, switching activity is reduced. And for temperature reduction, power density is reduced in the cost metric. NSGA-II provides the Pareto optimal solution set consisting of area best, power best, power density best and optimum solution considering all the three parameters. To obtain the actual silicon area (in μm^2), power dissipation (in nW) and absolute temperature (in ℃) Cadence (Genus and Innovus) and HotSpot tools are used. Cadence and HotSpot Electronics Design Automation (EDA) software packages are involved for simulating the digital and analog circuits. We have generated the graphical design specification information interchange (GDS-II) report after layout design for best and optimum solutions obtained using the proposed algorithm for each benchmark circuit, but there is no hardware implementation (chip fabrication) of the circuit. The total discussion of result is divided into two sections. The first section of result concerns the area, power and power density based result using NSGA-II approach. The next section briefly describes the implementation of physical design at 45 nm technology using Cadence Genus and Innovus Implementation tool. Then absolute temperature estimation using HotSpot tool is presented.

A. Result Based on NSGA-II

The '.pla' based circuits of MCNC and LGSynth93 benchmark suit are considered as an input circuits which are to be optimized in terms of area, power and temperature optimization. The circuits are decomposed into MPRM expansion based on input variable polarity

encoding as explained in chromosome structure. NSGA-II is used to find efficient chromosome polarity based on area, power, and power density. Twenty (20) benchmark circuits are tested for experimentation. Table II gives the parameters and evolution operator's settings for the proposed NSGA-II based approach.

TABLE II. Parameters and Evolution Operator's Settings for the Proposed NSGA-II Approach

Parameter	Value
No. of initial population	100.00
Total no. of iteration	200.00
Crossover probability	0.9
Mutation probability	0.1
Crossover operation	Two-point crossover method
Mutation operation	Bit mutation based on Roulette wheel criteria
Termination criterion	Max. no. of iteration

To verify the efficiency of the proposed approach using NSGA-II, the proposed best and optimized results of MPRM circuits are compared with previously published best and optimum results of FPRM [36], Shared Reed-Muller Decision Diagram (SRMDD) [22], MPRM [10], AND-Inverter Graphs (AIGs) [37] and GA based FPRM [38] decomposed circuits. A set of solutions (called Pareto optimal solutions) are obtained comprising of the individual best solutions ('Area Best', 'Power Best' and 'Power density Best') and 'optimum solution'. An area comparative study of the proposed approach with FPRM, SRMDD, MPRM, AIGs, and GA based FPRM is presented in Table III. For power comparison, the proposed method is compared with FPRM, AIGs and GA based FPRM solutions, which are reported in Table IV. For power density based comparison, the proposed power density solutions are compared with SRMDD and AIGs based solutions and reported in Table V. In Tables III, IV and V, the first column shows the circuit name with which experimentation is carried out. The second

and third columns of Tables III, IV and V represent the proposed best and optimum solution for area, power and power density, respectively. The "Save Best" and "Save Opt" columns in Tables III, IV and V shows the percentage savings of the proposed approach with respect to the existing works reported in the literature. The average percentage saving is calculated and reported in the last row of Tables III, IV and V. The percentage savings referred to in the column "Save Best" and "Save Opt" of Tables III, IV and V are calculated by the following Eq. (25) and (26).

$$\text{Save Best} = \left(\frac{\text{Existing}_{\text{solution}} - \text{Best solution}}{\text{Existing}_{\text{solution}}} X100 \right)\%$$
(25)

$$\text{Save Opt} = \left(\frac{\text{Existing}_{\text{solution}} \quad \text{Optimum solution}}{\text{Existing}_{\text{solution}}} X100 \right)\%$$
(26)

The percentage savings for the proposed best solution and proposed optimum solution are represented by "Save Best" and "Save Opt" as referred in Eq. (25) and (26), respectively. The "Best solution", "Optimum solution" and "Existing$_{\text{solution}}$" represent the proposed NSGA-II based best solution, NSGA-II based optimum solution and existing reported works of literature, respectively.

From Table III, it is observed that the chromosome with the proposed area best solution of NSGA-II save 28.61%, 22.85%, 29.45%, 30.45% and 35.19% area compared to that of FPRM, SRMDD, MPRM, AIGs, and GA based FPRM based results respectively. When the proposed optimum solution comparative study is performed with respect to FPRM, SRMDD, MPRM, AIGs, and GA based FPRM results then proposed optimum solution shows an area saving of 18.42%, 10.92%, 15.94%, 6.86% and 23.51% compared to that of FPRM, SRMDD, MPRM, AIGs, and GA based FPRM based results, respectively.

When power-based realization is compared in Table IV, it is observed that FPRM, AIGs, and GA based FPRM solutions consume

TABLE III. Area Comparative Study Of the Proposed MPRM Realization

Circuits	Proposed approach		FPRM [36]			SRMDD [22]			MPRM [10]			AIGs [37]			GA based FPRM [38]		
	Area Best	Optimum solution	Area	Save Best	Save Opt	Area	Save Best	Save Opt	Area	Save Best	Save Opt	Area	Save Best	Save Opt	Area	Save Best	Save Opt
5xp1	49	49	61	19.67	19.67	61	19.67	19.67	61	19.67	19.67	66	25.76	25.76	61	19.67	19.67
9sym	87	116	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
alu2	257	314	-	-	-	225	-14.22	-39.56	-	-	-	333	22.82	5.71	225	-14.22	-39.56
alu4	993	1485	-	-	-	-	-		2438	59.27	39.09	719	-38.11	-106.54	3683	73.04	59.68
clip	118	118	-	-	-	206	42.72	42.72	182	35.16	35.16	-	-	-	206	42.72	42.72
cm162a	25	25	25	0.00	0.00	25	0.00	0.00	-	-	-	38	34.21	34.21	25	0.00	0.00
cm163a	18	18	18	0.00	0.00	18	0.00	0.00	-	-	-	-	-	-	18	0.00	0.00
con1	9	10	-	-	-	-	-	-	14	35.71	28.57	-	-	-	-	-	-
cu	21	23	37	43.24	37.84	-	-	-	-	-	-	-	-	-	-	-	-
inc	34	37	106	67.92	65.09	48	29.17	22.92	34	0.00	-8.82	87	60.92	57.47	-	-	-
misex1	16	19	20	20.00	5.00	32	50.00	40.63	13	-23.08	-46.15	-	-	-	-	-	-
misex2	55	57	87	36.78	34.48	-	-	-	-	-	-	84	34.52	32.14	-	-	-
misex3c	296	757	-	-	-	-	-	-	1421	79.17	46.73	533	44.47	-42.03	1831	83.83	58.66
pm1	19	21	27	29.63	22.22	-	-	-	-	-	-	30	36.67	30.00	27	29.63	22.22
rd53	15	24	20	25.00	-20.00	20	25.00	-20.00	20	25.00	-20.00	-	-	-	20	25.00	-20.00
rd73	43	43	-	-	-	64	32.81	32.81	63	31.75	31.75	-	-	-	63	31.75	31.75
rd84	73	73	107	31.78	31.78	-	-	-	107	31.78	31.78	-	-	-	107	31.78	31.78
sao2	74	85	100	26.00	15.00	-	-	-	76	2.63	-11.84	-	-	-	-	-	-
table3	175	219	-	-	-	-	-	-	401	56.36	45.39	-	-	-	1945	91.00	88.74
x2	17	27	30	43.33	10.00	30	43.33	10.00	-	-	-	36	52.78	25.00	30	43.33	10.00
Av. % Savings w.r.t Proposed Solution				28.61	18.42		22.85	10.92		29.45	15.94		30.45	6.86		35.19	23.51

28.71%, 85.35% and 50.27% more power with respect to the proposed best power-based solutions respectively. When proposed optimum power-based solutions are compared, then the FPRM, AIGs, and GA based FPRM solutions consume 11.72%, 81.19%, and 41.96% more power than that of the proposed solutions, respectively. The best power density based solutions show 23.68% and 69.83% reduction in power density compared to that of SRMDD and AIGs based solutions in Table V. The optimum power density based solutions provide 10.90% and 70.02% better results than that of SRMDD and AIGs based solutions respectively. Fig. 4 shows the Pareto-optimal graph for "*table3*" benchmark circuit.

The solutions nearer to the origin form the Pareto-optimal front, which are optimal with respect to area, power, and power density. The solutions nearer to each axis represent the best solutions. The last column of Table V reports the total CPU time required (in CPU seconds) to execute the algorithm in an identical platform.
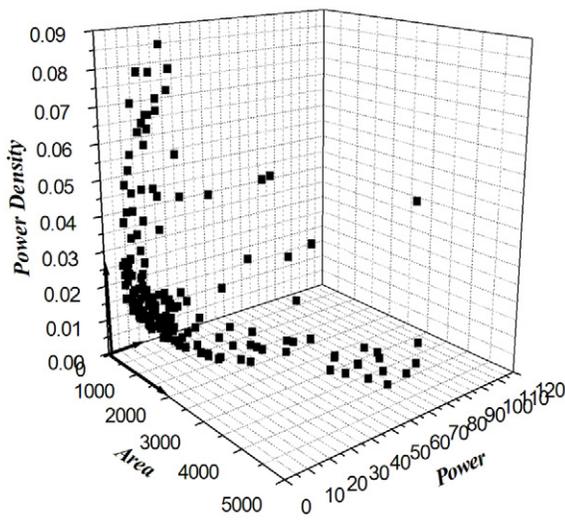
Fig. 4. Pareto optimal graph of '*table3*' benchmark circuit using NSGA-II.

B. Physical Design Implementation At 45Nm Technology

At logic level, evenly distributed or average power density is considered. NSGA-II algorithm is used to determine the optimum input variable polarity based on area, power, and power density. Initially, the dynamic power is estimated by calculating the switching activity, and the area is estimated by calculating the total number of product terms. To calculate the power density for a particular logic, the ratio of power to area is considered. Then the optimized realization is synthesized using Cadence Genus digital design platform. The synthesized netlist is implemented to have physical design at 45nm technology using Cadence Innovus platform. After physical design realization, Innovus generates the floorplan information (.flp file) and power profile (.pptrace file). In floorplan information, the synthesized logic is represented with different modules with their height, width, X and Y coordinates to allocate the position of a particular module within the chip. In the power profile, the power dissipation information of each module is given. The floorplan information and power profile are given as input to the HotSpot tool for generating the temperature profile. Based on floorplan information and power profile given, the HotSpot tool generates the temperature profile for each module in degree centigrades (°C). Fig. 5 shows the schematic flow-diagram of temperature generation using HotSpot tool. For an example case, the floorplan information and power profile of "rd53" benchmark circuit is shown in Fig. 6 and Fig. 7, respectively. The corresponding floorplan generation is shown in Fig. 8. The temperature profile generation using HotSpot tool using the floorplan information and power profile

is shown in Fig. 9.
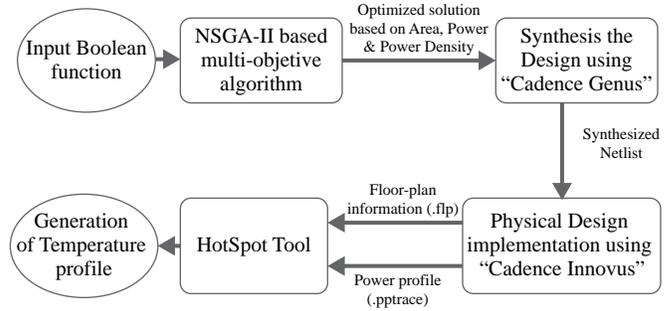
Fig. 5. Schematic Flow-diagram of temperature profile generation using the HotSpot tool.

Floorplan information
Line Format: <unit-name>   <width>   <height>   <left-x>   <bottom-y>
all dimensions are in meters

g458	0.0000540	0.0000501	0.0028400	0.0017100
g460	0.0000540	0.0000501	0.0020000	0.0028290
g459	0.0000540	0.0000501	0.0024800	0.0028290
g461	0.0000560	0.0000501	0.0028400	0.0024870
g462	0.0000540	0.0000501	0.0025600	0.0024870
g463	0.0000410	0.0000501	0.004800	0.0028290
g464	0.0000540	0.0000501	0.0025600	0.0017100
g467	0.0000440	0.0000501	0.0040000	0.0017100
g466	0.0000440	0.0000501	0.0068000	0.0028290
g469	0.0000620	0.0000501	0.0027600	0.0028290
g465	0.0000460	0.0000501	0.0022000	0.0028290
g468	0.0000440	0.0000501	0.0076000	0.0028290
g470	0.0000540	0.0000501	0.0012000	0.0017100
g473	0.0000480	0.0000501	0.0023200	0.0028290
g471	0.0000480	0.0000501	0.0016000	0.0024870
g472	0.0000460	0.0000501	0.0032000	0.0024870
g474	0.0000440	0.0000501	0.0044000	0.0024870

Fig. 6. Floorplan information of the "*rd53*" benchmark circuit generated by Cadence Innovus tool.

Power profile
Line Format: dissipation>
All power information's are in Watts

g458	g460	g459	g461	g462	g463	g464	g467	g466	g469	g465
	g468	g470	g473	g471		g472	g474			

0.449000	0.448000	0.873000	0.268000	0.558000	0.100000
	0.592000	0.078000	0.157000	0.664000	0.040000
	0.078000	0.441000	0.203000	0.187000	0.149000
	0.079000				

Fig. 7. Power profile of the "*rd53*" benchmark circuit generated by Cadence Innovus tool.
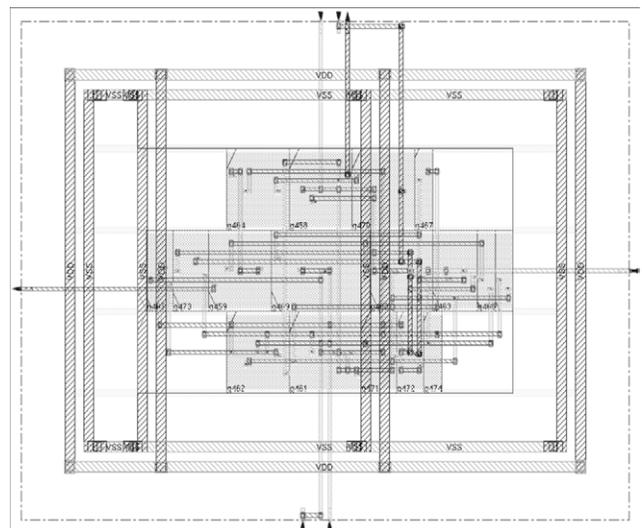
Fig. 8. Floorplan of the "*rd53*" benchmark circuit generated by the Cadence Innovus tool.

TABLE IV. Power Comparison Of the Proposed MPRM Realization

Circuits	Proposed approach		FPRM [36]			AIGs [37]			GA based FPRM [38]		
	Power Best	Optimum solution	Power	Save Best	Save Opt	Power	Save Best	Save Opt	Power	Save Best	Save Opt
5xp1	5.274	5.274	12.29	57.09	57.09	30.979	82.98	82.98	12.29	57.09	57.09
9sym	3.003	3.003	-	-	-	-	-	-	-	-	-
alu2	12.747	12.747	-	-	-	112.370	88.66	88.66	22.24	42.68	42.68
alu4	20.752	70.842	-	-	-	240.279	91.36	70.52	108.40	80.86	34.65
clip	4.568	4.568	-	-	-	-	-	-	18.85	75.77	75.77
cm162a	4.523	4.523	5.48	17.46	17.46	20.770	78.22	78.22	5.48	17.46	17.46
cm163a	3.599	5.094	5.09	29.29	-0.08	-	-	-	5.09	29.29	-0.08
con1	2.097	2.330	-	-	-	-	-	-	-	-	-
cu	7.175	7.197	4.99	-43.79	-44.23	-	-	-	-	-	-
inc	6.556	6.556	13.23	50.45	50.45	35.389	81.47	81.47	-	-	-
misex1	4.929	4.929	6.46	23.70	23.70	-	-	-	-	-	-
misex2	5.228	8.654	9.53	45.14	9.19	84	93.78	89.70	-	-	-
misex3c	5.037	8.910	-	-	-	162.090	96.89	94.50	67.42	92.53	86.78
pm1	4.288	5.899	6.56	34.63	10.08	15.900	73.03	62.90	6.28	31.72	6.07
rd53	4.580	4.580	5.61	18.36	18.36	-	-	-	5.61	18.36	18.36
rd73	7.764	7.764	-	-	-	-	-	-	13.48	42.40	42.40
rd84	12.459	12.459	20.18	38.26	38.26	-	-	-	20.18	38.26	38.26
sao2	1.825	4.654	2.49	26.71	-86.91	-	-	-	-	-	-
table3	5.425	5.703	-	-	-	-	-	-	26.75	79.72	78.68
x2	3.112	3.112	5.9	47.25	47.25	17.090	81.79	81.79	5.91	47.34	47.34
Av. % Savings of Proposed Approach				28.71	11.72		85.35	81.19		50.27	41.96

TABLE V. Power Density Comparison Of the Proposed MPRM Realization

Circuits	Proposed approach		SRMDD [22]			AIGs [37]			CPU time (s)
	Power density Best	Optimum solution	Pow_Den	Save Best	Save Opt	Pow_Den	Save Best	Save Opt	
5xp1	0.107	0.107	0.191	43.98	43.98	0.396	72.98	72.98	118.40
9sym	0.025	0.025	-	-	-	-	-	-	140.65
alu2	0.037	0.040	0.084	55.95	52.38	0.324	88.58	87.65	126.48
alu4	0.008	0.047	-	-	-	0.279	97.13	83.15	481.84
clip	0.038	0.038	0.092	58.70	58.70	-	-	-	78.82
cm162a	0.180	0.180	0.087	-106.90	-106.90	0.536	66.42	66.42	155.39
cm163a	0.199	0.283	0.283	29.68	0.00	-	-	-	148.26
con1	0.231	0.233	-	-	-	-	-	-	106.15
cu	0.247	0.312	-	-	-	-	-	-	117.16
inc	0.139	0.177	0.181	23.20	2.21	0.373	62.73	52.55	111.26
misex1	0.154	0.259	0.254	39.37	-1.97	-	-	-	147.24
misex2	0.149	0.151	-	-	-	0.333	55.26	54.65	29.92
misex3c	0.150	0.011	-	-	-	0.257	41.63	95.72	97.13
pm1	0.154	0.280	-	-	-	0.494	68.83	43.32	170.89
rd53	0.188	0.190	0.256	26.56	25.78	-	-	-	105.87
rd73	0.121	0.180	0.211	42.65	14.69	-	-	-	216.26
rd84	0.097	0.171	-	-	-	-	-	-	104.41
sao2	0.054	0.054	-	-	-	-	-	-	123.21
table3	0.016	0.026	-	-	-	-	-	-	805.14
x2	0.110	0.115	0.144	23.61	20.14	0.438	74.89	73.74	131.44
Av. % Savings of Proposed Approach				23.68	10.90		69.83	70.02	

Note: (–) indicates that the results are not reported or unavailable.

```
# Temperature profile
# Line Format:   <unit-name>        <temperature>
# All temperatures are in ºC


g458    g460    g459    g461    g462    g463    g464    g467    g466    g469    g465
        g468    g470    g473    g471            g472    g474


66.25   66.24   72.15   63.60   67.77   61.83   68.24   61.33   62.68   68.05   60.65
        61.33   66.14   63.18   62.93           62.43   61.35


Maximum Temperature = 72.15 ºC
Minimum Temperature = 60.65 ºC
```

Fig. 9. Temperature profile of the "*rd53*" benchmark circuit generated by the HotSpot tool.

Cadence (Genus and Innovus) and HotSpot tools are electronic design automation tools used for simulating the digital and analog circuits. We have generated the GDS-II report for best and optimum solutions for each benchmark circuit. Netlist, Synopsis Design Constraints (SDC) library and Library Exchange Format (LEF) files at 45nm technology are provided as input to the Cadence tool. The above process generates floor-plan information (.flp) and power profile (.pptrace), which act as input to the HotSpot tool for calculating absolute temperature profile. Thermal packaging used in HotSpot tool to generate temperature profile are ambient temperature (45.5 °C), chip thickness (0.15mm), convection capacitance (140.4 J/K), convection resistance (5 K/W), heat sink side (60mm), heat sink thickness (6.9mm), spreader side (30mm), spreader thickness (1mm), chip to spreader interface thickness (0.020mm). The dynamic thermal management (DTM) approach is applied to the proposed method by the HotSpot tool. The HotSpot tool has an in-built thermal management technique, where the threshold thermal value can be set to restructure the model to trim down the peak temperature. By default, it is 82°C, so we kept the threshold value as same for our realization. If for a particular placement of logic cells, depending on its power value and location of each cell, temperature increases beyond 82°C, then thermal model of HotSpot tool dynamically changes the relative placement of the cells such that temperature becomes below 82°C. This technique is called "dynamic thermal management" of HotSpot tool. NSGA-II provides the Pareto optimal solution set consisting of best and optimum solution based on area, power and power density. Only the solutions with the best area, best power, best power density and optimum solution consisting of area, power and power density (4   solutions) are processed further for physical design implementation using the Cadence tool. NSGA-II optimized circuits are driven into the physical design synthesis level and area (μm^2), power (nW) and temperature (°C) values are reported in Table VI.

Second, third and fourth columns with 'Best_area', 'Best_power' and 'Best_peak_Temp' report the standard cell area, power consumption and peak temperature generated by best area, best power and best power density solution of NSGA-II respectively. The next three columns report the same for the optimal solution. Comparative analysis with SRMDD and espresso decomposed AND-INVERTER GRAPH (AIG) structure [22], [37] is reported in Table VI. The last column of Table III, indicates the maximum CPU time (in seconds) to implement a benchmark circuit among all the cases (best solutions and optimal solution) in an identical platform. The average percentage savings referred in the last three rows of Table VI is calculated by the following equation.

$$Savings_{Average} = \left(\frac{Earlier_{solution} - Proposed_{solution}}{Earlier_{solution}} X100 \right)\% \qquad (27)$$

Average percentage savings is represented by "$Savings_{Average}$" as referred in Eq. (27). "$Proposed_{solution}$" and "$Earlier_{solution}$" represent the proposed approach based solution and earlier literature reported solutions, respectively.

Fig. 10, 11, and 12 show the average percentage improvement of the

proposed results reported in Table VI. Fig. 10, 11 and 12 depict that the best solution and optimum solution save 75.21% (76.20%) and 73.69% (74.82%) standard cell area than that of SRMDD-based best solution (SRMDD-based optimum solution), respectively. The best peak temperature and optimum peak temperature are reduced by 13.52% (17.06%) and 12.49%(16.08%) than that of SRMDD-based best solution (SRMDD-based optimum solution) respectively. Best area, best power and best peak temperature based solutions of MPRM expansion save 8.80%, 29.09% and 3.89% area, power, and peak temperature respectively when compared with espresso decomposed AIGs structure-based solutions. When optimal solution from Pareto-Optimal solution set is compared, it shows the average savings of 26.20% power and 2.70% peak temperature than that of espresso decomposed AIGs structure solutions at the cost of 4.39% increase in area.

	Area	Peak Temp
■ Best Solutions	75.21	13.52
□ Optimum Solution	73.69	12.49
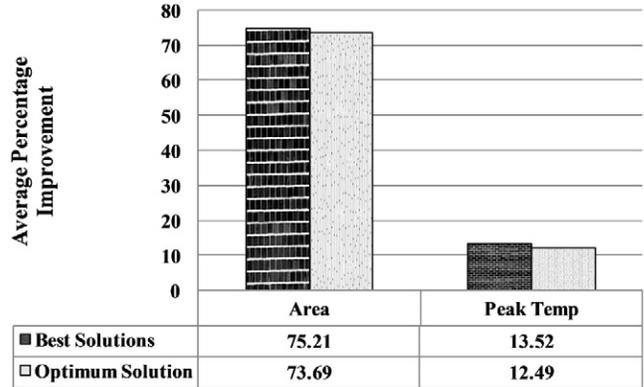
Fig. 10. Average percentage savings of the proposed approach w.r.t. SRMDD best solutions [22].

	Area	Peak Temp
■ Best Solutions	76.2	17.06
□ Optimum Solution	74.82	16.08
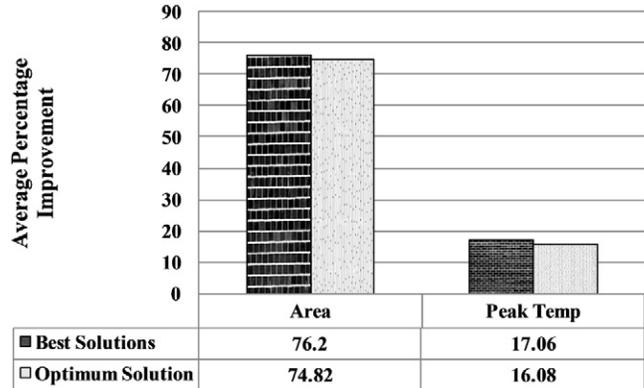
Fig. 11. Average percentage savings of the proposed approach w.r.t. SRMDD optimum solutions [22].

|	Area	Power	Peak Temp
■ Best Solutions	8.8	29.09	3.89
□ Optimum Solution	-4.39	26.2	2.7
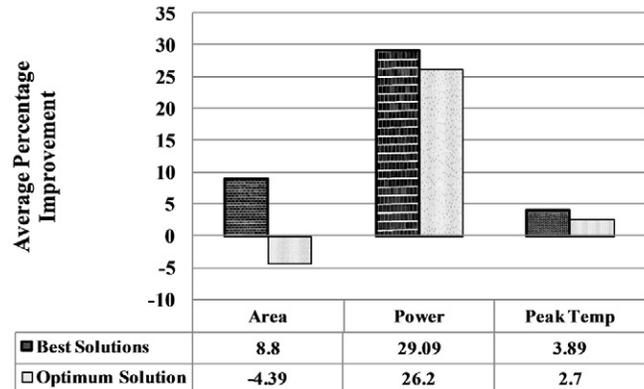
Fig. 12. Average percentage savings of the proposed approach w.r.t. AIGs circuit decomposition [37].

TABLE VI. Post Layout Analysis Of Floorplan Area (µm²), Power (nW), and Temperature (°C)

| Circuits | Proposed MPRM | | | | | | SRMDD [22] | | | | Espresso decomposed AIGs structure [37] | | | Max. CPU Time (s) |
| | Best solutions | | | Optimum solution | | | Best solutions | | Optimum solution | | | | | |
	Area (µm²)	Power (nW)	Peak Temp (°C)	Area (µm²)	Power (nW)	Peak Temp (°C)	Area (µm²)	Peak Temp (°C)	Area (µm²)	Peak Temp (°C)	Area (µm²)	Power (nW)	Peak Temp (°C)	
5xp1	145	1264.55	65.21	145	1264.55	65.21	559.88	76.70	559.88	82.12	91.31	1222.25	68.73	0.72
9sym	54.04	1121.23	67.95	89.94	1134.23	68.95	-	-	-	-	56.28	1223.30	72.29	0.60
alu2	250.34	2290.73	67.37	380.98	1942.06	69.27	2583.16	64.05	2583.16	72.57	289.67	2737.17	67.66	0.57
alu4	533.86	5093.27	66.22	676.47	5241.93	66.53	-	-	-	-	798.23	8104.59	71.06	0.89
clip	176.47	1364.12	67.90	176.47	1364.12	67.90	229.73	77.86	229.73	77.86	136.10	1385.40	70.50	0.64
cm162a	56.43	371.48	66.61	56.43	403.24	68.54	119.75	74.63	143.04	82.65	28.04	389.73	66.54	0.46
cm163a	35.91	187.24	62.95	35.91	209.69	63.41	169.65	73.98	179.83	78.91	27.92	294.60	64.50	0.40
con1	13.34	106.98	66.17	17.10	106.98	66.67	-	-	-	-	15.17	185.60	66.32	0.36
cu	9.23	88.91	63.81	14.12	88.91	65.32	-	-	-	-	24.26	186.45	65.17	0.31
inc	96.78	862.40	64.60	114.57	862.40	65.75	543.04	67.93	559.81	80.44	82.42	835.20	67.18	0.65
misex1	25.39	313.81	63.54	32.16	459.74	63.94	239.71	73.02	276.51	77.66	43.44	496.53	65.38	0.40
misex2	29.41	199.20	63.47	29.41	199.20	63.47	-	-	-	-	71.82	432.52	62.66	0.40
misex3c	285.57	1324.64	66.50	389.88	1631.50	68.15	-	-	-	-	406.30	3986.98	72.10	0.68
pm1	9.23	82.67	62.33	11.63	82.67	62.70	-	-	-	-	32.49	337.15	65.22	0.30
rd53	20.86	346.28	66.72	20.86	346.28	66.72	149.87	84.90	149.87	89.32	29.46	460.21	69.05	0.36
rd73	40.96	598.21	68.92	40.96	598.21	68.92	446.57	78.41	446.57	78.41	49.11	897.40	73.24	0.52
rd84	57.45	736.43	69.42	57.45	736.43	69.42	-	-	-	-	72.60	910.60	76.22	0.56
sao2	98.49	963.07	65.72	98.49	963.07	67.14	-	-	-	-	102.40	1094.18	68.53	0.62
table3	728.12	2343.71	65.67	771.55	2390.37	67.40	-	-	-	-	734.24	2843.53	69.10	1.00
x2	27.70	170.46	62.27	32.83	214.61	64.02	166.53	96.57	166.53	73.42	32.15	337.74	66.42	0.41
Average % savings w.r.t. SRMDD best solutions [22]	**75.21**		**13.52**	**73.69**		**12.49**								
Average % savings w.r.t. SRMDD optimum solutions [22]	**76.20**		**17.06**	**74.82**		**16.08**								
Average % savings w.r.t. AIGs Circuits [37]	**8.80**	**29.09**	**3.89**	**-4.39**	**26.20**	**2.70**								

VI. Conclusion and Future Works

This paper proposed an NSGA-II based input variable polarity selection of MPRM expansion for thermal aware realization. Area, power, and temperature are considered simultaneously as objective parameters. Product terms are considered as representative area, and switching activity is considered as the power consumption at logic level. Power per unit area (Power density) is taken as the temperature metric to estimate the effect of temperature. The input polarity of MPRM is chosen such that all the parameters are optimum. To find the non-dominated optimal solution based on input polarity of MPRM circuits, NSGA-II based approach is performed and Pareto optimal solution set is reported.

The proposed results are compared with FPRM, GA based FPRM, SRMDD, MPRM and AIGs based solutions; and significant reduction in area, power and power density generation is observed. Finally, NSGA-II based solutions are implemented using CADENCE tool at 45nm technology to obtain on-chip silicon area and power consumption. The floorplan information and power profile are used to get the absolute temperature generated by a particular logic circuit in degree Celsius using HotSpot tool. Maximum 76.20% saving in area, 29.09% saving in power and 17.06% reduction in peak temperature are observed using the proposed MPRM approach with respect to earlier reported works.

The future research is aimed to figure out the correlation between the ageing aware with the thermal aware design and to find an optimum solution to realize a circuit using MPRM expansion.

References

[1] T. Sasao, *Logic synthesis and optimization*, 1st ed. MA: Kluwer Academic Publishers, Springer US, 1993, ch. 13.

[2] J. Saul, *Logic synthesis for arithmetic circuits using the reed-muller representation*, in *Proc. of 3rd IEEE The European Conference on Design Automation,* Belgium, 1992, pp. 109-113.

[3] H. Rahaman, D.K. Das and B.B. Bhattacharya, "Testable design of AND–EXOR logic networks with universal test sets," *Comp. & Elec. Eng.*, Vol. 35, no. 5, pp. 644-658, Sept. 2009.

[4] S. Chattopadhyay, S. Roy and P.P. Chaudhuri, "Synthesis of highly testable fixed-polarity AND-XOR canonical networks-A genetic algorithm-based approach," *IEEE trans. on comps.*, Vol. 45, no. 4, pp. 487-490, Apr. 1996.

[5] T. Villa, T. Kam, R. K. Brayton and A. L. Sangiovanni-Vincentelli, *Synthesis of finite state machines: logic optimization*, Springer Science & Business Media, Dec. 2012, ch. 10.

[6] W. Pengjun, and L. Hui, "Low power mapping for AND/XOR circuits and its application in searching the best mixed-polarity,' *J. of Semiconductors*, Vol. 32, no. 2, pp. 025007, 2011.

[7] U. Narayanan and C.L. Liu, "Low power logic synthesis for XOR based circuits,"in *Proc. of the IEEE/ACM International conference on Computer-aided design*, USA, 1997, pp. 570-574.

[8] Z. He, L. Xiao, L. Ruan, F. Gu, Z. Huo, G. Qin, M. Zhu, L. Zhang, R. Liu, and X. Wang, "A Power and Area Optimization Approach of Mixed Polarity Reed-Muller Expression for Incompletely Specified Boolean Functions," *J. of Comp. Sci. and Techn.*,Vol. 32, no. 2, pp. 297-311, Mar. 2017.

[9] T. Sasao, *Switching theory for logic synthesis*, 1st ed. Springer Science & Business Media, 2012, ch. 3.

[10] B.A. Al Jassani, N. Urquhart and A.E.A. Almaini, "Manipulation and optimisation techniques for Boolean logic," *IET Comp. & Dig. Techn.,* Vol. 4, no. 3, pp. 227-239, may 2010.

[11] Y. S. Xia, L.Y. Wang, Z.G. Zhou, X.E. Ye and J.P. Hu, "Novel synthesis and optimization of multi-level mixed polarity Reed-Muller functions," *J. of Comp. Sci. and Techn.*, Vol. 20, no. 6, pp. 895-900, Nov. 2005.

[12] Z. He, L. Xiao, Z. Huo, T. Wang, and X. Wang, "Fast Minimization of Fixed Polarity Reed-Muller Expressions," *IEEE Access*, Vol. 7, pp. 24843-24851, 2019.

[13] Z. He, L. Xiao, Z. Huo, C. Wang, J. Liu and X. Wang, "POA-FPRMs: Power Optimization Approach of Fixed Polarity Reed-Muller Expressions for Incompletely Specified Boolean Functions," Chinese Journal of Electronics, Vol. 28, no. 6, pp. 1144-1151, 2019.

[14] C. Chen, B. Lin, and M. Zhu, "Verification Method for Area Optimization of Mixed-Polarity Reed-Muller Logic Circuits," *Journal of Engg. Sci. & Techn. Rev.*, Vol. 11, no. 1, pp. 28-34, 2018.

[15] Z. He, L. Xiao, F. Gu, L. Ruan, Z. Huo, M. Li, M. Zhu, L. Zhang, R. Liu, and X. Wang, "EDOA: an efficient delay optimization approach for mixed-polarity Reed-Muller logic circuits under the unit delay model," *Frontiers of Comp. Sci.,* Vol. 13, no. 5, pp. 1102-1115, 2019.

[16] A. Das, and S. N. Pradhan, "Area-Power-Temperature Aware AND-XOR Network Synthesis Based on Shared Mixed Polarity Reed-Muller Expansion," *Int. J. of Intelligent Sys. and App.*, Vol. 10, no. 12, pp. 35-46, 2018.

[17] A. Das, A. Debnath and S.N. Pradhan, "Area, power and temperature optimization during binary decision diagram based circuit synthesis,"in *proc.IEEE International conference on Devices for Integrated Circuit*, India, 2017, pp. 778-782.

[18] The Temperature Ratings of Electronic Parts. In: Electronics COOLING [online]. 2004. Available: http://www.electronics-cooling.com/2004/02/the-temperature-ratings-of-electronic-parts/. [last accessed July 2018].

[19] L. Wang, "Fast algorithms for thermal-aware floorplanning,"*J. of Circ., Sys., and Comp.,* Vol. 23, no. 07, pp. 1450098, 2014.

[20] K. Sankaranarayanan, S. Velusamy, M. Stan, and K. Skadron, "A case for thermal-aware floorplanning at the microarchitectural level," J. of Instr.-Lev. Parallelism, Vol. 7, no. 1, pp. 8-16, May 2005.

[21] S. Gunther, F. Binns, D.M. Carmean and J.C. Hall, "Managing the impact of increasing microprocessor power consumption," *Intel Techn. J.*, Vol. 5, no. 1, pp. 1-9, 2001.

[22] A. Das, and S.N. Pradhan, "Shared Reed-Muller Decision Diagram Based Thermal-Aware AND-XOR Decomposition of Logic Circuits," *VLSI Design*, Vol. 2016, pp. 1-14, Mar. 2016.

[23] A. Das and S.N. Pradhan, "Thermal aware FPRM based AND-XOR network synthesis of logic circuits,"in *proc. 2nd IEEE International Conference on Recent Trends in Information Systems*, India, 2015, pp. 497-502.

[24] A. Das, and S.N. Pradhan, "Thermal aware output polarity selection of programmable logic arrays,"in *proc. IEEE International Conference on Electronic Design, Computer Networks & Automated Verification*, India, 2015, pp. 68-71.

[25] A. Das, S. R. Choudhury, B.K. Kumar and S.N. Pradhan, "An elitist area-power density trade-off in VLSI floorplan using genetic algorithm,"*in proc. 7th IEEE International Conference on Electrical and Computer Engineering*,Bangladesh, 2012, pp. 729-732.

[26] P. Choudhury, K. Manna, V. Rai, and S. N. Pradhan, "Thermal-Aware Partitioning and Encoding of Power-Gated FSM," J. of Circ., Sys.and Comp., Vol. 28, no. 9, pp. 1950144, 2018.

[27] M. Pedram, and S. Nazarian, "Thermal modeling, analysis, and management in VLSI circuits: Principles and methods,"*in Proc. of the IEEE*, Vol. 94, no. 8, pp. 1487-1501, Sept. 2006.

[28] L. Shang, and R.P. Dick, "Thermal crisis: challenges and potential solutions," *IEEE Potentials*, Vol. 25, no. 5, pp. 31-35, Sept. 2006.

[29] R. E. Ladner, "On the structure of polynomial time reducibility," *Journal of the ACM,* Vol. 22, no. 1, pp. 155-171, Jan. 1975.

[30] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, & M. R. Stan, "HotSpot: A compact thermal modeling methodology for early-stage VLSI design," *IEEE Trans. on Very Large Scale Integration Sys.*, Vol. 14, no. 5, pp. 501-513, Jul. 2006.

[31] Cadence Innovus Implementation System. Available: https://www.cadence.com/content/dam/cadence-www/global/en_US/documents/tools/digital-design-signoff/innovus-implementation-system-ds.pdf. [last accessed August 2018].

[32] S. N. Pradhan, M. T. Kumar, and S. Chattopdhyay, "AND-OR-XOR Network Synthesis with Area-Power trade-off" *Journal of Circ., Sys., and Comp.,* Vol. 20, no. 06,pp. 1019-1035, 2011.

[33] K. Deb, A. Pratap, S. Agarwal and T.A.M.T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. on Evo. Comp.* Vol. 6, no. 2, pp. 182-197, Apr. 2002.

[34] S. Chaudhury, S. Chattopadhyay, "Output phase assignment for area and power minimization in PLAs," *Journal of Indian Inst. Sci.*, Vol. 86, pp. 33-43, Jan-Feb., 2006.

[35] MCNC and LGSynth93 benchmarks. In: Collection of Digital Design Benchmarks [online]. Available: http://ddd.fit.cvut.cz/prj/Benchmarks/. [last accessed September 2018].

[36] S. Chaudhury, and S. Chattopadhyay, "Fixed polarity Reed-Muller network synthesis and its application in AND-OR/XOR-based circuit realization with area-power trade-off," *IETE J. of Resarch*, Vol. 54, no. 5, pp. 353-363, Sept. 2008.

[37] A. Das, and S. N. Pradhan, "Thermal-aware Output Polarity Selection Based on And-Inverter Graph Manipulation," *Rec. Adv. in Elec.& Electronic Engg.*, Vol. 12, no. 1, pp. 30-39, Feb. 2019.

[38] S. N. Pradhan, and S. Chattopadhyay, "Two-level AND-XOR networks synthesis with area-power trade-off," Int. J. of Comp. Sci. and Net. Sec.,Vol 8, no. 9, pp. 365-375, Sept. 2008.

Apangshu Das

Apangshu Das is an assistant professor in the Department of Electronics & Communication Engineering, National Institute of Technology Agartala. He received M.Tech degree in Microelectronics and VLSI Design from National Institute of Technology Agartala and B. E. degree in Electronics and Communication Engineering from Nagpur University, in 2012 and 2009 respectively. He is currently working toward the PhD degree on thermal-aware logic synthesis. His current research interest includes thermal aware logic synthesis, low power design, and multi-objective optimization techniques.

Sambhu Nath Pradhan

Sambhu Nath Pradhan is an associate professor in the Department of Electronics & Communication Engineering, National Institute of Technology Agartala. He received Ph.D. from Indian Institute of Technology Kharagpur and M.E. from Indian Institute of Engineering Science and Technology, Shibpur (Previously known as, Bengal Engineering and Science University) in 2010 and 2004 respectively. His research interest includes low power design and testing and thermal aware logic synthesis.

Chrome Layer Thickness Modelling in a Hard Chromium Plating Process Using a Hybrid PSO/RBF–SVM–Based Model

Paulino José García Nieto[1]*, Esperanza García-Gonzalo[1], Fernando Sánchez Lasheras[1], Antonio Bernardo Sánchez[2]

[1] Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo (Spain)
[2] Department of Mining Technology, Topography and Structures, University of León, 24071 León (Spain)

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

The purpose of chromium plating is the creation of a hard and wear-resistant layer of chromium over a metallic surface. The principal feature of chromium plating is its endurance in the face of the wear and corrosion. This industrial process has a vast range of applications in many different areas. In the performance of this process, some difficulties can be found. Some of the most common are melt deposition, milky white chromium deposition, rough or sandy chromium deposition and lack of toughness of the layer or wear and lack of thickness of the layer deposited. This study builds a novel nonparametric method relied on the statistical machine learning that employs a hybrid support vector machines (SVMs) model for the hard chromium layer thickness forecast. The SVM hyperparameters optimization was made with the help of the Particle Swarm Optimizer (PSO). The outcomes indicate that PSO/SVM–based model together with radial basis function (RBF) kernel has permitted to foretell the thickness of the chromium layer created in this industrial process satisfactorily. Thus, two kinds of outcomes have been obtained: firstly, this model permits to determine the ranking of relevance of the seven independent input variables investigated in this industrial process. Finally, the high achievement and lack of complexity of the model indicate that the PSO/SVM method is very interesting compared to other conventional foretelling techniques, since a coefficient of determination of 0.9952 is acquired.

I. Introduction

HARD chromium is an electroplating process that has been employed for many years. Its purpose is the creation of an extremely hard surface over a metal part [1].

According to the thickness of the chromium layer, there are two kinds of chromium plating processes. The decorative plating which makes use of layer thickness of up to 1.5 μm and the hard chromium plating whose thickness goes from 2.5 to 500 μm. Not only the thickness is the difference of decorative and hard chromium plating processes, but also that in the case of hard chromium plating the layer is applied over the metal part while in the case of decorative plating a coat of nickel or copper nickel alloy is employed.

In general, the hard chromium plating process requires of some initial baths that clear the piece and of an electropolishing phase. Afterwards the chrome plating operation is performed. The bath in which this operation is performed contains chromic and sulphuric acid in proportions of 100 to 1 respectively [2]. A chemical equilibrium is maintained of chromic and sulphuric acid that have a reversible chemical reaction, forming chrome metal, water and oxygen. In this electrolytic process the anode employed has not to be replaced as it is solid.

The presence of sulphuric acid is compulsory, as it is essential for the reaction and therefore, the coating, to take place but it must be controlled in order to maintain among certain limits [3].

There is a vast range of composition proportions of the bath. For example, the weight ratio of chromic acid to sulphate goes from 75 to 1 to 120 to 1 with an optimum value about 90 to 1.

The disadvantages of low chromic acid have been reported in the literature [4] and they mainly consist of the difficulties to obtain the required thickness for the layer, the high sensitivity to different bath concentration and the need of a higher voltage. Despite these, such kind of baths are employed as they work faster. When high concentrations are employed, the resulting bath is less affected by bath contamination and, also, it can operate with lower voltages.

Hard chromium plating process is mainly applied in those parts that require of a high resistance to wear. The hardness of chromium plating is between 800 and 1000 HV. Also, it has a low friction coefficient. For example, it is of 0.16 for chromium to steel while for steel to steel it is of 0.30 [5]. One of the main fields where this kind of plating is useful is in automotive industry, where these covers are employed to fight against wear. Also, in iron and steel industry hard chromium plating is a process employed for plating rolling mill cylinders.

---

* Corresponding author.

E-mail address: pjgarcia@uniovi.es

In the case of tools, the hard chromium plating recommendation is linked to the achievement of a higher wear, friction and tear resistance [6]. This treatment is also of interest for the corrosion reduction. It means than any tool with a chrome layer from 2.5 to 13 μm will have a longer life. Hard chromium plating is also employed for the rework of parts out of tolerance. For example, engine components like shafts and cylinders.

Also, it is remarkable that those rolling mills that are coated experiment a great increase in their average life [3]. From a mechanical point of view, the properties of electrodeposited chromium are like those of metal chromium. The most important of these are its great hardness, low friction coefficient, good corrosion resistance, low thermal expansion, and high conductivity.

For a good result of a chromium plating process, it is not only required the control of the plating parameters but also having the required kind of base material. A cleaning with electropolishing before the plating is in many cases required. The base material employed must be able to withstand the external forces applied without hardly any deformation. That is that the properties of the base material must be like those of the applied chromium layer. Also, the thickness of the chromium layer should be reduced to a minimum in those areas of the part that would suffer from high deformation values.
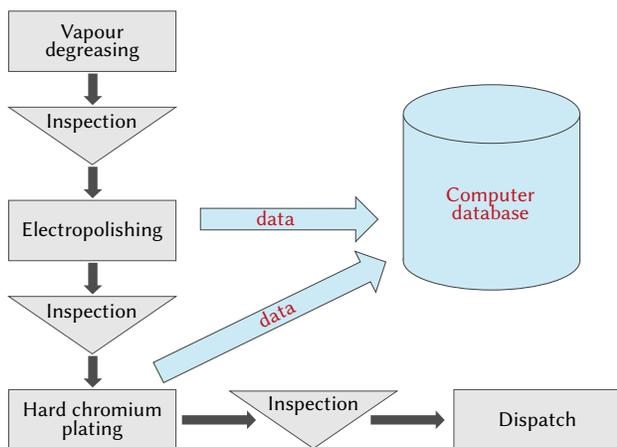

Fig. 1. The industrial process of hard chromium plating.

Fig. 1 shows the flowchart of the industrial hard chromium process. In a first stage, parts go through a vapour degreasing operation that cleans the part. It is followed by an electropolishing phase that prepares the part surface to receive the chromium layer. The next step is the proper chromium operation. This entire process requires of some intermediate inspection points and generates data that can be analyzed in order to adjust and improve the whole process.

The aim of this research is to assess the goodness of the support vector machines (SVMs) regression model [8]–[12] to predict the chromium layer thickness considering different variables from the industrial process. The purpose of the model developed in this research can overcome some problems found in industry and give to the researchers a more in-depth knowledge of the industrial process as a whole.

SVM are a kind of machine learning models that have shown a great performance for both regression and classification tasks [8]–[12]. In the case of the present research SVM models were trained to predict the thickness of the chromium plating layer. SVM models has proved to be successful in other disciplines such as the prediction of biological parameters in environmental problems like forest modeling [13], solar power generation prediction [14], [15] and air and water quality estimation [16]–[19].

Particle swarm optimization (PSO) technique is a metaheuristic evolutionary global method able to solve multidimensional optimization problems that makes use of continuous variables. It is a bio-inspired algorithm relied on swarm intelligence [20]–[25].

In summary, the main objective of the present study was to develop a hybrid algorithm with PSO optimizing SVM parameters (PSO/SVM–based model) to predict the thickness of the hard chromium layer from the seven physical-chemical input parameters from the hard chromium plating process.

II. Materials and Methods

A. The Chromium Plating Process and Experimental Datasets

The chromium plating process requires of different stages in order to achieve the required chromium layer. There are four main process stages: vapor degreasing, cleaning, electropolishing and hard chrome process.

Usually the degreasing in vapor phase is performed in a stainless steel tank. This process takes about 10 minutes as such time is enough in order to reach the temperature required. It can be verified that the required temperature has been reached because the steam does not condense on the lower surfaces of the pieces. The next step consists on a parts cleaning in order to remove all the residual traces of dirt and surface impurities, this operation has an important manual component.

The next stage is the Electropolishing: The electropolishing operation consists of the following phases, first of all, pieces are cleaned with acetone. Afterwards, laded in the frames where the electropolishing process is performed. For each electropolished load, the system records the minimum, maximum and average temperature values. The electropolishing time depends on the amount of material to be removed.

After the electropolishing the hard chromium plating stage of the process is performed. Before the chrome operation, pieces are cleaned with a pressurized water gun. Next, each piece is blown with pressurized air to remove the remains of water. The chromating of the pieces is carried out by immersion in a bath with a temperature between 53 and 56ºC. Once the chromium plating process has finished, pieces are removed from the frame, washed with water and introduced into an alkaline bath. The minimum time in this bath will be of three minutes. After this time, they are extracted from the bath and washed by hand with water.

The experimental dataset was collected using a database from the hard chromium plating of the factory belonging to the company Tecnocrom Industrial Ltd (Madrid, Spain). Eight process variables are collected from the hard chrome plating process including the dependent variable (thickness of the hard chrome layer). These seven input variables are briefly described in Table I. The total number of data used in this study is 37,360. The output variable (or dependent variable) considered in this study was the hard chromium plating thickness (in microns).

In this research relationship between the thickness of the hard chromium layer (dependent variable) in microns and some physical-chemical variables (independent variables) has been studied. The variables employed are listed below:

- Iron content of the electropolishing bath (mg/L): this variable measures the iron content in the electropolishing bath expressed in mg/L.

- Electropolishing time (minutes): this variable represents the amount of time, expressed in seconds that the part is submerged in the electropolishing bath.

- Electropolishing bath temperature (°C): this variable expresses the electropolishing bath temperature in Celsius degrees. It is well-know that when the higher the temperature value, the lower the process performance.
- Layer thickness removed by electropolishing (mm): the electropolishing process suposses the removal of a certain amount of material. This variable measures the amount of material removed.
- Chromic acid content (g/L): chromic acid content in the hard chromium bath expressed in g/L.
- Hard chrome plating time (minutes): this variable measures the time of the hard chromium plating operation.
- Hard chrome plating temperature (°C): this variable expresses the chrome plating bath temperature in Celsius degrees. As in the case of the electropolishing and in this case, due to the electrolytical process, the bath temperature increases. The bath temperature influences in the operation performance.

TABLE I. Set of Operation Physical input Variables Used in This Study and their Names Along With their Mean and Standard Deviation

Input variables	Variable name	Mean	Stand. dev.
Iron content of the electropolishing bath (mg/l)	iron_ep	0.12	0.09
Electropolishing time (minutes)	ep_time	4.50	0.91
Electropolishing bath temperature (°C)	ep_temp	53.60	4.62
Layer thickness removed by electropolishing (μm)	thick_electropol	305.9	34.46
Chromic acid content (g/l)	chrom_acid	240.6	8.71
Hard chromium process time (minutes)	hard_chrom_time	30.7	3.33
Hard chrome bath temperature (°C)	hard_chrom_temp	53.9	0.47

Fig. 2 shows an overview of the eight sample variables over time. The highly nonlinear behaviour of the output variables can be clearly appreciated.

B. Computational Procedures

The first step in this study is the determination of the correlation matrix of all the process variables (see Fig. 3). This matrix indicates that there is hardly any correlation among the independent variables and some correlation of the dependent variable thickness of the hard chrome layer (thickness_chrom) with input variables layer thickness removed by electropolishing (thick_electropol) and hard chromium process time (hard_chrom_time). This outcome stands out the trouble that we can encounter in acquiring foretold models using linear regression methods. As a consequence, it is required to deal with this complicated issue employing more developed nonlinear procedures relied on statistical machine learning such as the support vector machines (SVMs) together with the particle swarm optimizer (PSO) giving place to the PSO/SVM–relied model used in this study.
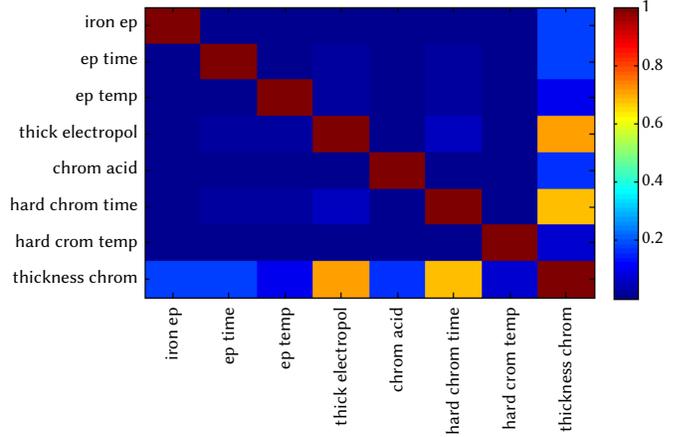
Fig. 3. Correlation matrix.

1. Support Vector Machines For Regression (SVR)

Initially, support vector machines (SVMs) arose to treat binary classification problems. Under these circumstances, it was quickly observed that the basic rules that support them could be used to tackle other kind of problems such as the regression problems. Certainly, in place of trying to sort new unnoticed variables $\mathbf{x}'$ into one of two classes $y' = \pm 1$, we are going now to foretell a real-valued output for y' from the training data $\{\mathbf{x}_i, y_i\}$, so that $i = 1, 2, ..., L$ with $y_i \in \Re$, $\mathbf{x}_i \in \Re^D$ [26]–[29] so that:

$$y_i = \mathbf{w} \cdot \mathbf{x}_i + b \qquad (1)$$

where:

- $\mathbf{w}$ is director vector of the hyperplane (normal to this hyperplane);
- $\dfrac{b}{\|\mathbf{w}\|}$ is the perpendicular (normal) distance from the hyperplane to the coordinates origin.

The support vector regression (SVR) approach employs a more
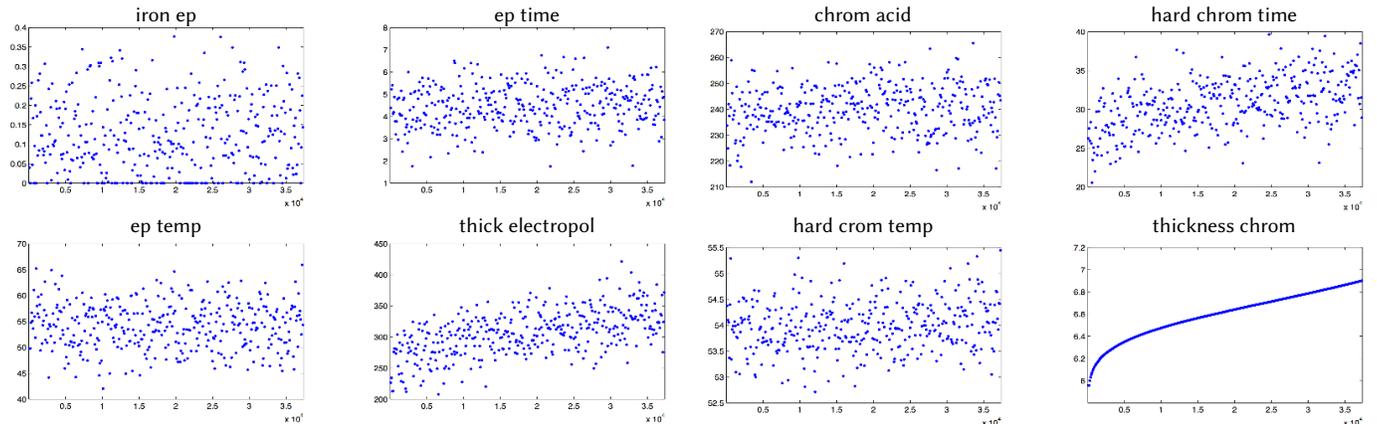
Fig. 2. Graphical overview of the process variables: variable value versus observation number.

cumbersome penalty function than in case of classification, not permitting a penalization if the foretold value y_i is less than a distance ε away from the observed value t_i, that is to say, if $|t_i - y_i|$ is fulfilled. Alluding to Fig. 4, the region enclosed by $y_i \pm \varepsilon$ for all i is termed an ε – insensitive tube. The another change of the penalization function is that output variables falling out of the tube receive one of two slack variable penalizations relying on whether they stay above (ξ^+) or below (ξ^-) the tube (so that $\xi_i^+ > 0$, $\xi_i^- > 0$ $\forall i$) [12], [13], [30], [31]:

$$t_i \le y_i + \varepsilon + \xi_i^+ \tag{2}$$

$$t_i \ge y_i - \varepsilon - \xi_i^- \tag{3}$$
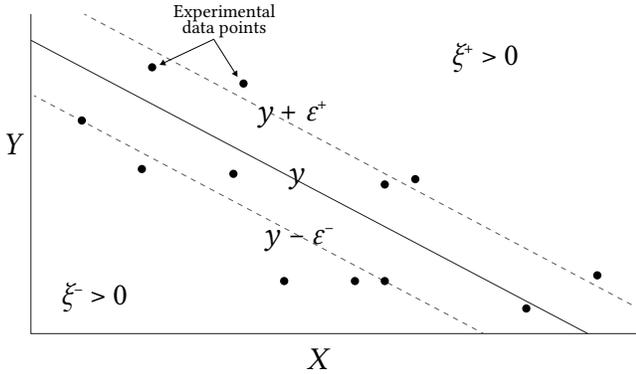
Fig. 4. Regression with ε – insensitive tube.

The error function for SVR approach is given by [9], [30]:

$$C \sum_{i=1}^{L} \left(\xi_i^+ + \xi_i^- \right) + \frac{\|w\|^2}{2} \tag{4}$$

In brief, the *primal problem* in the event of regression is described as [12] , [31], [32]:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{L} \left(\xi_i^+ + \xi_i^- \right)$$

$$\text{s.t.} \begin{cases} t_i - \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) - \varepsilon - \xi_i^+ \le 0 \\ t_i - \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) + \varepsilon + \xi_i^- \ge 0 \\ \xi_i^+, \xi_i^- \ge 0, \forall i = 1, 2, ..., n \end{cases} \tag{5}$$

Next, we minimize the error function supplied by Eq. (4) along with the constraints $\xi_i^+ \ge 0$, $\xi_i^- \ge 0$ $\forall i$ in addition to inequalities (2) and (3), which is expressed compactly by the Eq. (5). For the purpose of building the dual problem, the following steps are demanded [9], [12], [18], [30], [31]:

- Step 1: Lagrange function construction

To this end, we use the following Lagrange multipliers $\alpha_i^+ \ge 0$, $\alpha_i^- \ge 0$, $\mu_i^+ \ge 0$, $\mu_i^- \ge 0$ $\forall i$:

$$L_P\left(\mathbf{w}, b, \xi^+, \xi^-, \alpha^+, \alpha^-, \mu^+, \mu^-\right) = \frac{1}{2} \|\mathbf{w}\|^2$$

$$+ C \sum_{i=1}^{L} \left(\xi_i^+ + \xi_i^- \right) - \sum_{i=1}^{L} \left(\mu_i^+ \xi_i^+ + \mu_i^- \xi_i^- \right)$$

$$- \sum_{i=1}^{L} \alpha_i^+ \left[\varepsilon + \xi_i^+ + \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) - t_i \right] - \sum_{i=1}^{L} \alpha_i^- \left[\varepsilon + \xi_i^- - \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) + t_i \right] \tag{6}$$

- Step 2: Implementation of Karush–Kuhn–Tucker (KKT) optimality conditions

Carrying out the differentiation of L_P with respect to $\mathbf{w}$, b, ξ_i^+ and ξ_i^-, and matching the derivatives to zero, it turns out that:

$$\frac{\partial L_P}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) \mathbf{x}_i \tag{7}$$

$$\frac{\partial L_P}{\partial b} = 0 \Rightarrow \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) = 0 \tag{8}$$

$$\frac{\partial L_P}{\partial \xi_i^+} = 0 \Rightarrow C = \alpha_i^+ + \mu_i^+ \Leftrightarrow \mu_i^+ = C - \alpha_i^+ \tag{9}$$

$$\frac{\partial L_P}{\partial \xi_i^-} = 0 \Rightarrow C = \alpha_i^- + \mu_i^- \Leftrightarrow \mu_i^- = C - \alpha_i^- \tag{10}$$

$$\alpha_i^+ \left[\varepsilon + \xi_i^+ + \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) - t_i \right] = 0 \tag{11}$$

$$\alpha_i^- \left[\varepsilon + \xi_i^- - \left(\mathbf{w} \cdot \mathbf{x}_i + b\right) + t_i \right] = 0 \tag{12}$$

- Step 3: Establishment of the relations among the primal problem's variables ($\mathbf{w}$, b, ξ^+, ξ^-) and those of the dual problem($\boldsymbol{\alpha}^+$, $\boldsymbol{\alpha}^-$, $\boldsymbol{\mu}^+$, $\boldsymbol{\mu}^-$). For this, we make use of Eq. (7).

- Step 4: From the result obtained in previous step, it is possible to remove the primary variables of the Lagrangian function so that:

$$L_D\left(\alpha^+, \alpha^-\right) = \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) t_i - \varepsilon \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right)$$

$$- \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right)\left(\alpha_j^+ - \alpha_j^- \right) \mathbf{x}_i \cdot \mathbf{x}_j \tag{13}$$

Using $\mu_i^+ \ge 0$ and $\mu_i^- \ge 0$ together with Eqs. (9) and (10) means that $\alpha_i^+ \le C$ and $\alpha_i^- \le C$. Finally, the mathematical formulation of the dual problem can be expressed as [9], [12], [18], [30], [31]:

$$\max_{\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-} \left[\sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) t_i - \varepsilon \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) - \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right)\left(\alpha_j^+ - \alpha_j^- \right) \mathbf{x}_i \cdot \mathbf{x}_j \right]$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) = 0 \\ 0 \le \alpha_i^+ \le C, i = 1, ..., n \\ 0 \le \alpha_i^- \le C, i = 1, ..., n \end{cases} \tag{14}$$

Replacing Eq. (7) into Eq. (1), new forecasts y' can be calculated employing [12], [30]:

$$y' = \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) \mathbf{x}_i \cdot \mathbf{x}' + b \tag{15}$$

An ensemble S of support vectors $\mathbf{x}_S$ can be determined encountering the indices i such that $0 < \alpha < C$ and $\xi_i^+ = 0$ (or $\xi_i^- = 0$). This operation gives place to [12], [30]:

$$b = t_S - \varepsilon - \sum_{m=1}^{L} \left(\alpha_m^+ - \alpha_m^- \right) \mathbf{x}_m \cdot \mathbf{x}_S \tag{16}$$

Following the methodology, we average over all indices i in S so that b is defined correctly by [9], [12], [17]:

$$b = \frac{1}{N_S} \sum_{s \in S} \left[t_S - \varepsilon - \sum_{m=1}^{L} \left(\alpha_m^+ - \alpha_m^- \right) \mathbf{x}_m \cdot \mathbf{x}_S \right] \tag{17}$$

Additionally, some investigations have shown that some regression problems are not linearly solved in the space of the inputs $\mathbf{x}$, but are solvable in a higher-dimensional space called the *feature space* by using a suitable mapping $\mathbf{x} \rightarrow \Phi(\mathbf{x})$. This indicates that if the functions can be brought into a space of higher dimensionality by some potentially nonlinear mapping $\mathbf{x} \rightarrow \Phi(\mathbf{x})$, it is only necessary

to determine the scalar products of the inputs mapped in the feature space without having to explicitly calculate $\Phi(\mathbf{x})$. For the purpose of using an SVM approach to solve a regression problem with nonlinear separable data, firstly it is necessary to select a kernel together with its relevant parameters so that we can map the nonlinear separable data into a *feature space* where they are linearly separable (*kernel trick*).

Hence, the first step is the selection of a suitable kernel that gives place to a mapping $\mathbf{x} \rightarrow \Phi(\mathbf{x})$. Next, it is mandatory to follow the steps indicated below [9], [12], [30]:

- Selection of the appropriate values for the hyperparameters C and ε, that is to say, the weight of misclassifications and the size of the insensitive loss region.

- Determination *of* $\boldsymbol{\alpha}^+$ and $\boldsymbol{\alpha}^-$ by means of:

$$
\max_{\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-} \left[\sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) t_i - \varepsilon \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) \right.
$$

$$
\left. - \frac{1}{2} \sum_{i=1}^{L} \sum_{j=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) \left(\alpha_j^+ - \alpha_j^- \right) k \left(\mathbf{x}_i, \mathbf{x}_j \right) \right]
$$

$$
\text{s.t.} \begin{cases} \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) = 0 \\ 0 \leq \alpha_i^+ \leq C, i = 1, ..., n \\ 0 \leq \alpha_i^- \leq C, i = 1, ..., n \end{cases}
$$

(18)

such that $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ is termed the kernel function. Moreover, the kernel function must be symmetric and semidefinite positive according to the Moore–Aronszajn theorem. Likewise, the solution of the dual problem (15) is carried out employing a quadratic programming (QP) solver.

- Calculation of $\mathbf{w} = \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) \Phi(\mathbf{x}_i)$.
- Determination of the ensemble of *Support Vectors S* encountering the indices i such that $0 < \alpha \leq C$ and $\xi_i^+ = 0$ (or $\xi_i^- = 0$).
- Calculation of b:

$$
b = \frac{1}{N_S} \sum_{s \in S} \left[t_s - \varepsilon - \sum_{m=1}^{L} \left(\alpha_m^+ - \alpha_m^- \right) k \left(\mathbf{x}_m, \mathbf{x}_s \right) \right]
$$

(19)

- All novel point $\mathbf{x}'$ is encountering by means of the expression:

$$
y' = \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}') + b
$$

$$
= \sum_{i=1}^{L} \left(\alpha_i^+ - \alpha_i^- \right) k \left(\mathbf{x}_i, \mathbf{x}' \right) + b
$$

(20)

Several usual functions used as kernels in the scientific literature [10], [12], [30]–[32] are formulated as:

- Radial basis function termed RBF kernel:

$$
k \left(\mathbf{x}_i, \mathbf{x}_j \right) = e^{-\sigma \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2}
$$

(21)

- Polynomial kernel:

$$
k \left(\mathbf{x}_i, \mathbf{x}_j \right) = \left(\sigma \mathbf{x}_i \cdot \mathbf{x}_j + a \right)^b
$$

(22)

- Sigmoid kernel:

$$
k \left(\mathbf{x}_i, \mathbf{x}_j \right) = \tanh \left(\sigma \mathbf{x}_i \cdot \mathbf{x}_j + a \right)
$$

(23)

so that a, b and σ are hyperparameters indicating the kernel's functioning.

Moreover, representative parameters of the SVM approach can be summarized as [12], [30], [32]:

- Regularization constant (C): also term *cost function*. This factor defines the trade-off between the margin (i.e., model flatness) and the relevance of the slack variables (related to the training error). Furthermore, this constant C must be chosen a priori, being a parameter of the machine learning.

- ε parameter: this factor controls the width of the error margin allowed. The second term of the objective function (see Eqs. (4) and (5)) is known as empirical error determined by means of the $\varepsilon-$ insensitive loss function, which indicates that it does not disregard errors below ε (i.e. to a distance ε of the true value).

- a, b and σ: these factors determine the expression of the different kernels in the subsequent model.

Therefore, it is appropriate to utilize some mathematical technique that determines the previous hyperparameters accurately enough. These parameters determine the ensemble of support vectors and their influence in the subsequent regression model, that is to say, the precision and robustness of the last model. Moreover, the particle swarm optimizer (PSO) explained in more detail below was employed [20], [33] with triumph in this work.

2. The Particle Swarm Optimization (PSO) Algorithm

Particle swarm optimizer (PSO) [20]–[22], [33] is a heuristic kind of optimizer aimed at finding global minimums or maximums. Its operation is motivated by the habits of groups of birds or ensembles of fish in which, the movement of each individual (direction, speed, acceleration, etc.), is the result of combining the individual decisions of each with the behaviour of the rest.

Although there are variations, in general terms, the structure of a PSO algorithm to optimize (maximize or minimize) a function with one or multiple variables follows the following steps:

1. Create an initial swarm of n random particles. Each particle consists of 4 elements: a position that represents a certain combination of variable values, the value of the goal function at the position where the particle is located, a velocity that indicates how and where the particle moves, and a record of the best position in which the particle has been so far.

2. Evaluate each particle with the objective function.

3. Update the position and speed of each particle. This is the part that gives the algorithm the ability to optimize. This process is described in detail in the step *Move the particle* below.

4. If a stopping criterion is not fulfilled, return to step number 2.

In the following steps, each stage of the process is implemented to finally combine them all into a single function:

- *Create a particle*

Each particle is determined by a position, speed and value that vary as the particle moves. In addition, it also stores the best position in which the particle has been so far. When a new particle is created, only information about its position and velocity (normally initiated as zero) is available and the rest of the values are not known until the particle is evaluated.

- *Evaluate the particle*

Evaluating a particle consists in calculating the value of the goal function in the position that the particle occupies in that moment. The position with the best value in which the particle has been so far is stocked up on. In order to identify if a novel location is better than the previous ones, it is necessary to know if it is a problem of minimization or maximization.

- *Move the particle*

Moving a particle means updating its speed and position. This step is the most important since it gives the algorithm the ability to optimize.

The speed of each particle in the ensemble is upgraded employing the next equation:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 \left[\hat{x}_i(t) - x_i(t) \right]$$
$$+ c_2 r_2 \left[g(t) - x_i(t) \right] \qquad (24)$$

where:

- $v_i(t+1)$: particle velocity i at time $t+1$, that is, the new velocity.
- $v_i(t)$: particle velocity i at time t, that is, the current velocity.
- ω: inertia coefficient, which reduces or increases at the speed of the particle.
- c_1: cognitive coefficient.
- r_1: vector of random values between 0 and 1 of length equal to that of the velocity vector.
- $\hat{x}_i(t)$: best position in which the particle i has been so far.
- $x_i(t)$: position of the particle i at time t.
- c_2: social coefficient.
- r_2: vector of random values between 0 and 1 of length equal to that of the velocity vector.
- $g(t)$: position of the whole swarm at time t, that is, the best overall value.

To understand how this equation relates to the motion of the particle, it is useful to differentiate three parts:

- $\omega v_i(t)$ is the inertia constituent, in charge of remaining the particle advancing in the direction in which it has been doing so far. The recommended value of the inertia constituent ω is usually ranging from 0.8 to 1.2. If $\omega < 1$, the particle slows down as the iterations progress. This means less exploration but a faster convergence towards the optimum. If $\omega > 1$, the particle is accelerating, allowing more areas of the function space to be explored, but this makes convergence difficult.
- $c_1 r_1 [\hat{x}_i(t) - x_i(t)]$ is the cognitive constituent, responsible for the particle tendency to advance towards the position where it has obtained better results so far. The cognitive coefficient c_1 is usually bounded in the range [0, 2], where 2 is the recommended value. r_1 is a vector of random values between 0 and 1 (one value for each dimension) that provides some stochastic behavior to the movement of the particles, thus improving the ability to escape from local minima.
- $c_2 r_2 [g(t) - x_i(t)]$ is the social constituent, responsible for the particle tending to advance towards the best position found by the flock up until now. It can be interpreted as *collective knowledge*. The value of the social coefficient c_2 is usually limited in the range [0, 2], where 2 is the recommended value. r_2 is a vector of random values between 0 and 1 (one value for each dimension) that provides some stochastic behavior to the movement of the particles, thus improving the ability to escape from local minima.

The relative magnitude between the cognitive component and the social component allows regulating the exploratory behaviour of the algorithm. The higher the value of c_1 with respect to c_2, the greater independence of movement each particle has, which allows greater exploration but slower convergence. On the contrary, the higher the value of c_2 with respect to c_1, the more obliged the particles are to advance towards the most excellent zone found up until now, which reduces exploration but accelerates convergence.

Once the new velocity is calculated, the position of the particle can be updated with the expression:

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (25)$$

Finally, the parameters ω, c_1 and c_2 must be selected and control the behaviour and efficacy of the PSO method. Moreover, the Standard PSO 2011 [34] has been employed in this work. It considers some betterments from the point of view of its implementation [20]–[22], [33], [34] and besides the PSO parameters are calculated according to the next expressions:

$$\omega = \frac{1}{2 \ln 2} \quad \text{and} \quad \varphi_p = \varphi_g = 0.5 + \ln 2 \qquad (26)$$

3. Accuracy of this Approach

Seven input variables already described previously in subsection 2.1 were employed in this study to build this novel PSO/SVM–relied method applying a regularized method [8]. As it is also known, the hard chrome layer thickness (HCLT) is the dependent variable that we want to foretell. In order to forecast HCLT from the seven remaining input variables with sufficient security, it is mandatory to select the best model fitted to the observed dataset. Although several possible statistics can be used to ascertain the goodness–of–fit, the rule used in this study was the coefficient of determination R^2 [8], [35]. The coefficient of determination is a statistic used in the context of a statistical model to foretell future results or to test a hypothesis. Next, we will call the observed values t_i versus the values predicted by the model y_i. Then, it is possible to define the following sums of squares given by [35]:

- $SS_{tot} = \sum_{i=1}^{n} \left(t_i - \bar{t} \right)^2$: is the overall sum of squares, proportional to the sample variance.
- $SS_{reg} = \sum_{i=1}^{n} \left(y_i - \bar{t} \right)^2$: is the regression sum of squares, also termed the explained sum of squares.
- $SS_{err} = \sum_{i=1}^{n} \left(t_i - y_i \right)^2$: is the residual sum of squares.

where $\bar{t}$ is the mean of the n observed data:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i \qquad (27)$$

Considering the former sums, the coefficient of determination is given by the following equation:

$$R^2 \equiv 1 - \frac{SS_{err}}{SS_{tot}} \qquad (28)$$

The closer the R^2 statistic is to the value 1.0, the smaller the difference between the observed and predicted data.

Two additional criteria considered in this study were the root mean square error (RMSE) and mean absolute error (MAE) [35]. These statistics are also used frequently to evaluate the forecasting capability of a mathematical model. Indeed, the root mean square error (RMSE) and mean absolute error (MAE) are given by the expressions [8], [35]:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(t_i - y_i \right)^2}{n}} \qquad (29)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| t_i - y_i \right|}{n} \qquad (30)$$

If the root mean square error (RMSE) has a value of zero, it means that there is no difference between the predicted and observed data. Mean Absolute Error (MAE) is the average vertical distance between each point and the identity line. MAE is also the average horizontal distance between each point and the identity line. MAE has a clear interpretation as the average absolute difference between t_i and y_i.

III. Results and Discussion

The hybrid model built here used as input variables seven distinct input operator parameters. All these variables were previously given in Table I. The hard chrome layer thickness (HCLT) is taken as output variable of this suggested new hybrid PSO/SVM–relied model. The forecast carried out from the seven self-reliant variables [8], [26] was appropriate as it was already exposed before, the SVM approximation is affected by the assortment of the SVM hyperparameters such as the regularization constant C, the width ε of the insensitive tube, and the parameters a, b and σ which condition the shape of the distinct kernels in the ultimate model. Usually, the traditional way of performing hyperparameter optimization in most computational codes has been *grid search*, or a *parameter sweep*, which is simply an exhaustive searching through a manually specified subset of the hyperparameter space of a learning algorithm. Indeed, the grid search is a *brute force* method and, as such, almost any optimization method improves its efficiency [10], [12], [30], [32].

The determination of the SVM optimal parameters was carried out assisted by the metaheuristic optimizer termed as *particle swarm optimizer* (PSO) [20]–[22], [34] which showed a good performance. In this sense, we have built a novel hybrid model called as PSO/SVM–relied model, which utilizes as output (dependent) variable HCLT in the hard chromium plating process [4].

For the purpose of studying variables effect, the coefficient of determination was calculated. The flowchart indicating the way of implementation of the present novel model is made visible in Fig. 5.
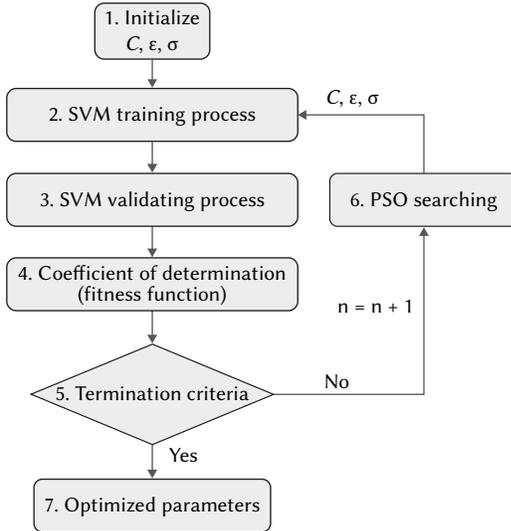
Fig. 5. Flowchart of the novel hybrid PSO/SVM–relied model.

Next, Fig. 6 makes visible those terms that make up of the best-fitted PSO/SVM–relied model for prediction of the hard chrome layer thickness (HCLT).

This graphical chart permits us to know the relevance of the distinct input variables entailed in the model. From the previous chart, Fig. 6 (a) shows the hard chrome layer thickness (Y–axis) as a function of hard chromium process time (X–axis), keeping as constant the six

remaining input variables. Similarly, Fig. 6 (d) shows the hard chrome layer thickness as a function of hard chromium process time and the thickness of the layer removed by electropolishing, keeping constant the five remaining input variables.
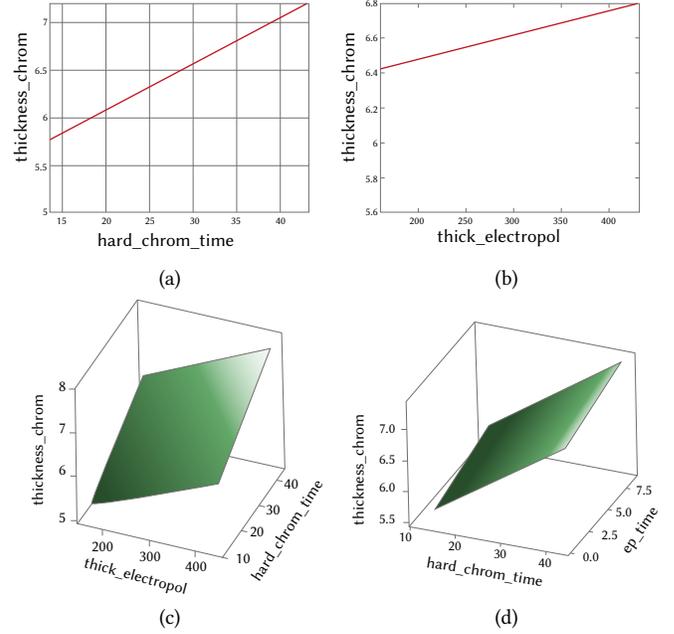
Fig. 6. Graphical chart of the terms that make up the hybrid PSO/SVM model for the hard chrome layer thickness (HCLT): (a) hard chromium process time first-order term; (b) layer thickness removed by electropolishing first-order term; (c) electropolishing time first-order term; (d) hard chromium process time and layer thickness removed by electropolishing second-order term; (e) hard chromium process time and electropolishing time second-order term; and (f) layer thickness removed by electropolishing and electropolishing time second-order term.

The value of R^2 was calculated assisted by the K–fold cross-validating procedure [36]. The mean of the term R^2 values obtained is taken into account to be the cross-validating R^2. The LIBSVM library of the statistical software R [37], along with the PSO technique utilizing the standard PSO 2011 form [34], were employed to build the ultimate regression model. Table II indicates the ranges of the space of solutions utilized for the metaheuristic PSO optimizer.

TABLE II. Intervals of the Solutions Space for Each of the RBF–SVM Parameters in the PSO Tuning Procedure

SVR parameters	Lower limit	Upper limit
C	10^{-4}	10^1
ε	10^{-6}	10^0
σ	10^{-4}	10^0

Hence, the PSO algorithm was employed to find the optimal SVM parameters utilizing the cross-validation error in each interaction. The particles $\mathbf{x}_i$ are represented by vectors that include the parameters to tune: for example, $\mathbf{x}_i = (C_i, \varepsilon_i, \sigma_i)$ for the RBF kernel employed here. Moreover, we have employed 20 particles in this work. Next, we initialized them randomly in the first iteration. Then, the particles for the next iterations were calculated according to the PSO algorithm. The value of the objective function for the particles was calculated in each stage. In this sense, the objective function value was computed as the cross-validation coefficient of determination for each particle. When the stopping criteria were fulfilled, the global best $\mathbf{x}_i$ included the optimized parameters. Additionally, the search space is three-dimensional and Table III points out the optimal parameters of the

best-fitted PSO/SVM–relied model with RBF kernel encountered with the PSO optimizer.

TABLE III. Optimal Parameters of the Best-fitted RBF–SVM Model Encountered With the PSO Optimizer

RBF/SVR hyperparameters	Optimal values
C	0.99900×10^1
ε	2.972757×10^{-3}
σ	6.396450×10^{-1}

Moreover, Table IV shows the determination and correlation coefficients for the novel PSO/SVM–relied approach with RBF kernel adjusted for the hard chrome layer thickness.

TABLE IV. Coefficient of Determination (R^2), Correlation Coefficient (R), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the Novel PSO/SVM–relied Model With RBF Kernel Adjusted for the Hard Chrome Layer Thickness

Model	R^2/ R	RMSE	MAE
PSO/RBF–SVM	0.9952/0.9975	0.01571	0.006050

Taking into account the outcomes accomplished, the SVM technique together with the PSO optimizer is able to build a novel model with a high efficiency for the forecasting of the hard chrome layer thickness. The R^2 of the fitted SVM model turned out to be 0.9952 and a correlation coefficient of 0.9975. A computer with a CPU Intel Core i7-4770 @ 3.40 GHz with eight cores and 15.5 GB RAM memory was used, taking 333 seconds (approximately 6 min) to obtain the hard chrome layer thickness (HCLT) model.

Table V indicates the weights of the seven input variables entailed in the best-fitted PSO/SVM–relied model with RBF kernel for the hard chrome layer thickness (HCLT) forecast. These weights in absolute value indicate the relevance of the distinct independent input variables within the model. Indeed, the greater the weight in absolute value, the more significant is the variable. According to these rules, hard chromium process time is the most significant input variable in the HCLT forecast. The next input variable in relevance is the layer thickness removed by electropolishing. Similarly, the electropolishing time and chromic acid content are the third and fourth most significant variables in the HCLT forecast, respectively. Then, the iron content of the electropolishing bath holds the fifth place in relevance and, finally, the two last places are occupied by electropolishing bath temperature and hard chrome bath temperature (see Table V and Fig. 7).

TABLE V. Relevance Ranking for the input Variables Entailed in the Best-fitted PSO/SVM–relied Model With RBF Kernel for the Hard Chrome Layer Thickness (HCLT) forecast

Input variable	Weight
hard_chrom_time	0.9683
thick_electropol	0.9129
ep_time	0.2493
chrom_acid	0.2327
iron_ep	0.2025
hard_chrom_temp	0.1246
ep_temp	0.1220

Furthermore, the model demonstrated its ability to forecast the HCLT. Fig. 8 makes a comparison between the observed and predicted values of the HCTL by employing this novel hybrid PSO/SVM–relied model. Therefore, a convenient way to overcome the difficulties due to the problem complexity as a result of nonlinearities is the use of this PSO/SVM–relied model.
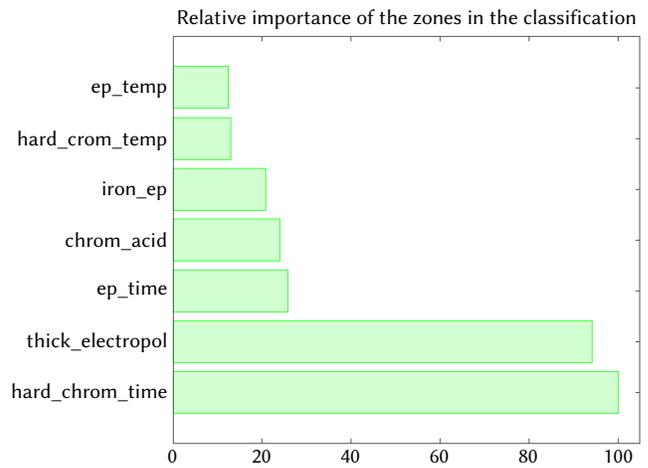
Fig. 7. Relative relevance of the input operation variables to forecast the hard chrome layer thickness (HCLT).
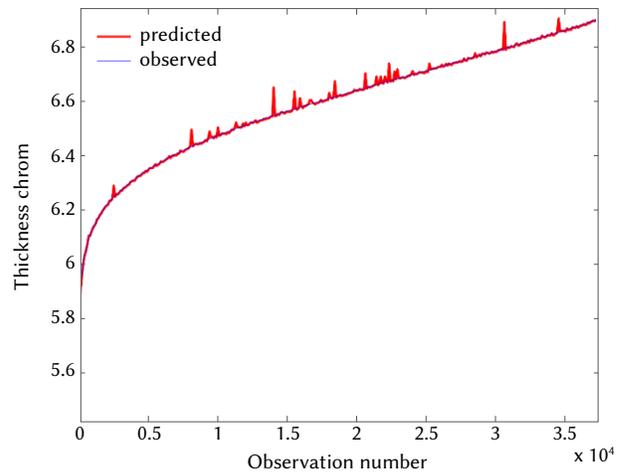
Fig. 8. Comparison between the HCTL values observed and predicted by employing the PSO/SVM–relied model (R^2 = 0.9952).

In summary, this novel hybrid model exposed here permits to assess the HCTL conforming to the actual experimental values observed employing the PSO/SVM–relied model with enough precision and success. Therefore, it is very appropriate the use of a RBF–SVM model with a PSO–relied optimizer for the purpose of accomplishing the best effective approximation in this regression problem. Obviously, these results have concordance again with the statistical criterion based on 'the goodness-of-fit' (R^2), since the PSO/SVM–relied model with RBF kernel achieves an outstanding fitting.

IV. Conclusions

Relied on the former results, several core discoveries of this study can be drawn and indicated as follows:

- A PSO/SVM–relied model is an accurate tool in order to predict the hard chrome layer thickness (HCLT).

- In this study, we have predicted the HCLT from the measured seven independent variables. This type of data-driven models are serviceable for industry for the purpose of decreasing costs in the setup of new industrial processes.

- The predicted HCLT values match with the observed ones since this novel PSO/SVM model gives place to a high coefficient of determination equal to 0.9952.

- The assessment of the input variables relevance in this industrial process was possible. Indeed, the hard chromium process time and the variable layer thickness removed by electropolishing are the most significant variables.

- The PSO/SVM–relied model used the LIBSVM library of the free software environment for statistical computing R [37] in combination with the PSO optimizer employing the standard PSO 2011 form [34].

- According to the results obtained it has been tested that PSO/SVM–relied model with RBF kernel improves the generalization capability reached employing only the standard support vector machines (SVMs) for regression.

To sum up the current research, it can be affirmed that this new approximation can be put into practice in other similar plating industrial processes. From our point of view and to conclude, this effective PSO/SVM–relied approach with RBF kernel is a convenient and pragmatic solution to the problem of the forecasting of the hard chrome layer thickness obtained from the hard-chromium plating industrial process.

Acknowledgment

References

[1] C. G. Fink, U.S. Patent 1,581,188, April 20, 1926.

[2] J. K. Dennis, T. E. Such, *The nickel and chromium plating*, New York, USA: Woodhead Publishing, 1994.

[3] F. Loffler, "Methods to investigate mechanical properties of coatings," *Thin Solid Films*, vol. 339, no. (1–2), pp. 181–186, 1999.

[4] R. K. Guffie, *The handbook of chromium plating*, New York, USA: Gardner Publications Ltd, 1986.

[5] J. H. Lindsay, "Decorative and hard chromium plating," *Plating and Surface Finishing*, vol. 90, pp. 22–24, 2003.

[6] M. Schlesinger, M. Paunovich, *Modern electroplating*, New York, USA: Wiley-Interscience, 2000.

[7] B. Irving, R. Knight, R. W. Smith, "The HVOF process - The hottest topic in the thermal spray industry," *Welding Journal*, vol. 72, no. 7, pp. 25–30, 1993.

[8] M. Kuhn, K. Johnson, *Applied predictive modeling*, New York, USA: Springer, 2013.

[9] B. Schölkopf, A. J. Smola, R. Williamson, P. Bartlett, "New support vector algorithms," *Neural Computation.*, vol. 12, no. 5, pp. 1207–1245, 2000.

[10] T. Hansen, C. J. Wang, "Support vector based battery state of charge estimator," *Journal of Power Sources*, vol. 141, pp. 351–358, 2005.

[11] X. Li, D. Lord, Y. Zhang, Y. Xie, "Predicting motor vehicle crashes using Support Vector Machine models," *Accident Analysis and Prevention*, vol. 40, pp. 1611–1618, 2008.

[12] I. Steinwart, A. Christmann, *Support vector machines*, New York, USA: Springer, 2008.

[13] C. Ordóñez, F. Sánchez Lasheras, J. Roca-Pardiñas, F. J. de Cos Juez, F. J., "A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines". *Journal of Computational and Applied Mathematics*, vol. 346, pp. 184–191, 2019.

[14] V. H. Quej, J. Almorox, J. A. Arnaldo, L. Saito, "ANFIS, SVM and ANN soft–computing techniques to estimate daily global solar radiation in a warm sub-humid environment," *Journal of Atmospherical and Solar-Terrestrial Physics*, vol. 155, pp. 62–70, 2017.

[15] F. J. de Cos Juez, F. Sánchez Lasheras, N. Roqueñí, J. Osborn, "An ANN-Based Smart Tomographic Reconstructor in a Dynamic Environment", *Sensors*, vol. 12, no. 7, pp. 8895–8911.

[16] A. Suárez Sánchez, P. J. García Nieto, P. Riesgo Fernández, J. J. del Coz Díaz, F. J. Iglesias–Rodríguez, "Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain)," *Mathematical and Computer Modelling*, vol. 54, no. (5–6), pp. 1453–1466, 2011.

[17] P. J. García Nieto, E. F. Combarro, J. J. del Coz Díaz, E. Montañés, "A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study," *Applied Mathematics and Computation*, vol. 219, no. 17, pp. 8923–8937, 2013.

[18] P. J. García Nieto, J. R. Alonso Fernández, F. Sánchez Lasheras, F. J. de Cos Juez, C. Díaz Muñiz, "A new improved study of cyanotoxins presence from experimental cyanobacteria concentrations in the Trasona reservoir (Northern Spain) using the MARS technique", *Sciences of the Total Environment*, 430, pp. 88–92, 2012.

[19] X. Xu, Y. Liu, S. Liu, J. Li, G. Guo, K. Smith, "Real-time detection of potable-reclaimed water pipe cross-connection events by conventional water quality sensors using machine learning methods," *Journal of Environmental Management*, vol. 238, pp. 201–209, 2019.

[20] R. C. Eberhart, Y. Shi, J. Kennedy, *Swarm intelligence*, San Francisco, USA: Morgan Kaufmann, 2001.

[21] M. Clerc, *Particle swarm optimization*, London, UK: Wiley-ISTE, 2006.

[22] K. E. Parsopoulos, M. N. Vrahatis, *Particle swarm optimization and intelligence: advances and applications*, Hershey, USA: IGI Global, 2010.

[23] B. K. Panigrahi, Y. Shi, M.–H. Lim, *Handbook of swarm intelligence: concepts, principles and applications*, Berlin: Springer, 2011.

[24] D. Simon, *Evolutionary optimization algorithms*, New York, USA: Wiley, 2013.

[25] X.–S. Yang, Z. Cui, R. Xiao, A. H. Gandomi, M. Karamanoglu, *Swarm intelligence and bio-inspired computation: theory and applications*, London, UK: Elsevier, 2013.

[26] V. Vapnik, *Statistical learning theory*, New York, USA: Wiley-Interscience, 1998.

[27] M. Pal, A. Goel, "Estimation of discharge and end depth in trapezoidal channel by support vector machines," *Water Resources Management*, vol. 21, no. 10, pp. 1763–1780, 2007.

[28] J.–L. Chen, G.–S. Li, S.–J. Wu, "Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration," *Energy Conversion and Management*, vol. 75, pp. 311–318, 2013.

[29] M. Abbaszadeh, A. Hezarkhani, S. Soltani–Mohammadi, "Proposing drilling locations based on the 3D modeling results of fluid inclusion data using the support vector regression method," *Journal of Geochemical Exploration*, vol. 165, pp. 23–34, 2016.

[30] N. Cristianini, J. Shawe–Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge, UK: Cambridge University Press, 2000.

[31] T. Gu, W. Lu, X. Bao, N. Chen, "Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors," *Solid State Sciences*, vol. 8, pp. 129–136, 2006.

[32] J. Shawe–Taylor, N. Cristianini, *Kernel methods for pattern analysis*, Cambridge, UK: Cambridge University Press, 2004.

[33] J. Kennedy, R. Eberhart, "Particle Swarm Optimization," in *Proceedings of the Fourth IEEE International Conference on Neural Networks, Perth, Australia, IEEE Publishing*, 1995, vol. 4, pp. 1942–1948.

[34] M. Clerc, "Standard particle swarm optimisation: from 2006 to 2011," Technical Report, Accessed: Nov. 5, 2015. [Online]. Available: http://clerc.maurice.free.fr/pso/SPSO_descriptions.pdf/.

[35] D. Freedman, R. Pisani, R. Purves, *Statistics*, New York, USA: WW Norton & Company, 2007.

[36] R. Picard, D. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 575–583, 1984.

[37] C.–C. Chang, C.–J. Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 1–27, 2011.

Paulino José García Nieto

Paulino J. García Nieto was born in Oviedo, Spain, in1965. He received the Bachelor's degree in mining engineering with specialization in fuels and energy and the Master's and Ph.D. degrees in mining engineering from the University of Oviedo, in 1989, 1990, and 1994, respectively. Since 1996, he has been with the Department of Mathematics, University of Oviedo, where he is currently a Full Professor. He was involved in research on numerical simulation, as well as the application of finite-element methods in numerous physical and engineering problems. He is currently involved in the application of statistical learning and data mining to several biological, electrical, and physical data sets to obtain predictive models.

Esperanza García-Gonzalo

Esperanza García Gonzalo was born in Soria, Spain, in 1964. She received the M.Sc. degree in mining engineering in 1990 and the Ph.D. degree in mathematics from Oviedo University, Asturias, Spain, in 2010. She is an associate professor at the Mathematics Department (Oviedo University. Her research interests include global stochastic optimization methods, evolutionary algorithms, particle swarm optimization and machine learning.

Fernando Sánchez Lasheras

Fernando Sánchez Lasheras was born in Oviedo, Spain, in 1975. He received the M.Sc. and Ph.D. degrees in industrial engineering from the University of Oviedo, in 2000 and 2008, respectively. In 2017, he joined the Department of Mathematics of Oviedo University. His current research interests include applied mathematics, machine learning and artificial intelligence, with about a hundred of technical publications.

Antonio Bernardo Sánchez

Antonio Bernardo Sánchez was born in Oviedo, Spain, in 1961. He received the M.Sc. degree in mining engineering in 1988 and the Ph.D. from Oviedo University, Asturias, Spain, in 1994. Since 1990 to 2014, he was an associate professor at the Construction and Manufacturing Engineering Department (Oviedo University). Since 2014, he is an associate professor at the Department of Mining Technology, Topography and Structures, University of León, (Spain). His research interests include energy storage, mining technology, materials and the application of numerical simulation and data mining in several engineering problems.

Rumour Source Detection Using Game Theory

Minni Jain*, Aman Jaswani, Ankita Mehra, Laqshay Mudgal

Delhi Technological University, Delhi (India)

Abstract

Social networks have become a critical part of our lives as they enable us to interact with a lot of people. These networks have become the main sources for creating, sharing and also extracting information regarding various subjects. But all this information may not be true and may contain a lot of unverified rumours that have the potential of spreading incorrect information to the masses, which may even lead to situations of widespread panic. Thus, it is of great importance to identify those nodes and edges that play a crucial role in a network in order to find the most influential sources of rumour spreading. Generally, the basic idea is to classify the nodes and edges in a network with the highest criticality. Most of the existing work regarding the same focuses on using simple centrality measures which focus on the individual contribution of a node in a network. Game-theoretic approaches such as Shapley Value (SV) algorithms suggest that individual marginal contribution should be measured for a given player as the weighted average marginal increase in the yield of any coalition that this player might join. For our experiment, we have played five SV-based games to find the top 10 most influential nodes on three network datasets (Enron, USAir97 and Les Misérables). We have compared our results to the ones obtained by using primitive centrality measures. Our results show that SV-based approach is better at understanding the marginal contribution, and therefore the actual influence, of each node to the entire network.

I. Introduction

RUMOUR Source Detection (RSD) aims to identify the most powerful nodes that are the primary sources of rumour propagation within a network. Social networking has become a modern tool for people to connect and spread the news with the development of science and technology. Diffusion of information in a social network can occur at lightning speeds and more often than not, this is considered a boon when it comes to relevant and correct information being spread. But at the same time, these networks can also be used to spread false or unverified information, either deliberately or by mistake. Therefore, rumours spread quickly and widely, and they have a great power of destruction. It is therefore of great theoretical and practical importance to decide whether there is an influential spreader and to recognize who is the influential spreader in the process for prevention and control of rumour propagation. This task is considered to be challenging due to the high speed of diffusion of information, and also because of the continuously evolving and dynamic nature of these social networks.

The most common approaches to finding the most influential node used in the past include single centrality [1] and group centrality [2] measures. The four major centrality measures are as follows. First, Degree Centrality (DC) refers to the number of associations that a node has with other nodes in a network. For an undirected graph, it is taken equal to the number of nodes to which a node is directly connected. For a directed graph, we need to compute the in-degree as well as the out-degree for each node. Second, Eigen-Vector Centrality

(EVC) considers the relative power or significance of the nodes. Here, each node is assigned a value representing its relative significance considering the fact that nodes which are connected to high-power nodes have a stronger influence over the network in comparison to those which are connected to low-power nodes. Third, Betweenness Centrality (BC) measures how strongly two nodes are connected via a given node. It is estimated as the ratio of the aggregate of shortest distances between any two nodes in the network, on which the node lies, to the shortest path between the two nodes considered. Finally, Closeness Centrality (CC) measures how quickly rumour can be spread from one node to all the other nodes in a network. It is measured as the inverse of the total sum of all shortest path distances between a given node and all other nodes in a network. For more insight into centrality measures along with mathematical derivations, refer to [3].

But these measures have a lot of disadvantages as different measures are based on different concepts and emphasize upon different topological properties of the network. For instance, DC gives the same weight to all the neighbours of a node when computing its importance. It would be more intuitive to give higher weights to nodes that are themselves important. In EVC, most of the weights get concentrated in a relatively smaller subgraph and therefore, all nodes are not quantified as they should be [4]. The remaining measures do not tend to capture the flow of information in the graph. Moreover, single centrality measures suffer from an inevitable disadvantage due to the failure to recognize the effects when considered in groups on node functionality. Group centrality measures were created to overcome this barrier and place great focus on operating in groups of nodes and not on their individual functionalities. Nonetheless, group centrality also suffers from a drawback as it focuses on a-priori-determined node groups and contributes to confusion when prioritizing individual nodes within the network.

* Corresponding author.

E-mail address: minnijain@dtu.ac.in

We aimed to work on game-theoretical algorithms to explore different strategies and metrics to assess the root cause of the rumour spread. Game-Theory is a significant paradigm that finds its applications in various fields. It is used in statistics and business analytics for prototyping the interactivity among participating agents [5]. Game-Theory has helped us to improve our presentiment, allowing for a logical analysis of various ideas which can be implemented in tandem with decision theory. Game-Theory has been widely used in the field of natural language processing. One of its most prominent applications is finding the most influential node within a network, which is relevant to our problem statement. We also do not face any of the above-mentioned disadvantages in this approach. Typical social network analysis cannot capture the dynamics of strategic interactions among the individuals in the network. Our proposed model is based on cooperative game-theory that solves this issue [6]. The elemental constituents of intricate interactivities in a network can be efficiently processed using a rich class of games, called influence games, as has been demonstrated in [7].

Shapley Value (SV) algorithm is a game-theoretic approach that has been explored in the past for finding the most influential nodes in a graphical network [8], [9], [10], but not for RSD problem specifically. The strategic issues in the Gale-Shapley model and its applications have been discussed in [11]. On the basis of the concept of marginal (or borderline) contribution, an important solution concept was proposed. Player i's SV, denoted by $SV_i(v)$, is equal to the weighted mean of i's borderline contributions to each coalition C, to which the player may belong.

$$SV_i(v) = \frac{1}{n!}\sum_{\pi \in \pi(I)}\{v(C_\pi(i)\cup\{i\}) - v(C_\pi(i))\} \qquad (1)$$

In (1), the aggregate count of players is given by 'n' while $\pi(I)$ gives the set of all permutations with 'n' players. This concept is based on cooperative game-theory - an aspect of game-theory which encourages players to form coalitions to maximize their yield in the game. Coalitions are gatherings of players that form the essential or fundamental elements of decision making. These are assumed to uphold cooperative conduct which makes it reasonable to view these games as a contest between alliances of participants and not between separate players. The core assumption here is that as the game proceeds, an eminent alliance or coalition comprising all participants will manifest eventually. The theory of cooperative games provides a high-level approach as it describes only the coalitions' structure, strategies and benefits. More insight into the SV algorithm and its derivation can be found in [8].

We have used SV-based centrality algorithm that is based on the key idea of a game-theoretic network which means defining a cooperative game across a network in which agents are nodes, coalitions are node groups, and coalition payoffs are defined to meet the requirements of a given application. The main contribution of our work is that we have explored the power of five different variations of the SV algorithm on various social networks that can be used for the purpose of spreading rumours.

We also used main centrality measures to identify the prominent top-k nodes to demonstrate a distinct and detailed contrast between our game-theoretical approach and the measures of prime centrality. Such a good analogy helped to portray the game-theoretical algorithm's aspects and accuracy vividly.

Section II gives a detailed study of various works done in the related field. Section III explains the datasets used and the algorithmic flow. Section IV describes the results obtained and the evaluations performed. Section V discusses the results and gives a theoretical explanation for the obtained results. Section VI concludes the research work with an insight into its future scope.

II. Related Work

One of the fundamental research discussions in the literature on network analysis is the topic of connectivity. The first to experiment to detect the primary top-k nodes were Domingos and Richardson [12]. They developed an algorithmic model to address this problem by modelling social media network as Markov random fields which mathematically characterized the probability of occurrence of an event.

Chen and Teng [1] explained that single node centrality measures are suitable for assessing individual influence in isolation while Shapley centrality assesses individuals' performance in group influence settings. Wei et. al. [2] explored the need to learn distributed vector representation for each vertex in a network. They laid emphasis on node classification and link prediction. An interesting approach to discover influential nodes in a network by formulating a target set selection problem has been discussed in [9]. Here, the problem comprises two main steps – the first step deals with finding a set of 'k' key nodes that would maximize the number of nodes being influenced in the network, while the second step is based on the λ-coverage problem.

We further investigated various kinds of centrality measures used for finding the most influential nodes in a network. DC, discussed by Gao et. al. [13], is used to efficiently measure the significance of nodes. However, it suffers from a severe disadvantage which is that it does not take into consideration the overall, detailed anatomy of the network. EVC, according to Stephenson and Zelen [14], overcomes the defects associated with DC. It takes into account the influence of neighbours of the node in consideration. BC, as explored by Freeman [15], learns topology-related data of networks in advance. Al-Garadi et. al. [16] describes how CC can be efficiently used to identify multiple influential spreaders. We also investigated the disadvantages associated with using centrality measures to find the most influential node in [1], [17], [18], which have been discussed in section I.

An attempt has been made to find the most influential node in a network using mapping entropy (ME) that reflects the correlation between a node and its neighbours [18]. We particularly inspected the application of ME using ENRON email dataset which is commonly used for the study of social networks [19]. ME recognizes the significance of a node in a complex network based on the knowledge of degree of the node and degrees of its neighbours. This technique for network attack helps to identify the node to attack, thereby saving valuable resources. However, the game-theoretic approach, that has been proposed, is able to capture and take into account the interactivity and dynamics of strategic interactions in a network, not only with immediate neighbours, but also with a larger subset of relationships in the graphs. Thus, we chose an SV-based algorithm to find the most influential nodes in a cooperative game.

Previous research by Tan et. al. [20] on spreading rumours focused primarily on communities' viral epidemics. The normal (and somewhat standard) model for viral epidemics is called the restored or SIR model that is susceptible-infected- recovered. There are three types of nodes in a typical rumour propagation model: i) vulnerable nodes capable of infection, ii) infected nodes capable of further spreading the virus, and iii) recovered nodes that are healed and no longer capable of infection. The most influential spreaders of rumour are identified. Various methods have been defined for the same including weighted k-core decomposition method [15] and rumour centrality with a mass centre technique [20]. An advanced form of this model, called the SEIR model, was also studied. Zhou et. al. [21] considered the graph topology and observed snapshots in a network to identify the single rumour source by formulating the nodes in a network into four possible states: susceptible (S), exposed (E), infected (I), and recovered (R).

We studied about Explosion-Trust (ET) Game Model by referring to [22]. It remarkably explains how a rumour spreading model can be constructed using game-theory by considering two very significant factors – rumour explosion degree and trust degree of the source node. In [23], a unique Belief-Propagation-based (BP) algorithm has been discussed that computes the joint likelihood function of the source location and the spreading time for the general continuous-time to detect the rumour source in a network. In [24], the concept of influence maximization has been explained from a game-theoretical perspective. A Coordination Game (CG) model, in which every individual node makes its decision based on the benefit of coordination with its network neighbours, has been proposed. SV or other game-theory solution theories can be applied to other network-related issues as well, for example, to the cost allocation problem in the electric market transmission system, and for each application, the mathematical aspects of the problem should then be addressed.

The original SV algorithms that have been implemented using Monte-Carlo simulations in the past are computationally expensive and may not arrive at an exact answer. Michalak et. al. [8] developed approximate analytical formulas for these simulations that run in polynomial time. They discuss five characteristic functions, each of which tries to convey a certain centrality concept. We have taken inspiration from their work and worked with five SV games that focus upon one characteristic function each. Furthermore, to show the comparisons of our work with existing literature, we have taken the works of Qiao et. al. [25], Hardin et. al. [26] and Munjal et. al. [27]. We found very few works that list out the top 10 most influential nodes on one of the datasets that we used in our study, with the help of primitive centrality measures. Hence, we have used these three works for our comparative study. Qiao et. al. [26] explored an entropy-based centrality measure along with the primitive centrality measures and tested it on the USAir97 dataset [28]. Hardin et. al. [26] studied the relationships in the Enron dataset [29], [30] using six centrality measures. Finally, Munjal et. al. [27] found the most influential nodes from the Les Misérables dataset [31]. We have performed our experiments on these three datasets and compared the top 10 most influential nodes obtained by using our five SV games, with the top nodes listed in these works. More details about the datasets used are given in section III.A.

III. Proposed Method For RSD

Section III.A explains the datasets used and their importance. Section III.B explains the algorithmic flow used in detail.

A. Datasets

This section gives an elaborate description of the datasets that have been used for our implementation. For our experiments, we required undirected, positive weighted-graphs that could be expressed as social networks, the top 10 most influential nodes of which were already known (so that we could compare our results with these already known influential nodes). We have used three major datasets which satisfy these criteria and they have been described below.

Unweighted Graph

An unweighted graph can be technically defined as a graph $G(N, E)$ having 'n' nodes represented by set N and 'e' edges represented by set E consisting of unordered pairs, such that $(n_1, n_2) = (n_2, n_1)$ and $(n_1, n_2) \in E$ and $n_1, n_2 \in N$. Games 1 and 2 are played by creating an unweighted network from the datasets.

Weighted Graph

A weighted graph can be technically defined as a graph $G(N, E)$ having 'n' nodes represented by set N and 'e' edges represented by set E consisting of ordered pairs, such that $(n_1, n_2) \neq (n_2, n_1)$ and $(n_1, n_2) \in E$ and $n_1, n_2 \in N$. Games 3 – 5 are played by creating a weighted network from the datasets.

1. Enron Dataset

The CALO Project (A Cognitive Assistant that Learns and Organizes) compiled and planned this dataset [29], [30]. It contains data from about 150 users, belonging to the Enron organization, grouped into files, mainly senior Enron executives. There are a total of about 0.5 M messages in the corpus. We used a subset of this dataset, containing 143 nodes (people from the Enron organization) and 1800 edges (an edge exists between two people if they have communicated with each other via email). Edges are weighted with the frequency of email exchanges between two users. This dataset can act as a social network which can be used to spread rumours within the members of the organization. Hence, we can identify the important nodes and assign them labels that symbolize their relative network value. This dataset has been commonly used for the study of social networks as well as for finding the most influential nodes [19], [26], [32], and so we have compared the results of our algorithm with other studies that used the same dataset [26].

2. Les Misérables

This is a co-occurrence graph for the characters that appear in the novel 'Les Misérables' by Victor Hugo [31]. This dataset consists of 77 nodes and 254 edges where a node represents a character and an edge between two nodes shows that these two characters appear in the same chapter of the book. The weight of each link indicates how often such a co-appearance occurs. This dataset too can act as a social network for the spread of a rumour. We have compared the results obtained by our SV-based approach with those obtained by various centrality measures used in [27].

3. USAir97

USAir97 dataset [28] has been transformed into an undirected network, created by 332 nodes, where one airport represents a node, and 2126 edges, with each edge reflecting a direct airline between two American airports if any. Here, weights represent the normalized distance between two airports. This dataset is not particularly useful for the purpose of rumour spreading but due to lack of supervised datasets with their most influential nodes known to us, we have included this dataset to test the results of our approach with the most influential nodes obtained by various centrality measures, as in [25].

B. Algorithm

Focusing on Game-Theory's Shapley algorithm, we referred to the algorithms described in Michalak and Szczepański's work [8]. In both weighted and unweighted networks, the exact analytical formulae for SV-based centrality were established. The SV-based centrality polynomial-time algorithms have been developed.

1. Creation of Weighted and Un-Weighted Network Graphs

Graphs were created by using the *networkx* library in Python for all three datasets. Games 1 and 2 require unweighted graphs whereas the remaining games require weighted graphs.

2. Coalition Games Based on Shapley Algorithm

SV is the average marginal cost contribution across all potential coalitions of the function value. The Shapley algorithm was applied carefully and it tries to find the top-k nodes that might be the most prominent nodes.

Specifically, we concentrated on five underlying network-defined coalition games that vary in degree and centrality of the network. Each game has a certain characteristic function v(C) which represents how prominent a particular node is to a given coalition C.

For more insight into the working of these games and their underlying mathematics, refer to [8]. The game descriptions are as follows:

a) Game 1: In this game, we considered all the permutations of all the nodes that are immediately reachable, by one hop to the node $n_i \in$ N(G). Let each random permutation be denoted by P_i, the neighbours of node n_i, in the graph G(N, E) be denoted by $n_i$neighbours and the degree of node n_i, be denoted by n_i.degree.

Algorithm 1 describes the procedure involved in SV calculation.

Algorithm 1: SVs for Game 1

Input: An unweighted graph G(N, E)

Output: SVs of all nodes in G

Initialise: $\forall\, n_i \in$ N(G) set SV[n_i] ← 0

for each $n_i \in$ N(G) **do**

  SV[n_i] ← 1/(1 + n_i.degree)

  for each $u_i \in n_i$.neighbours **do**

  SV[n_i] ← SV[n_i] + $\frac{1}{(1 + u_i.\text{degree})}$

  end for

end for

return SV

---

b) Game 2: In many real-life social scenarios, often taking into account nodes that are directly attached to each other is not enough. A rumour source will, more often than not, affect farther nodes.

For the purpose of taking relationships with farther nodes into account, and generalising the game, we introduced a value, p, depicting the number of agents that the node is adjacent to in a coalition. In this game, a node is considered 'influenced' if at least p of its neighbours are influenced. We divided the analysis using this game into two parts, first, where the degree of the node is less than p and second where the degree is more than p.

Algorithm 2 describes the procedure involved in SV calculation.

Algorithm 2: SVs for Game 2

Input: An unweighted graph G(N, E) and a positive integer p

Output: SVs of all nodes in G

Initialise: $\forall\, ni \in$ N(G) set SV[ni] ← 0

for each $ni \in$ N(G) **do**

  SV[ni] ← min(1, $\frac{p}{(1 + n_i.\text{degree})}$)

  for each $ui \in ni$.neighbours **do**

  SV[n_i] ← SV[n_i]+ max(0, $\frac{u_i.\text{degree} - p + 1}{u_i.\text{degree} * (1 + u_i.\text{degree})}$)

  end for

end for

return SV

---

c) Game 3: In this game, we introduced the concept of weighted graph networks. This game is an extension of game 1; it uses the Dijkstra Algorithm to compute the distance between 2 nodes. The cutoff value, d, is the maximum permissible distance of a node from any member in a given coalition.

The extended degree is defined as the size of the set of all nodes that are at most distance 'd' away from the node n_i.

Algorithm 3 describes the procedure involved in SV calculation.

Algorithm 3: SVs for Game 3

Input: A weighted graph G(N, E, W) and a positive cut-off value d

Output: SVs of all nodes in G

Initialise: $\forall\, n_i \in$ N(G) set SV[n_i] ← 0

for each $n_i \in$ N(G) **do**

  Distance_Vector D ← Dijkstra(n_i)

  extended_neighbours ← empty 2D array

  extended_degree[n_i] ← 0

  for each $u_i \in$ N(G) such that $u_i \neq n_i$ **do**

    if D[u_i] ≤ d **then**

    extended_neighbours[n_i].add(u_i)

    extended_degree[n_i]++

    end if

  end for

end for

for each $n_i \in$ N(G) **do**

  SV[n_i] ← $\dfrac{1}{1+\text{extended\_degree}[n_i]}$

  for each $u_i \in$ extended_neighbours[n_i] **do**

  SV[n_i] ← SV[n_i] + $\dfrac{1}{(1 + \text{extended\_degree}[u_i])}$

  end for

end for

return SV

---

d) Game 4: This is a generalization of game 3. Here we worked with the assumption that a node closer to a coalition will have a greater effect on it than some other node farther away, even if both

nodes satisfy the cut-off criteria as in game 3.

For this purpose, we introduced a positive-valued decreasing function f(x). f(d) refers to the function which has a directly proportional effect on SV of the coalition which is 'd' units away from a node.

The marginal contribution of each node n_i through node $n_i \neq n_j$, for each coalition C_i gives SV, as shown in Algorithm 4.

Algorithm 4: SVs for Game 4

Input: A weighted graph G(N,E,W) and function $f : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$

Output: SVs of all nodes in G

Initialise: $\forall\, n_i \in$ N(G) set SV[n_i] ← 0

for each $n_i \in$ N(G) **do**

  [Distance D , Nodes w] ← Dijkstra(n_i)

  sum ← 0, index ← |N|-1 , prev_dist ← -1, prevSV ← -1

  while index > 0 **do**

    if D(index) == prev_dist **then**

      currSV = prevSV

    else

      currSV = $\dfrac{f(D(index))}{1+index}$ - sum

    end if

    SV[w(index)] ← currSV + SV[w(index)]

    sum ← sum + $\dfrac{f(D(index))}{1+\text{index}}$

    prev_dist = D(index) , prevSV = currSV

    index ← index - 1

  end while

  SV[n_i] ← SV[n_i] + f(0) – sum

end for

return SV

---

e) Game 5: This is a generalization of game 2 in case of weighted networks. Here, we have defined a cut-off value (n_i) for each $n_i \in$ N(G). $d(n_j, C) = \sum_{nj \in \ ni.neighbours} W(n_i, n_j)$ for every coalition C, where $W(n_i, n_j)$ is the weight of the edge between nodes n_i and n_j (0 if no edge exists).

A node n_i marginally contributes node $n_j \in n_i$. neighbours to the value of coalition C_i if and only if $n_j \notin C_i$ and $d(n_j)$ - $W(n_i, n_j) \leq W(C_i, n_j) < d(n_j, C)$.

Algorithm 5 describes the procedure for calculating the SVs.

Algorithm 5: SVs for Game 5

Input: A weighted graph G(N, E, W) and cut-offs $Wcutoff(n_i)$ for each $n_i \in$ N(G)

Output: SVs of all nodes in G

Initialise: $\forall \ n_i \in$ N(G) set SV[n_i] $\leftarrow$ 0

for each $n_i \in$ N(G) **do**

    compute and store α_i and β_i

end for

for each $n_i \in$ N(G) **do**

    for each m in 0 to n_i. degree **do**

        compute $\mu \leftarrow \mu(X^{ii}_m)$, $\sigma \leftarrow \sigma(X^{ii}_m)$

        compute $p \leftarrow$ Pr $\{\mathcal{N}(\mu, \sigma^2) < Wcutoff(n_i)\}$

        SV[n_i] $\leftarrow$ SV[n_i]$+ \dfrac{p}{1 + n_i.degree}$

    end for

    for each $v_j \in n_i$. neighbours n_i. neighbours **do**

        p $\leftarrow$ 0

        for each m in 0 to n_i. degree **do**

            compute $\mu \leftarrow \mu(X^{ij}_m)$, $\sigma \leftarrow \sigma(X^{ij}_m)$

            compute $z \leftarrow \mathcal{Z}^i_m$

            p $\leftarrow$ p $+ \dfrac{z * (v_i.degree-m)}{v_j.degree * (v_j.degree+1)}$

        end for

        SV[n_i] $\leftarrow$ SV[n_i] + p

    end for

end for

return SV

---

3. Estimating Centrality Measures

After working on the five coalition games, we introduced multiple centrality measures to determine the network's most powerful node with the highest scope or effect. To generate an elaborate comparison, various network centrality measures such as DC, EVC, BC, CC, have been used.

IV. Results

We experimented on three real-world network datasets - USAir97 dataset [28], Enron email dataset [29], [30] and Les Misérables dataset [31], and then compared the results of five coalition games defined previously, with the results obtained using the four aforementioned centrality measures. Qiao et. al. [25] has applied these centrality algorithms using the USAir97 network to assess the performance of network centrality model. Table I accurately shows for USAir97 dataset, the comparison between the top-k (k=10) nodes identified by our model for all the five coalition games and those identified by various centrality models employed in [25]. Also, Table II shows for Les Misérables dataset, the comparison between the top-k (k=10) nodes identified by our model for all the five coalition games, and those identified by various centrality models employed in [25].

We observed that the number of common items between the top-10 nodes found using coalitional game 1 and those found using DC, BC, and CC measures are nine, nine and four, respectively. The most significant observation is that the top-10 nodes are the same for both the coalitional game 1 and EVC measure. For Les Misérables dataset, we observed that node 11 was recognized as the most influential node in all five coalitional games and also using DC, BC, and CC measures. We noticed an overlap of six nodes in the observations of game 3, game 5 and CC measure.

TABLE I. Comparison Between Shapley and Centrality Values Using USAIR97 Dataset

DC	BC	CC	EVC	Proposed Model				
				Game 1	Game 2	Game 3	Game 4	Game 5
118	118	118	118	118	261	261	118	118
261	8	261	261	261	118	118	261	261
255	261	67	255	255	152	152	182	67
182	201	255	182	166	182	182	152	255
152	47	201	152	152	255	255	201	201
230	182	182	230	182	230	230	255	166
166	255	47	112	230	201	201	230	293
67	152	248	67	67	8	8	8	248
112	313	166	166	147	166	166	67	47
201	13	112	147	112	67	67	166	182

TABLE II. Comparison Between Shapley and Centrality Values Using Les Misérables Dataset

DC	BC	CC	EVC	Proposed Model				
				Game 1	Game 2	Game 3	Game 4	Game 5
11	11	11	11	11	11	11	11	11
49	1	56	49	2	55	56	2	56
56	49	28	56	49	49	28	49	26
28	56	26	59	28	43	26	28	49
26	24	49	65	24	44	49	56	28
24	26	59	63	56	73	27	26	27
59	28	27	28	26	32	25	24	70
63	52	65	26	52	51	59	27	69
65	59	69	66	27	57	65	25	71
64	17	70	66	25	40	69	52	42

Similarly, we referred to the work of Hardin, Sarkis and Urc [26] to compare the efficiency of our model using Enron email dataset. Table III shows the results obtained for the same.

We observed that Philip K. Allen, the Managing Director of Trading, appeared in the results of all the coalition games. Mike Grigbsy, VP of Trading, is also an important figure who is present in the results of three of the five games. Found in results of four coalition games, Barry Tycholiz is also the VP of Trading. Another person who can be identified as a prominent figure is Director for State Government, Jeff Dasovich. Game 5 recognizes Louise Kichen – the president of Enron – as one of the most significant nodes.

To get a better numerical understanding of our results, we used a comparison metric – *The Jaccard* Index, also known as the Union Intersection and the *Jaccard* Similarity Coefficient – which is used to calculate the similarity and diversity of sample sets.

TABLE III. Comparison Of Most Important Nodes Using Enron Dataset Based On Various Centrality Measures

DC	BC	EVC	CC	Game 1	Game 2	Game 3	Game 4	Game 5
Jeff Dasovich	Louise Kitchen	Tana Jones	Robert Benson	Scott Neal	Phillip K. Allen	Kevin Presto	Phillip K. Allen	Scott Neal
Mike Grigsby	Mike Grigsby	Sara Shackleton	Mike Grigsby	Phillip K. Allen	Scott Neal	James D. Steffes	Scott Neal	Mike Grisby
Tana Jones	Susan Scott	Stephanie Panus	Louise Kitchen	Mike Grisby	Mike Grisby	Phillip K. Allen	Mike Grisby	John Arnold
Sara Shackleton	Jeff Dasovich	Marie Heard	Kevin M. Presto	Barry Tycholiz	Barry Tycholiz	Mark Haedick	Barry Tycholiz	John Lavorato
Richard Shapiro	Mary Hain	Susan Bailey	Susan Scott	Sally Beck	Sally Beck	Steven J. Kean	Sally Beck	Joe Quenet
Steven J. Kean	Sally Beck	Kay Mann	Scott Neal	John Lavorato	John Lavorato	Mike Swerzbin	John Lavorato	Phillip Allen
Louise Kitchen	Kenneth Lay	Louise Kitchen	Barry Tycholiz	Susan Scott	Mark Haedick	Jeff Dasovich	Mark Haedick	Barry Tycholiz
Susan Scott	Scott Neal	Elizabeth Sager	Greg Whalley	Kim Ward	Susan Scott	Richard Sanders	Richard Sanders	Sally Beck
Michelle Lokay	Kate Symes	Jason Williams	Phillip K. Allen	Mark Haedick	Richard Sanders	Doug Gilbert-Smith	Kim Ward	Louise Kitchen
Chris Germany	Cara Semperger	Jeff Dasovich	Jeff Dasovich	Bill Williams	Kim Ward	Richard Shapiro	Kevin Presto	David Delainey

The coefficient of Jaccard measures similarity between finite sample sets and is defined as the intersection size divided by the size of the union of sample sets which is shown in (2).

$$i(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

We compared the intersection similarity of the most significant nodes from each coalition game, with the results of the proposed model. Finally, for holistic comparison, we took the mean overall intersections, as shown in (3).

$$I_{centrality} = \frac{i_1 + i_2 + i_3 + i_4 + i_5}{5} \quad (3)$$

$I_{centrality}$ depicts the mean of all intersections between sets over the five coalition games, where *centrality* denotes the centrality model used, and *ij* represents the intersection similarity between centrality measures with game j. The results are displayed in Table IV.

TABLE IV. Jaccard Similarity Coefficient Values

MEASURE	INTER-SECTION OF SETS		
	USAIR97	ENRON EMAIL	LES MISÉRABLES
I_{dc}	0.762	0.112	0.368
I_{ec}	0.685	0.010	0.354
I_{bc}	0.475	0.156	0.389
I_{cc}	0.514	0.304	0.490

Fig. 1 shows the comparison of the *Jaccard* Indices measured with various degree centralities. I_{dc}, I_{ec}, I_{bc} and I_{cc} denote Degree Centrality, Eigen-vector Centrality, Betweenness Centrality and Closeness Centrality, respectively. USAir97 shows the maximum similarity with I_{dc}, whereas Enron Email and Les Misérables dataset show maximum similarity with Closeness Centrality. Thus we were able to show the comparison of the Shapely algorithm considering EC, DC, BC and CC as benchmarks for all three datasets.
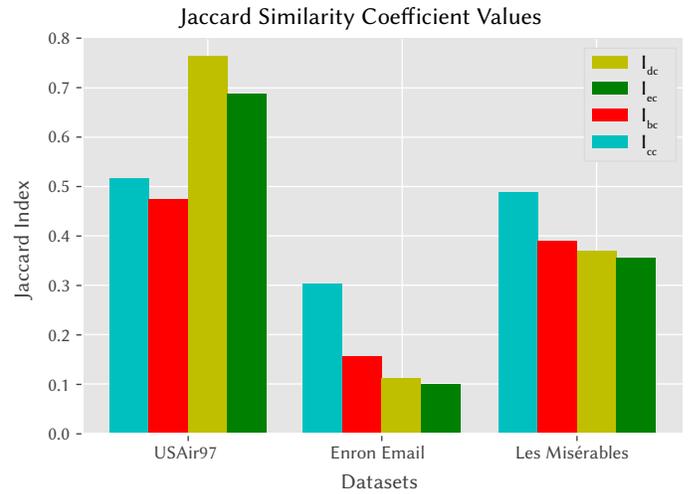
Fig. 1. Comparison of *Jaccard* Indices.

V. Discussion

We had observed many disadvantages in primitive centrality measures that had been used in the past for finding the most influential node, including putting too much focus on the individual node and not on the neighbours of the node. An elaborate description of these disadvantages is mentioned in section II. Game-theoretic approaches like the SV algorithm, take into consideration the marginal contribution of a node to every coalition that it is a part of. This approach has also not been specifically used in the past for RSD problem. For this reason, we aimed to explore the effectiveness of this approach for the purpose of RSD. Our results show a good similarity score (*Jaccard* Index) with the previous studies that used primitive centrality measures.

But as discussed, there were numerous disadvantages with these measures that our SV-based approach tried to overcome. Hence, we observe a slight difference between the most influential nodes found by our approach and those found by the earlier studies conducted on the same datasets.

VI. Conclusion

Sometimes the propagation of rumours on online social networks can lead to serious social problems. It is known to be of great value to accurately identify them from regular comments. Social media rumours have recently become a major concern, especially as people are aware of their ability to influence society. Rumours can not only cause social hysteria in all sorts of crises, but can also cause mass events that are unpredictable and threaten social stability.

We tried to introduce a game-theoretical algorithm in our research work in order to detect the origin of rumour in a complex network. The algorithm used is the algorithm of Shapley. We compared the performance of our game-theoretic approach with prime centrality measures. We also sought to locate prominent top nodes to catch and record multiple potential gossip sources, rather than concentrating discreetly on a single source. The most influential node identified is assumed to be the rumour source in the network.

To evaluate our algorithm on various real-world scenarios, we examined five different game situations, thereby taking into consideration various approaches to determine the most influential nodes in a given dataset. This helped us to gain a deeper and holistic understanding of the game-theoretical algorithm. The *Jaccard* Index has been used as a metric of comparison for our proposed method. The model has shown significant success as the most prominent nodes are successfully identified for both the datasets used.

We are currently working on expanding the theory of Shapley algorithm to consider each person's impact in a social network and thus determine the most serious cause of rumours. We plan to extend the idea of finding the most powerful node in social networks to numerous other similar applications for future work, such as the Internet, or urban networks, and involving a given node in disease dynamics. This will help us understand our algorithm's efficiency and accuracy in multiple applications in the real world.

Further, various optimisation techniques on the SV algorithm, for example, Fuzzy Logic will be implemented for mining much larger social networks and to improve accuracy and other relevant metrics of the project. Fuzzy-based implementation will solve various complexities and limitations that we are currently encountering.

Acknowledgement

References

[1] W. Chen, and S.-H. Teng, "Interplay between social influence and network centrality: a comparative study on shapley centrality and single-node-influence centrality," in *WWW '17: Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 967–976.

[2] H. Wei, Z. Pan, G. Hu, L. Zhang, H. Yang, X. Li, and X. Zhou, "Identifying influential nodes based on network representation learning in complex networks," *PLoS ONE*, vol. 13, 2018, doi: 10.1371/journal.pone.0200091.

[3] J. Zhan, S. Gurung, and S. P. K. Parsa, "Identification of top-k nodes in large networks using katz centrality," *Journal of Big Data*, vol. 4, no. 16, 2017, doi: 10.1186/s40537-017-0076-5.

[4] F. A. Rodrigues, "Network centrality: an introduction," *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*, Nonlinear Systems and Complexity, Springer, vol. 22, pp. 177-196, 2019, doi: 10.1007/978-3-319-78512-7_10.

[5] B. A. Bhuiyan, "An overview of game theory and some applications," *Philosophy and Progress*, vol. 59, no. 1-2, pp. 111–128, 2018, doi: 10.3329/pp.v59i1-2.36683.

[6] N. Yadati, and R. Narayanam, "Game theoretic models for social network analysis," in *Proceedings of the 20th International Conference Companion on World Wide Web – WWW' 11*, 2011, pp. 291-292.

[7] M. T. Irfan, and L. E. Ortiz, "A game theoretic approach to influence in networks," in *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI'11)*, AAAI Press, 2011, pp. 688-694.

[8] T. P. Michalak, K. V. Aadithya, B. Ravindran, N. R. Jennings, and P. L. Szczepanski, "Efficient computation of the shapley value for game-theoretic network centrality," *Journal of Artificial Intelligence Research*, vol. 46, no. 1, pp. 607-650, 2013, doi: 10.5555/2512538.2512553.

[9] R. Narayanam, and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," in *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 1, 2011, pp. 130–147, doi: 10.1109/tase.2010.2052042.

[10] R. Narayanam, and Y. Narahari, "Determining the top-k nodes in social networks using the shapley value," *AAMAS*, pp. 1509-1512, 2008, doi: 10.1145/1402821.1402911.

[11] C.-P. Teo, J. Sethuraman, and W.-P. Tan, "Gale-shapley stable marriage problem revisited: strategic issues and applications (extended abstract)," *Integer Programming and Combinatorial Optimization Lecture Notes in Computer Science*, vol. 1610, pp. 429-438, 1999, doi: 10.1007/3-540-48777-8_32.

[12] P. Domingos, and M. Richardson, "Mining the network value of customers," in *KDD '01: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 57–66.

[13] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Physica A: Statistical Mechanics and its Applications*, vol. 403, pp. 130-147, 2014, doi: 10.1016/j.physa.2014.02.032.

[14] K. Stephenson, and M. Zelen, "Rethinking centrality: methods and examples," *Social Networks*, vol. 11, no. 1, pp. 1-37, 1989, doi: 10.1016/0378-8733(89)90016-6.

[15] L. C. Freeman, "Centraity in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215-239, 1978-79, doi: 10.1016/0378-8733(78)90021-7.

[16] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, "Identification of influential spreaders in online social networks using interaction weighted k-core decomposition method," *Physica A: Statistical Mechanics and its Applications*, vol. 468, pp. 278-288, 2017, doi: 10.1016/j.physa.2016.11.002.

[17] H.-L. Liu, C. Ma, B.-B. Xiang, M. Tang, and H.-F. Zhang, "Identifying multiple influential spreaders based on generalized closeness centrality," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 2237-2248, 2018, doi: 10.1016/j.physa.2017.11.138.

[18] T. Nie, Z. Guo, K. Zhao, and Z.-M. Lu, "Using mapping entropy to identify node centrality in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 453, pp. 290-297, 2016, doi: 10.1016/j.physa.2016.02.009.

[19] J. Shetty, and J. Adibi, "Discovering important nodes through graph entropy: the case of enron email database," in *KDD '05: Proceedings of the 3rd International Workshop on Link Discovery*, 2005, pp. 74–81.

[20] C. W. Tan, P.-D. Yu, L. Zheng, C.-K. Lai, W. Zhang, and H.-L. Fu, "Optimal detection of influential spreaders in online social networks," in *2016 Annual Conference on Information Science and Systems (CISS)*, 2016, pp. 145-150.

[21] Y. Zhou, C. Wu, Q. Zhu, Y. Xiang, and S. Loke, "Rumour source detection in networks based on the seir model," in *IEEE Access*, vol. 7, 2019, pp. 45240-45258.

[22] F. Liu, and M. Li, "A game theory-based network spreading model: based on game experiments," *International Journal of Machine Learning and Cybernetics*, vol. 10, pp. 1449–1457, 2019, doi: 10.1007/s13042-018-0826-5.

[23] T.-H. Fan, and I.-H. Wang, "Rumor source detection: a probabilistic perspective," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 4159-4163.

[24] Y. Zhang, and Y. Zhang, "Top-k influential nodes in social networks: a game perspective," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1029–1032, doi: 10.1145/3077136.3080709.

[25] T. Qiao, W. Shan, and C. Zhou, "How to identify the most powerful node in complex networks? a novel entropy centrality approach," *Entropy*, vol. 19, pp. 614, 2017, doi: 10.3390/e19110614.

[26] J. Hardin, G. Sarkis, and P. C. Urc, "Network analysis with the enron email corpus," *Journal of Statistics Education*, vol. 23, 2015, doi: 10.1080/10691898.2015.11889734.

[27] P. Munjal, N. Arora, and H. Banati, "Dynamics of online social network based on parametric variation of relationship," in *2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, Kolkata, 2016, pp. 241-246.

[28] Bureau of Transportation Statistics. Accessed: July 10, 2020. [Online]. Available: https://transtats.bts.gov/

[29] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, "Simplicial closure and higher-order link prediction," *PNAS*, vol. 115, no. 48, pp. E11221-E11230, 2018, doi: 10.1073/pnas.1800683115.

[30] email-Enron Dataset. Accessed: Jan. 7, 2020. [Online]. Available: https://www.cs.cornell.edu/~arb/data/email-Enron/index.html

[31] Donald E. Knuth, "The stanford graphbase: a platform for combinatorial computing," Association for Computing Machinery, New York, NY, USA, 1993.

[32] H. Yang, J. Luo, Y. Liu, M. Yin, and D. Cao, "Discovering important nodes through comprehensive assessment theory on enron email database," in *3rd International Conference on Biomedical Engineering and Informatics*, Yantai, 2010, pp. 3041-3045.

Minni Jain

Minni Jain is currently working as an Assistant Professor in Delhi Technological University, Delhi, India. She obtained her M. Tech. degree in Information Security and B. Tech. degree in Information Technology. Her major research interests include Natural Language Processing, Sentiment Analysis, Information Security, Neural Networks and Fuzzy Logic.

Aman Jaswani

Aman Jaswani obtained his B. Tech. degree in Software Engineering from Delhi Technological University, Delhi, India. He has worked on numerous projects in the field of Machine Learning and Natural Language Processing. He plans to pursue a Masters Degree in a similar field. His areas of research include Game Theory, Machine Learning, Natural Language Processing and Data Science.

Ankita Mehra

Ankita Mehra obtained her B. Tech. degree in Software Engineering from Delhi Technological University, Delhi, India. She is a final year student. Her areas of research include Game Theory, Machine Learning, Deep Learning, Computer Vision and Natural Language Processing.

Laqshay Mudgal

Laqshay Mudgal obtained his B. Tech. degree in Software Engineering from Delhi Technological University, Delhi, India. His major research interests include Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, and Data Science.

Multi Layered Multi Task Marker Based Interaction in Information Rich Virtual Environments

Inam Ur Rehman[1*], Sehat Ullah[1], Dawar Khan[2,3,4]

[1] Department of CS and IT, University of Malakand, Chakdara (Pakistan)
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055 (China)
[3] Department of Information Technology, University of Haripur, Haripur (Pakistan)
[4] Interactive Media Design Lab, Nara Institute of Science and Technology, Nara 630-0192 (Japan)

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

Simple and cheap interaction has a key role in the operation and exploration of any Virtual Environment (VE). In this paper, we propose an interaction technique that provides two different ways of interaction (information and control) on complex objects in a simple and computationally cheap way. The interaction is based on the use of multiple embedded markers in a specialized manner. The proposed marker like an interaction peripheral works just like a touch paid which can perform any type of interaction in a 3D VE. The proposed marker is not only used for interaction with Augmented Reality (AR), but also with Mixed Reality. A biological virtual learning application is developed which is used for evaluation and experimentation. We conducted our experiments in two phases. First, we compared a simple VE with the proposed layered VE. Second, a comparative study is conducted between the proposed marker, a simple layered marker, and multiple single markers. We found the proposed marker with improved learning, easiness in interaction, and comparatively less task execution time. The results gave improved learning for layered VE as compared to simple VE.

Keywords

I. Introduction

REALISTIC interaction in an effective manner is essential for the exploration of the physical or behavioral characteristics of a VE along with its spatial objects [1]. VEs can be viewed as an innovative model for human computer interaction which not only allows external examination of the virtual entities but interactively involve the user in the 3D virtual space [2], [3]. The virtual environment provides a base for the integration of different VEs methodologies and information visualization which is carried out for the discovery of the association between entities, environment, and information [4]. Different studies addressed the same concepts in different areas such as Scientific visualization [5], VEs [6]-[8], Psychologies of Perception [9]-[12], and Information Visualization [2], [13]. Information Rich Virtual Environment (IRVE) is a combination of accurate 3D information with improved mental and non-spatial information [4]. The objective of an IRVE is to provide systematic multi-form representations [14] that build precise ideas [15], [16] and the cognitive representations of the system [17]. VEs are successfully applied in different areas ranging from engineering design, information visualization, and educational training [3].

Different IRVEs have been developed such as Venue [2], Habitat [15], HUD [18], etc., for the delivery of information related to the VEs in different shapes. In this regard to achieve maximum information

* Corresponding author.

E-mail address: inam.btk@gmail.com

delivery along with limited cognitive load on users, efficient human computer interaction is needed [3]. There is a need to develop interaction techniques that provide simple access to information along with control over the system [2].

AR markers have been used for different purposes in AR applications. For interaction in AR environments, fiducial markers are effectively used due to its low cost and flexible nature. Rehman et al. [19]-[22] used multiple single markers for interaction in VEs.

Teteno et al. [23] proposed a nested marker with a hierarchically structured nature. It is used for increasing the range of viewpoint movement. Rabbi et al. [24] proposed a layered marker to cover the tracking distance of large indoor AR spaces. These markers have been limited to mono functional capabilities i.e. they are unable to perform other interaction tasks such as selection, navigation, and manipulation, etc. They are unable to work with complex AR environments where multiple functionalities are desired.

In this paper, we propose a new interaction technique with two-fold functional nature i.e. realistic interaction along with information delivery in a simple manner using a newly designed Multi Layered Multi Task (MLMT) marker. It is used for selection, navigation, rotation, and scaling of virtual objects as well as the provision of information in a systematic manner. The MLMT marker is used for interaction with AR and VR. In the first phase, we used MLMT for different types of interaction i.e. visualization, scaling, and rotation of AR 3D objects. In the second phased this marker is used for interaction with complex 3D objects in VR.

Section II presents related work, section III is about MLMT marker,

System architectural model is presented in section IV, section V describes Biological VR application, section VI is about experiments and evaluation, and finally section VII is about conclusion.

II. Related Work

The use of IRVEs in learning, training, and other related fields resulted in a valuable response in these areas [16]. Bowman et al. [2], in Virtual Venue, inserted different types of audio, textual, special animation, imagery, and empirical information in the virtual environment. A comparative study was carried out between IRVE and traditional hypermedia or paper-based information system. Interaction with the system is made using a pen and tablet and hand menu. For visualization and interaction with the system, costly devices such as HMD, Joystick, and pen and tablet were used [2]. AnthroGloss, a desktop base VE was developed by [18]. The human anatomy related information was displayed in textual form in the system. For joining the perceptual information and textual labels various tags were used. In Mobile Augmented Reality Systems (MARS), information was visualized over the real environment [25]. A virtual zoo exhibit, was developed by Bowman et al. [15]. The system was used to educate students in the designing of exhibit. In the comparative study with traditional lectures, students got improved learning. Chen et al. [3] developed an immersive IRVE, where they compared two navigation techniques i.e. GoGo [26] and HOMER [26] for search, and exploration tasks. The previous systems mostly stressed the delivery of abstract information of simple virtual objects/concepts but they were unable to provide systematic delivery of information concerning complex objects.

Fiducial markers are used in various types of AR systems. Marker based tracking is carried out using different types of toolkits such as ARToolKit [28], ARToolKitPlus [29], ARTag [29], and ALVAR [30]. These toolkits use different types of markers placed in real world scenes and tracked by the AR systems. For the development of AR applications, these toolkits provide the basic framework. ARToolKit uses square shaped markers placed in the 3D space [31].

Jun et al. [32] used a large room space for the fiducial marker tracking to avoid occlusion for multiple markers. Khan et al. [33], [34] identified different factors affecting the fiducial marker tracking. Rehman et al. [19]-[22] used multiple single markers for interaction such as navigation, selection, and manipulation of objects in a virtual assembly environment and interactive writing board. Azhar et al. [35] used a single marker for interaction in a biological IRVE. Tateno

et al. [23] used a hierarchical structured nested marker to extend the range of viewpoint movement. They used four markers inside a marker and each of them was further consisted of four other markers making a total of three layers. A nested marker may lead to inter marker confusion between inner and outer layer markers. Due to limited hierarchical structure, it can cover limited tracking distance, therefore, can't be used in large indoor applications. Recently Rabbi et. al [24] proposed a layered marker for extending the tracking distance of the fiducial marker. The use of these markers is limited to single functional capabilities.

We propose MLMT marker as an interaction tool for Augmented and VEs i.e. a single marker with multiple functional capabilities at the same time. We also propose a new interaction technique for IRVEs which provide textual information related to complex objects having multiple sub-parts in a simple, easy, and interest-oriented manner. The information delivery is based on an interest-based step by step/layered manner.

III. MLMT MARKER

The newly designed MLMT marker (see Fig. 1), is an ARToolKit [31] marker but we have extended its functionality to use it for different purposes. In its design, multiple markers are placed in a nested/layered fashion. Each layer itself represents a single marker with its unique pattern. While the innermost layer consists of two different markers. The complete description of each layer along with its selection procedure is given in Table I.

TABLE I. Description and Selection Procedure of Each Layer of MLMT Marker

| S/No.	MLMT layers	Definition of layer
1.	Layer1	Layer1 is the outermost layer of the MLMT marker.
2.	Layer2	Layer2 is the second layer of the MLMT marker.
3.	Layer3	Layer3 refers to the third layer of the MLMT marker.
4.	LayerN-1	LayerN-1 refers to the second last layer of the MLMT marker.
5.	Layer N	LayerN refers to the inner most or last layer of the MLMT marker.
6.	Marker M and E	The individual markers placed inside the innermost layer (LayerN) of the MLMT.
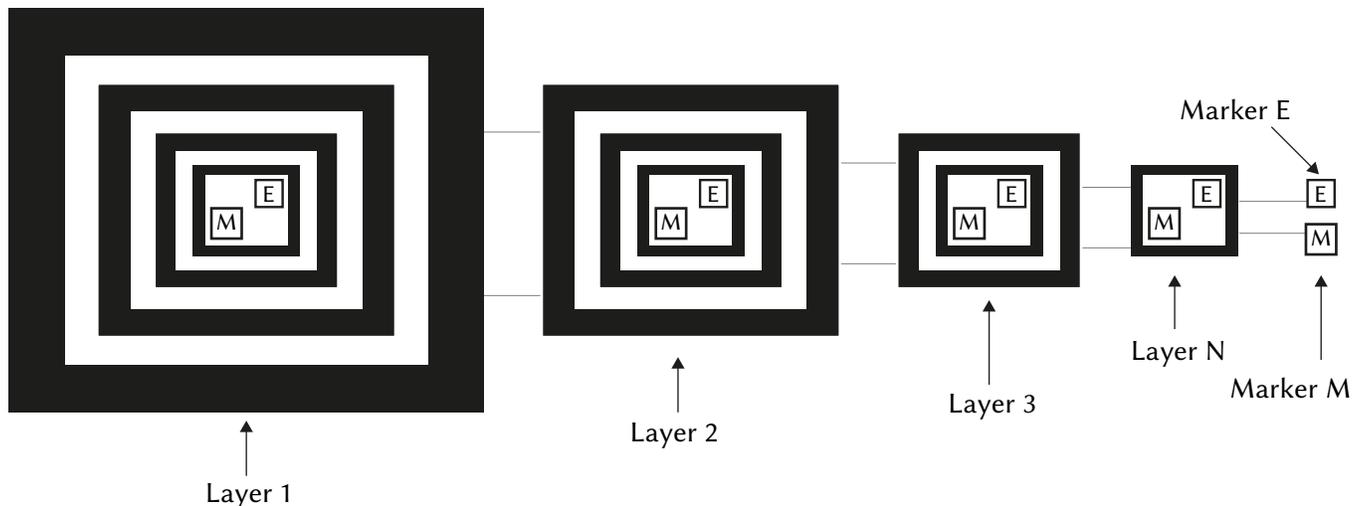
Fig. 1. The MLMT marker design.

The MLMT marker can be used for different tasks such as visualization, rotation, scaling, etc. at the same time needless of any extra marker i.e. we can call it "all in one". In this study, we have used the MLMT marker as an interaction tool that can perform navigation, selection, rotation, and exploration.

The addition of individual markers (M and E) in the innermost layer enables the MLMT marker to perform different operations dynamically such as visualization, rotation, scaling, etc. at the same time and there is no need for extra markers. These markers bring dynamicity to the functionality of the MLMT marker, enables it to perform differently via changing its mode through these markers.

A. Applications of MMLMT Marker in AR

The use of MLMT is very simple and easy. Each layer is responsible for performing a specific operation. An inner marker can be visualized by occluding the upper layer above it by simply putting a finger over a section of the upper layer (see Fig. 2). If the marker is completely visible to the camera, it means that it is the first layer. Occlusion of the first layer leads to the visualization of the second layer and so on up to the innermost layer. A detailed description of the use of MLMT is given in Table I.

There are various applications of MLMT in different fields. Few applications of MLMT in AR are given below:

1. Multitasking Operation

The main purpose of the designing of the MLMT marker is to achieve multiple tasks performance capability using a single marker. The same marker can be used for performing different types of operations such as visualization, scaling, rotation, selection, and navigation, etc.

a) Visualization

Visualization of different 3D objects in AR can be done via a single MLMT marker. visualization of each layer of the MLMT marker displays a different 3D object. The first layer displays a cube, the second a sphere, the third a teapot, while the fourth layer displays a cone as shown in Fig. 2.
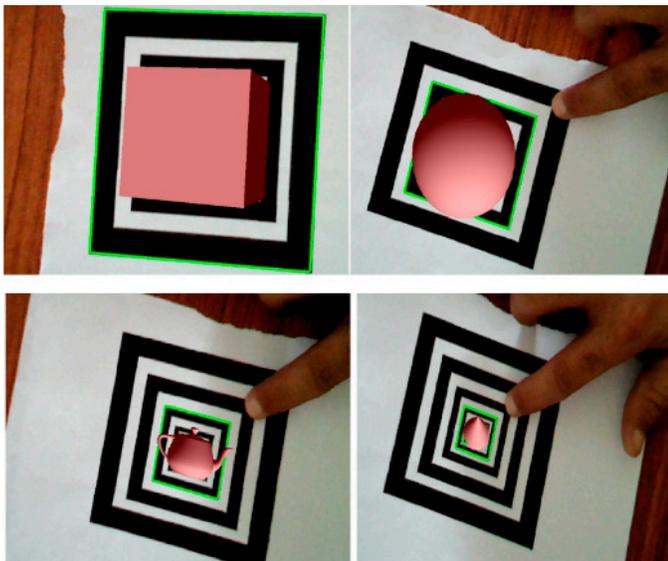
Fig. 2. Visualization of different 3D objects using the MLMT marker. Visualization of Layer1 displays a 3D cube (Upper left). Visualization of Layer2 displays a 3D sphere (Upper right). Visualization of Layer3 displays a 3D teapot (Lower left) and visualization of the Layer4 displays a 3D cone object (lower right).

b) Task Selection

The selection of different tasks of objects can also be done in AR via the MLMT marker. The innermost layer of the MLMT marker consists of two independent markers M and E. Both markers are used to perform the selection of different tasks (see Fig. 3). Visualization of marker M makes the MLMT marker to perform rotation while marker E, to perform scaling.
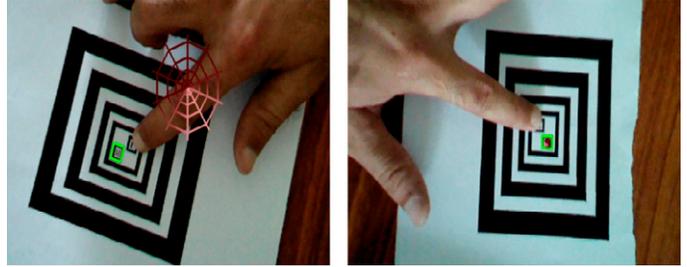
Fig. 3. Selection of innermost marker (E or M) for a specific Task selection using MLMT marker. Visualization of innermost marker M by occluding marker E with a finger (left) and visualization of E (right).

c) Rotation

The third application of MLMT is the rotation of any 3D object via the MLMT marker. Visualization of M shifts the MLMT marker to perform rotation task. In this mode, at each layer of the MLMT marker performs the rotation of a 3D object in a specific axis. The 3D object rotates in the x-axis when layer 1 of the marker is visualized to the camera. Visualization of the second layer leads to the rotation of the object in the y-axis, 3rd layer in the z-axis, while the 4th layer in the xy-axis (see Fig. 4).
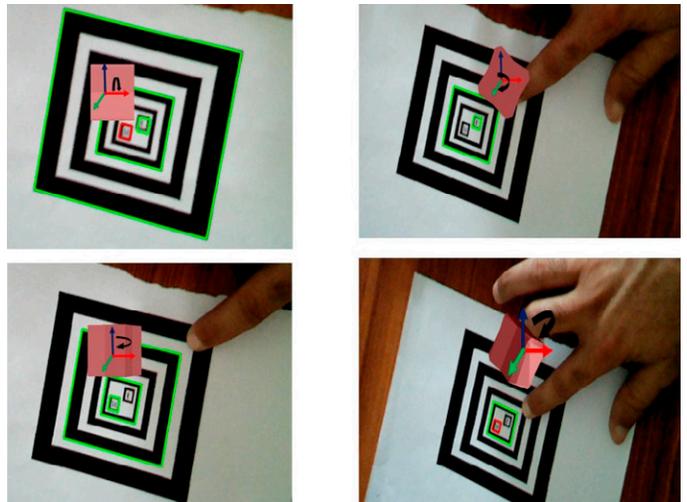
Fig. 4. Selection Object rotation with each layer of MLMT marker. Visualization of Layer1 leads to rotation of the object in the x-axis (upper left). Visualization of Layer2 leads to rotation of the object in the z-axis (upper right). Visualization of Layer3 leads to rotation of the object in the y-axis (lower left) and visualization of Layer4 leads to rotation of the object in the xy-axis (lower right).

d) Scalin

The fourth application of MLMT in AR is the scaling of 3D objects. Visualization of marker E leads the MLMT marker to perform scaling operations with each layer. Visualization of the outermost layer leads to the object scaling with the biggest size, while the second scale down its size to a medium size and so on. Scale down operation is performed by moving from bigger to smaller layers while scaling up in the reverse direction (see Fig. 5).
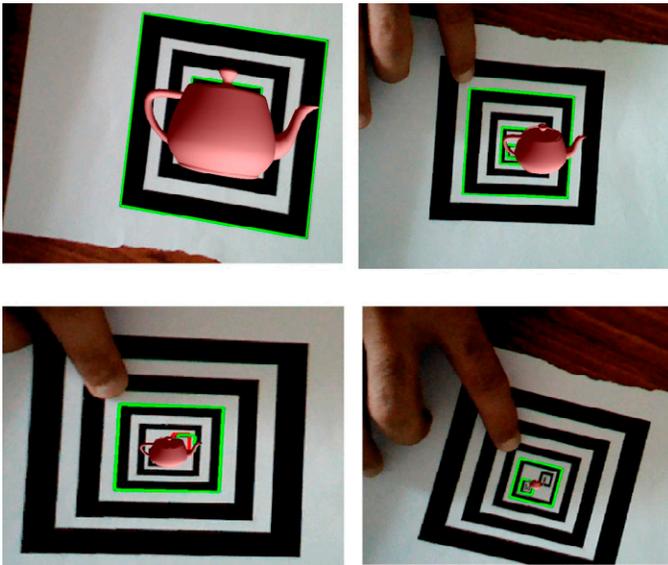
Fig. 5. Scaling operation using MLMT marker. Visualization of MLMT marker's Layer1 leads to 4x scaling of the 3D object (upper left). Visualization of Layer2 leads to 3x scaling of the 3D object (upper right). Visualization of Layer3 leads to 2x scaling of the 3D object (lower left) while the visualization of Layer4 leads to 1x scaling of the 3D object.

IV. System Architectural Model

The proposed system is a mixture of AR and VR where AR works as a backend tool and VR as frontend (see Fig. 6). AR deals with marker detection, marker ID, and pose calculation. VR is responsible for interaction with VE.

AR uses ARToolKit [31] for the detection and processing of fiducial markers from the real scene taken using a camera. First of all, a video path is initialized for the calculation of camera parameters, then the pattern file database is searched for the marker patterns. The marker detection is carried out when the marker is visualized to the camera. Then in the pattern matching phase, the pattern file of the specific marker layer is searched in the database. After identifying a particular layer, the system then calculates the position, orientation, and ID of the layer.
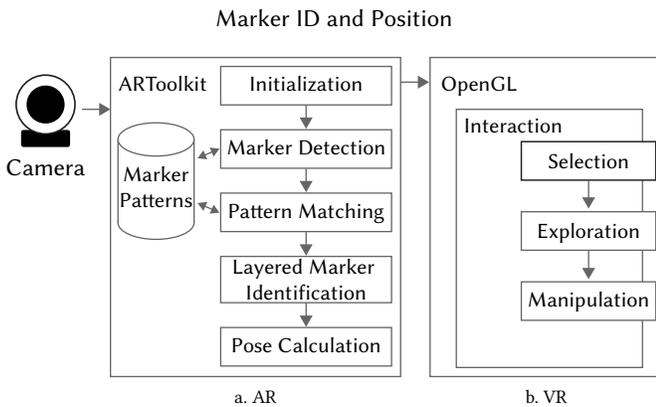
Fig. 6. System architecture model description (a). Describe the backend processing of the system, while (b). describes the front end VE.

In normal situations, ARToolKit [28] recognizes all registered and visible markers at the same time. In most situations, there is a need to identify and use only one specific marker. The multiple marker identification problems also arise when using MLMT. The MLMT

consists of multiple layers each represents a unique marker, so we need a single marker each time to deal with. To cope with multiple marker identification problems, we proposed and implemented a new algorithm described with a flowchart (see Fig. 7) which identifies a specific marker while ignoring all others. The algorithm simply selects the uppermost visible layer of the MLMT marker while ignores all the other markers.
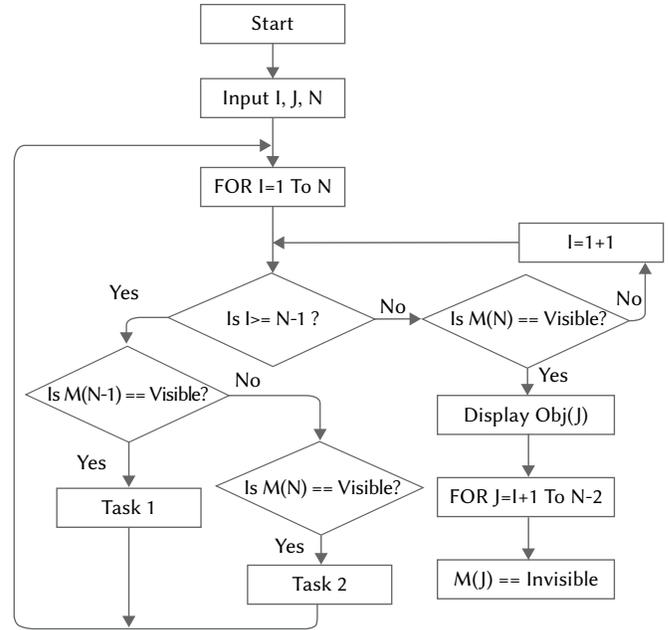
Fig. 7. Procedure for identification of specific MLMT marker layer.

In the first phase, all the layers of the MLMT marker are searched in the library. The first marker (I=1) represents the outermost layer (L1) while the innermost marker (I=N) is represented by the LN (Nth layer). If the outermost layer is visible, the algorithm identifies this marker and performs the associated task while ignoring all other inner layers (i.e. L2 to LN). If the first layer is not visible, the algorithm searches for the second layer (L2), if it is visible, its associated task is performed while ignoring layers. The algorithm repeats the whole process for all N layers. If the visible marker is LN, then the algorithm keeps track of the two markers. If marker M is visible and E is occluded, the task (new scenario) associated with M is performed and vice versa.

The VR system performs different interaction tasks in the VE based on acquired marker ID and position. OpenGL library is used for the creation and realistic interaction in the VE.

A. Mode of Interaction

Individual markers M and E in the innermost layer enable the MLMT marker to perform different operations dynamically such as visualization, rotation, scaling, etc. at the same time without needing any extra markers.

The system allows interaction with VE in two different modes. These modes are represented via 3D interactive labels i.e. Explore and Manipulate (see Fig. 8).

The selection of the interaction mode is done using the MLMT's innermost layer markers. Visibility of marker M leads to Manipulation mode while the visibility of marker E to Exploration mode.

Fig. 8. Scenario of the proposed VR learning environment. Consists of a 3D virtual human skull along with Explore and Manipulate interactive 3D label labels.

1. Exploration Mode

The selection of the exploration label leads the system to the exploration mode in the VE. This mode is responsible for the delivery of object-related textual information to the user. Exploration of the object is carried out in a step by step fashion.

a) Textual Information

The system displays information about the object at each layer in textual form. The first layer displays just the name, the second layer displays some detail such as the function of the object, and so on. So, the delivery of information depends on the interest of the user. If the user is more interested, he can move towards inside, and so he receives more and more information about the object.

2. Manipulation Mode

The selection of the manipulation mode allows the user to perform some manipulation operation on each layer. The first layer of the marker rotates the object at 90^0, the second at 180^0, and so on, in this way up to 360^0. So as the user moves inside by selecting layers, the system rotates the object accordingly. Upon visualization of the innermost layer, the system displays the sub-objects of the parent object.

B. Interaction With Objects

When the camera observes the innermost layer of the marker, the system displays sub-objects of the parent object. A complex parent object may have more than one sub-object. A specific object can be selected simply by intersecting the virtual pointer with it. After the selection of an object, all other objects disappear from the screen. Now the user can interact with that object either in Exploration or Manipulation mode. The Exploration mode can be activated by simply occluding the marker M, and vice versa for Manipulation. Upon selecting the Exploration mode, the user can explore object related information in a step by step fashion. While the selection of Manipulation mode allows the user to manipulate the object via visualizing different layers of the marker. In both modes upon approaching the innermost layer, the system enters the next phase, i.e. displays sub-objects of this object.

V. Biological VR Application

A detailed description of the proposed VR system is shown in Fig. 9. The VE displays a complex virtual object and allows users to interact with it. The object has further internal subobjects. Interaction in the VE consists of selection, manipulation, exploration, and object inside/internal navigation in a layered fashion. First of all the system allows us to select an interaction mode and after that, the visualization

and identification of each layer of the marker are carried out by the system, and finally, the task associated with each layer is performed. At the innermost (Nth) layer, the system visualizes sub-objects of the parent and allows users to select any one of them. After the selection of a specific sub-object, the above whole process is repeated for the selected sub object. We have developed a biological application for evaluation and experimental purposes. This VE is a room like structure where a human skull is visualized as the main/parent object as shown in Fig. 8. A human skull consists of various human organs where eyes and brain are the most important. So we will analyze these objects in our study. The VE also consists of Explore and Manipulate objects used for the selection of interaction mode.

The VE when loaded contains only the parent object to explore and manipulate labels. Interaction with the system is carried out using the MLMT marker. The virtual pointer follows the movement of a completely visible MLMT marker. First of all the user can select the explore or manipulate mode. Visualization of object related information at different layers, sub-objects of the parent object, and rotation of object during manipulation are shown in Fig. 10, 11, and 12 respectively.
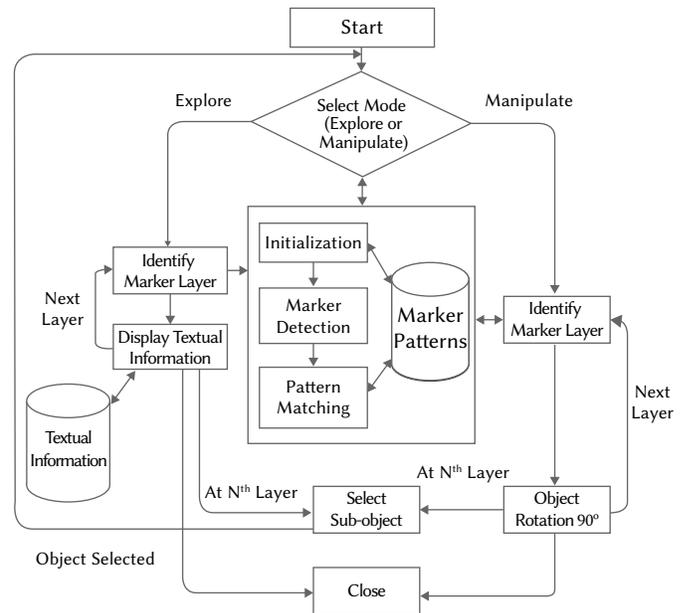
Fig. 9. Detailed System architecture model.
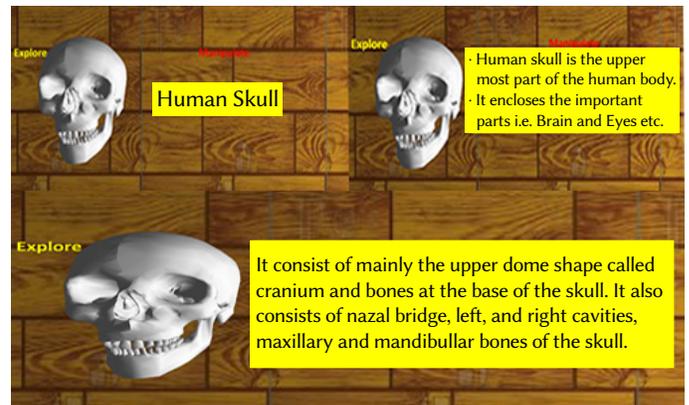
Fig. 10. Provision of layered based object related information. Upper left (information displayed with Layer 1), upper right (information displayed with Layer 2), and lower (information displayed with Layer 3).

Fig. 11. Visualization of 3D objects in VE at the inner most layer.

Fig. 12. Object anti-clock wise rotation at each layer of the MLMT marker. Visualization of 3D human skull with Layer1 (upper left). Visualization of Layer2 leads to rotation of 3D skull 90^0 (upper right). Visualization of Layer3 leads to rotation of 3D skull 180^0 (lower left) and Layer4 leads to rotate the skull to 270^0 (lower right).

VI. Experiments and Evaluation

In the experimental section, we will perform two types of studies. The experimental setup is shown in Fig. 13.
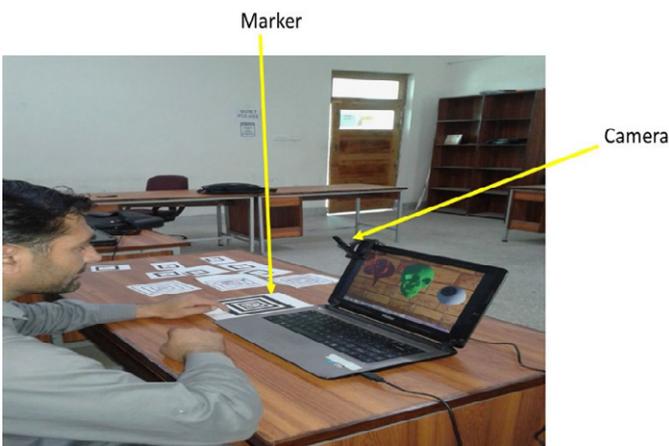
Fig. 13. Experimental setup.

A. Comparison of Simple VE Vs Layered VE

To perform a comparative study, we have designed two VEs i.e. a Simple VE and a Layered VE. In Simple VE three different 3D human organs i.e. human skull, eye, and brain are visualized (see Fig. 14).

Fig. 14. Scenario of simple VE. Visualization of all the three objects at the same time.

To interact with these virtual objects in Simple VE, we designed three different markers as shown in Fig. 15. Firstly, all three markers are visualized to the camera. Occlusion of specific marker results in information visualization of the related virtual object i.e. occlusion of the Skull marker displays information related to human skull object while Brain and Eye markers visualize information related to the human brain and eye. Upon occlusion of a specific marker, the system displays complete information related to that concerned virtual object as one big chunk as shown in Fig. 16.
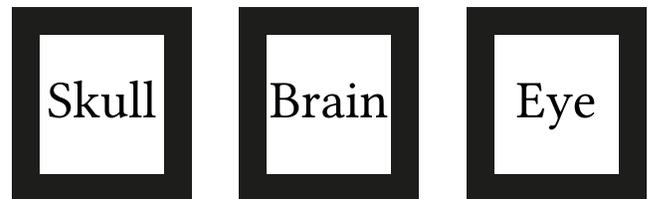
Fig. 15. Markers used for interaction with Simple VE.

The layered VE consists of a 3D human skull as shown in Fig. 8. MLMT marker is used for interaction with the VE in a layered fashion i.e. the Layer1 displays name, Layer2 displays some details and so on up to the LayerN-1, which displays detailed information related to the skull. The LayerN displays the subparts i.e. human Brain and Eye. Brain or Eye can be selected via visualization of one marker and occlusion of the markers i.e. E or M. Occlusion of E leads to the visualization of the human eye and occlusion of M displays 3D human brain. After the selection of the brain/eye, the system allows the user to display related information via the interaction of MLMT marker in a layered fashion as discussed above.
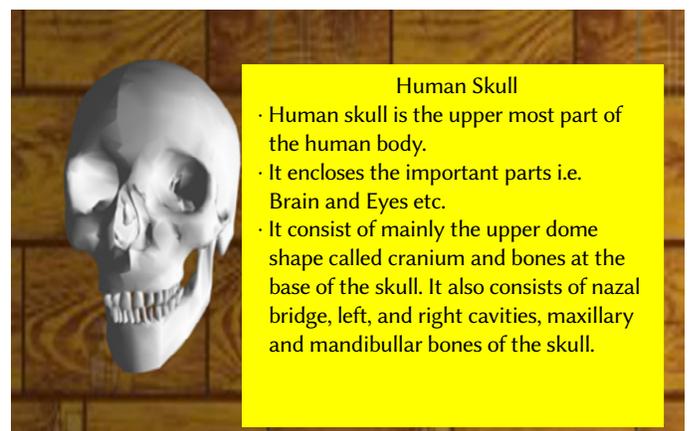
Fig. 16. Information visualization related to objects in simple VE.

We will experimentally examine the learning effect and task execution time in both VEs.

1. Protocol

To investigate the learning improvement, task execution time, and easiness in interaction, we have randomly selected thirty (30), participants. All the participants were SSC (Secondary School Certificate) level science (biology) students from three different schools. The topic was included in their course work. The students had no previous experience with VR. The study was designed in a manner to use all the VE features.

2. Task

All the students were divided into two groups (G1 and G2) each of 15 students. All the students were demonstrated and trained about the use of VE. The task was to interact and study each object in a sequence (explore and manipulate).

Students of G1 performed three trails on the task in the simple VE. While that of G2 performed three trails on the same task in the layered VE. The task execution time was recorded for both groups. After that, they filled a questionnaire to evaluate their learning enhancement.

3. Results Analysis

In this section, we will analyze the questionnaire filled by both groups. These questions aimed to assess their learning enhancement and ease of interaction. The students have to answer the questions related to the learning. For the objective analysis, the task execution time for both groups was also recorded.

a) Learning

The analysis of variance (ANOVA) concerning the learning enhancement for G1 and G2 is significant ($F(1,28) = 35.087$, $p < 0.05$). So, there is a significant difference between G1 and G2. The mean and standard deviation (SD) of both groups (G1 (48.80, 11.143) and G2 (68.8, 9.096)) is shown in Fig. 17. The results show that students of G1 got more knowledge as compared to G2. The main reason for improved learning of G1 may be the provision of information in small chunks in a stepwise manner. As the small amount of information is easy to learn as compared to a big chunk of information.
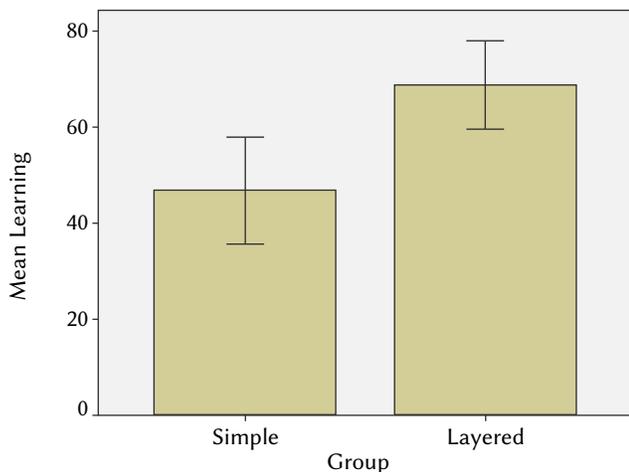
Fig. 17. Mean Learning and SD of two groups.

b) Task Execution Time

The ANOVA for task performance of both groups is significant ($F(1,28) = 60.222$, $p < 0.05$). The mean and standard deviation (SD) for G1 is (161.6, 24.20) and G2 is (218.93, 15.25), as shown in Fig. 18. It means that the students of G1 who used simple VE completed the task fast as compared to G2 who used the layered VE. The reason behind the good performance of G2 was the simple selection of a marker among three markers placed in front of them while in case of layered VE, users need more cognitive and physical work i.e. selection of different layers in a sequence and selection of objects, etc. which took more time.
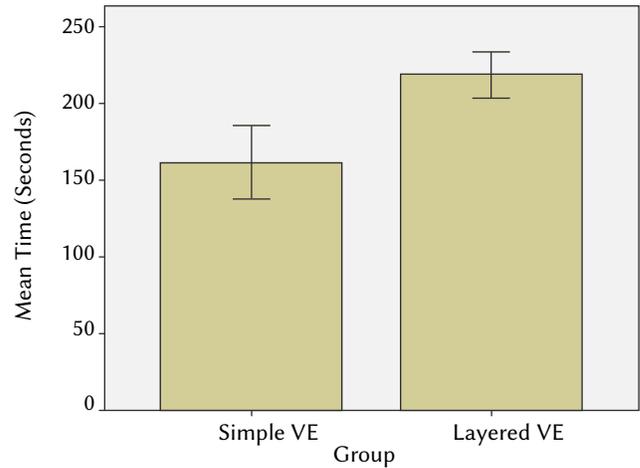
Fig. 18. Mean task execution time of two groups.

B. Comparison of MLMT With Single and Multi Layered Markers

In this study, we evaluated the effect of interaction tool on student's task execution time, learning, and usability in the VE. For this purpose, we compared our proposed interaction tool (i.e. MLMT marker) with multiple single markers [31], and multi layered markers [24].

1. Single Marker

Every single marker is uniquely designed for each type of function/operation as shown in Fig. 19. For example, a single marker may perform the rotation of an object or display the name of an object, etc. We have designed 24 unique markers i.e. 8 markers for the human skull (4 for 4 layers of exploration, 4 for rotation), 8 for the human brain (4 for 4 layers of exploration, 4 for rotation), and 8 for the human eye (4 for 4 layers of exploration, 4 for rotation).

2. Multi Layered Marker

We have designed six different multi layered markers i.e. two for skull exploration and manipulation, two for brain exploration and manipulation, and two for eye exploration and manipulation as shown in Fig. 20. Each multi layered marker consists of 4 layers while the last/inner layer is a marker with some letters. Each layer performs its specific function. In the case of Skull exploration, the first layer is used for the visualization of simple information (i.e. Human Skull), Layer2 visualizes some detail, and so on, the LayerN displays in-depth information of the skull. While in case of manipulation, the visibility of each layer leads to some type of manipulation task e.g. rotation, scaling, etc. These markers can perform a single task and thus they have a lack of dynamicity. Interaction with VE can be performed via occluding each layer one by one from outer to the inner layer.

3. MLMT Marker

We designed an MLMT marker for interaction with VE, as shown in Fig. 21. This single marker is responsible for interaction which includes navigation, selection, and manipulation in the layered VE. The MLMT marker is a multi layered marker, consisted of four nested layers while the innermost layer consisted of two unique markers M and E. The Addition of markers M and E extend its capabilities to operate dynamically in different situations.

SR1	SR2	SR3	SR4
SE1	SE2	SE3	SE4
BR1	BR2	BR3	BR4
BE1	BE2	BE3	BE4
ER1	ER2	ER3	ER4
EE1	EE2	EE3	EE4
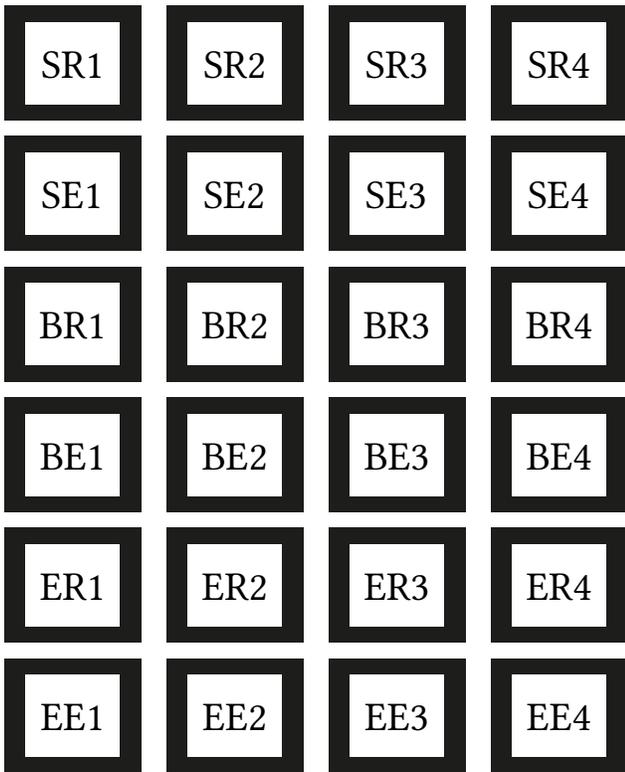
Fig. 19. Single ARToolKit markers including markers with BE (Brain Explore-Layer1 to Layer4), BR (Brain Rotate- Layer1 to Layer4), EE (Eye Explore-Layer1 to Layer4), ER (Eye Rotate- Layer1 to Layer4), and SE (Skull Explore-Layer1 to Layer4), SR (Skull Rotate- Layer1 to Layer4).
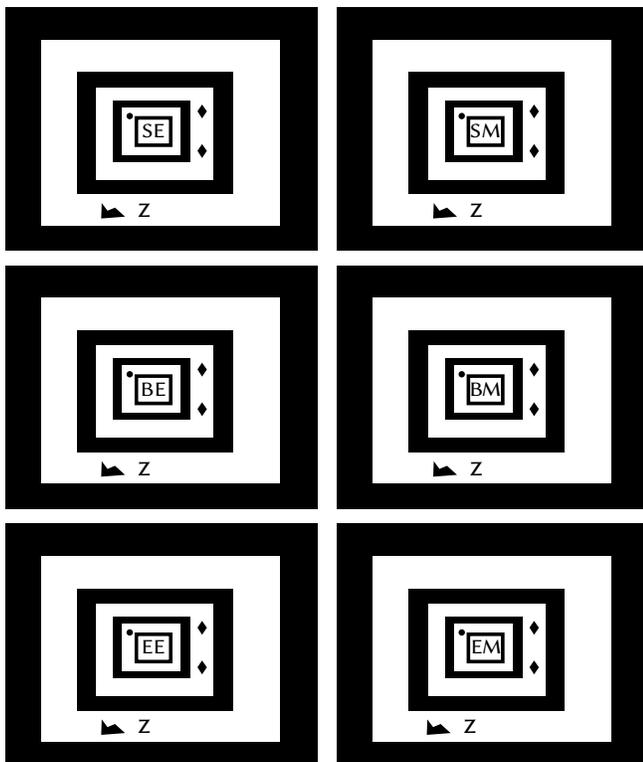
Fig. 20. Multi layered markers. Markers with letter SM and SE represent layered markers for Skull Manipulation and Exploration of (top). Markers with letter BM and BE represent layered markers for Brain Manipulation and Exploration (middle). While markers with letter EM and EE represent layered markers for Eye Manipulation and Exploration (bottom).
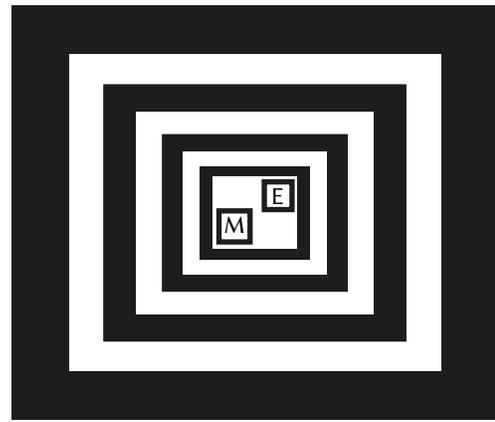
Fig. 21. Design of MLMT marker.

4. Protocol and Task

We selected another group of 30 students for the experimental study. We randomly divided these students into three groups (G1, G2, and G3). G1 is assigned to use a single marker while G2 and G3 used layered and MLMT markers for interaction with layered VE. All the students were first briefed regarding the use of their assigned marker. After that, they used the VE for 10 minutes before the actual experiment. After training, they performed the experimental task. The task was to interact and explore all the three 3D objects i.e. skull, brain, and heart.

5. Results Analysis

In this section, we performed both the objective and subjective analysis of the three groups. In the objective analysis, we compared the task execution time of the three groups. In the subjective analysis, we first used a questionnaire to assess learning enhancement using their assigned system. After that, we used the System Usability Scale (SUS) [36] for evaluation based on the student's opinions.

a) Task Execution Time

The ANOVA concerning task execution time for all groups i.e. G1, G2, and G3 is significant ($F(2,27) = 60.289$, $p < 0.05$). The mean and SD for G1, G2, and G3 is (55.00, 9.274), (37.20, 6.630), (20.50, 41.16) as shown in Fig. 22. It means that the group G3 completed the task in less time as compared to the G1 and G2. The reason for the low performance of using single markers is the searching of specific markers among multiple markers for each task.
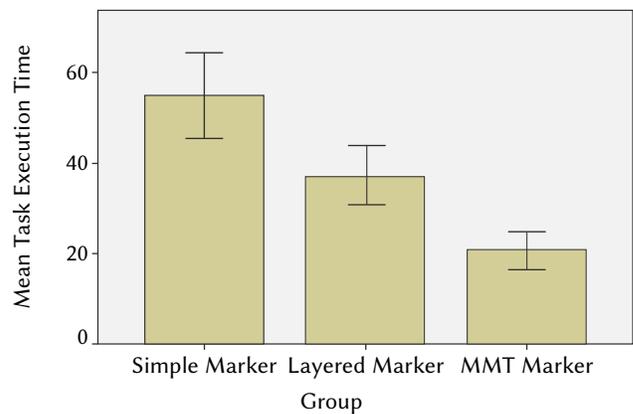
Fig. 22. Task execution time of the three groups.

b) Learning

To assess students learning, the questionnaire consisted of

different questions about the information given in the biological application. The ANOVA related to students learning for all groups i.e. G1, G2, and G3 is significant ($F_{(2,27)} = 14.152$, $p < 0.05$). The mean and SD for G1 is (82.60, 6.569), G2 is (69.40, 9.383), and G3 is (60.80, 11.153), as shown in Fig. 23. From the above results, we can conclude that G3 is comparatively better in learning enhancement than G1 and G2. The provision of step-wise information in small chunks (easy to read) using MLMT improves learning as compared to others. The reason of reduce learning for G1 and G2 may be the cognitive work required for identification and selection of a specific marker among multiple markers.
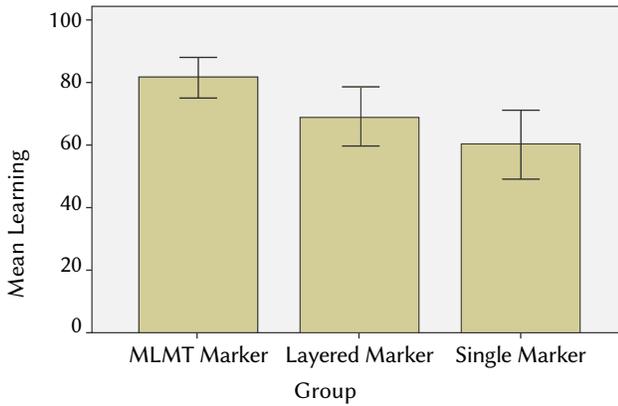
Fig. 23. Mean and SD of students learning of the three groups.

C. Usability

We use a standard usability test to evaluate these interaction tools i.e. markers based on the student's opinions. The SUS consists of ten questions that define the ease of use, learnability, efficiency, effectiveness, and user satisfaction. The SUS consists of 10 questions each with options ranges from strongly disagree (1) to strongly agree (5). The score for each odd number question is measured as scale value minus 1 i.e. strongly disagree has score 0 and strongly agree has 4. The score of even number questions is calculated as subtracting from each value from 5 i.e. strongly disagree has a score 5-1 = 4 and strongly agree has 5-5=0 score. For example, in Table II, for question 1, 7 students opted for Strongly agree and 3 for Agree option. The average score of ten students for question 1 is calculated as $(((5-1) \times 7 + (4-1) \times 3)/10) = 3.7$. For question 2, three students selected the option Strongly disagree and 7 opted for Disagree. The average score of ten students for question 2 is $(((5-1) \times 3 + (5-2) \times 7)/10) = 3.3$.

SUS questionnaire results show that the overall results for Layered markers are good as students opted to 80.25 SUS score (see Table III). As the students are satisfied with their assigned tool but having less score in terms of consistency and user-friendliness. The SUS questionnaire results of students who used multiple single markers show an average SUS score of 70 (see Table IV). Results of the SUS questionnaire (see Table II) shows that students selected the best options in favor of the MLMT marker which got SUS usability score 88. Students' opinions regarding questions 1 and 9 show that all students are satisfied with the proposed interaction tool i.e. MLMT marker. Results of questions 2, 3, 7, and 8 show that the proposed marker has user-friendly nature.

TABLE II. SUS RESULTS OF G3 WHO USED MLMT FOR INTERACTION WITH SYSTEM. AVERAGE SUS SCORE IS 35.2 AND TOTAL SUS SCORE IS 35.2*2.5= 88

S.No	Concerned Statements	Strongly Disagree 1	2	3	4	Strongly Agree 5	Average score
1.	I think I would like to use this system frequently.	0	0	0	3	7	**3.7**
2.	I found the system unnecessarily complex.	3	7	0	0	0	**3.3**
3.	I thought the system was easy to use.	0	0	0	2	8	**3.8**
4.	I think that I would need the support of a technical person to be able to use this system.	8	2	0	0	0	**3.8**
5.	I found the various functions in this system were well integrated.	0	0	0	6	4	**3.4**
6.	I thought there was too much inconsistency in this system.	5	5	0	0	0	**3.5**
7.	I imagine that most people would learn to use this system very quickly.	0	0	0	1	9	**3.9**
8.	I found the system very cumbersome to use.	9	1	0	0	0	**3.9**
9.	I felt very confident using the system.	0	0	0	2	8	**3.8**
10.	I needed to learn a lot of things before I could get going with this system.	7	3	0	0	0	**3.7**

TABLE III. SUS RESULTS OF G2 WHO USED LAYERED MARKER FOR INTERACTION WITH SYSTEM. AVERAGE SUS SCORE IS 32.1 AND TOTAL SUS SCORE IS 32.1*2.5= 80.25

S.No	Concerned Statements	Strongly Disagree 1	2	3	4	Strongly Agree 5	Average score
1.	I think I would like to use this system frequently.	0	0	2	3	5	**3.5**
2.	I found the system unnecessarily complex.	4	2	4	0	0	**3**
3.	I thought the system was easy to use.	0	0	3	4	3	**3.3**
4.	I think that I would need the support of a technical person to be able to use this system.	6	3	1	0	0	**3.5**
5.	I found the various functions in this system were well integrated.	0	0	2	5	3	**3.3**
6.	I thought there was too much inconsistency in this system.	2	5	3	0	0	**2.9**
7.	I imagine that most people would learn to use this system very quickly.	0	0	2	3	5	**3.5**
8.	I found the system very cumbersome to use.	0	6	3	1	0	**2.5**
9.	I felt very confident using the system.	0	0	4	2	4	**3.4**
10.	I needed to learn a lot of things before I could get going with this system.	4	4	2	0	0	**3.2**

TABLE IV. SUS Results of G1 Who Used Single Markers for interaction With System. Average SUS Score Is 28 and total SUS Score Is 28*2.5= 70

S.No	Concerned Statements	Strongly Disagree 1	2	3	4	Strongly Agree 5	Average score
1.	I think I would like to use this system frequently.	0	0	6	4	0	**3**
2.	I found the system unnecessarily complex.	0	2	4	4	0	**2.6**
3.	I thought the system was easy to use.	0	2	3	5	0	**2.6**
4.	I think that I would need the support of a technical person to be able to use this system.	2	6	2	0	0	**3**
5.	I found the various functions in this system were well integrated.	0	2	3	5	0	**2.6**
6.	I thought there was too much inconsistency in this system.	1	5	4	0	0	**2.7**
7.	I imagine that most people would learn to use this system very quickly.	0	0	3	5	2	**3.2**
8.	I found the system very cumbersome to use.	0	5	5	0	0	**2.5**
9.	I felt very confident using the system.	0	0	5	4	1	**3.1**
10.	I needed to learn a lot of things before I could get going with this system.	2	3	5	0	0	**2.7**

Results of questions 5 and 6 show that the proposed system is well integrated while the results of questions 4 and 10 show that systems learnability is very good.

VII. Conclusion and Future Work

The provision of information easily and effectively is the most important prerequisite in any information rich virtual environment. We propose a novel interaction technique for the selection, manipulation, and exploration (textual information delivery) of complex objects in virtual environments. Exploration consists of an interest-based, step by step (layered based) information delivery to users. A newly designed MLMT fiducial marker is used for interaction with virtual objects. The MLMT marker was used for navigation, selection, and manipulation of virtual objects. A biological virtual learning application was used for evaluation and experimental purposes. We performed a comparative study between the proposed MLMT marker, simple layered marker, and multiple single markers. The experiments resulted in improved learning, easiness in interaction, and comparatively less task execution time using the MLMT marker.

In the future, we will use the proposed marker in different areas such as for interaction with interactive writing boards [22] and interactive games. We also plan to work on the occlusion of the MLMT marker, i.e., to differentiate between intentional and unintentional hiding of the markers.

References

[1] I. U. Rehman, S. Ullah, and M. Raees, "Two Hand Gesture Based 3D Navigation in Virtual Environments," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, pp. 128-140, 2019.

[2] D. A. Bowman, L. F. Hodges, and J. Bolter, "The virtual venue: User-computer interaction in information-rich virtual environments," Presence, vol. 7, pp. 478-493, 1998.

[3] J. Chen, "Effective interaction techniques in information-rich virtual environments," Proceedings of the Young VR, 2003.

[4] N. F. Polys, D. A. Bowman, and C. North, "The role of depth and gestalt cues in information-rich virtual environments," International journal of human-computer studies, vol. 69, pp. 30-51, 2011.

[5] C. Ware, Information visualization: perception for design: Elsevier, 2012.

[6] D. A. Bowman, J. Chen, C. A. Wingrave, J. F. Lucas, A. Ray, N. F. Polys, et al., "New directions in 3D user interfaces," IJVR, vol. 5, pp. 3-14, 2006.

[7] A. Munro, R. Breaux, J. Patrey, and B. Sheldon, "Cognitive aspects of virtual environments design," Handbook of virtual environments: Design, implementation, and applications, pp. 415-434, 2002.

[8] M. C. Salzman, C. Dede, R. B. Loftin, and J. Chen, "A model for understanding how virtual reality aids complex conceptual learning," Presence: Teleoperators & Virtual Environments, vol. 8, pp. 293-316, 1999.

[9] R. E. Mayer, "Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction," New directions for teaching and learning, vol. 2002, pp. 55-71, 2002.

[10] J. Saiki, "Spatiotemporal characteristics of dynamic feature binding in visual working memory," Vision Research, vol. 43, pp. 2107-2123, 2003.

[11] H. Tardieu and V. Gyselinck, Working memory constraints in the integration and comprehension of information in a multimedia context: na, 2003.

[12] E. K. Vogel, G. F. Woodman, and S. J. Luck, "Storage of features, conjunctions, and objects in visual working memory," Journal of Experimental Psychology: Human Perception and Performance, vol. 27, p. 92, 2001.

[13] S. Card, J. Mackinlay, and B. Shneiderman, "Information visualization," Human-computer interaction: Design issues, solutions, and applications, vol. 181, 2009.

[14] M. Peter and M. Keller, "Visual Cues Practical Data Visualization," IEEE Computer Society, Los Alamitos, 1993.

[15] D. A. Bowman, L. F. Hodges, D. Allison, and J. Wineman, "The educational value of an information-rich virtual environment," Presence: Teleoperators & Virtual Environments, vol. 8, pp. 317-331, 1999.

[16] D. A. Bowman, C. North, J. Chen, N. F. Polys, P. S. Pyla, and U. Yilmaz, "Information-rich virtual environments: theory, tools, and research agenda," in Proceedings of the ACM symposium on Virtual reality software and technology, 2003, pp. 81-90.

[17] N. F. Polys, D. A. Bowman, and C. North, "Information-rich virtual environments: challenges and outlook," in NASA Virtual Iron Bird Workshop, 2004.

[18] S. Ressler, "A Web-based 3D glossary for anthropometric landmarks," in Proceedings of HCI International, 2001, pp. 5-10.

[19] I. Rehman and S. Ullah, "The effect of Constraint based Multi-modal Virtual Assembly on Student's Learning," Sindh University Research Journal-SURJ (Science Series), vol. 48, 2016.

[20] I. U. Rehman, S. Ullah, and I. Rabbi, "The effect of semantic multi-modal aids using guided virtual assembly environment," in 2014 International Conference on Open Source Systems & Technologies, 2014, pp. 87-92.

[21] I. U. Rehman, S. Ullah, and I. Rabbi, "Measuring the student's success rate using a constraint based multi-modal virtual assembly environment," in International Conference on Augmented and Virtual Reality, 2014, pp. 53-64.

[22] D. Khan, I. Rehman., S. Ullah, W. Ahmad, Z. Cheng, G. Jabeen, & H. Kato, "A Low-Cost Interactive Writing Board for Primary Education using Distinct Augmented Reality Markers", Sustainability, vol. 11, no. 20, pp. 5720, 2019.

[23] K. Tateno, I. Kitahara, and Y. Ohta, "A nested marker for augmented reality," in 2007 IEEE Virtual Reality Conference, 2007, pp. 259-262.

[24] I. Rabbi and S. Ullah, "Extending the tracking distance of fiducial markers for large indoor augmented reality applications," Advances in Electrical and Computer Engineering, vol. 15, pp. 59-65, 2015.

[25] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, "Exploring

MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system," Computers & Graphics, vol. 23, pp. 779-785, 1999.

[26] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: non-linear mapping for direct manipulation in VR," in ACM Symposium on User Interface Software and Technology, 1996, pp. 79-80.

[27] D. A. Bowman and L. F. Hodges, "An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments," SI3D, vol. 97, pp. 35-38, 1997.

[28] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top AR environment," in Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 2000, pp. 111-119.

[29] M. Fiala, "Comparing ARTag and ARToolkit Plus fiducial marker systems," in IEEE International Workshop on Haptic Audio Visual Environments and their Applications, 2005, p. 6 .

[30] K. Rainio and A. Boyer, "ALVAR–A Library for Virtual and Augmented Reality User's Manual," VTT Augmented Reality Team, 2013.

[31] O. S. A. R. SDK, "ARToolKit. org," ed.

[32] J. Jun, Q. Yue, and Z. Qing, "An extended marker-based tracking system for augmented reality," in 2010 Second International Conference on Modeling, Simulation and Visualization Methods, 2010, pp. 94-97.

[33] D. Khan, S. Ullah, and I. Rabbi, "Factors affecting the design and tracking of ARToolKit markers," Computer Standards & Interfaces, vol. 41, pp. 56-66, 2015.

[34] D. Khan, S. Ullah, D.-M. Yan, I. Rabbi, P. Richard, T. Hoang, et al., "Robust tracking through the design of high quality fiducial markers: an optimization tool for ARToolKit," IEEE access, vol. 6, pp. 22421-22433, 2018.

[35] M. Azhar, S. Ullah, I. U. Rahman, and S. Otmane, "Multi-Layered Hierarchical Bounding Box based interaction in virtual environments," in Proceedings of the 2015 Virtual Reality International Conference, 2015, p. 14.

[36] Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. 1996 Sep 1;189(194):4-7.

Inam Ur Rehman

He is currently pursuing Ph. D. degree from the University of Malakand, Pakistan. He received his BS degree in Computer Science and M. Phil in Virtual and Augmented Reality from University of Malakand in 2006 and 2016 respectively. His research interests include 3D interaction, virtual and augmented reality.

Sehat Ullah

He is currently working as Associate Professor in the department of Computer Science & IT, University of Malakand, Pakistan. He received his PhD degree in Computer Science in specialty in Robotics from University of Evry Val d'Essonne, France in 2011. He also earned MS degree of Computer Science with specialty in Virtual Reality and Intelligent Systems from the same University in 2007. He received MSc degree of Computer Science and BSc with major in Mathematics and Computer Science from University of Peshawar in 2001 and 1998 respectively. Dr. Sehat Ullah has research interest in virtual and augmented reality systems, machine learning, image processing and haptic systems.

Dawar Khan

He is working as an Assistant Professor at the Interactive Media Design Lab, Nara Institute of Science and Technology (NAIST), Japan. He completed his PhD in Computer Science from the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, University of Chinese Academy of Sciences (UCAS), Beijing China, in 2018. During his PhD study, he was supported by the Chinese Government Scholarship, and the National Natural Science Foundation of China. He was awarded with the 2018-Excellent International Graduates Award of the UCAS. Prior to that, he did his BS and MS in Computer Science from University of Malakand Pakistan, in 2011 and 2014 respectively. His research interests include Computer Graphics, Computational Geometry, Mesh Processing, and Augmented Reality.

Multi-sense Embeddings Using Synonym Sets and Hypernym Information from Wordnet

Krishna Siva Prasad Mudigonda*, Poonam Sharma

Department of Computer Science and Engineering, Visvesvaraya National Institute of Technology, Nagpur (India)

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

Word embedding approaches increased the efficiency of natural language processing (NLP) tasks. Traditional word embeddings though robust for many NLP activities, do not handle polysemy of words. The tasks of semantic similarity between concepts need to understand relations like hypernymy and synonym sets to produce efficient word embeddings. The outcomes of any expert system are affected by the text representation. Systems that understand senses, context, and definitions of concepts while deriving vector representations handle the drawbacks of single vector representations. This paper presents a novel idea for handling polysemy by generating Multi-Sense Embeddings using synonym sets and hypernyms information of words. This paper derives embeddings of a word by understanding the information of a word at different levels, starting from sense to context and definitions. Proposed sense embeddings of words obtained prominent results when tested on word similarity tasks. The proposed approach is tested on nine benchmark datasets, which outperformed several state-of-the-art systems.

I. Introduction

Classical problem like semantic analysis continues to grab the attention of researchers since it is important to different fields of study. Semantic similarity has gained importance in fields like information retrieval [1], bio-medical domains [2], creating knowledge graphs [3], sentence clustering [4], cross-lingual text similarity [5]. Primitive semantic similarity measures are based on the distance between the concepts. Measures like Jaccard distance [6], Euclidean distance [7], and dice coefficient [8] are examples of it. These measures are not suitable for natural language processing activities as there is no prior knowledge taken into account for assessing the similarity. Researchers use the measures which consider background knowledge and the relation between the concepts to assess semantic similarity.

Knowledge-based measures, which are one of the kinds of semantic similarity measures use relations between the concepts. State-of-the-art knowledge-based measures use a lexical database like Wordnet [9] to measure semantic similarity. Wordnet is a structured organization of terms, referred to as synonym sets or concepts. Knowledge-based measures are formulated based on distance, depth or information content of concepts. Measures proposed by Rada [14], Wu and Palmer [17], Leacock and Chodorow [11], Li [12] specify different ideas of using the distance between the concepts to calculate the semantic similarity. The measure proposed by Li [12] uses distance and depth between the concepts as metrics to evaluate the similarity between the concepts. Resnik [15] calculates similarity by using the information content of the lowest common subsumer between the concepts. Lin [13], Jiang and Conrath [10] also use the information content between the concepts. Zhu [16] proposed a hybrid measure by combining the Resnik measure [15] and path measure [14]. Zhu [16] extracted semantic similarity by using knowledge-based measures on knowledge graphs like YAGO [38]. These measures do not discuss vector representations. These measures are not substantially strong to perform present-day NLP activities. Hence, the measures which address the distributional semantics of words are required in the current era.

In recent years, distributional semantic representations of words gained popular attention. Distributional representations of words well known as word embeddings set the state-of-art systems for core natural language processing activities (NLP). Word embeddings [18], [19], represent words or phrases with vectors to extract relationships between them. Word embedding techniques improved the efficiency of various NLP tasks, like word similarity [20], sentiment analysis [21], [22], text classification [23], question answering tasks [24].

Even though word embedding techniques improved the accuracy of core NLP tasks, these approaches fail to address the polysemy problems [25]. Research related to polysemy, dealing context, and senses of words is needed in the present scenario. Based on the word's sense and context, recent research is trying to improve the semantic representations. Word embedding techniques generally neglect to explore lexical structures with valuable semantic relations from lexical databases like Wordnet [9]. To improve the efficiency of word embeddings, word vectors should consider different contexts and multiple senses of words. Multi-sense embeddings require a well-formed semantic network to extract multiple senses of the word, alongside word vector representations.

---

* Corresponding author.

E-mail address: krishnasivaprasad536@gmail.com

Semantic similarity requires the knowledge of concepts that are obtained from background sources. Well defined structures like Wordnet are limited, yet this knowledge cannot be ignored. Hence, sources like Wikipedia and Wordnet can be combined to obtain the context. Wordnet establishes the relationship between the concepts using hierarchical relations. Each concept in Wordnet has synonym sets, which help to predict the multiple senses of words. Hence, this paper develops semantic vector representations of words considering their context, sense, definitions, and hypernym path using the lexical structure of Wordnet and word embedding techniques like Word2vec models.

The following are the contributions of this paper:

- This paper uses the rich semantic structure of Wordnet to generate the senses and definitions of each word.
- Regarding each word, suitable contexts are generated from a large source of Wikipedia text.
- Multiple sense based word embeddings are generated.
- This paper aims at developing parts-of-speech related synonym set embeddings to carry out word similarity tasks.
- Detailed comparison and analysis of state-of-art techniques are mentioned.
- Nine benchmark datasets are compared to portray the significance of the proposed approach.

The paper is organized as follows. Section II gives the related work, which covers the existing word embedding techniques. Section III gives the proposed approach and covers the senses extracted from Wordnet, context from Wikipedia and word vector representations. In section IV, experimentation details of existing measures and the proposed measure are mentioned, followed by discussion in section V. Section VI gives the conclusion of the work done and recommends some suggestions for future work.

II. Related Work

This section discusses the existing distributed representations of words. Distributed representations of words proposed by Mikolov et al, [18], [19] become popular among NLP related tasks. Working with contexts has its roots with the distributional hypothesis proposed by Harris [26]. After that, the bag-of-words [27] approach also discussed the distributed representations of words, but these approaches suffer from drawbacks like data sparsity, not maintaining word order, and dimensionality related issues. Approaches based on language prediction also exist in the literature of the NLP [28]. The language models are transformed with continuous bag-of-words and skip-gram models.

Mikolov et al, [18], [19] proposed Word2vec with continuous skip-gram and continuous bag-of-words (CBOW) which portrayed the importance of word embeddings. This approach gained importance due to its efficient log-linear neural network language model, low-dimensionality vector representations. Skip-gram and CBOW models produce vector representations of words. Word embedding representations like SENNA [29], GloVe [30], and fastText [31] exist in the literature. All these are single vector representations. These models fail to address the polysemy of the words. Word vector representations can be enhanced by combining multiple senses of words. Examining the sense, context, and definitions of words while deriving word embeddings improves the efficiency of the NLP tasks. The proposed approach is developed using the Word2vec model to train the corpus and later the vectors are used based on context, sense, and definitions of each word.

Researchers after the invention of distributional semantic representations of words developed sense embeddings [25], [32], [33],

[34], [35] to perform NLP tasks like word similarity. Interestingly, Li [25] presented the idea of developing separate vectors for each sense. The approach mentioned by Iacobacci et al, [32] obtains sense embeddings using lexical resource BabelNet [52] for measuring semantic similarity. This approach is an effective measure for word similarity tasks, yet this can be enhanced by making the model understand the context of words. Chen et al. [33] presented a model for word sense representation. This model considers a single representation per single sense. This model is considered as the basis for the innovation of presenting the word sense disambiguation with embedding techniques for similarity tasks. This model does not learn the relationships like hypernyms and hyponyms between the concepts from Wordnet. The model presented by Oele et al. [34], combines word-sense, context, and word-definition to develop embeddings. This approach [34] developed lexeme embeddings for senses using the Lesk algorithm [36] and an AutoExtend [37] training procedure. The approach mentioned by Ruas et al, [35] disambiguates the text using a context window. The authors [35] explain the limitations of single vector representations and the advantages of multi-sense embeddings. This model [35] derived synonym set based embeddings by integrating Wordnet synonym sets and Word2vec model [18] model.

Word embeddings mechanisms with single vector representations present all the senses in a single vector. There is limited work done in this area. Hence, this paper focuses mainly on generating vector representations for each sense and the context of words is selected from top-n context-rich sentences from Wikipedia. Word embedding models capture the taxonomic information of words but fail in capturing hyponymy and entailment relations. To understand the relation between two words, the word`s synonyms and hyponyms play a crucial role. The above mentioned works in the literature [32-35] developed multi-sense embeddings but with little attention paid on synonym and hypernym relations. This paper projects the importance of the Wordnet hierarchy for understanding synonym sets, hypernyms, lowest common subsumer of concepts to generate multi-sense embeddings for word similarity tasks.

The next section portrays the proposed approach for generating multi-sense embeddings based on the word's synonyms, parts-of-speech, hypernyms. In the next section, a detailed exploration of Wordnet is presented at the start, followed by the explanation of working with multi-sense embeddings.

III. Proposed Methodology

In this section, the main idea of multi-sense embeddings using Wikipedia data and lexical database Wordnet is discussed. This section discusses developing Wikipedia corpus at the start, which is used by the proposed algorithm. The methodology presented in this section has two main tasks: (i) exploring Wordnet lexical structure (ii) using knowledge of synonym sets, hypernyms to produce multi-sense embeddings. The last subsection applies the multi-sense embeddings developed to perform the word similarity tasks.

A. Pre-Processing of Wikipedia Articles

The initial process is to transform the Wikipedia articles into a corpus and then map the words with synonym sets of Wordnet. Initially, the Wikipedia dump is preprocessed to form the corpus. The latest Wikipedia dump is around 15.7 GB of size and has 4,677,566 documents with more than 19,508,987 articles. The articles in the dump are in the XML format, which needs to be converted to text format. This research converts the Wikipedia dump into text format. Articles shorter than 50 words, article name starting with numbers, articles which are not standard words are pruned. From the text, punctuation and other unknown symbols are removed. The tokens which occurred

more than ten times in the articles are preserved, and rest is removed since it is challenging to obtain their context. After cleaning the entire text, more than 2 million unique tokens are formed.

Later, these tokens are trained using the Word2vec model. To train the model, this paper uses the Word2vec with default hyperparameters. CBOW model of Word2vec [18] is the training algorithm, with a window size of 15 and vectors of size 300 dimensions (300 d) are the parameter specifications to generate embeddings. After this, all the possible senses obtained from Wordnet also are assigned vectors from the trained Word2Vec model. For each sense of the word, 300 dimensional vectors are generated. This paper proposes an efficient and simple training phase with minimal hyperparameters to understand the sense embeddings. Once this training phase is completed, Wordnet is explored to understand to generate all possible senses of a given word.

B. Exploring Wordnet Hierarchy

One of the main relationships that can exist among the concepts in Wordnet is synonymy. Wordnet has 117,000 synonym sets (synsets), and each of these is linked to one another using conceptual relations. A word may have many forms, every form has a distinct synset and these are uniquely represented in Wordnet. The relation between the synsets is an ancestor-child relation, this kind of relationship is referred to as hyponymy or IS-A relation.

Polysemous words can be used in many senses; the synonym set of a concept refers to a particular sense. If two words are similar, then they share a common synset in the network. Traditional knowledge-based measures [11]-[16] handle the polysemy and synonymy of the words using the taxonomic structure of the semantic network. These measures derive the synonym sets of the words using the hierarchical semantic relationships. Any model developed using ontologies for deriving semantic similarity should be capable of exploring different kinds of relations like hierarchical (IS-A), has-a, hypernyms, etc. The hierarchy between the concepts is a significant relationship as it maps the category of objects into a taxonomy.

The hierarchical structure of the concepts is provided by taxonomy. Each taxonomy concept is represented as a node, and the nodes are connected based on hierarchical relations in the network. The edges of the concepts represent the semantic relationship. WordNet [12] is a semantic network of English words. In this network, each word is represented as a synset. The noun network in WordNet was developed very richly using hyponymy/hypernymy hierarchy.

The maximum depth of the noun hierarchy in this network is 16 nodes. The noun network includes nine types of relations; synonymy relation, hyponymy (IS-A) relation, and its inverse hypernymy, and six meronymic (PART-OF) relations. Synonymy relations account for 80 percent of the relations. Hyponymy relation between the words articulates IS-A relationship between two words and the inverse of this relation is hypernymy. COMPONENT-OF, MEMBER-OF, SUBSTANCE-OF and their inverses are meronymic relations.

Fig. 1 shows an example of Wordnet's "is-a" relation structure. The concepts in Wordnet are arranged in the hierarchical structure. From Fig. 1, it can be noticed that the root of all concepts is 'Entity.' The figure shows only a fragment of Wordnet structure. The leaves 'mammal,' 'tree,' of the tree is under 'organism.' The leaf node 'mammal' is-a 'vertebrate' and 'vertebrate,' is-a 'chordate.' The concept 'chordate,' is-a 'animal.' Similarly, 'carrot' is-a 'root,' and 'root' is-a 'plant-organ,' which is-a 'plant-part.' This organized representation in Wordnet helps in obtaining various senses of concepts. Fig. 1 represents the basic hierarchical cover of three concepts 'tree,' 'mammal,' and 'carrot.' This figure does not cover the various classifications under the concepts like 'organism' or under 'plant.' Once again, the concept like 'mammal' has various subconcepts like cat, dog etc. All these are well organized in the Wordnet.
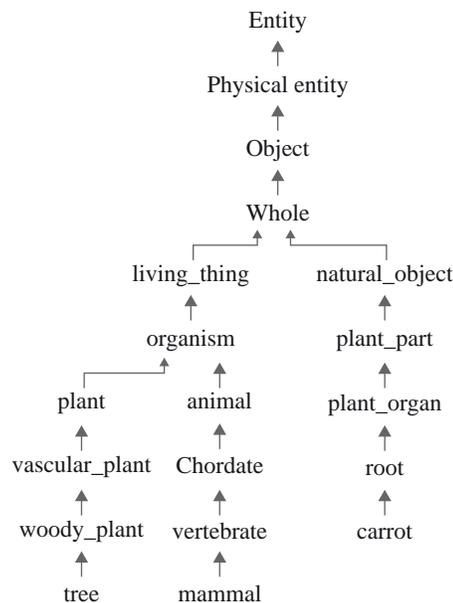
Fig. 1. Wordnet is-a relation example.

A concept is a synset of Wordnet that exists in the form as 'concept. pos.number.' The first part in the synset represents the concept, pos represents the possible parts-of-speech (noun, verb, adjective and adverb), and the number indicates the number of applicable parts-of-speech that exist for the concept. Consider the concept 'advance,' the possible synsets for the concept is shown in Fig. 2. From Fig. 2, it can be noted that the concept 'advance' is having a relationship with six noun synsets, twelve verb synsets, and two adjective synsets. Each synonym set in Fig. 2 is a different sense of the word 'advance.'

Hypernyms in Wordnet specify the hierarchical structure of a concept and from the hypernyms of the two concepts, common ancestors can be checked. Checking the path of the hypernyms of

```
Synset('progress.n.03'), Synset('improvement.n.01'),  Synset('overture.n.03'),
Synset('progress.n.02'), Synset('advance.n.05'),  Synset('advance.n.06'),

Synset('advance.v.01'), Synset('advance.v.02'),  Synset('boost.v.04'),
Synset('promote.v.02'), Synset('advance.v.05'),  Synset(gain.v.05'),
Synset('progress.v.01'), Synset('advance.v.08'),  Synset('promote.v.02'),
Synset('advance.v.10'), Synset('advance.v.11'),  Synset('advance.v.12'),

Synset('advance.s.01'), Synset('advance.s.02')
```

Fig. 2. Possible synonym sets for the concept 'advance' obtained from Wordnet.

```
Synset('act.v.01')          Level 1    Synset('act.v.01')
Synset('interact.v.01')                Synset('interact.v.01')
Synset('communicate.v.02')             Synset('communicate.v.02')
Synset('inform.v.01')                  Synset('inform.v.01')
Synset('tell.v.02')         Level 5    Synset('tell.v.02')
Synset('impart.v.01')                  Synset('impart.v.01')
Synset('convey.v.01')                  Synset('convey.v.01')
Synset('express.v.01')                 Synset('express.v.01')
Synset('stress.v.01')       Level 9    Synset('stress.v.01')
                                       Synset('bring_out.v.04')
Hypernyms of the concept "accentuate"  Synset('foreground.v.01')


                              Hypernyms of the concept "highlight"
```

Fig. 3. Hypernym paths for concepts 'accentuate' and 'highlight'.

two concepts helps in recognizing additional senses of word pairs. Hypernyms of two concepts 'accentuate' and 'highlight' are shown in Fig. 3. From the figure, it can be noticed that the concepts have more common ancestors. The length of the hypernym path for the concept 'accentuate' is 9, and that of 'highlight' is 11. Ancestors of concept 'accentuate' are up to level 9, the depth of the concept from the root is 9, and that of 'highlight' is 11 from the root. Both concepts have the same root up to level 9. Hence, these concepts are more similar. Wordnet provides different modules to obtain synonym sets, hypernyms, and other relations between the concepts. The proposed approach uses the information of synonym sets of a word, hypernym path of the words to generate the appropriate synset, hypernym path based sense embeddings for evaluating word similarity. Another important information this paper considers to derive multi-sense embeddings while performing the word similarity task is least common subsumer (LCS). Least common subsumer (LCS) is the least possible ancestor of the two concepts in the taxonomy. With the concepts represented in Fig. 1, it is not possible to understand entirely about LCS. There are many underlying concepts under a category, hence a more complex tree structure is shown in Fig. 4, to understand the LCS of the concepts.

Fig. 4 is an instance of Wordnet taxonomy. In Fig. 4, A is the root node and every node is a concept (represented as the synonym set) of Wordnet. Distance between two nodes is the shortest path possible from one node to another. Consider nodes H and L, the length between H and L is 6, least common subsumer of H, L is A. Similarly, LCS(F, L) is C and LCS(E, I) is B.
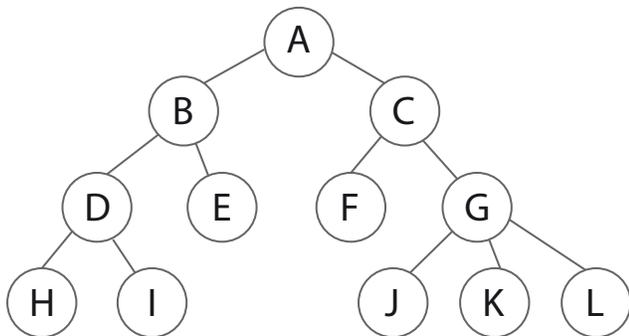
Fig. 4. A sample Tree in WordNet.

C. Methodology

In this subsection, the process of generating the multi-sense embeddings using synonym sets and hypernyms is presented. Fig.

5 shows the block diagram of the proposed approach. Multi sense embeddings are generated using the Wordnet and Wikipedia text. Extracting multiple senses of a word is the crucial step in the entire process. In order to extract the senses, synonym sets and hypernym path of the word from Wordnet is used.
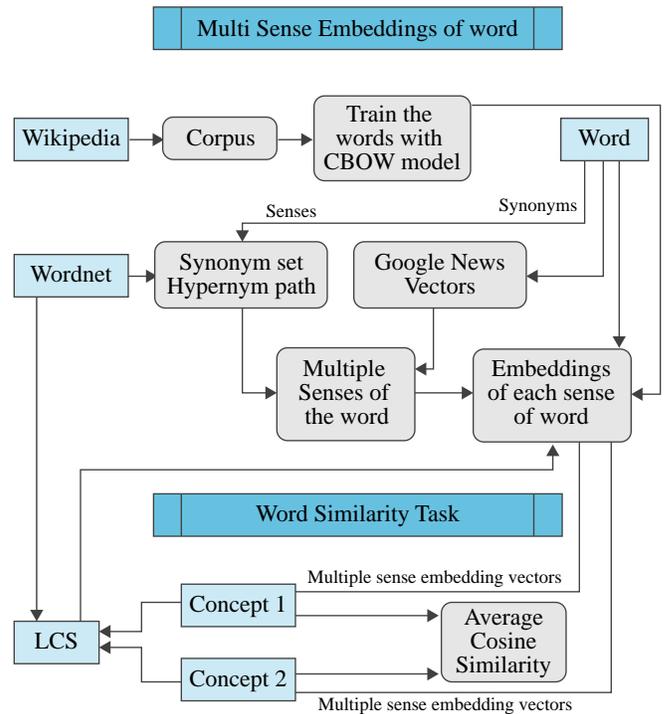
Fig. 5. Block Diagram of generating Multi-sense Embeddings and performing Word similarity task.

There are some concepts with fewer or zero synsets in Wordnet. This paper handles the words with fewer synsets using Google News vectors[1] to derive synonyms.

The similarity of the synonyms with the respective word is calculated. Synonyms which are in certain threshold are considered as possible senses of the respective word. Once the senses are obtained, each sense is converted into a 300-dimensional vector from the pre-trained Wikipedia CBOW model. The words in the corpus are trained on Wikipedia text using the CBOW model as it is an efficient model and deals with the context of words in more massive datasets

1 https://code.google.com/archive/p/word2vec/

quite well. The block diagram also shows how to perform the word similarity task. The similarity of the two words is obtained using their multi-sense embedding vectors. Since there are multiple sense vectors for each word, the average cosine similarity between them is calculated to get the final similarity. This paper also inspects another similarity metric called Tanimoto similarity [42] along with cosine similarity. The discussion about these metrics is mentioned in the next subsection. The block diagram shows how to obtain the multi-sense embeddings of a single word and similarity between two concepts. For the entire corpus with more than 2 million tokens, the same procedure is applied.

Once the information of synonym sets and hypernyms path are extracted from Wordnet, the next step is to generate the sense embeddings. The approach mentioned in this section is not explored in any of the previous works, the results obtained with this approach highlights the efficiency of this work. Algorithm I, Sense Embedding using Wordnet Synonym sets, gives the idea of generating sense embedding of a word C_i.

Algorithm I: Sense Embedding using Wordnet Synonym sets: Procedure SenseEmbed(C_i)

Input: Word (C_i), preprocessed Wikipedia Corpus, Wordnet, context of the word.
Output: Multi-Sense embeddings of the word
Step 1: Generate the synonym sets for concept C_i,
   $SS(C_i) = [s_1, s_2, s_3,....s_n]$, *n* synonym sets for C_i
   SS_Preserve(C_i)={}, empty set
Step 2: Group synonym sets based on parts-of-speech.
Step 3: Preserve the most suitable sense by checking the definitions of each synonym set in Wordnet
   pos ← POS(word (C_i)) from context
   SS_POS ← pos(SS(C_i))
   for j in 1 to len(SS_POS):
      if def(SS_POS[j]) matches context:
         SS_Preserve(C_i) ← SS_POS[j]
         SS_Preserve(C_i) ← lemmas(SS_POS[j])
      end if
   end for
Step 4: Generate the hypernym path of each synonym set to add appropriate senses.
   for j in 1 to n:
      Hypernyms = synsets(hypernym_path(s_j))
      if hypernym_path(s_j) exists:
         for q in 1 to len(Hypernyms):
            if Similar(Hypernyms[q], word(C_i)) >= δ
               Add Hyper to SS_Preserve(C_i)
            end if
         end for
      end if
   end for
Step 5: Extract top-m related synonyms of C_i
   top_m(Syn)<-Google News vectors
Step 6: Update SS(C_i) of C_i
   for k in 1 to m:
      if top_m(Syn) not in SS(C_i)[k]
         Add item to the candidates of SS_Preserve(C_i)
      end if
   end for
Step 7: Generate word vector representations for each sense in *SS_Preserve(C_i)* using trained Word2vec representations.
   for k in 1 to len(SS_Preserve(C_i)):
      SEV(C_i) ← Word2Vec(SS_Preserve(C_i)[k]
   end for

Algorithm I, SenseEmbed(C_i), takes a word (C_i), preprocessed Wikipedia corpus, Wordnet as input. The output of the algorithm is multi-sense embeddings $SEV(C_i)$ of the word C_i. The articles in Wikipedia are preprocessed by removing punctuations, HTML tags, followed by removing stop words from the document. The words are then lemmatized to their root form. A corpus of Wikipedia words is formed after this. Wikipedia corpus is an input for the algorithm to check the word for which multi-sense embeddings need to be generated is present or not. Lexical database Wordnet and the context of the word are other important inputs to the algorithm. Context usually is a sentence or a sequence of words in which the required word (C_i) is present. Based on context, different related synonyms of the sense are extracted. Algorithm I generates multi-sense embedding vectors as output for the given input word (C_i) based on seven steps.

The first part of the algorithm from steps 1 to 4 majorly relies on Wordnet to extract multiple senses of the word. After step 1 in the algorithm, the possible synsets of the word are in $SS(C_i)$. Appropriate synsets of the words are stored in $SS\_Preserve(C_i)$, which is empty at the start and is updated in steps 3, 4 and 6. In step2, the synonym sets are grouped based on their parts-of-speech and the tagging of synonym sets is removed. Grouping synonym sets based on their parts-of-speech helps to obtain synsets required for the context.

In step 3, POS(word (C_i)) is an internal procedure that returns the parts-of-speech of the input word. Based on this, the synsets of the word are selected and stored in SS_POS. The synsets definitions are matched with the context of the word to update $SS\_Preserve(C_i)$. To explain in further consider the words shown as examples in Table I. Table I shows the information of two words, 'tiger' and 'sofa.' For the word 'tiger,' a context, synsets and definitions of synsets are mentioned in the table. Similarly, for the word 'sofa,' synsets and lemmas are mentioned in Table I. The context of 'tiger,' matches the definition of 'tiger.n.02', and the lemmas of this sense are added to $SS\_Preserve(C_i)$. The lemmas of a synset are the synonyms of the word in a particular sense. There are certain cases where the algorithm has to handle the words with no context. For these words, the algorithm considers all the synsets and their lemmas. The second word 'sofa' in the table has only one synset and the context of this word is not given. The lemmas of the words are added to $SS\_Preserve(C_i)$.

TABLE I. Synsets, Definitions and Lemmas for Example Words 'Tiger' and 'Sofa' From Wordnet

Word ------- 'tiger'	
Context	'tiger of snows'
Synsets	Synset('tiger.n.01'), Synset('tiger.n.02')
Definition of 'tiger.n.01' Definition of 'tiger.n.02'	'a fierce or audacious person' 'large feline of forests in most of Asia having a tawny coat with black stripes; endangered'
Lemmas of 'tiger.n.02'	'tiger', 'Panthera_tigris'
Hypernyms of 'tiger.n.02'	Synset('big_cat.n.01')
Hypernym path of 'tiger.n.02'	'entity.n.01'→'physical_entity.n.01'→ 'object.n.01' → 'whole.n.02' → 'living_thing.n.01' → 'organism.n.01' → 'animal.n.01' → chordate.n.01' → 'vertebrate.n.01'→ 'mammal.n.01' → 'placental.n.01' → 'carnivore.n.01' → 'feline.n.01'→ 'big_cat.n.01'→'tiger.n.02'
Word ------ 'sofa'	
Context	----
Synsets	Synset('sofa.n.01')
Definition of 'sofa.n.01'	'an upholstered seat for more than one person'
Lemmas of 'sofa.n.01'	'sofa', 'couch', 'lounge'

Algorithm II : Word Similarity using Sense Embed vectors and Wordnet information:: WordSimilarity(C_i,C_j)

Input: Words C_i, C_j

Output: Similarity of C_i, C_j

Step 1: Multi sense vectors for C_i, C_j

    SEV(C_i) ← SenseEmbed(C_i)

    SEV(Cj) ← SenseEmbed(C_j)

Step 2: Update *SEV(C_i)*, *SEV(C_j)* using LCS information from Wordnet, hypernym path of both concepts

    if *LCS(C_i, C_j) is TRUE*:

        Synsets(C_i, LCS) ← *Path(C_i,LCS)*

        for j in 1 to len(Synsets(C_i, LCS)):

            if Similar(Synsets(C_i, LCS) [j], C_i) >= δ

                Add Vec(Synsets(C_i, LCS) [j]) to SEV(C_i)

            end if

        end for

    end if

    // Repeat above for SEV(C_j), C_j

        else

    path1← *hypernym_path(C_i)*

    path2 ← *hypernym_path(C_j)*

    for k in len(path1):

        for l in len(path2):

            if path1[k]== path2[l]

                if path1[k] not in SEV(C_i)

                    update SEV(C_i)

                end if

                if path1[k] not in SEV(C_j)

                    update SEV(C_j)

                end if

            end if

        end for

    end for

Step 3: Similarity of the concepts C_i, C_j

    for p in 1 to len(SEV(C_i)):

        for q in 1 to len(SEV(C_j)):

            sim[p][q] = M(SEV(C_i)[p], SEV(C_j)[q])

            //Where M is either Cosine or Tanimoto similarity

        end for

    end for

    //average of similarity

There are certain words with fewer or no synsets in Wordnet, for these words hypernyms and hypernym paths are considered to extract possible senses. Step 4 discusses these details. The word 'tiger' in Table I, has only two senses with only one sense is matching the context. At the same time, the lemmas (synonyms) of the word are also two. Hence, navigating through the hypernym path of such words, helps in increasing the multiple senses. The hypernym path of synset 'tiger.n.02' is shown in Table I, from the path it is observed that the head portion of the path contains more generalized synsets. But the synsets which are two or three levels up the synset 'tiger.n.02' are more specific to the synset. Hence, these senses are also examined to update them in *SS_Preserve(C_i)*. To determine, whether these can be added or not, cosine similarity between the synset in the hypernym path and word are determined. If the similarity is greater than δ, the synset is added to *SS_Preserve(C_i)*. The value this paper considers for the δ is 0.5.

Steps 5 and 6 in Algorithm I handle the words with no synsets and hypernym paths in Wordnet by extracting synonyms of the word using pre-trained Google News vectors. From step 5, top-m related synonyms of Ci are extracted. The value of m chosen is 10. In step 6, these are added to *SS_Preserve(C_i)*.

After step 6, *SS_Preserve(C_i)* contains synsets, lemmas (synonyms) and certain synsets from hypernym path or synonyms from Google News vectors. As the synonym sets are in the form 'concept.pos. number,' it is not possible to generate the embeddings for words with these representations hence tagging is removed. For each sense present in *SS_Preserve(C_i)*, a vector of 300 dimensions is obtained from the trained model discussed in subsection A. The context sliding window in this approach is similar to the one used in CBOW [18]. The word vector representations for each sense of *SS_Preserve(C_i)* are stored in *SEV(C_i)*.

The vector representations stored in *SEV(C_i)* are based on the synsets, lemmas (synonyms), hypernym path information from Wordnet, and synonyms extracted from Google News vectors. These vectors can be used in word similarity, analogy, entailment tasks. In the next subsection, the multi-sense embeddings are used to carry out word similarity tasks.

D. Word Similarity Task Using Multi-Sense Embeddings

In the previous subsection, the procedure to obtain multi-sense embedding representations of a word is presented. This subsection uses the methodology of multi-sense embeddings to carry out word similarity tasks. For two concepts or words, C_i, C_j, Algorithm I generates the multi-sense vector representations as,

$$SEV\left(C_i\right)=\left[s_{11}, s_{12}, \dots s_{1n}\right]$$
$$SEV\left(C_j\right)=\left[s_{21}, s_{22}, \dots s_{2m}\right]$$

Where C_i, C_j represents words, s_{1i} represents sense embedding vector 'i' of word C_i and s_{2j} represents the sense embedding vector 'j' of C_j. In the above representations, the number of senses for concepts C_i and C_j are n and m, and both greater than one.

Algorithm II obtains the similarity of two words C_i, C_j, using *SEV(C_i), SEV(C_j)* vectors. The initial step in the algorithm is to generate Multi sense embedding vectors of the concepts using Algorithm I. The next step is to update *SEV(C_i), SEV(C_j)*. The idea here is to utilize all the related senses of the words to derive similarity. Wordnet is searched for the least common subsumer (LCS) of words C_i, C_j. LCS is the shared ancestor of two concepts.

Using LCS information of C_i and C_j, *SEV(C_i), SEV(C_j)* are updated. Step 2 in Algorithm II updates the sense vectors by checking the path in between LCS to C_i. The list *Synsets(C_i, LCS)* in the algorithm holds all the synsets in the path from C_i to LCS. The synsets which are similar to the word C_i are added to *SEV(C_i)*, by calculating the *Similar(Synsets(C_i, LCS) [j], C_i)*. If this value is greater than δ, then the synset is converted into a vector of 300 dimensions from the pretrained CBOW model on Wikipedia text. The same procedure is repeated to update *SEV(C_j)*. If LCS of concepts does not exist, the hypernym path of each concept is examined. It is observed from the experimentations that this step hardly updates the sense vectors, since individual hypernym paths are checked in Algorithm I. But this is done not to miss out on any possible senses. Algorithms I and II check all the possible senses, which increases the overall efficiency of the proposed approach.

Step 3 in Algorithm II calculates the average similarity of the two multi vectors, *SEV(C_i), SEV(C_j)*. MSSA approach [35] gives two metrics for evaluating the approaches, namely average similarity and maximum similarity. If the maximum similarity is considered,

similarity obtained will be of only one sense in each word. Hence, this paper uses average similarity of all the senses. Given $SEV(C_i)$, $SEV(C_j)$, the similarity is calculated using the following equation,

$$Similarity(C_n, C_m) = avg(M(s_{1i}, s_{1j})),$$
$$i \in (1, n), j \in (1, m)$$

(1)

Where $M(s_{1i}, s_{1j})$ gives the similarity of sense embedding vectors s_{1i}, s_{1j}, M is either Cosine or Tanimoto similarity metric for assessing the vectors. If the words are represented as vectors, the task of finding the similarity between them is simpler. There are many measures that calculate the similarity between the vectors, like Cosine, Tanimoto [42], etc. The most used metric for Word2vec models is cosine similarity. The vectors with the same orientation are more similar and tend to have a value of 1 with cosine similarity. The results of state-of-art approaches are based on the cosine similarity between the vectors. To estimate the strength of the proposed approach with state-of-art techniques cosine similarity is considered. Tanimoto metric [42] is mainly used to calculate the similarity of vectors with binary data. But this metric can be easily enhanced to apply on vectors with continuous data. Tanimoto [42] metric for continuous data is used in this paper. Moreover, this metric is analogous to cosine similarity. Hence, this metric is also considered for measuring the similarity.

$$Cosine(V_1, V_2) = \frac{V_1 \star V_2}{|V_1||V_2|}$$

(2)

$$Tanimoto(V_1, V_2) = \frac{V_1 \star V_2}{|V_1^2 + V_2^2 - V_1 \star V_2|}$$

(3)

This section proposes multi-sense embeddings using Wikipedia data and lexical database Wordnet. Multiple senses of words are extracted from the synsets, definitions, hypernym path and Google News vectors. These vectors are enhanced using LCS information of concepts to perform word similarity tasks. The next section discusses the experimentation details, datasets and models compared with the proposed approach.

IV. Results

In this section, experiments related to word similarity task are presented. The proposed algorithm and its comparison against other approaches are shown in detail.

A. Experimentation Setup

The proposed methodology in this paper is implemented using Python 3.6 with NLTK 3.4[2] and gensim 3.4[3] libraries. NLTK is used to get the information of Wordnet's synsets, definitions, lemmas, LCS of concepts and hypernym path. This paper builds a corpus from Wikipedia dump[4] which is open-source. CBOW Word2vec [18] is used to train the corpus of words. The hyperparameters of the model are as follows: CBOW model for training, a window size of 15, vector size of 300 dimensions. Google News Vectors[5] used in Algorithm I is a pre-trained model with tokens and respective word vectors.

B. Datasets

The benchmark datasets used in this paper are described as follows:

1. MC dataset [43]: MC dataset is a subset of the RG dataset [44],

---

2 https://www.nltk.org/

3 https://radimrehurek.com/gensim/models/word2vec.html

4 https://dumps.wikimedia.org/enwiki/latest/

5 https://code.google.com/archive/p/word2vec/

this dataset has 30 word pairs. Two word pairs are not present in Wordnet and hence the researchers work on 28 word pairs. The rating for each word pair in the dataset is in the range from 0 to 4, higher rating indicates word pairs are more similar.

2. RG dataset [44]: This dataset contains 65 word pairs and rating is in the range 0 to 4.

3. WS-353 dataset [45]: This dataset has 353 noun pairs, the rating of word pairs are in a range from 0 to 10. The dataset is divided into two sets. The first set is with 153 word pairs and the later with 200 word pairs. The complete version is considered in experimentation.

4. MTurk771 dataset: Halawi et al, [46] proposed this dataset with 771 word pairs and each word pair is given a rating in the range 0 to 5.

5. MTurk287 dataset [47]: This dataset has 287 word pairs, the rating of each word is in the range 0 to 5. All the word pairs are noun pairs.

6. SCWS (Stanford Context Word Similarity) dataset [48]: This dataset has 1997 word pairs, for each word in the word pair context is mentioned in the dataset. Word pairs are given a rating in range from 0 to 10.

7. Rare words dataset [49]: This dataset has 2034 word pairs and each word pair is given a rating in the range 0 to 4.

8. Men dataset [50]: This dataset has 3000 word pairs. This dataset has word pairs of noun, verb, and adjectives. The rating of each word ranges from 0 to 1.

9. Simlex999 dataset [51]: This dataset has 999 word pairs, out of these 999 word pairs, 666 word pairs are nouns, 222 are verb pairs and 111 adjective pairs. For experimentation purposes, 666 noun pairs (Simlex666) are considered.

C. Models Compared

The following models are compared with the proposed approach on benchmark datasets,

1. CBOW: Mikolov et al. [18] proposed Word2vec with two word embedding models, namely Skip-gram and CBOW. Skip-gram aims to predict a context word in a window around a target word. CBOW predicts the current target word using the representations for its context words. It is mentioned in several works that CBOW is more efficient, hence this approach is selected as one of the state-of-art methods for comparison.

2. GloVe: Pennington et al. [30] proposed a log-bilinear model called GloVe for learning word vectors. This is based on global matrix factorization and local context window methods. Local window methods like Word2vec[18], fail to capture global statistical information in the corpus, GloVe overcomes this drawback.

3. fastText [31]: Word2vec [18] and GloVe [30] ignore morphology of words. Hence, these approaches produce more out of vocabulary words. The fastText approach overcomes this problem by representing each word as a bag of character n-grams.

4. SenseEmbed [32]: SensEmbed uses BabelNet [52] for word sense disambiguation. This model trains a word2vec model (400 dimensions). This approach integrates structural knowledge from BabelNet and the distributional sense representations.

5. Chen et al. [33]: This approach performs disambiguation of words to learn word sense representations. It has two sub-tasks, namely L2R (left to right) and S2C (simple to complex). L2R disambiguates words from left to right and S2C selects senses of the word. S2C selects only those senses with a certain threshold.

6. MSSA: MSSA approach [35], mentions that Chen et al. [33] considers only specific senses leading to poor representations.

Hence, MSSA explores all senses available for a word. MSSA-T and MSSA-D approaches mentioned in this work are taken for comparison.

7. Paragram : Wieting et al. [42] proposed a word embedding model based on paraphrase phrases in PPDB database [53].

8. Counter-fitting [41]: This approach considers antonymy and synonymy constraints from the PPDB database and Wordnet.

9. Attract-Repel: Mrksic et al. [39] proposed Attract–Repel approach with synonymy and antonymy constraints to develop embeddings. In this approach, semantic relations are taken from BabelNet [52].

D. Metrics

This paper uses two evaluation metrics, the Spearman rank correlation, the Pearson correlation factor. Pearson correlation was first given in the studies of synonymy between words by Rubenstein and Goodenough [43]. A majority of the researchers in their works evaluated their approaches with Pearson correlation. Hence, this paper uses this evaluation metric to compare the results with state-of-the-art approaches. Another metric that is used by all the researchers for word similarity tasks is Spearman correlation. Hence, this metric is also used for comparing the results. Pearson correlation coefficient measures the strength and direction of the linear relationship between the variables. Whereas, the Spearman coefficient measures the correlation between the ranks of two variables. There are fewer publications that reported the results of both Spearman and Pearson correlation coefficient values. The unavailable data are not mentioned in the results.

The experimentations of the approaches Word2vec [18], GloVe [30], fastText [31] are implemented by us. The results of the approaches [39], [41], [42] are obtained from the HESML library[6]. The results of the approaches [32] and [33] for some datasets are reported in [35]. The MSSA [35] approach reported results for only Spearman correlation. The results reported in the tables for [32], [33] are taken from MSSA [35], as it became difficult to set up the same environment and to produce results.

Table II presents the results of the proposed approach against several state-of-the-art distributional approaches on the MC dataset [43]. On this dataset, Attract-Repel, SenseEmbed, and proposed approaches have obtained the highest Spearman correlation results. The proposed method obtained the highest results for the Pearson correlation, which is 0.851.

TABLE II. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on MC-28 Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
SenseEmbed [32]	0.880	--
MSSA-T [35]	0.796	--
MSSA-D [35]	0.835	--
Attract-Repel [39]	**0.884**	0.837
Counter-fitting [41]	0.857	0.806
Paragram [42]	0.824	0.796
CBOW [18]	0.781	0.796
GloVe [30]	0.862	0.845
FastText [31]	0.845	0.842
Proposed approach (cosine)	0.865	**0.851**
Proposed approach (Tanimoto)	0.860	0.845

6 https://github.com/jjlastra/HESML

Table III shows the results of the proposed approach against several models for the RG [44] benchmark dataset. In this experiment, SensEmbed and proposed approach obtained the highest results for the Spearman correlation. The approach SenseEmbed [32] develop multi-sense vectors using BabelNet [51]. BabelNet is a lexical database composed of different resources. This approach trains the Word2vec model with 400 dimensions. This approach obtained better results on smaller datasets when the size of the dataset increases, the results of SenseEmbed approach are not promising as they are for MC and RG datasets.

TABLE III. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on RG-65 Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
SenseEmbed [32]	**0.871**	--
MSSA-T [35]	0.776	--
MSSA-D [35]	0.801	--
Attract-Repel [39]	0.825	0.840
Counter-fitting [41]	0.808	0.806
Paragram [42]	0.813	0.810
CBOW [18]	0.760	0.772
GloVe [30]	0.755	0.770
FastText [31]	0.801	0.793
Proposed approach (cosine)	0.863	**0.851**
Proposed approach (Tanimoto)	0.865	0.847

MC [43] benchmark is a subset of the RG [44] benchmark dataset. The words in the MC benchmark and RG are nouns. Algorithm I to generate the multi-sense embeddings of these datasets uses noun synsets from Wordnet. Some of the word pairs in these datasets are having very few noun synsets. For the words with one or two synsets, lemmas (synonyms) are generated. For words with fewer synonyms tracing the hypernym path gives multiple senses. There is minimal variation in the Spearman and Pearson correlation for cosine and Tanimoto metrics because both are analogous. The human ratings of half of the word pairs in MC benchmark are meager, suggesting that these word pairs are less similar. Out of 28 word pairs in the dataset, the human ratings of 10 word pairs are below 1.0, the scale of the ratings is in the range [0,4]. Hence, the approaches like Attract-Repel [39], Counter-fitting [41] handle the antonym relations work well on this dataset. The multi-sense embeddings approaches also worked well on both MC and RG benchmarks since all the senses are covered. The SenseEmbed [32] approach uses BabelNet [52], which has a wide coverage of synsets. Hence it is able to achieve better results on these two benchmarks and also on other datasets.

Table IV presents the Spearman and Pearson correlation values of the approaches on the WS-353 dataset. From the results, it can be observed that the proposed approach outperformed all the other state-of-art approaches. WS-353 dataset is an interesting one because the dataset is used for carrying out semantic relatedness and semantic similarity tasks. The dataset has a collection of nouns, verbs, and adjectives. Algorithm I groups the synsets based on the parts-of-speech while generating the synsets, and hence, the proposed approach is able to achieve better results for both cosine and Tanimoto similarities.

TABLE IV. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on WS-353 Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
SenseEmbed [32]	0.779	--
MSSA-T [35]	0.694	--
MSSA-D [35]	0.708	--
Attract-Repel [39]	0.666	0.608
Counter-fitting [41]	0.680	0.615
Paragram [42]	0.764	0.679
CBOW [18]	0.603	0.642
GloVe [30]	0.716	0.713
FastText [31]	0.738	0.698
Proposed approach (cosine)	**0.817**	**0.836**
Proposed approach (Tanimoto)	0.814	0.832

Tables V and VI represent the results on Turk771 and Turk287 datasets. Semantic relatedness tasks use these two datasets. The difference between semantic similarity datasets and relatedness datasets is in the former, the concepts are more specific, and in the later, the concepts are more general. The proposed approach handles both the types because it observes both specific (using synonyms) terms, and general (using hypernyms) terms efficiently. The proposed approach obtained the highest results for Spearman correlation on both benchmarks. The single vector representation model GloVe achieved the highest results for the Pearson correlation.

TABLE V. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on Turk771 Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
Attract-Repel [39]	0.599	0.590
Counter-fitting [41]	0.701	0.666
Paragram [42]	0.745	0.704
CBOW [18]	0.672	0.698
GloVe [30]	0.715	**0.749**
FastText [31]	0.661	0.728
Proposed approach (cosine)	**0.761**	0.733
Proposed approach (Tanimoto)	0.757	0.727

TABLE VI. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on Turk287 Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
Attract-Repel [39]	0.606	0.618
Counter-fitting [41]	0.639	0.630
Paragram [42]	0.699	0.701
CBOW [18]	0.674	0.650
GloVe [30]	0.724	**0.749**
FastText [31]	0.709	0.728
Proposed approach (cosine)	**0.726**	0.736
Proposed approach (Tanimoto)	0.724	0.732

SCWS dataset has 1997 word pairs. Each word in the word pair has a context sentence, parts-of-speech of the word is mentioned for each word. SCWS dataset has four different sets of word pairs. The majority of the word pairs in the dataset have the same parts-of-speech, and some word pairs are with different parts-of-speech.

- The dataset has 1323 pairs, which are nouns.
- The dataset has 399 verb pairs.
- The dataset has 97 adjective pairs.
- Remaining 178 word pairs are with different parts-of-speech like noun-verb, verb-noun, verb-adjective, etc.

The results of the SCWS dataset are given in Table VII. The state-of-art methods produced competitive results with Spearman correlation but failed to produce the same with Pearson correlation. But the proposed approach is consistent in obtaining the highest results on this dataset also. Algorithm I handles words with different parts-of-speech, that enabled the proposed approach to achieve better results on this dataset in comparison with state-of-art methods.

TABLE VII. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on SCWS Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
SenseEmbed [32]	0.624	--
MSSA-T [35]	0.649	--
MSSA-D [35]	0.640	--
Chen et al. [33]	0.662	--
Attract-Repel [39]	0.587	0.113
Counter-fitting [41]	0.611	0.114
Paragram [42]	0.691	0.115
CBOW [18]	0.643	0.105
GloVe [30]	0.624	0.106
FastText [31]	0.652	0.106
Proposed approach (cosine)	**0.693**	**0.658**
Proposed approach (Tanimoto)	0.690	**0.658**

Table VIII shows the results of the approaches to the Rare words dataset. This dataset has nouns, verbs and adjective word pairs. The proposed approach has obtained the highest results of Spearman and Pearson correlation on this dataset also because Algorithm I handles different parts-of-speech efficiently.

TABLE VIII. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on Rare Words Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
Attract-Repel [39]	0.273	0.319
Counter-fitting [41]	0.207	0.288
Paragram [42]	0.536	0.505
CBOW [18]	0.456	0.438
GloVe [30]	0.451	0.440
FastText [31]	0.464	0.432
Proposed approach (cosine)	**0.543**	**0.544**
Proposed approach (Tanimoto)	0.535	0.535

Table IX shows the Spearman and Pearson correlation values on Men dataset. This dataset is a collection of 3000 word pairs, with all the parts-of-speech included in it. The proposed approach continues to outperform all the state-of-art approaches on this dataset also. The proposed approach obtained the highest values of *0.836, 0.815* for Spearman and Pearson correlation, respectively. From the results obtained on all the datasets, it is observed that the proposed approach performed well on different benchmarks. It is because of the procedure followed to generate the multi-sense embeddings.

TABLE IX. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on Men Dataset, Highest Results Highlighted In Boldface

Method	Spearman correlation	Pearson correlation
SenseEmbed [32]	0.805	--
MSSA-T [35]	0.769	--
MSSA-D [35]	0.768	--
Chen et al. [33]	0.620	--
Attract-Repel [39]	0.709	0.655
Counter-fitting [41]	0.741	0.680
Paragram [42]	0.799	0.754
CBOW [18]	0.732	0.723
GloVe [30]	0.801	0.800
FastText [31]	0.762	0.755
Proposed approach (cosine)	**0.836**	**0.815**
Proposed approach (Tanimoto)	0.831	0.810

Table X shows the analysis on Simlex666 dataset, the results on this dataset for the proposed approach and other multi-sense approaches are marginally less when compared with Attract-Repel [39], counter-fitting [41], paragram [42]. These three approaches Attract-Repel, counter-fitting, paragram outperformed other single vector and multi-vector representations on the Simlex666 dataset. The reasons are mentioned in the discussion section.

TABLE X. Spearman and Pearson Correlations Of Proposed and Existing Word Embedding Approaches on Simlex666 Dataset, Highest Results Highlighted in Boldface

Method	Spearman correlation	Pearson correlation
MSSA-T [35]	0.460	--
MSSA-D [35]	0.425	--
Chen et al. [33]	0.430	--
Attract-Repel [39]	0.690	0.691
Counter-fitting [41]	**0.698**	**0.697**
Paragram [42]	0.645	0.662
CBOW [18]	0.454	0.461
GloVe [30]	0.429	0.467
FastText [31]	0.410	0.411
Proposed approach (cosine)	0.522	0.531
Proposed approach (Tanimoto)	0.521	0.527

V. Discussion

In the results section, nine benchmark datasets are analyzed with the state-of-art systems and the proposed approach. The state-of-the-art systems compared in this paper are categorized as follows:

1. Symmetric pattern vector representations: Approaches like CBOW of Word2vec [18], GloVe [30], fastText [31] fall in this category. These approaches use shallow linguistic information like word and context co-occurrences within a single window.

2. Inject approaches : The approaches such as Attract-Repel [39], counter-fitting [41], Paragram [40] fall in this category. In these approaches, word embeddings are enriched with synonymy and antonymy constraints and paraphrase relations. These three inject approaches use synonymy and antonymy relations of the words.

3. Multi-sense representations: Approaches like SenseEmbed [32], Chen et al. [33], MSSA [35], fall in this category. These approaches use multi-sense vector representations to carry out various NLP tasks.

On MC and RG datasets, there is a little variation in the results of all the approaches. This is because of the less number of word pairs in the datasets. Most of the researchers use these two benchmarks as these are more popular datasets in word similarity tasks. From, the results it is observed that the proposed approach performed consistently on all the datasets except Simlex666 dataset. The proposed approach obtained the first or second highest results of Spearman and Pearson correlations. The single vector representations CBOW [18], GloVe [30], fastText [31] when compared with multi-sense representations obtained lesser results. The CBOW [18], GloVe [30], fastText [31] approaches laid the basis of modern research for carrying out multiple NLP tasks. As the proposed approach generates all the multiple senses based on context, definitions, hypernym path, and LCS, the results are consistent on all the datsets.

Inject approaches like Attract-Repel [39], counter-fitting [41], paragram [40] performed well on WS-353 and Simlex666 datasets. The reason why these methods obtained better results on these datasets is because of using synonymy and antonymy relations. The multi-vector representations performed well on the WS-353 dataset but failed to perform well on the Simlex666 dataset. Compared to the WS-353 dataset, the Simlex666 dataset has more antonym word pairs, the inter-annotator agreement is also less in this dataset. As the multi-vector representations do not consider dealing with antonym relations, injection approaches produced better results on these datasets compared to multi-sense representations. Turk771 and Turk287 datasets have only noun pairs and the proposed approach obtained better results on these two datasets. These datasets are used mostly in semantic relatedness purposes. The proposed approach obtained the highest results on the WS-353 dataset because of handling synsets of different parts-of-speech.

On larger datasets like SCWS, Men and Rare words, the performance of the multi-vector representations and the proposed approach are consistent. These three datasets have different parts-of-speech word pairs. Algorithm I generates synsets with respect to their parts-of-speech to produce suitable multi-sense vector representations of the word. On larger datasets, the proposed multi-vector representation model outperformed the other state-of-art systems. The overall performance of the proposed approach is stable on all the datasets.

VI. Conclusion

This paper discusses the importance of multi-sense embedding for carrying out word similarity tasks. The proposed approach overcomes the limitations of single vector representations. This paper understands

the importance of synonym sets, hypernyms, LCS, and definitions of words to generate multi-sense embeddings. This inspection of Wordnet for various parameters is a novel idea, which is not explored by the NLP community. With improved results of word similarity, this paper shows the significance of multi-sense representations.

The proposed approach performs disambiguation of words using Wordnet. A library of words from Wikipedia is created, and for each word, multi-sense embeddings are generated. Sense embedding vectors are enriched by using the LCS, hypernym information between the words. A comparison with recent state-of-art methods confirmed the efficiency of the proposed approach.

The proposed multi-sense representation produced state-of-art results on nine benchmark datasets. With these considerations, the next challenge is to see how multi-sense embeddings help to improve the efficiency of other NLP tasks like text summarization and document classification.

References

[1] X. Ji, A. Ritter, P. Y. Yen, "Using ontology-based semantic similarity to facilitate the article screening process for systematic reviews," *Journal of biomedical informatics,* 2017, vol. 69, pp. 33-42.

[2] M. B. Aouicha, and M. A. H. Taieb, "Computing semantic similarity between biomedical concepts using new information content approach," Journal of biomedical informatics, 2016, vol. 59, pp. 258-275.

[3] G. Zhu, and C. A. Iglesias, "Computing semantic similarity of concepts in knowledge graphs," *IEEE Trans. Know. Data Engg.,* 2016, vol. 29(1), pp. 72-85.

[4] A. Skabar, and K. Abdalgader, "Clustering Sentence-Level Text Using a Novel Fuzzy Relational Clustering Algorithm," *IEEE Trans. Know. Data Engg.*, 2013, vol. 25(1), pp. 62-75.

[5] G. Glavas, M. Franco-Salvador, S. P. Ponzetto, and P. A. Rosso, "A resource-light method for cross-lingual semantic textual similarit," *Knowl.-Based Syst.*, 2018, vol. 143, pp. 1–9.

[6] Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin de la Société Vaudoise des Sciences Naturelles,* 1945, vol. 37, pp. 547-579.

[7] V. Levenshtein, "Binary Codes capable of correcting deletions, insertions, and reversals," *Cyber. and Control Theory,* 1966, vol.10, pp. 707–710.

[8] Dice, "Measures of the amount of ecologic association between species," *Ecology*, 1945.

[9] C. Fellbaum, "WordNet: An Electronic Lexical Database," *Cambridge, MA: MIT Press*, 1998.

[10] Jiang, D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," *arXiv preprint cmp-lg/9709008,* 1997.

[11] C. Leacock, M. Chodorow, "Combining local context and WordNet similarity for word sense identification," *WordNet: An electronic lexical database,* 1998, vol. 49, pp. 265-283.

[12] Li, McLean, "An approach for measuring semantic similarity between words using multiple information sources,'' *IEEE Trans. on know. and data eng.,* 2003, vol. 15, pp. 871-882.

[13] Lin, Dekang, "An information-theoretic definition of similarity," *Proc. Of Icml,* 1998.

[14] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and application of a metric on semantic nets," *IEEE Trans. Syst. Man. Cyber.,* 1989, vol. 19(1), pp. 17-30.

[15] Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *arXiv preprint cmp-lg/9511007*, 1995.

[16] G. Zhu, C. A. Iglesias, "Computing semantic similarity of concepts in knowledge graphs," *IEEE Trans. Knowl. and Data Eng.,* 2017, vol. 29, pp. 72-85.

[17] Z. Wu, M. Palmer, "Verbs semantics and lexical selection," *In Proceedings of the 32nd annual meeting on ACM,* 1994, pp. 133-138.

[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR . ArXiv: 1301.3781,* 2013.

[19] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and Dean, "Distributed representations of words and phrases and their compositionality," *in NIPS'13,* 2013, pp. 3111–3119.

[20] T. Chen, R. Xu, and Y. He, and X. Wang, "Improving distributed representation of word sense via wordnet gloss composition and context clustering," *International Joint Conference on Natural Language Processing*, 2015, pp. 15-20.

[21] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Systems with Applications,* 2019, vol. 117, pp. 139-147.

[22] T. Khai Tran, and T. Thi Phan, "Deep Learning Application to Ensemble Learning—The Simple, but Effective, Approach to Sentiment Classifying," *Applied Sciences,* 2019, vol. 9(13).

[23] F. Li, Y. Yin, J. Shi, X. Mao, and R. Shi, "Method of feature reduction in short text classification based on feature clustering," *Applied Sciences,* 2019, vol. 9(8).

[24] D. Dimitriadis, and G. Tsoumakas, "Word embeddings and external resources for answer processing in biomedical factoid question answering," *Journal of biomedical informatics*, 2019, vol. 92, pp. 103-118.

[25] J. Li, D. Jurafsky, "Do multi-sense embeddings improve natural language understanding?" *Proceedings of Empirical methods in natural language processing,* 2015, pp.1722–1732.

[26] Z. Harris, "Distributional structure," *Word*, 1954, vol. 10(23), pp. 146–162.

[27] G. Salton, A. Wong, C. S. A. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, 1975, vol. 18(11), pp. 613–620.

[28] P. D. Turney, and P. Pantel, "From frequency to meaning: Vector space models of semantics," *CoRR . ArXiv: 1003.1141,* 2010.

[29] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research,* 2011, vol. 12, pp. 2493–2537.

[30] J. Pennington, and R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," *Proceedings of Empirical Methods in Natural Language Processing*, 2014, pp. 1532–1543.

[31] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, 2017, vol. 5, pp. 135–146.

[32] I. Iacobacci, M. T. Pilehvar, and R. Navigli, "Sensembed: Learning sense embeddings for word and relational similarity," *International Joint Conference on Natural Language Processing*, 2015, pp. 95-105.

[33] X. Chen, Z. Liu, M. A. Sun, "A unified model for word sense representation and disambiguation," *Proceedings of Empirical Methods in Natural Language Processing*, 2014, pp.1025-1030.

[34] D. Oele, G. van Noord, "Simple Embedding-Based Word Sense Disambiguation," *Proceedings of the 9th Global WordNet Conference*, 2018.

[35] T. Ruas, W. Grosky, and A. Aizawa, "Multi-Sense embeddings through a word sense disambiguation process," *Expert Systems with Applications*, 2019 (In press).

[36] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone," *Proceedings of the 5th annual international conference on systems documentation SIGDOC' 86*, 1986, pp. 24–26.

[37] S. Rothe, and H. Schutze, "Autoextend: Extending word embeddings to embeddings for synsets and lexemes," *arXiv preprint arXiv:1507.01127,* 2015.

[38] M. Fabian, M. F. Suchanek, G. Kasneci, G. Weikum, "Yago: A Core of Semantic Knowledge," *16th International Conference on the World Wide Web*, 2007, pp. 697–706.

[39] N. Mrksic, I. Vulic, D. O. Seaghdha, I. Leviant, R. Reichart, M. Gašić, A. Korhonen, and S.Young, "Semantic specialisation of distributional word vector spaces using monolingual and cross-lingual constraints," *Trans. ACL.*, 2017, vol. 5, pp. 309–324.

[40] J. Wieting, M. Bansal, K. Gimpel, K. Livescu, and D. Roth, "From paraphrase database to compositional paraphrase model and back," *Trans. ACL.,* 2015, vol. 3, pp. 345–358.

[41] N. Mrksic, D. O Seaghdha, et al., "Counter-fitting word vectors to linguistic constraints," *In Proceedings of NAACL-HLT,* 2016, pp.142–148.

[42] T. Tanimoto, "An Elementary Mathematical theory of Classification and Pre-diction," *IBM Internal Report 17th IBM,* 1957.

[43] G. A. Miller, W. G. Charles, "Contextual correlates of semantic similarity," *Lang. Cogn. Process.,* 1991, vol. 6 (1), pp. 1–28.

[44] H. Rubenstein, J. B. Goodenough, "Contextual correlates of synonymy," *Commun. ACM,* 1965, vol. 8(10), pp. 627–633.

[45] L. Finkelstein, et al., "Placing search in context: The concept revisited," *ACM Trans. Inf. Syst.,* 2002, vol. 20(1), pp. 116–131.

[46] G. Halawi, G. Dror, E. Gabrilovich, and Y. Koren, "Large-scale learning of word relatedness with constraints," *In Proc. of ACM SIGKDD.* 2012, pp. 1406–1414.

[47] K. Radinsky, et al., "A word at a time: computing word relatedness using temporal semantic analysis," *In Proc. of the Intl. Conf. on WWW. ACM,* 2011, pp.337–346.

[48] E. H. Huang, R. Socher, C. D. Manning, and A. Y. Ng, "Improving word representations via global context and multiple word prototypes," *In Proc. of the Annual Meeting of the ACL,* 2012, pp. 873–882.

[49] T. Luong, R. Socher, C. D. Manning, "Better word representations with recursive neural networks for morphology," *In: Proc. of CoNLL.,* 2013, pp. 104–113.

[50] E. Bruni, N. K. Tran, and M. Baroni, "Multimodal distributional semantics," *Journal of Artificial Intelligence Research,* 2014, vol. 49(1), pp. 1–47.

[51] F. Hill, R. Reichart, A. Korhonen, "SimLex-999: Evaluating semantic models with (genuine) similarity estimation," *Comput. Linguist.,* 2015, vol. 41 (4), pp. 665–695.

[52] R. Navigli, S. P. Ponzetto, "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network," *Artificial Intelligence*, 2012, vol. 193, pp. 217–250.

[53] J. Ganitkevitch, B. Van Durme, C. Callison-Burch, "PPDB: The paraphrase database," *In: Proc. of HLT-NAACL,* 2013, pp. 758–764.

Krishna Siva Prasad Mudigonda

Krishna Siva Prasad Mudigonda is presently a Research Scholar in Computer Science and Engineering department at Visvesvaraya National Institute Technology, Nagpur, India. He received his M.Tech and B.Tech degrees from Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India. His research area includes natural language processing, deep learning and text processing.

Poonam Sharma

Poonam Sharma is presently Assistant Professor in Computer Science and Engineering department at Visvesvaraya National Institute Technology, Nagpur, India. She completed her PhD degree from Maulana Azad National Institute of Technology, Bhopal, India. Her research area includes Biometrics, Image processing, soft computing and pattern recognition. She is a recipient of Visvesvaraya young faculty fellowship from the government of India. She has completed a project sponsored by Department of Science and Technology, New Delhi.

Assessed by Machines: Development of a TAM-Based Tool to Measure AI-based Assessment Acceptance Among Students

J.C. Sánchez-Prieto[1*], J. Cruz-Benito[2*], R. Therón[1], F. J. García-Peñalvo[1]

[1] GRIAL Research Group. University of Salamanca (Spain)
[2] IBM Quantum, IBM Research. IBM T.J. Watson Research Center. Yorktown Heights, NY 10598 (USA)

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

In recent years, the use of more and more technology in education has been a trend. The shift of traditional learning procedures into more online and tech-ish approaches has contributed to a context that can favor integrating Artificial-Intelligence-based or algorithm-based assessment of learning. Even more, with the current acceleration because of the COVID-19 pandemic, more and more learning processes are becoming online and are incorporating technologies related to automatize assessment or help instructors in the process. While we are in an initial stage of that integration, it is the moment to reflect on the students' perceptions of being assessed by a non-conscious software entity like a machine learning model or any other artificial intelligence application. As a result of the paper, we present a TAM-based model and a ready-to-use instrument based on five aspects concerning understanding technology adoption like the AI-based assessment on education. These aspects are perceived usefulness, perceived ease of use, attitude towards use, behavioral intention, and actual use. The paper's outcomes can be relevant to the research community since there is a lack of this kind of proposal in the literature.

I. Introduction

WILL be your next teacher a neural network model? What would be your feelings related to being evaluated by a non-human entity in such a human process like learning? In the last years, we have been experiencing a rising in integrating Artificial Intelligence (AI) applications in our daily lives [1]-[3].

People are beginning to be aware that they may interact with AI applications when interacting with mobile devices, computers, or in general, in technology-surrounded environments. What is less known, or the people are less aware, is that AI applications surround us even in more traditional contexts, even in those that typically were pure human-based and did not involve any intelligent software stakeholder [4]-[6]. Many previously pure human-based contexts are now intermediated by intelligent systems or capable decision-making software that affect how we experience our lives or interact with other people.

As commented in the literature [7], [8], one of the fields in which we can observe the introduction of more sophisticated software programs is education. As [7] outlines, in recent years, we have experienced the inclusion of virtual teaching assistants in classrooms (for example, Jill Watson, deployed at the Georgia Institute of Technology),

technology to study and improve students' learning performance, AR/VR technologies, teaching robots, etc. The trend of including more technology in different educational environments is even more significant today, during the current global COVID-19 pandemic. Since the beginning of the pandemic, many people have used virtual learning and teaching and different approaches to improve those online experiences. By using video-meeting resources and online educational platforms, many teachers have solved part of the issue; they have given their lectures and provided learning materials.

However, what happens with other fundamental aspects of the learning processes like the assessment? Are there intelligent systems being used to evaluate students learning? Are the students OK with being potentially evaluated by software artifacts based on AI algorithms or any other similar approach? According to our previous research [8]-[10], there is a lack of studies on this topic, especially in studying acceptance among students and teachers.

Based on the literature gap detected and related to our previous research, this paper presents a novel TAM-based tool to measure students' acceptance of being assessed by intelligent software artifacts (AI-based ones). The purpose of this tool is to serve as the foundation to develop further empiric studies that could enrich our understanding of how people (and students) interact with intelligent software, in particular in contexts like education, and even more when the AI-based software is in a power position like in the role of an evaluator.

The paper includes the following sections: Section I introduces the problem and some relevant research. Section II presents the

* Corresponding author.

E-mail addresses: josecarlos.sp@usal.es (J. C. Sánchez-Prieto), juan.cruz@ibm.com (J. Cruz-Benito).

theoretical background of the paper. Section III describes the TAM-based model we propose to study the AI-based assessment acceptance among students, while section IV presents the specific instrument we propose to research that question. Section V discusses the outcomes of the paper and its future implications. Finally, Section VI presents some conclusions.

II. Theoretical Background

The possibilities entailed by using AI-based systems in the educational field constitute a topic of growing attention among the scientific community, based on the fast-development of these technologies and the vivid social debate regarding the risks ethical considerations on the use of AIs [8].

This way, along the last decade, we can observe an increasing number of initiatives that explore the different uses of these technologies to support the teaching-learning process that can be classified into three groups: use of AIs to analyze human behavior, use of AIs as didactic tools and use of AIs as assessment tools.

In the first group, we can find investigations that use AIs as tools to perform complex statistical analysis in data-driven approaches with large groups of data to detect behavioral patterns of teachers and students susceptible to intervention. The initiatives in this group are mainly focused on students' elements such as their learning strategies [11], [12], although we can find some examples of its use on investigations with teachers [13].

On the other hand, the second group is focused on the didactic potential of AIs and examines its application in the creation of virtual persons to help students during their learning [14], the development of virtual learning environments [15] or even the use of social robots [16].

Finally, the third group is interested in the possibilities derived from using AI-based systems on the students' assessment, which constitutes one of the students' primary concerns and a key competence of the teachers [17].

The study of the use of AIs for evaluation is still a very young field in an early stage of development in which most research is concentrated in the last decade, mainly in North America[8].

Even so, we can find exciting initiatives in this area developed in the context of higher education. These investigations can be grouped into three categories:

- **Use of AIs for the assessment of behaviors**: These investigations are focused on the analysis of student behavior at different circumstances of the teaching-learning process.

  This way, we can find initiatives that evaluate through AIs the interactions of students in online courses to predict their academic performance [18], [19] or their behavior during activities to personalize their learning experience or develop adaptive processes [20], [21].

- **Use of AIs for the assessment of feelings**: Experiences that apply AI-based tools to determine the emotions experienced by students during the development of different educational processes

  The most common application of these technologies is the analysis of students' satisfaction in online courses [22], [23], although we can find research that uses the analysis of feelings with other objectives such as predicting the success of MOOCs [24].

- **Use of AIs for the assessment of academic performance**: In this group of investigations, the most numerous of the three, the AI-based tools are used to automate the correction process of both exams (multiple choice or short answers) [25], [26] and other types of student productions [27].

The application of these tools is intended to lighten teachers' workload and improve the fairness of their grading [28].

Despite this growing presence of AIs in the educational field and the interest among the scientific community and organizations in developing this practice, the study of the adoption of these technologies among the educational agents constitutes an underexplored research area.

However, we can find examples of research focused on this topic in other fields, such as the development of automated vehicles, banking, or e-commerce [29]-[31] that indicate the critical role played by factors such as perceived usefulness, confidence, or social pressure in the decision to use AI-based tools.

The educational agents' willingness is a fundamental element to guarantee the success of any technological innovation process. Therefore, the development of adoption models to analyze the factors that condition such adoption can provide essential information for developing of new initiatives in this field.

III. Model Development

As a starting point for developing the model, we reference the technology acceptance model (TAM) proposed by Davis [32]. This theory understands the process of adoption of a specific technology based on five factors:

- **Perceived usefulness (PU)**: The subject's perception of the effect that the tool's use has on their work performance.

- **Perceived ease of use (PEU)**: The person's assessment of the degree of effort required to use the system.

- **Attitude towards use (AT)**: The feelings, opinions, and favorable or unfavorable assessments about the use of technology.

- **Behavioral intention (BI)**: The level of willingness of the user to use the tool.

- **Actual use (AU)**: The frequency of use of the technology by the subject.

PU, and PEU are the main antecedents of the model. In Davis's proposal, these two factors condition the attitude towards the user's use, which in turn, conditions BI which would be the factor leading to the actual use of technology (Fig.1).
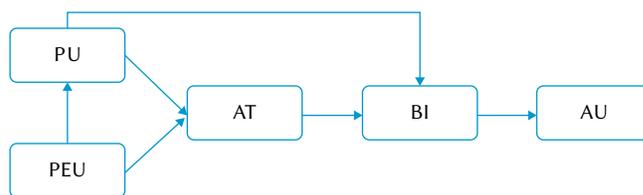
Fig. 1. TAM model [32].

TAM's main advantages are its parsimony and flexibility, which allows the application of the model to a wide variety of contexts and technologies [33].

In the educational field, enhanced by the fast technological development, TAM based models expanded with constructs from other theories have been widely applied to analyze the adoption process of technologies such as mobile devices [34], LMS [35], or AR [36] both between students [37] and teachers [38].

However, since the use of AI for the assessment is still in a very initial stage of development, the development of TAM-based models for the analysis of these technologies' adoption process among the educational agents constitutes an unexplored line of research.

Following this line of development, this model includes the six hypotheses proposed by Davis [32] applied to the object of study:

H1 PU is positively related to the intention to participate in the AI-driven assessment activities of the students.

H2 PU is positively related to the attitude towards the participation in AI-driven assessment activities of the students.

H3 PEU is positively related to the attitude towards the participation in AI-driven assessment activities of the students.

H4 PEU is positively related to the students' usefulness in the implementation of AI technologies for their assessment.

H5 AU is positively related to the intention to participate in AI-driven assessment activities of the students.

H6 BI is positively related to the participation in the AI-driven assessment activities of the students.

One of the most common practices during the process of adapting TAM to a new field of technology is the expansion of the original proposal and the inclusion of constructs from other theories pertinent for the object of study, with the intention to increase the percentage of variance explained by the model [39].

Following this line of reasoning, we decided to expand TAM with three additional constructs to measure the influence of environmental pressure, the feeling of distrust towards AIs, and the natural opposition to the individual's changes.

Firstly, to measure the effect of the individual's perception of the social discourse on the use of AIs for evaluation on their adoption of this technology, this model includes subjective norm (SN).

This construct refers to the perceived social pressure towards the performance of a particular behavior, the belief that a substantial group or individual approves or disapproves a given action [40], in this case, the use of AIs in the evaluation of the students.

Since its formulation in the theory of reasoned action (TRA) [41] and the theory of planned behavior (TPB) [42] this construct has been widely incorporated in the design of technology adoption models such as TAM2 [43], TAM3 [44] and expanded TAM models [45].

In the educational field, we can find a wide variety of studies that incorporate SN to measure the social influence on the technology adoption of the students [46] confirming its effect on the behavior of the subjects through its influence on other TAM constructs such us PU and BI.

Given the heated debate surrounding the use of AIs in education, and the lack of information and experience of the students regarding this practice, SN should have an important influence on both, the perception of the individual of AIs and their disposition towards its use.

Therefore, the model includes the following two hypotheses for SN based on the proposals by Venkatesh and Davis [43] and Venkatesh and Bala [44] :

H7 SN is positively related to the intention to participate in AI-driven assessment activities of the students.

H8 SN is positively related to the usefulness perceived by the students in the implementation of AI technologies for their assessment.

One of the main problems when dealing with the implementation of the AI-based systems is the lack of user confidence in both the handling of the user's personal information and the resolution of complex problems [47].

This issue has been acquiring greater importance due to the proliferation and development of these systems and the interest of organizations and administrations in their incorporation [48] .

Trust is a variable often included in technology adoption studies, especially in contexts such as online banking [49], e-commerce [50] or eGovernment [51] defined as the willingness of the individual to rely on the other part [52], establishing its relationship with other variables from TAM like PU or BI.

In the educational field, the effect of this dimension is still little explored, although we can find some examples of its application in research on the adoption of technologies such as cloud computing [53].

In this case, trust (TR) plays a special role since, although they may be perceived as agents, AIs have no conscience or morals [54], thus in the context of this study, users will perceive AIs to be useful and be more inclined to be evaluated by them if they trust both the teachers that use them and the AIs themselves. In consequence, for this research, we propose the following hypotheses based on the works of [49]- [51]:

H9 TR is positively related to the intention to participate in the AI-driven assessment activities of the students.

H10 TR is positively related to the attitude towards the participation in AI-driven assessment activities of the students.

H11 TR is positively related to the usefulness perceived by the students in the implementation of AI technologies for their assessment.

Incorporating the use of AI in education implies a series of changes in the teaching-learning process, including the attribution of one of the most important tasks of teachers, as is the assessment, to a technological tool.

This change can cause a feeling of anxiety among the students in the face of the expectation of not being evaluated by a human being, which may generate a negative predisposition towards participation in educational activities with this modality of evaluation. In order to measure this effect, the model includes the construct of resistance to change (RC).

RC was born in the field of organizational sciences to analyze the organizational elements that generate an attitude of resistance in the individual, although this concept was also applied to study factors related to the individual [55], [56].

In this proposal, RC is modeled after the construct developed by Guo et al [57] and defined as the opposition of the individual to the rupture of the status quo produced by the use of AIs for their evaluation. This way, RC would be an inhibiting factor that will negatively affect the students' technology adoption [58].

In the educational field, RC is recently being incorporated in expanded TAM models to study students' adoption of ICTs with promising results [59] supporting is the effect in the three main dimensions of TAM; namely, PU, AT, and BI. Considering the importance of the changes that the implementation of this technology entails for an area that constitutes one of the students' main concerns, RC is expected to affect the aforementioned three constructs. Therefore, the model is completed (Fig. 2) with the following three hypotheses:

H12 RC is positively related to the intention to participate in AI-driven assessment activities of the students.

H13 RC is positively related to the attitude towards the participation in the AI-driven assessment activities of the students.

H14 RC is positively related to the usefulness perceived by the students in the implementation of AI technologies for their assessment.
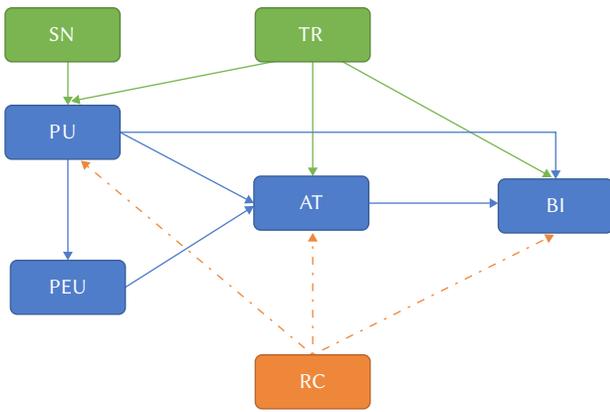
Fig. 2. Research model.

IV. Instrument

The instrument proposed to measure the seven constructs of our model is divided into three sections. Firstly, taking into account that the use of these technologies is still in the early stages of development, it is very likely that students will not have a clear idea of what constitutes the use of AIs for assessment, the instrument to perform the data gathering process includes a brief introduction text explaining the fundamentals of this topic.

The second section is dedicated to gathering the participants' identification data regarding four variables that may affect the adoption process, namely, gender, age, experience with AIs, and the branch of knowledge.

The third section is composed of 30 Likert-type items adapted from previous works to measure the object of study. Firstly, the items to assess the constructs from the original TAM were adapted from the proposals by Davis [32] and Venkatesh and Bala [44] (table I):

TABLE I. Items to Measure the Tam Constructs

Item	Text
AU1	Using AIs to support the assessment of the students is a good idea.
AU2	Using AI-based tools to assess the academic achievement of the students is a smart choice.
AU3	I like the idea of using AIs for the assessment of the students.
AU4	I have positive feelings about implementing AI-based tools for the assessment of the students.
BI1	I would participate in AI-based assessment processes.
BI2	I plan to participate in AI-based assessment processes often.
BI3	I hope to participate in AI-based assessment processes in the future.
PEU1	My interaction with AI-based tools is clear and understandable.
PEU2	I find it easy to get AIs to do what I want them to do.
PEU3	I find AI-based systems easy to use.
PEU4	I can easily learn how to use AI-based tools for my self-assessment.
PU1	Using AI-based tools for my assessment allows me to learn effectively.
PU2	Using AI-based tools for my assessment improves my academic productivity.
PU3	Using AI-based tools increases my assessment opportunities.
PU4	Incorporating AI-based technologies for my assessment is useful.

Secondly, to measure the new constructs added to the model, we took as reference different expanded TAM models (Table II).

TABLE II. Tems to Measure the Expanded Constructs

Item	Text
SN1	People who are important to me think AIs should be used to assess the students.
SN2	My classmates think teachers should use AIs to assess their students.
SN3	In the universities, teachers are expected to use AI-based tools to assess the academic achievement of the students.
SN4	People who influence my behavior think I should participate in AI-based assessment processes often.
SN5	People who are important to me think I should participate in AI-based assessment processes often.
RC1	I would like AIs to change the way assessment is carried out.
RC2	I want AIs to change teacher-student interaction during the assessment.
RC3	I would like AIs to change the way I am assessed.
RC4	I agree with the changes entailed by the use of AI for the assessment.
TR1	AIs are trustworthy.
TR2	I tend to trust AIs.
TR3	I trust AIs even though I have little knowledge of them.
TR4	AIs can provide an accurate assessment of the students.
TR5	AIs can provide a reliable assessment of the students.
TR6	AIs can provide a convenient assessment of the students.

In the case of SN, the five items for this dimension were adapted from the works of Venkatesh and Bala [44] and Taylor and Todd [60], [61] to measure both the general sources of SN and the specific sources related to the educational environment of the student.

The RC items were developed based on the proposals by Guo et al. [57] and Sánchez-Prieto et al. [59] that refer both to elements related to the teaching-learning process and with the interaction between teacher and student.

Finally, the items to measure the trust of the students in AI and their capability to assess them were adapted from the versions of Gefen et al. [52].

V. Discussion and Conclusion

As we have shown, AI-based evaluation and the integration of algorithmic solutions in the learning process have been a trend in recent years. It is increasing its relevancy at a rapid pace. The shift of traditional learning procedures into more online and tech-*ish* approaches contributes to algorithmic systems' implantation. Even more, with the current acceleration because of the COVID-19 pandemic, more and more learning processes are becoming online and are incorporating technologies like those described relative to assessment. The opportunity of going online is a powerful temptation in the current situation or, in some cases, is the only way to continue learning. Related to that, the use of new assessment technologies based on AI or any other algorithmic approaches correlates perfectly with that trend of going online. Since most of the interaction is online and produces a digital footprint, we can think that we can use, easily, automated ways of evaluating the interaction. That is partially true. It is right in the sense that many digital footprints can be analyzed automatically. We can indeed make simple analytics in almost any digital data. We can probably even introduce improvements in the education process that can favor students and teachers, fostering their performance, opportunities, and engagement [62]. What is not exact is

that we can introduce a relevant change like AI-based or algorithmic-based assessment of learning without paying much attention to some different aspects. Before going forward, we must evaluate the real need to use this kind of procedure, the level of detail required for the analysis, the interaction of humans in the process, the ethical issues related, or, as underlies in this paper, the perception of being evaluated (or helped in evaluation) by a non-conscious software entity that enables decisions or takes them in our behalf or while considering our human performance.

Furthermore, considering most of the existing algorithms and systems cannot grasp the human processes' entropy and their multiple dimensions and subtle details completely, leading to harmful results for the humans involved. For these reasons, there are needed studies that research the human perceptions of these procedures driven by algorithms.

This article presents a theoretical model of technology adoption that aims to explain students' adoption using AI-based tools for their assessment. The model is based on TAM and expanded with the most relevant constructs in adopting this technology: perceived usefulness, perceived ease of use, attitude towards the use, behavioral intention, trust, resistance to change, and subjective norm.

With the inclusion of these variables, the model proposed can address both the effect of the perceived advantages and disadvantages and the factors related to the trust issues and the reluctance of the subjects to be assessed by these tools. The resulting model includes 14 relational assumptions that determine the effects of the different variables.

To measure these constructs and aim to guarantee the proposal's parsimony, we have designed a questionnaire composed of 30 Likert-type items adapted from previous theories, which allows for the empirical testing of the model and its expansion and modification.

This way, this model constitutes a first step in the study of the adoption of AI-based assessment among students, which opens the door to further investigations. Firstly, although the items are adapted from previous investigations and have been exhaustively tested in other adoption models with good results, the instrument could be subjected to a content validation process to confirm the items' quality. In this line, a pilot study is currently in development to test the proposal's statistical validity in an empirical setting.

Finally, this model is also susceptible to further modification, including its expansion with additional constructs from other theories that may increase the understanding of the adoption process of these technologies.

References

[1] S. Makridakis, "The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms," *Futures*, vol. 90, pp. 46-60, 2017.

[2] I. Roll and R. Wylie, "Evolution and revolution in artificial intelligence in education," *International Journal of Artificial Intelligence in Education*, vol. 26, no. 2, pp. 582-599, 2016. https://doi.org/10.1007/s40593-016-0110-3.

[3] A. Abdul, J. Vermeulen, D. Wang, B.Y. Lim and M. Kankanhalli, "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, 2018, pp. 1–18.

[4] European Commission, "Smart lie-detection system to tighten EU's busy borders,"European Commission Website, 2018. https://ec.europa.eu/research/infocentre/article_en.cfm?id=/research/headlines/news/article_18_10_24-3_en.html%3Finfocentre&item=Infocentre&artid=49726&pk_campaign=rss_page

[5] J. Cook, "Amazon scraps "sexist AI" recruiting tool that showed bias against women," *The Telegraph* 2018. https://www.telegraph.co.uk/technology/2018/10/10/amazon-scraps-sexist-ai-recruiting-tool-showed-bias-against/.

[6] "China's hi-tech war on its muslim minority," *The Guardian* 2019. http://www.theguardian.com/news/2019/apr/11/china-hi-tech-war-on-muslim-minority-xinjiang-uighurs-surveillance-face-recognition.

[7] J. Kim, K. Merrill Jr., K. Xu and D.D. Sellnow, "My teacher is a machine: Understanding students' perceptions of AI teaching assistants in online education", International Journal of Human–Computer Interaction, pp. 1-10, 2020. https://doi.org/10.1080/10447318.2020.1801227.

[8] J.C. Sánchez-Prieto, A. Gamazo, J. Cruz-Benito, R. Therón and F.J. García-Peñalvo, "AI-Driven Assessment of Students: Current Uses and Research Trends" in *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences. HCII 2020. Lecture Notes in Computer Science*, 2020, pp. 292-302.

[9] J.C. Sánchez-Prieto, J. Cruz-Benito, R. Therón and F.J. García-Peñalvo, "How to Measure Teachers' Acceptance of AI-Driven Assessment in ELearning: A TAM-Based Proposal," in *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, León, Spain, 2019, pp. 181–186.

[10] J. Cruz-Benito, J.C. Sánchez-Prieto, R. Therón and F.J. García-Peñalvo, "Measuring Students' Acceptance to AI-Driven Assessment in eLearning: Proposing a First TAM-Based Research Model," in *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences. HCII 2019. Lecture Notes in Computer Science*, 2019, pp. 15-25.

[11] P. García, S. Schiaffino and A. Amandi, "An enhanced bayesian model to detect students' learning styles in web-based courses," *J.Comput.Assisted Learn.*, vol. 24, no. 4, pp. 305-315, 2008. https://doi.org/10.1111/j.1365-2729.2007.00262.x.

[12] Y. Wei, Q. Yang, J. Chen and J. Hu, "The exploration of a machine learning approach for the assessment of learning styles challenges," *Mechatronic Systems and Control*, vol. 46, no. 3, pp. 121-126, 2018.

[13] C. Verma, A. S. Tarawneh, Z. Illes, V. Stoffova and S. Dahiya, "Gender Prediction of the European School's Teachers Using Machine Learning: Preliminary Results," in *- 2018 IEEE 8th International Advance Computing Conference (IACC)*, 2018, pp. 213-220.

[14] D.B. Chin, I.M. Dohmen, B.H. Cheng, M.A. Oppezzo, C.C. Chase and D.L. Schwartz, "Preparing students for future learning with teachable agents," *Educational Technology Research and Development*, vol. 58, no. 6, pp. 649-669, 2010.

[15] D. Xu and H. Wang, "Intelligent agent supported personalization for virtual learning environments," *Decis.Support Syst.*, vol. 42, no. 2, pp. 825-843, 2006.

[16] T.B. Sheridan, "Human–Robot interaction: Status and challenges," *Hum.Factors*, vol. 58, no. 4, pp. 525-532, 2016. https://doi.org/10.1177/0018720816644364.

[17] R.J. Stiggins, N.F. Conklin and N.J. Bridgeford, "Classroom assessment: A key to effective education," *Educational Measurement: Issues and Practice*, vol. 5, no. 2, pp. 5-17, 1986.

[18] J. Ma, J. Kang, E. Shaw and J. Kim, "Workflow-Based Assessment of Student Online Activities with Topic and Dialogue Role Classification" in *Artificial Intelligence in Education. AIED 2011. Lecture Notes in Computer Science*, 2011, pp. 187-195.

[19] S. Singh and S. P. Lal, "Educational courseware evaluation using Machine Learning techniques," in *- 2013 IEEE Conference on e-Learning, e-Management and e-Services*, 2013, pp. 73-78.

[20] R. Zatarain-Cabada, M.L. Barrón-Estrada and J.M. Ríos-Félix, "Affective Learning System for Algorithmic Logic Applying Gamification" in *Advances in Soft Computing. MICAI 2016. Lecture Notes in Computer Science*, 2017, pp. 536-547.

[21] K. Kuk, I.Z. Milentijević, D. Ranđelović, B.M. Popović and P. Čisar, "The design of the personal enemy - MIMLebot as an intelligent agent in a game-based learning environment," *Acta Polytechnica Hungarica*, vol. 14, pp. 121-139, 2017.

[22] D. Dudek, "Survey Analyser: Effective Processing of Academic Questionnaire Data," in *Information Systems Architecture and Technology:*

Proceedings of 39th International Conference on Information Systems Architecture and Technology – ISAT 20, 2019, pp. 245-257.

[23] A. Tzacheva, J. Ranganathan and R. Jadi, "Multi-Label Emotion Mining From Student Comments," in *Proceedings of the 2019 4th International Conference on Information and Education Innovations*, Durham, United Kingdom, 2019, pp. 120-124.

[24] L. Wang, G. Hu and T. Zhou, "Semantic analysis of learners' emotional tendencies on online MOOC education," *Sustainability*, vol. 10, pp. 1-19, 2018.

[25] M. Wang, C. Wang, C. Lee, S. Lin and P. Hung, "Type-2 fuzzy set construction and application for adaptive student assessment system," in *- 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2014, pp. 888-894.

[26] T. Luchoomun, M. Chumroo and V. Ramnarain-Seetohul, "A Knowledge Based System for Automated Assessment of Short Structured Questions," in *- 2019 IEEE Global Engineering Education Conference (EDUCON)*, 2019, pp. 1349-1352.

[27] M. Floryan, T. Dragon, N. Basit, S. Dragon and B. Woolf, "Who Needs Help? Automating Student Assessment Within Exploratory Learning Environments," in *Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science*, 2015, pp. 125-134.

[28] I.A. Hameed, "A Fuzzy System to Automatically Evaluate and Improve Fariness of Multiple-Choice Questions (MCQs) Based Exams," in *Proceedings of the 8th International Conference on Computer Supported Education*, Rome, Italy, 2016, pp. 476–481.

[29] D. Gursoy, O.H. Chi, L. Lu and R. Nunkoo, "Consumers acceptance of artificially intelligent (AI) device use in service delivery," *Int.J.Inf. Manage.*, vol. 49, pp. 157-169, 2019.

[30] T. Zhang, D. Tao, X. Qu, X. Zhang, J. Zeng, H. Zhu and H. Zhu, "Automated vehicle acceptance in china: Social influence and initial trust are key determinants," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 220-233, 2020.

[31] M. Beynon, M.M.H. Goode, L.A. Moutinho and H.R. Snee, "Modelling satisfaction with automated banking channels,", vol. 26, no. 4, pp. 77-94, 2005. https://doi.org/10.1300/J396v26n04_05.

[32] F.D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319-340, 1989. http://www.jstor.org/stable/249008.

[33] W.R. King and J. He, "A meta-analysis of the technology acceptance model," *Information & Management*, vol. 43, no. 6, pp. 740-755, 2006.

[34] M. Al-Emran, V. Mezhuyev and A. Kamaludin, "Towards a conceptual model for examining the impact of knowledge management factors on mobile learning acceptance," *Technology in Society*, vol. 61, pp. 101247, May 2020.

[35] S.S. Al-Gahtani, "Empirical investigation of e-learning acceptance and assimilation: A structural equation model," *Applied Computing and Informatics*, vol. 12, no. 1, pp. 27-50, 2016.

[36] E. Ibili, D. Resnyansky and M. Billinghurst, "Applying the technology acceptance model to understand maths teachers' perceptions towards an augmented reality tutoring system," *Education and Information Technologies*, vol. 24, no. 5, pp. 2653-2675, 2019. https://doi.org/10.1007/s10639-019-09925-z.

[37] W. M. Al-Rahmi, N. Yahaya, A. A. Aldraiweesh, M. M. Alamri, N. A. Aljarboa, U. Alturki and A. A. Aljeraiwi, "Integrating technology acceptance model with innovation diffusion theory: An empirical investigation on students' intention to use E-learning systems," *IEEE Access*, vol. 7, pp. 26797-26809, 2019.

[38] T. Teo, F. Huang and C.K.W. Hoi, "Explicating the influences that explain intention to use technology among english teachers in china," *Interactive Learning Environments*, vol. 26, no. 4, pp. 460-475, 2018. https://doi.org/10.1080/10494820.2017.1341940.

[39] J.C. Sánchez-Prieto, S. Olmos-Migueláñez and F.J. García-Peñalvo, "Informal tools in formal contexts: Development of a model to assess the acceptance of mobile technologies among teachers," *Comput.Hum.Behav.*, vol. 55, Part A, pp. 519-528, 2 2016.

[40] H.I. Jeong and Y. Kim, "The acceptance of computer technology by teachers in early childhood education," *Interactive Learning Environments*, vol. 25, no. 4, pp. 496-512, 2017. https://doi.org/10.1080/10494820.2016.1143376.

[41] M. Fishbein and I. Ajzen, *Belief, attitude, intention, and behavior : an introduction to theory and research.* Reading, Massachusets, Addison-Wesley Pub. Co., 1975.

[42] I. Ajzen, "From Intentions to Actions: A Theory of Planned Behavior," in *Action Control. SSSP Springer Series in Social Psychology*, J. Kuhl and J. Beckmann, Eds: Springer Berlin Heidelberg, 1985, pp. 11-39.

[43] V. Venkatesh and F.D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management Science*, vol. 46, no. 2, pp. 186-204, 2000.

[44] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," *Decision Sciences*, vol. 39, no. 2, pp. 273-315, 2008.

[45] K. Zarafshani, A. Solaymani, M. D'Itri, M.M. Helms and S. Sanjabi, "Evaluating technology acceptance in agricultural education in iran: A study of vocational agriculture teachers," *Social Sciences & Humanities Open*, vol. 2, no. 1, pp. 100041, 2020.

[46] M. Ismail, E. Çelebi and H. Nadiri, "How student information system influence students' trust and satisfaction towards the university?: An empirical study in a multicultural environment," *IEEE Access*, vol. 7, pp. 111778-111789, 2019.

[47] O. Gillath, T. Ai, M. Branicky, S. Keshmiri, R. Davison and R. Spaulding, "Attachment and trust in artificial intelligence," *Comput.Hum.Behav.*, pp. 106607, 2020.

[48] E. Glikson and A.W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *ANNALS*, vol. 14, no. 2, pp. 627-660, 2020. https://doi.org/10.5465/annals.2018.0057.

[49] S.K. Sharma and M. Sharma, "Examining the role of trust and quality dimensions in the actual usage of mobile banking services: An empirical investigation," *International Journal of Information Management*, vol. 44, pp. 65-75, 2019.

[50] Z. Shao, L. Zhang, X. Li and Y. Guo, "Antecedents of trust and continuance intention in mobile payment platforms: The moderating effect of gender," *Electronic Commerce Research and Applications*, vol. 33, pp. 100823, 2019.

[51] M. Warkentin, S. Sharma, D. Gefen, G.M. Rose and P. Pavlou, "Social identity and trust in internet-based voting adoption," *Government Information Quarterly*, vol. 35, no. 2, pp. 195-209, 2018.

[52] D. Gefen, E. Karahanna and D.W. Straub, "Trust and TAM in online shopping: An integrated model," *MIS Quarterly*, vol. 27, no. 1, pp. 51-90, 2003. http://www.jstor.org/stable/30036519.

[53] A.A. Almazroi, H. Shen and F. Mohammed, "The Impact of Trust on the Adoption of Cloud Computing Services by University Students," in *Recent Trends in Data Science and Soft*, 2019, pp. 902-911.

[54] F. Qin, K. Li and J. Yan, "Understanding user trust in artificial intelligence-based educational systems: Evidence from china," *Br J Educ Technol*, vol. 51, no. 5, pp. 1693-1710, 2020. https://doi.org/10.1111/bjet.12994.

[55] L.M. Arciniega and A. Maldonado, "Assessing the impact of dispositional resistance to change on organizational attraction," *The Spanish journal of psychology*, vol. 14, no. 02, pp. 798-807, 2011.

[56] I.B. Saksvik and H. Hetland, "Exploring dispositional resistance to change," *Journal of Leadership & Organizational Studies*, vol. 16, no. 2, pp. 175-183, 2009.

[57] X. Guo, Y. Sun, N. Wang, Z. Peng and Z. Yan, "The dark side of elderly acceptance of preventive mobile health services in china," *Electronic Markets*, vol. 23, no. 1, pp. 49-61, 2013. http://dx.doi.org/10.1007/s12525-012-0112-4.

[58] R.T. Cenfetelli, "Inhibitors and enablers as dual factor concepts in technology usage," *Journal of the Association for Information Systems*, vol. 5, no. 11, pp. 473-492, 2004.

[59] J.C. Sánchez-Prieto, F. Huang, S. Olmos-Miguéláñez, F.J. García-Peñalvo and T. Teo, "Exploring the unknown: The effect of resistance to change and attachment on mobile adoption among secondary pre-service teachers," *Br J Educ Technol*, vol. 50, no. 5, pp. 2433-2449, 2019. https://doi.org/10.1111/bjet.12822.

[60] S. Taylor and P. Todd, "Assessing IT usage: The role of prior experience," *MIS Quarterly*, vol. 19, no. 4, pp. 561-570, 1995. http://www.jstor.org/stable/249633.

[61] S. Taylor and P.A. Todd, "Understanding information technology usage: A test of competing models," *Information Systems Research*, vol. 6, no. 2, pp. 144-176, 1995. http://dx.doi.org/10.1287/isre.6.2.144.

[62] J. Cruz-Benito, R. Therón, F.J. García-Peñalvo and E. Pizarro Lucas, "Discovering usage behaviors and engagement in an educational virtual world," *Comput.Hum.Behav.*, vol. 47, pp. 18-25, 2015.

José Carlos Sánchez-Prieto

José Carlos Sánchez-Prieto received a bachelor's degree in Pedagogy and a master's degree in ICT applied in education from the University of Salamanca (Spain) where he also presented his PhD on the adoption of mobile technologies among teachers within the Programme on Education in the Knowledge Society. Currently he is an associate professor at the Faculty of Education of said university. His area of research is the assessment of attitudes among in-service and pre-service teachers and students. He is the author of several articles on the subject published in prestigious international journals. You can contact him at: josecarlos.sp@usal.es.

Juan Cruz-Benito

Juan Cruz-Benito received a Ph.D. in Computing Engineering from the University of Salamanca, Spain, in 2018. Previously, he received the M.Sc. degree in Intelligent Systems, the B.Sc. in Computer Systems, and the Technical Engineering Degree in Computer Systems from the same university in 2013, 2012, and 2011. He is currently Senior Software Engineer at IBQ Quantum, IBM Research's T.J. Watson Research Center, NY, USA. Before it, he worked for five years in research & development projects at the GRIAL Research Group, University of Salamanca, Spain, where he participated as a software engineer, researcher, and developer in many European and Spanish R&D projects (public and private ones). His research interests involve Human-Computer Interaction, Machine Learning, Data Science, and technologies for educational purposes. He is the co-author of more than 80 publications, serves as a reviewer for many high-respected journals and conferences, and is Associate Editor of the Journal of Global Information Management (JGIM) and the Journal of Information Technology Research (JITR). His research activity conducted him to receive the Research Award "Scientific Computer Society of Spain and BBVA Foundation" as one of the best young researchers in Computer Science in Spain in 2019. As a software engineer and developer, he is involved in many different technical communities and Open Source initiatives. Contact him at https://juancb.es/ or juan.cruz@ibm.com

Roberto Therón

Roberto Therón Sánchez is currently the Manager of the VisUSAL Group (within the Recognized Research Group GRIAL), University of Salamanca, Salamanca, Spain, which focusses on the combination of approaches from computer science, statistics, graphic design, and information visualization to obtain an adequate understanding of complex datasets. He received the Diploma degree in computer science from the University of Salamanca, the B.S. degree from the University of A Coruña, the B.S. degree in communication studies and the B.A. degree in humanities from the University of Salamanca, and the Ph.D. degree from the Research Group Robotics, University of Salamanca. His Ph.D. thesis was on parallel calculation of the configuration space for redundant robots. He has authored more than 100 articles in international journals and conferences. In recent years, he has been involved in developing advanced visualization tools for multidimensional data, such as genetics or paleoclimate data. In the field of visual analytics, he develops productive collaborations with groups and institutions internationally recognized as the Laboratory of Climate Sciences and the Environment, France, or the Austrian Academy of Sciences, Austria. He was the recipient of the Extraordinary Doctoral Award for his Ph.D. thesis. Contact him at theron@usal.es.

Francisco José García-Peñalvo

Francisco José García-Peñalvo is a Full Professor in the Department of Computer Science and Automation at the University of Salamanca (USAL), with three six-year periods of research, one six-year period of transferring and innovation, and four five-year periods of recognized teaching. He received the Beatriz Galindo award for teaching excellence in 2019. He was also a Distinguished Professor at the School of Humanities and Education at the Tecnológico de Monterrey, Mexico and a Researcher of International Impact at the Universidad Nacional San Agustín, Arequipa, Peru. Since 2006 he is the head of the Research Group Recognized by the USAL GRIAL (research GRoup on InterAction and eLearning), a group that is a Consolidated Research Unit of the Junta de Castilla y León Government (UIC 81). He has been Vice-Dean of Innovation and New Technologies of the Faculty of Sciences of the USAL between 2004 and 2007 and Vice-Rector of Technological Innovation of this University between 2007 and 2009. He is currently the Rector's Delegate for Virtual Teaching and the Coordinator of the Doctorate Programme in Education in the Knowledge Society at USAL. He is Editor-in-Chief of the journals Education in the Knowledge Society and Journal of the Information Technology Research, and Associate Editor of many journals, with a special mention to the journal Computers in Human Behavior Reports. He has published more than 100 research papers in JCR-indexed journal (51 Q1). For more detailed information on the publications, these are the public links to the profiles in Google Scholar (http://goo.gl/sDwrr0), Publons (https://bit.ly/2u2FN5l) and ORCID (http://orcid.org/0000-0001-9987-5584).

A Fine-Grained Model to Assess Learner-Content and Methodology Satisfaction in Distance Education

Magdalena Cantabella, Raquel Martínez-España*, Belén López, Andrés Muñoz

Escuela Politécnica, Universidad Católica de Murcia (Spain)

unir
LA UNIVERSIDAD EN INTERNET

Abstract

Learning Management System (LMS) platforms have led to a transformation in Universities in the last decade, helping them to adapt and expand their services to new technological challenges. These platforms have made possible the expansion of distance education. A current trend in this area is focused on the evaluation and improvement of the students' satisfaction. In this work a new tool to assess student satisfaction using emoticons (smileys) is proposed to evaluate the quality of the learning content and the methodology at unit level for any course and at any time. The results indicate that the assessment of student satisfaction is sensitive to the period when the survey is performed and to the student's study level. Moreover, the results of this new proposal are compared to the satisfaction results using traditional surveys, showing different results due to a more accuracy and flexibility when using the tool proposed in this work.

Keywords

I. Introduction

IN the new era of Education, the use of Information Technologies (IT) has already consolidated itself as a fundamental asset. The application of these technologies has allowed, among others, the integral development of distance education, especially at the higher education level. One of the tools that has enabled this development are the Learning Management System (LMS) platforms. These platforms allow students to work independently, facilitating the interaction with other users by means of collaborative tools and providing new methods for resource management that help to strengthen new teaching and learning models. Moreover, the LMS platforms act as content managers that provide students with a wide variety of resources.

A current trend regarding distance education is concerned with evaluating the quality in this methodology [1]-[3]. The main items that are considered in this process are the quality of the available resources, the quality of student-student and student-teacher interactions and the flexibility and ease of use of the LMS. A conventional approach to obtaining the data to evaluate these items is through surveys, interviews and/or focus groups. However, these methods do not seem totally well-suited for distance education methodology due to absence of objective measures, lack of evaluation of specific distance education items or imposing a period to complete the survey while using a virtual environment which is available 24/7, among others [4]-[5]. An alternative approach is based on the use of Learning Analytics (LA) methods [6]-[7]. This strategy focuses on analyzing the information about the users' activity that is recorded in the LMS, which allows to establish behavioral profiles not only for students but also for lecturers [8]-[11].

While both approaches foster the gathering of complementary information to assess the quality of teaching, it is detected a lack of fine-grained information about the student's satisfaction with respect to the content provided by the lecturer. In other words, there is a gap in the methods applied to obtain detailed information about the quality of each specific resource in each lesson of any subject, namely text documents, multimedia contents, self-assessment tests, assignments, etc. In this paper it is proposed a simple and effective method based on the use of smileys integrated into an LMS to collect these data. As a result, the evaluation of these substantial data can help lecturers to detect not only content that should be improved but also what type of content is the most enjoyed by the students. As a proof of concept, we evaluate the application of this method in a case study at the Catholic University of Murcia (UCAM), Spain.

The main objective of this study is to develop a method to measure the student's satisfaction respect to the available resources and methodology in each of the units of each subject. With these assessments an evaluation will be carried out to, on the one hand, analyze the students' behavior with respect to this method depending on factors such as their study level or time period when they perform the proposed evaluation method and, on the other hand, compare with data from traditional surveys answered by the same group of students so as to check if there are significant differences among the results of both methods.

The rest of the paper is structured as follows. Related works in the area are reviewed in Section II. The methodology followed in this work and the method proposed for student satisfaction evaluation is explained in Section III. Section IV discusses the results obtained in the case study developed to evaluate our proposal. Finally, Section V summarizes the findings obtained in this work.

* Corresponding author.

E-mail address: rmartinez@ucam.edu

II. Related Work

In order to evaluate the quality of the methodologies applied in distance education, it is necessary to distinguish the most relevant factors that allow an integral analysis of the main milestones in such methodologies.

Initially, the evaluation applied to distance learning only focused on evaluating the quality of the software or the LMS technologies in use [12]-[13], leaving aside the evaluation of the methodology and contents of the courses, being both crucial aspects for measuring the quality. Later on, Plomp and Ely [14] included a more complete evaluation of the quality of distance education courses by suggesting four categories to evaluate: (1) course design, (2) resource selection, (3) methodology and (4) software training for lecturers. One of the main criticisms of this work was that it proposed the instructors as the only ones responsible for reviewing and verifying the quality of the contents, being questionable whether other agents were needed in this process or whether the instructors had enough time for this laborious procedure [15].

There have also been efforts in defining standards and frameworks to evaluate the quality of distance education, for example by adapting software evaluation standards such as the ISO9241-210 "Ergonomics of Human System Interaction", higlighting the fact that the user experience has a fundamental role to evaluate the functionality, reliability, usability, efficiency, portability and maintainability of the LMS [16]-[17]. On the other hand, the SCORM (Sharable Content Object Reference Model) proposal [18] includes a set of rules for the reuse of content between LMSs in order to achieve a learning process with a common structure. It uses a set of standards and specifications that analyze the relationships and levels of granularity between the materials of different units in order to automatically manage the content of those units and reuse it between different platforms. A third proposal reflects on how to describe, capture, and communicate more effectively the complex and iterative nature of data visualization design throughout the research, design, development, and deployment of visualization systems and tools [19].

Recent studies have identified that the quality of the available services and the students' satisfaction are of great relevance for measuring the quality of distance education. Many researchers agree that student satisfaction is an important factor to be valued because it is in many cases linked to their academic performance and university experience [20]-[23]. Now with the use of LMS platforms, this student satisfaction in higher education, transformed into their university experience, is considered as a key component, since if the student is not satisfied with some component of the online course, he/she has a greater probability to transfer to other institutions [24]-[25]. The lecturer's feedback is also a very important factor, as shown by several research works stating that developing new tools in the LMS or updating them without taking into account the satisfaction of the instructors negatively affects the results of the distance learning course [26]-[28].

Different models, surveys and questionnaires have been used to measure student satisfaction, see for example [24], [29]-[31] to name just a few. Despite the differences in such evaluation items, it is possible to identify a common set of main factors affecting student satisfaction. These factors include student-lecturer interaction, student-student interaction, the learning content (resources) and system flexibility and support [24]. In this regard, the use of LMS positively impacts student satisfaction, higlighting the availability of resources, system accessibility, and its tools as the determinants factor of LMS self-efficacy [31]. The inclusion of the learning content among these factors is noteworthy, since it emphasizes that the effective configuration of curriculum content and pedagogical content is necessary to create an effective learning experience. In this work we investigate how to evaluate in a more detailed manner this learning content.

New assessment models have also been considered in this study to identify student satisfaction. Currently it is a challenge to develop new tools that specifically assess the resources available to students in LMS. Findings on this research line will help institutions by providing them with psychometric properties that add pedagogical value to distance learning. In [32], a framework has been designed to guide institutions to better improve student's satisfaction and further strengthen their e-learning implementation. Authors have shown that the satisfaction can be predicted mostly by student's interactions. Another interesting project can be found elsewhere [33], where authors introduce an intelligent classroom system that is able to classify student's satisfaction by examining the parameters of the physical environment obtained with different intelligent devices. Whilst these works focus on particular interaction and physical context information, they do not evaluate the resources and methodology of the course.

Following this research line on measuring the student's satisfaction, the work proposed in this paper advances in the evaluation of learning content by providing a method to assess different types of specific content in a more fine-grained level compared to the works reviewed above.

III. Materials and Methods

This work proposes a new method to evaluate students' satisfaction with respect to specific elements of learning content in distance courses. A case study to evaluate this method has been developed in the Catholic University of Murcia (UCAM), where several blended and online courses are available. In particular, Sakai[1] has been used as the LMS platform for this study, since it is the one adopted in UCAM. A regulation for these blended and online modalities exists to ensure that lecturers use a common framework with the aim of, on the one hand, providing students with quality resources and, on the other hand, ensuring that the students continue engaged in the course and therefore reducing drop-out rates. To achieve these two objectives, the regulation establishes certain parameters that have been designed by the Vice-Rectorate of Virtual Education at UCAM. Among the most important parameters, for lecturers there is a certain maximum number of days to correct tasks or to answer forums or private messages. This allows the students to know beforehand what the waiting times are and thus they are able to do a better planning. To keep track of these answer times and the compliance of the lecturers with the regulation, the university utilizes a tool called Online Data [34], which is integrated within the Sakai LMS.

This LMS also provides a tool for organizing content called "Lesson Builder" that allows students to browse learning content of various types organized by topics or units. This content includes every available resource from text material to audiovisual material, as well as direct access to assignments, forums, videoconferences or self-assessments. However, the Lesson Builder tool does not allow gathering students' opinions about each of the provided content. For this reason, a specific tool based on the representation of smileys has been developed in this work to allow students to evaluate such contents organized in their corresponding units.

Fig. 1 shows an example of the smileys tool integrated into a Lesson Builder unit. As can be observed, at the top of each unit it is displayed three smiley emoticons. Each of these emoticons allows student to express their satisfaction with the learning content of that unit. The students can express their satisfaction related to the 5 dimensions that are directly related to each unit: media resources,

---

[1] https://sakaiproject.org/
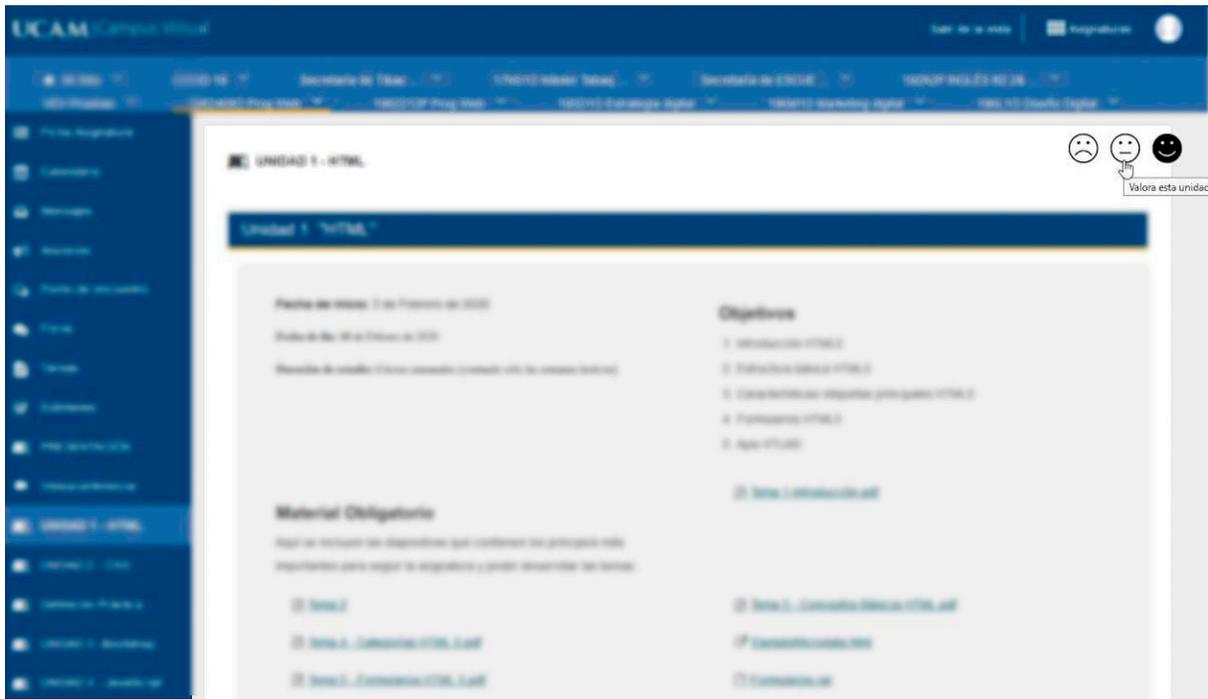
Fig. 1. A screenshot of a Lesson Builder unit integrating the smileys tool (upper right corner). Content is intentionally blurred for the shake of lecturer's privacy.

text resources, assignments, self-assessment and methodology. These five dimensions are grouped into two general categories, namely "Resources" and "Methodology". The Resources category groups the first 4 dimensions, while the Methodology category is composed of the dimension with the same name. The Resources category aims to analyze student satisfaction with respect to the available resources, assignments and self-assessment items. The methodology category evaluates the follow-up of the lecturer for that unit. Each dimension is assessed using a Likert-type scale from 1 (strong negative perception) to 5 (strong positive perception). Thus, when the student clicks on any of the smileys a satisfaction evaluation screen is displayed (see Fig. 2).

Fig. 2. A screenshot showing how the student's satisfaction is gathered for the learning content of each unit once a smiley emoticon has been clicked on.

If the chosen option is the sad smiley, all the dimensions are marked with one star. If it has been the neutral smiley, three stars are marked for each dimension (exactly as it appears in Fig. 2). If the student selects the happy smiley, then the five stars will be highlighted for each dimension. Regardless of the selected smiley, the student could modify the satisfaction for each item individually and he/she

may even write comments to justify the evaluation or suggestions to improve the contents. Students can anonymously evaluate each unit at any time during the academic year. Evaluations can be updated at any time; however, the system only stores the most recent evaluation. Observe also that a student can evaluate separately the different units of a subject, therefore the same subject can have several evaluations from the same student (one for each unit in the subject).

Once the data is gathered using the smileys tool, a statistical analysis will be performed to find if there are any significant differences in the results of student satisfaction according to the students' study levels and according to the different periods of time of an academic year (divided into months). These analyses will distinguish between student satisfaction with respect to the Resources category and the Methodology category. The non-parametric statistical test of Kruskal-Wallis, the Dunn-Bonferroni test post hoc and the Kolmogorov-Smirnov test will be used for statistical analysis.

On the other hand, prior to the implementation of this new tool based on smileys, only traditional surveys had been used at UCAM to evaluate student satisfaction (an example of the questionnaire used in this survey can be found in Annex I). In these traditional surveys, students are asked to evaluate four dimensions with respect to the lecturers' performance in the LMS: Methodology, Planning, Resources and General Overview. These dimensions are graded by the students following a Likert-type scale from 1 (strong negative perception) to 5 (strong positive perception). The activation of these traditional surveys takes place only in the last month of each academic quarter prior to the final exams. Responses to these traditional surveys are also anonymous and a student can only take one survey per subject (differently from the smileys method, where the evaluation takes place at the unit level and there could be more than one response by the same student for the same subject). In order to search for differences between the results of both types of evaluation of student satisfaction, a preliminary comparative study between the results for the Resources and Methodology dimensions of each method will be analyzed in this work using the visualization tool QlikSense [35].

To perform the case study a total of 245 students have participated

from 49 different blended and online courses during the 2018/19 academic term, classified in four study levels: Degree, Master, Ph.D. and Own Degree (i.e., university specific degrees endorsed by the prestige of the university but without the official recognition of the State as in the rest of the degrees). For the evaluation of the learning content one subject for each one of the 49 courses has been selected. Table I shows the number of subjects and students involved in the case study classified by study level along with the number of responses gathered in each level through the smileys tool (Resp. Smileys) and the number of responses obtained in traditional surveys (Resp. Traditional). It should be noted that the students who have participated in both surveys are the same, but since the surveys are anonymous, it is not possible to match the students' answers in both surveys in order to make an association between the two responses (i.e., they cannot be considered as samples of related responses). For this reason, the available data set is considered a set of independent samples and hence the justification for the application of the Kruskal-Wallis statistical test for the analysis of results.

TABLE I. Number of Subjects, Students and Responses Involved in Each Study Level. The Study Has Been Performed During the 2018/19 Academic Year

Level	Degree	Master	Ph.D.	Own Degree	Total
Students	50	100	20	75	245
Subjects	10	20	4	15	49
Resp.Smileys	582	211	63	63	919
Resp.Traditional	49	97	19	75	240

The data shown in Table I comprises the dataset for the study performed in this work. Specifically, the dataset consists of the following attributes:

- Type of survey: Traditional or smileys method.
- Subject: Subject that receives the evaluation.
- Study level: Level of studies to which the subject corresponds.
- Assessment for methodology: Assessment on the scale of 1 to 5 made for the methodology section.
- Assessment for resources: Assessment on the scale of 1 to 5 made for the resource items.
- Date: Date of the survey.

Using this dataset, in this paper we consider the following research questions:

- RQ1. Are there differences in the resource evaluation according to the students' study level when using the smileys tool?
- RQ2. Are there differences in the methodology evaluation according to the students' study level when using the smileys tool?
- RQ3. Are there differences in the resource evaluation according to the period of the survey response when using the smileys tool?
- RQ4. Are there differences in the methodology evaluation according to the period of the survey response when using the smileys tool?
- RQ5. Are there differences in students' satisfaction results about resources and methodology depending on the type of survey (traditional vs. smileys)?

The first four RQs perform a study focused only on the proposed smileys tool. The RQ1 studies the differences in the evaluation of the resources dimension depending on the level of studies that the student is taking. RQ2 studies the differences in the methodology dimension depending on the level of studies the student is taking. RQ3 studies if there are differences in the evaluation of the resources dimension

depending on the academic period in which the student answers the survey. RQ4 studies if there are differences in evaluation in the methodology dimension depending on the academic period in which the student responds to the survey. Finally, RQ5 analyzes at a global level if there are differences in the evaluation of the resource and methodology dimensions when the evaluation is done by means of the traditional surveys compared to the new smileys tool.

IV. Results

This section explores the results of the analysis proposed in Section III. Firstly, a general view of the results obtained by means of the smileys tool is displayed in Fig. 3. It shows the average values obtained for each dimension evaluated through this tool (see Section III) along the academic year for all the study levels. Although the data are shown from January to December, it is important to bear in mind that Fig. 3 reflects the two academic quarters in the Spanish academic year: the first quarter begins in mid-September and ends in January (with the final exams for this period) and the second quarter begins in mid-February and ends at the end of June (again with the final exams for this second period). September was the month for the remedial exams at UCAM in the 2018/19 academic year.

Analyzing the results obtained for each dimension, in general they follow the same trend. The best valued dimension is the media resources followed by the text resources, while the worst valued one is the self-assessment content. The increase in better average values coincides with the initial months of each quarters, namely September and February, and with the period of examinations corresponding to the months of January, May and the end of August/beginning of September. The lowest averages are shown for the Christmas period (December) and summer (June, July and August), both of them included in the holiday periods at UCAM.

Next, it is performed a statistical analysis of the results of the student satisfaction gathered by means of the smileys tool to evaluate research questions RQ1-RQ4. For all of them, the Kolmogorov-Smirnov test is applied in the first place to check for the normality of data. The test returns a p-value = 0.0 for each research question, therefore it can be stated with a 95% confidence level that the data do not follow a normal distribution for any research question. Therefore, non-parametric tests must be applied and the Kruskal-Wallis test is used as the non-parametric alternative to the One-Way ANOVA. In order to adjust the p-value and get the significant differences at a general level, the Dunn-Bonferroni post-hoc test is applied when necessary.

A. Research Questions RQ1 & RQ2

The null hypothesis to tackle RQ1 indicates that there are no significant differences in the student satisfaction regarding the Resource category for any study level (with a 95% confidence level). The Kruskal-Wallis test returns a p-value=0.048, and therefore the null hypothesis is rejected. Hence, there are significant differences in the student satisfaction regarding the Resource category depending on the study level with a 95% confidence level.

Table II shows the p-value and the adjusted p-value obtained after using the Dunn-Bonferroni post hoc adjustment test for RQ1. This table only shows combinations of pairs that have significant differences in the p-value with a confidence value of 95%. As can be seen on an individual level (p-value without adjusting), there are significant differences between the student satisfaction related to resources of own-degree and master's degree, as well as between the satisfaction related to resources of degree and master's degree. As can be seen in Table IV, Own-degree students are more satisfied with the available resources (slightly above 4 points) than the Master's students, whose satisfaction is lower and stands at 3.4 points. This table also shows
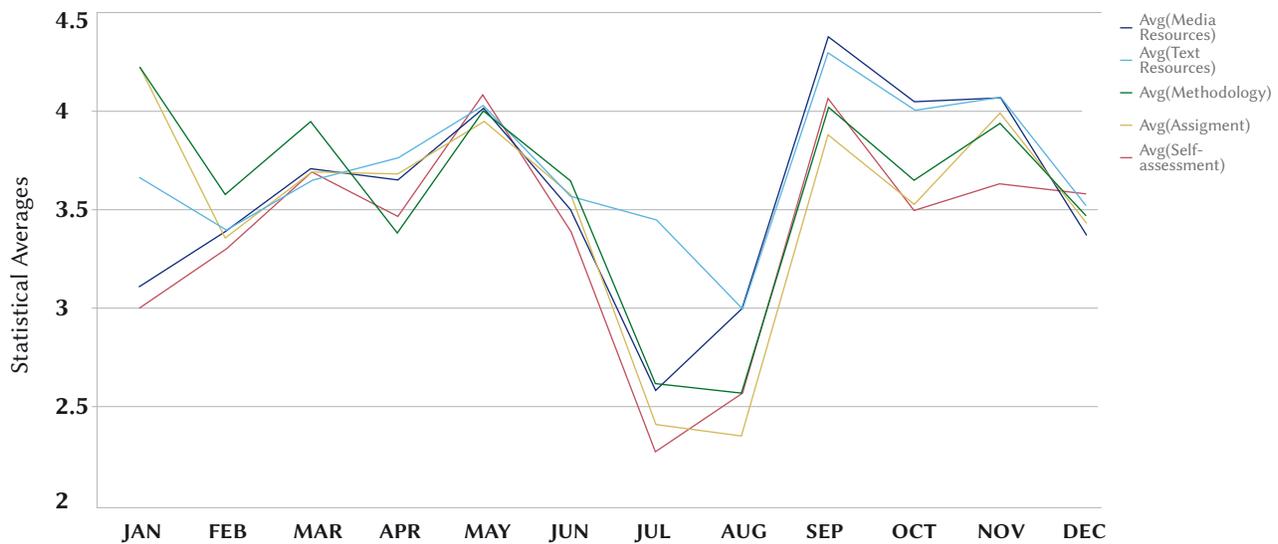
Fig. 3. Average values for each dimension (media resources, text resources, assignments, self-assessment tests and methodology) evaluated through the smileys tool along the academic year.

the difference in satisfaction about resources between Master's and Degree students. In this case, the difference is somewhat smaller, since the average satisfaction of the latter is 3.7 points. At the general level (adjusted p-value) there are no significant differences in satisfaction between Degree and Master's students, but there are still significant differences between the satisfaction in the Own-degree and Master's students with a confidence level of 93%.

TABLE II. P-Value Results With 95% Confidence Level Obtained With Kruskal Wallis Test and Adjusted P-Value Obtained With Dunn-Bonferroni Post Hoc Test for the Student's Satisfaction Regarding the Resource Category Grouped By Study Levels

Pairs of study Level	Own-degree / Master	Degree / Master
P-value	0,011	0,048
Adjusted p-value	0,065	0,288

Regarding RQ2, the null hypothesis indicates that there are no significant differences in student satisfaction regarding the Methodology category for any study level (with a 95% confidence level). The Kruskal-Wallis test returns a p-value=0.01, and therefore the null hypothesis is rejected. Hence, there are significant differences in the student satisfaction regarding the Methodology category depending on the study level with a 95% confidence level.

Table III indicates the p-value and the adjusted p-value obtained after using the Dunn-Bonferroni post hoc adjustment test for RQ2. This table only shows combinations of pairs that have significant differences in the p-value with a confidence value of 95%. In the comparisons between pairs, there are significant differences regarding the satisfaction about the methodology in the study levels of Own-Degree and Degree, Degree and Master and Degree and Ph.D. Note that at the general level they are not significant, as demonstrated through the adjusted p-value.

TABLE III. P-value Results With 95% Confidence Level Obtained With Kruskal Wallis Test and Adjusted P-value Obtained With Dunn-Bonferroni Post Hoc Test for the Student's Satisfaction Regarding the Methodology Category Grouped By Study Levels

Study Level	Own-degree / Master	Degree / Master	Degree/Ph.D.
P-value	0,035	0,039	0,017
Adj. p-value	0,209	0,234	0,101

For the smileys tool, the values of Table IV show the mean values (standard deviation in sub-index) for the resource and methodology evaluation, only showing the pairs of study levels for which there are significant differences.

TABLE IV. Average Results of the Student Satisfaction Regarding Resources and Methodology Levels Grouped By Study Levels (Data Gathered Through the Smileys tool). The Sub-Indexes of Each Mean Value indicate Its Standard Deviation

Study Level	Degree	Master	Ph.D.	Own Degree
Resources	3.71.8	3.41.7	3.61.7	4.01.3
Methodology	3.81.8	3.51.9	3.32.0	3.61.9

B. Research Questions RQ3 & RQ4

The null hypothesis related to RQ3 indicates that there are no significant differences in the student satisfaction for any month of the year regarding the Resource category (with a 95% confidence level). The Kruskal-Wallis test returns a p-value=0.0, and therefore the null hypothesis is rejected. Thus, there are significant differences in the student satisfaction with respect to the Resource category depending on the month of the year in which the evaluation is conducted.

Table V shows the p-value and the adjusted p-value obtained after using the Dunn-Bonferroni post hoc adjustment for RQ3. The first column shows the number of the month instead of the name, indicating by a '/' which pairs of months have significant differences. This table only shows combinations of pairs that have significant differences in the p-value with a confidence value of 95%. Among individual pair comparisons (according to the p-value) there are significant differences between July and the following months: April, May, June, September, October and November; and between the months of June and September and May and June. In a general interpretation of the results (observing the adjusted p-value), there are only significant differences between the month of July and the following months: April, May, September, October and November. This result coincides with the data displayed in Fig. 4, being July the month when students give a worse evaluation of resources compared to the other months.

Finally, the null hypothesis related to RQ4 indicates that there are no significant differences in the student satisfaction for any month of the year regarding the Methodology category (with a 95% confidence
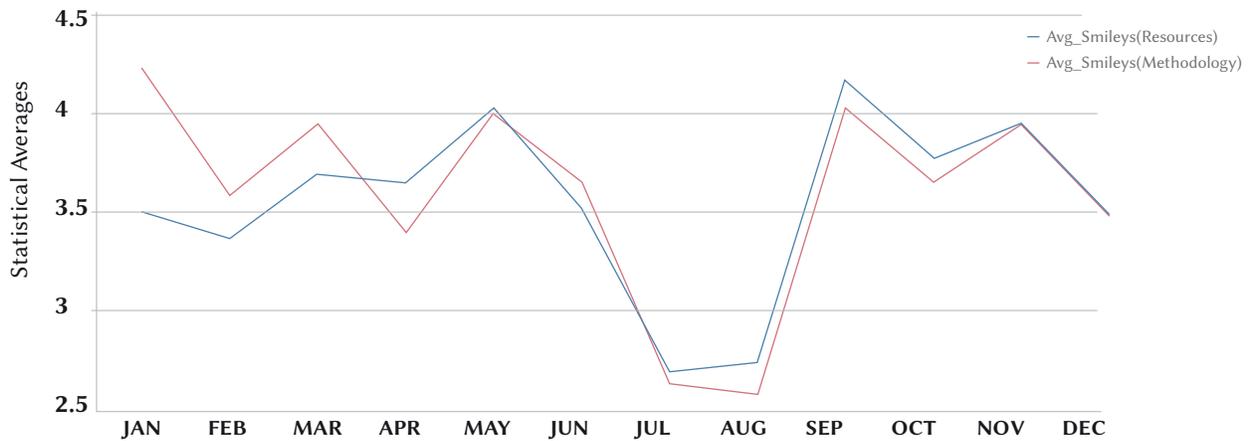
Fig. 4. Average values for the Resource and Methodology categories throughout the academic year (in months) obtained from the smileys tool.

level). The Kruskal-Wallis test returns a p-value=0.0, and therefore the null hypothesis is rejected. Thus, there are significant differences in the student satisfaction with respect to the Methodology category depending on the month of the year in which the evaluation is conducted.

TABLE V. P-Value Results With 95% Confidence Level Obtained With Kruskal Wallis Test and Adjusted P-Value Obtained With Bonferroni Post Hoc Test for the Student's Satisfaction in the Resource Category Throughout the Academic Year

Pairs of Months	P-value	Adjusted p-value
7/4	0,004	0,03
7/5	0,0	0,001
7/6	0,009	0,09
7/9	0,0	0,002
7/10	0,0	0,004
7/11	0,0	0,0
6/9	0,042	0,4
6/5	0,01	0,1

TABLE VI. P-value Results With 95% Confidence Level Obtained With Kruskal Wallis Test and Adjusted P-value Obtained With Bonferroni Post Hoc Test for the Student's Satisfaction in the Methodology Category Throughout the Academic Year

Pairs of Months	P-value	Adjusted p-value
7/1	0,03	0,2
7/2	0,004	0,03
7/3	0,0	0,001
7/4	0,01	0,1
7/5	0,0	0,0
7/6	0,001	0,01
7/9	0,0	0,001
7/10	0,001	0,007
7/11	0,0	0,0
7/12	0,01	0,1
4/11	0,023	0,2

Table VI presents the p-value and the adjusted p-value obtained after using the Dunn-Bonferroni post hoc adjustment test for RQ4. Again, the first column shows the month number instead of the name. This table only presents combinations of pairs that have significant differences in the p-value with a confidence value of 95%. In an individual comparison of pairs, it can be seen that there are significant differences in the assessment of student satisfaction in the Methodology dimension between the month of July and the following

months: January, February, March, April, May, June, September, October, November and December. There are also significant differences between June and May and between June and September. This can be seen in Fig. 4 since in June the satisfaction values regarding the methodology are lower than in May and September. Analyzing the global results with the adjusted p-value, there are only significant differences globally between the month of July and the following months: February, March, May, June, September, October and November. These differences are appreciated visually in Fig. 4.

In summary, the significant differences in the student satisfaction value for the Resources and Methodology dimensions are different with respect to the months in the academic year, with July being the worst rated month in terms of satisfaction in both the Resources and Methodology categories.

C. Research Question RQ5

Fig. 5 shows a bar graph which compares the results obtained by the two methods of satisfaction evaluation presented in Section III, namely traditional surveys and the smileys tool. The error bars represent the 95% confidence interval for the average values. The results are analyzed with respect to the two comparable categories of each method (Resources and Methodology) and grouped by the different study levels. For the traditional survey method, the blue color has been selected to represent the average value for Resources and the yellow color for the average value of Methodology. For the smileys tool method, the green color has been selected to represent the average value of Resource and the red color for the average value of Methodology. A more in-depth statistical analysis cannot be applied at the moment as the disaggregated data for the traditional survey method have not been made available, but only the mean values and standard deviation of each evaluated category for each study level.

All study levels have been considered for this study. However, it is important to highlight that, at the time of this study, the university had not implemented yet quality assessment processes in Ph.D. courses by means of traditional surveys due to their recent implementation (from 2015). In this case it has not been possible to perform a comparative study with the results obtained by the smileys tool method.

It is observed that, in general, the results obtained through traditional surveys show a greater student satisfaction than when using the smileys tool. This fact may be caused due to several factors: Firstly, traditional surveys are designed to evaluate the learning content and the methodology of any subject in a global manner, whereas the smileys tool allows for a more fine-grained evaluation. Therefore, if the content or the methodology of a specific unit is evaluated not as positive as the rest of the units, it may affect the global evaluation.

Secondly, traditional surveys are activated during a limited period of time and in a specific interval of dates. On the other hand, the smileys tool allows students to evaluate (and revaluate) any content at any time, including the period after examinations. In this way, students have more occasions to reflect on their satisfaction about the learning content or the methodology. A last factor might be related to the fact that, according to some authors, memory is more important than actuality [36]. Thus, the results of traditional survey may be influenced by a general positive memory of the course combined with the temporal distance (which usually highlights positive memories and occludes negative ones) between this memory and the moment when the survey is answered, usually two or three weeks after the course has been completed. Contrarily, the smiley approach proposed in this work does not benefit from this memory factor, since it allows students to express their satisfaction at any time.

Studying the graph among study levels, the results obtained by the traditional surveys do not present significant differences between the average values of Resources and Methodology dimensions for the different study levels. However, according to the results obtained by means of the smileys tool it can be observed that the resources are better valued than the methodology in Own Degree and Ph.D. courses and the opposite case is found for Master and Degree courses.

D. Discussion and Limitations

After analyzing the quantitative results of our study, some reflections and limitations are now discussed. In the first two RQs, we differentiate between the assessments in the dimensions of resources and methodology carried out by the students with the smileys tool depending on their level of study. The result shows that students at the grade level perform higher assessments of resources and methodology than students at other levels of study. This is assumed to be due to the fact that the students of Master and Ph. D. are more demanding students, who have a student profile that is already less time consuming and who are looking for more elaborated materials. However, these students are provided with materials such as scientific articles to encourage research activities and these resources along with the methodology followed take longer than expected by the students. With respect to RQs 3 and 4, related to the period in which the surveys are answered, the results indicate that July is the month with the worst results. Analyzing this result and consulting the statistics of the number of students who enter the subject in July and do not do so in the rest of the year, the number of students who make the most negative assessments coincide. This implies that these students have

not followed the subject and do not have a full perspective of it, which causes the evaluations to fall both in resources and in the methodology of the subject. Finally, regarding RQ5, where the results are compared between the traditional surveys and the new proposal of the smileys tool, it is concluded that the latter are more effective and the student expresses his/her opinion at any time, obtaining a finer granularity of the methodology and resources of a subject, by carrying out unit by unit evaluations and not in a global way.

Regarding the limitations of the study, they are mainly derived from the comparison between the two survey methods studied. Despite the fact that similar criteria are measured in both methods, we found differences in the results that may be caused by several limiting factors. On the one hand, the assessment of each unit using the smileys tool does not allow to have a global vision of the subject, and there may be significant differences in each one of them, based on objective aspects such as the fact that the lecturer may not dedicate the same time to the elaboration of the materials in one unit or in another or even has more or less knowledge of certain topics of the subject. On the other hand, the general overview provided by traditional surveys may be influenced by factors such as student qualifications, quality of the correction of assignments as perceived by the student, or level of difficulty of the exam (partial or final) and depends on the time at which the survey is conducted. Another limitation in this comparison is given by the anonymous method of data collection, which while we believe to be the correct manner of collecting students' opinions, does not allow us to make a more accurate study using related samples as we cannot match the answers of the same student in both methods of assessing his/her satisfaction.

V. Conclusion and Future Work

According to the current trends on quality evaluation applied to distance education, there exists a main focus on the assessment of student satisfaction, differentiating four categories to be evaluated: (1) student-student and student-lecturer interactions, (2) learning resources, (3) methodology and (4) flexibility and ease of use of the Learning Management System software. The main methods to perform the evaluation of these categories are based on surveys containing general questions and activated during a specific and limited time. Although the data obtained using these surveys are valuable for a global overview on the quality in distance courses, there is a lack of fine-grained methods for obtaining the student satisfaction on specific learning contents and methodologies in these courses.
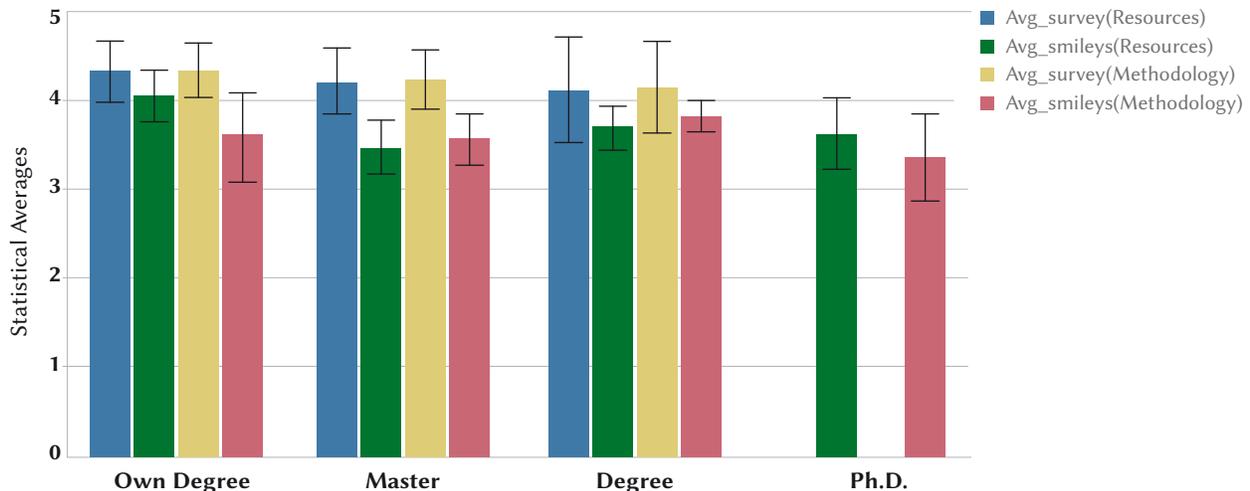

Fig. 5. Comparison of average student satisfaction results obtained through traditional surveys and smileys tools with respect to Resource and Methodology categories and grouped by study level. The error bars represent the 95% confidence interval for the average values.

This work proposes a new evaluation model through a tool based on the representation of smileys, which allows student to evaluate in a simple and intuitive manner the learning resources and the methodology applied at each unit of any subject in any course and at any time during the academic year. A case study has been carried out to evaluate this proposal and the results demonstrate that both the student's study level and the period when the satisfaction evaluation is performed are sensitive factors to take into account when interpreting the evaluation results. Hence, it has been observed that better qualifications on the learning content are obtained during non-holiday periods. Likewise, it has been detected that the resources and methodologies related to two specific study levels, namely Degree and Own-degree levels, are better valued than the rest. Finally, by comparing student satisfaction results obtained from traditional surveys with the results obtained through the tool proposed in this paper, it can be observed that the latter seems to be more accurate since the way of gathering data is more specific and flexible.

As a future work, we are investigating the use of semantic analysis to detect the most highlighted topics and sentiments in the opinions written by the students when using the smileys tool. It is also important to design a process to transfer the results obtained with this new tool to the lecturers in a simple and effective manner, with the aim of improving the lecturer's awareness about the quality of these elements. We are also planning to analyze students' changes of opinion throughout the course. To do this we will extend the smiley tool so that it can store a log of the changes in the evaluation performed by the students.

Annex I. Questionnaire for Traditional Survey

UCAM | **UNIVERSIDAD CATÓLICA SAN ANTONIO**   **STUDENT SATISFACTION SURVEY**

Please rate your satisfaction with the following aspects with a score of 1 to 5:
(from 1: Strongly disagree to 5: Strongly agree)

STUDENT SATISFACTION SURVEY - LECTURER ASSESSMENT					
PLANNING	**1**	**2**	**3**	**4**	**5**
1. The planning (date, duration, etc.) of the activities in the Teaching Guide (Syllabus) seems to me to be adequate and useful for the development of the subject					
2. The development of the course programme is in accordance with the commitments made in the Teaching Guide (Syllabus)					
3. The lecturer encourages self-learning, guides me in task planning and gives me correct guidance in the development of tasks					
METHODOLOGY	**1**	**2**	**3**	**4**	**5**
4. The lecturer motivates active participation and generates interest in the subject					
5. The lecturer organizes, structures and clearly explains the content in his/her classes					
6. The lecturer encourages the development of the capacity for reflection, analysis, synthesis and reasoning					
7. The tutoring of the subject bt the lecturer is adequate					
8. The lecturer applies the evaluation systems set out in the course's Teaching Guide					
9. The lecturer promotes teamwork to develop communication and relationship skills					
RESOURCES	**1**	**2**	**3**	**4**	**5**
10. The teaching resources (audiovisual media, virtual campus material, etc.) used by the lecturer are adequate to facilitate learning					
11. The study materials (books, articles, electronic resources, etc.) used in the course are appropiate					
GENERAL OVERVIEW	**1**	**2**	**3**	**4**	**5**
12. The lecturer is an expert in the subject					
13. Evaluate in a global view work developed by the lecturer in the subject, considering all the previous aspects					

REFERENCES

[1] A. Corbi, and D. Burgos, "Review of current student-monitoring techniques used in e-learning focused recommender systems and learning analytics. the experience api & lime model case study", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, no. 7, pp.44-52, 2014.

[2] S. K. Parahoo, M. I. Santally, Y. Rajabalee, and H. L. Harvey, "Designing a predictive model of student satisfaction in online learning". *Journal of Marketing for Higher Education*, vol. 26, no. 1, pp. 1-19, 2016.

[3] S. N. Uribe and M. Vaughan, "Facilitating student learning in distance education: a case study on the development and implementation of a multifaceted feedback system", *Distance Education*, vol. 38, no. 3, pp. 288-301, 2017.

[4] M. A. O'Neill and A. Palmer, "Importance-performance analysis: a useful tool for directing continuous quality improvement in higher education", *Quality Assurance in Education*, vol. 12, no. 1, pp. 39-52, 2004.

[5] V. Teeroovengadum, T. Kamalanabhan and A. K. Seebaluck, "Measuring service quality in higher education: Development of a hierarchical model (HESQUAL)", Quality *Assurance in Education*, vol. 24, no. 2, pp. 244-258, 2016.

[6] R. Ferguson, "Learning analytics: drivers, developments and challenges", International *Journal of Technology Enhanced Learning*, vol. 4 no. 5/6, pp. 304–317, 2012. Retrieved from http://oro.open.ac.uk/36374/

[7] V. A. Secades, and O. Arranz, "Big Data & eLearning: A Binomial to the Future of the Knowledge Society", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 6, pp. 29-33, 2016.

[8] M. Cantabella, B. López, A. Caballero and A. Muñoz, "Analysis and evaluation of lecturers activity in Learning Management Systems: Subjective and objective perceptions", *Interactive Learning Environments*, vol. 26, no. 7, pp. 911-923, 2018.

[9] M. Cantabella, R. Martínez-España, B. Ayuso, J. A. Yáñez, and A. Muñoz, "Analysis of student behavior in learning management systems through a Big Data framework", Future Generation Computer Systems, vol. 90, pp. 262 – 272, 2019

[10] A. Espasa, and J. Meneses, "Analysing Feedback Processes in an Online Teaching and Learning Environment: An Exploratory Study", *Higher Education*, vol. 59, no. 3, pp. 277–292, 2010.

[11] B. E. Shelton, J-L. Hung and P. R. Lowenthal, "Predicting student success by modeling student interaction in asynchronous online courses", Distance Education, vol. 38, no. 1, pp. 59-69, 2017.

[12] W. S. Cheung, "A new role for teachers: software evaluator. In Proceedings of the IFIP TC3/WG3." in *Proceedings of 5th International Working Conference on Exploring a New Partnership: Children, Teachers and Technology, 1994,* pp. 191–199.

[13] D. Hawkridge, "Software for schools: British reviews in the late 1980s", *Computers and learning*, pp. 88–108, 1990.

[14] T. Plomp and D. P. Ely, "International encyclopedia of educational technology", ERIC, 1996

[15] V. F. Sharp, "Computer education for teachers: Integrating technology into classroom teaching", Wiley, 2008.

[16] W. Nakamura, L. Marques, L. Rivero, E. Oliveira and T. Conte, "Are generic UX evaluation techniques enough? A study on the UX Evaluation of the Edmodo Learning Management System", *In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE),* vol. 28, no. 1, p. 1007, 2017.

[17] G. Ssekakubo, H. Suleman and G, Marsden, "Designing mobile LMS interfaces: Learners' expectations and experiences", *Interactive Technology and Smart Education*, vol. 10, no. 2, pp. 147-167, 2013.

[18] C. Fallon and S. Brown, S., "E-learning standards: a guide to purchasing, developing, and deploying standards-conformant e-learning". CRC Press, 2016.

[19] S. P. McKenna, "The Design Activity Framework: Investigating the Data Visualization Design Process" Doctoral dissertation, The University of Utah, 2017.

[20] H. Al-Samarraie, B. K. Teng, A. I. Alzahrani and N. Alalwan, "E-learning continuance satisfaction in higher education: a unified perspective from instructors and students", *Studies in Higher Education*, vol. 43, no. 11, pp. 2003–2019, 2018.

[21] T. Markova, I. Glazkova and E. Zaborova, "Quality issues of online distance learning", *Procedia-Social and Behavioral Sciences*, vol. 237, pp. 685-691, 2017.

[22] Q. T. Pham and T. P. Tran, "Impact factors on using of e-learning system and learning achievement of students at several universities in vietnam", In *O. Gervasi et al. (Eds.), International Conference on Computational Science and Its Applications – ICCSA 2018,* pp. 394–409, Springer International Publishing, 2018.

[23] I. S. Weerasinghe and R. L. Fernando, "Students' satisfaction in higher education", *American Journal of Educational Research*, vol. 5, no. 5, pp. 533-539, 2017.

[24] E. Martínez-Caro, J. G. Cegarra-Navarro and G. Cepeda-Carrión, "An application of the performance-evaluation model for e-learning quality in higher education", *Total Quality Management & Business Excellence*, vol. 26, no. 5-6, pp. 632–647, 2015.

[25] W. Sahusilawane and L. S. Hiariey, "The Role of Service Quality Toward Open University Website on The Level of Student Satisfaction", *Journal of Education and Learning*, vol. 10, no. 2, pp. 85-92, 2016.

[26] I. Almarashdeh, I., "Sharing instructors experience of learning management system: A technology perspective of user satisfaction in distance learning course", *Computers in Human Behavior*, vol. 63, pp. 249-255, 2016.

[27] N. C. Gee, "The Impact Of Lecturers' Compeencies On Students' Satisfaction", *Journal Of Arts and Social Science*, vol 1, no. 2, pp. 74-86, 2018.

[28] M. Kangas, P. Siklander, J. Randolph and H. Roukema, "Teachers' engagement and students' satisfaction with a playful learning environment", *Teaching and Teacher Education*, vol. 63, pp. 274-284, 2017.

[29] R. Yilmaz, "Exploring the role of e-learning readiness on student satisfaction and motivation in flipped classroom", *Computers in Human Behavior*, vol. 70, pp. 251–260, 2017.

[30] X. Zhai, J. Gu, H. Liu, J. C. Liang and T. Chin-Chung, "An experiential learning perspective on students' satisfaction model in a flipped classroom context", *Journal of Educational Technology & Society*, vol. 20, no. 1, 198-210, 2017.

[31] R. Prifti, "Self–efficacy and student satisfaction in the context of blended learning courses", *Open Learning: The Journal of Open, Distance and e-Learning*, pp. 1-15, 2020.

[32] M. Asoodar, S. Vaezi and B. Izanloo, "Framework to improve e-learner satisfaction and further strengthen e-learning implementation", *Computers in Human Behavior*, vol. 63, pp. 704-716, 2016.

[33] A. Uzelac, N. Gligorić and S. Krčo, S., "System for recognizing lecture quality based on analysis of physical parameters", *Telematics and Informatics*, vol. 35, no. 3, pp. 579-594, 2018.

[34] M. Cantabella, B. López-Ayuso, A. Muñoz and A. Caballero, A., "Una herramienta para el seguimiento del profesorado universitario en entornos virtuales de aprendizaje", *Revista Española De Documentación Científica*, vol. 39, no. 4, pp. 153, 2016.

[35] M. García and B. Harmsen, B., Qlikview 11 for developers. Packt Publishing Ltd, 2012.

[36] D. A. Norman, "The way I see it. Memory is more important than actuality", Interactions, vol. 16, no. 2, pp. 24-26, 2009.

Magdalena Cantabella

Magdalena Cantabella obtained her B.S. in Computer Science at the Catholic University of Murcia in 2008 and her M.S. in New Technologies in Computer Science applied to Biomedicine AT the University of Murcia in 2012, she obtained his PhD in Computer Science in 2018 at the University at the Catholic University of Murcia. Since 2010 she is an associate professor in the Polytechnic School within the Department of Degree in Computer Engineering of the Catholic University of Murcia. Her areas of research include massive statistical analysis of data, e-learning and definition of user profiles.

Raquel Martínez-España

Raquel Martínez-España is an associated professor in the Technical School at the Catholic University of Murcia (UCAM), Spain. She obtained her M.S. in Computer Science in 2009 and her PhD in Computer Science in 2014 at the University of Murcia. She has worked on several research projects in artificial intelligence and education. Raquel has participated in various academic and industry projects. Her research interests include data mining, big data, soft computing, artificial intelligence and intelligent data analysis.

Belén López

Belén López obtained her M.S. in Computer Science from the University at Murcia and her PhD in Computer Science at the same University. She has 18 years of experience in teaching, both in Degree and Master courses at University Level, include e-learning methodology. She has participated in several educational innovation projects from which publications in the area of educational innovation have been obtained. At the moment she is the Dean of the Degree in Computer Engineering of the Catholic University of Murcia and Heat of the Online Department at this University. Her areas of research include teaching assessment and e-learning methodology evaluation.

Andrés Muñoz

Andrés Muñoz is a senior lecturer in the Technical School at the Catholic University of Murcia (UCAM), Spain. He obtained his PhD in Computer Science in 2011 at the University of Murcia. He has worked on several research projects in artificial intelligence and education. His main research interests include argumentation in intelligent systems, Semantic Web technologies and Ambient Intelligence and Intelligent Environments applied to education.

The Possibilities of Classification of Emotional States Based on User Behavioral Characteristics

M. Magdin*, D. Držík, J. Reichel, Š. Koprda

Constantine the Philosopher University in Nitra, Faculty of Natural Sciences, Department of Informatics, Tr. A. Hlinku 1, 949 74 Nitra (Slovakia)

Abstract

The classification of user's emotions based on their behavioral characteristic, namely their keyboard typing and mouse usage pattern is an effective and non-invasive way of gathering user's data without imposing any limitations on their ability to perform tasks. To gather data for the classifier we used an application, the Emotnizer, which we had developed for this purpose. The output of the classification is categorized into 4 emotional categories from Russel's complex circular model - happiness, anger, sadness and the state of relaxation. The sample of the reference database consisted of 50 students. Multiple regression analyses gave us a model, that allowed us to predict the valence and arousal of the subject based on the input from the keyboard and mouse. Upon re-testing with another test group of 50 students and processing the data we found out our Emotnizer program can classify emotional states with an average success rate of 82.31%.

Keywords

I. Introduction

THE classification of user's emotional state belongs to the most debated topics in the areas of pedagogy, psychology and computer science. In computer science, the problem of classification is mostly encountered with cognitive systems or human-computer interaction (HCI). In the recent years, multiple methods have arisen to describe and recognise the emotional states of human subjects [1]. These methods are not only based on visual cues such as the facial expressions or gestures captured by a webcam [2], auditory cures – pitch, and tone of voice [3], physiological measurements (skin resistance, heartbeat rate, body temperature), but also on behavioural characteristics. Standard methods usually represent an invasive way of gathering data, during these the user is aware of the ways the data is gathered, which can distort the results. As an example of this influence we can mention face expression recognition, the extraction of a region of interest and subsequent classification of emotional state based on instructed expressions. Some authors have already pointed out this phenomenon [4], [5].

Methods of classifying the user's emotional state based on behavioural characteristics are relatively new and very topical area of scientific research even nowadays. Students of applied informatics focus their expertise mainly on the field of programming. The speciality of teaching programming (methodology) is that they must not only be able to program a specific application from the ground up but they must also be able to read the code, look for errors and correct them. Based on pedagogical research (interviews with students and questionnaires) we found out that reading and correcting the wrong code is often a stressful factor for some students that evokes different emotions (amount of anger, joy, frustration, etc.). In order to help them with the educational process and set a suitable learning style for them, we programmed and implemented the Emotnizer application into the educational process.

We have divided the paper into several sections. The Related Work section presents the theoretical bases of emotional state classification, basic models of emotions and their authors, as well as an analysis of the current state. We examined the researchers' focus on emotions based on behavioural characteristics, namely when using the keyboard and mouse of a computer. The Material and Methods section describes our methods used in Emotnizer, an application which collects raw data by capturing real-time keyboard and mouse events using the JNativeHook library. The application groups this data, and stores it for further use using a relational database system contained in a small SQLite library. In the Experiment section, we propose creating a reference database for classifying the emotional state based on user's behavioural characteristics. In the Results section, we present the results of the experiments and a description of the calculation of the overall success of the classification of the emotional state. The rationale for the results and their confrontation with the results published by other authors is given in the Discussion and Conclusion section.

II. Related Work

Emotions are complex psychological processes caused by chemical changes in the brain associated with subjective experiences (the way we experience emotions), physiological reactions (the way our bodies react to emotions) and behavioural or expressive reactions (the way we respond to emotions) [6].

* Corresponding author.

E-mail address: mmagdin@ukf.sk

The classification of emotions is how we can distinguish one emotion from another. There are two unique models for expressing emotions:

- categorical model - in which it is usually mandatory to select one emotion (or emotional category) from the set of emotions, which is best indicated by the mediated feeling. In these models, the emotions are precisely marked - categorised. The most significant categorical model is the Ekman model of classification [7], [8], which is currently referred to by many authors [2], [9], [10].

- dimensional model – definitions of human emotions are usually based on two or three emotional dimensions. Among the majority of these models there is a degree of valence involved as well (pleasurable - unpleasant) and a degree of arousal (arousing - subduing). The most influential model currently used is Russel's complex circuit model [11]. Dimensional emotional models suggest that all affective states are caused by the neurophysiological system [12].

An interesting issue that stands out in emotional classification is the usage of currently available technology [13]. Most commonly the technology used to collect emotional data is facial expression recognition using widely available web-cameras. With this approach, the recognition is realised in three basic phases:

1. face detection,
2. extraction of the region of interest,
3. classification of emotional state.

This procedure was established by Kanade in 1973 and it is still in use to this day [14]–[17].

The facial expression recognition technology is based on the knowledge that the classification of emotional state based on the visible facial expressions of emotion predicts the emotional reactions of people induced by presented imagery. Psychological studies show that experienced emotions can be influenced by various visual stimuli [18]. These influenced emotions are set on a gradient of emotional valence (low level - negative emotion, high level - positive emotion). By visual stimuli we understand the level of contrast, colour, texture, the position of an object on the scene, etc. [19], [20].

Other technologies for emotional recognition are physiological data acquisition using sensors, such as electrocardiograph (ECG), blood volume pulse (BVP), electroencephalograph (EEG) and other electrical signals, e.g. GSR - galvanic skin response, for detection and recognition of emotional states. It is known that the sources of the most reliable physiological signals for emotional recognition and its processing are EEG and GSR [18], [21], [22].

The main hurdle in current attempts of emotional state classification using subject's speech pattern is not only the lack of reference databases (such as RAVDESS, SAVEE, EMO-DB) but mainly the variability of the pitch and intensity, which can differ greatly across various human cultures [23], [24]. The standard emotion classification from speech typically employs only these two approaches: discrete categories or emotional dimensions. Morgan determines that also gender was found to impact the ratings of emotional speech and must be considered alongside stimulus factors in the design of future studies of emotion [25].

Nonverbal communication is a manifestation of human behaviour and involves various ways of transferring information, such as human activity, hand movements, facial expressions, etc. It is often used as an essential means of expressing attitudes, feelings and emotions. It plays an important role in multimodal interactions between the human and computer systems [26].

The issue of the classification of emotional state based on behavioural characteristics is related to the current state of professional literature which would be focused on this area. All research is based mainly on Epp, Lippold and Mandryk [27], who were trying to recognise emotional states based on the keyboard typing dynamics, by logging the duration of keypresses and typing latency. Based on the user's computer activity level, the program encouraged the user to self-evaluate throughout the day. The emotional state questionnaire contained several 5-point questions based on Likert scale regarding the current emotional state of the user for each of the 15 recognised emotional states. The best results of this experiment include two-stage classifiers for self-confidence, hesitancy, nervousness, relaxation, sadness and fatigue with an approximate accuracy of 77.4%. The results of the experiment also show a degree of recognition of anger and arousal with an accuracy of 84% [27].

A similar experiment was carried out in 2015 by Lee using the knowledge gained from Epp, Lippold and Mandryk [27]. The experiment was carried out on a group of 52 people, aged 20 to 26 years (44 men and 8 women). During the experiment subjects wore headphones and were ordered to write "748596132" immediately after hearing each recording from the International Affective Digitized Sounds 2nd edition (IADS-2). The experiment was conducted based on a simple dimensional view of emotions, which assumes that emotion can be defined by matching values on two different strategically chosen dimensions - valence and arousal. The Self-Assessment Manikin - SAM method was used to assess these two dimensions of emotional space. The results of this fixed target text experiment support the hypothesis that keystroke duration and latency are affected by arousal. A shorter key press time (106.70 ms ± 23.80) is shown when arousal is high compared to the keypress duration when arousal is low (108.76 ms ± 24.52). This result indicates an increase in the duration of key presses when people are tired, sad or bored [28].

In his work, Pentel stated that there are several studies on mouse movement and emotions that suggest the existence of a link between the two. He carried out 2 experiments to confirm his hypothesis. In his first experiment, he created a simple online computer game that collected data from a user's mouse. The screen was filled with random buttons labelled with numbers from 1 to 24. The user's task was to click all the buttons in the correct order as quickly as possible. As it turned out, some buttons were harder to find and caused emotions of confusion in users, thereby allowing the identification of mouse patterns associated with this mental state. It was found that users who reported a confusion rate in the Self-Assessment Scale of 5 to 7 (their confusion rate was high) scored lower in locating buttons speed than those who reported values of 1 to 3 on the Likert Scale. The machine learning algorithm model C4.5 predicted confusion using this data with an accuracy of 84.47%. The result of this experiment was that about half of the users were unable to specify exactly what they were feeling [29].

For this reason, in his second experiment, Pentel used the same data collection procedure as in the previous experiment. Besides, he has collected new data from 400 more game playthroughs. In total, the experimental group consisted of 282 individual users between the ages of 12 and 52. There were up to 21984 collected records representing mouse movement between two button clicks. In this experiment, three types of data were stored (coordinates, button sizes, and colours). Users have reported a few different kinds of emotions: confusion, frustration, shame and satisfaction. The strongest emotions were confusion and frustration, and these were associated with tasks during which the user was unable to find another button for a long time. Pentel has tested four popular machine learning algorithms: Logistic Regression, Support Vector Machine, Random Forest and C4.5. The most successful algorithm was Random Forest with a 93% success rate [29].

III. Material and Methods

The main purpose of this paper is to describe the development of the software providing automatic recognition of the user's emotional state based on the keyboard and mouse input. The Emotnizer application can recognize 4 basic types of emotions according to the dimensions of arousal and valence. The application was programmed in Java. For data collection from keyboard and mouse, it takes advantage of:

JNativeHook - Java library, which allows the program to globally read all the keyboard and mouse events.

SQLite - small, fast, standalone, highly reliable and full-featured relational database system contained in a relatively small library written in C.

JFreeChart - graphing library for Java. Supports the creation of pie charts (2D and 3D), bar charts (horizontal and vertical, regular and stacked), line charts, point charts, etc.

If we had aimed to collect and save all the available data from every mouse event, in an unsynchronized way, we would have gathered an incredible amount of data that would have no informative value. It is raw data, which needs to be grouped by action type (scrolling, mouse-click, etc.) and unit of time (data collection period).

A. Mouse Events

Mouse click frequency – using the `nativeMousePressed` method (from the JNativeHook library), which calls our thread method at the moment of pressing a mouse button and increases the number of the variable by one. As this thread sleeps for 5,000 milliseconds (5 seconds), during this time it only counts the number of mouse clicks. When the thread wakes up it calculates the difference of the number of mouse clicks and the amount of time the thread slept (5 seconds).

Mouse button hold duration – time difference between a mouse button is pressed and ends at its release. A couple of methods are being used - `nativeMousePressed` and `nativeMouseReleased`. At the moment of the mouse button press the `nativeMousePressed` method calls our thread method and passes the button code as a method parameter (left, right or scroll-wheel button) and the current time in milliseconds. At the release of the mouse button, the `nativeMouseReleased` is invoked, which calls the thread method to calculate the mouse button press duration. It also has 2 parameters: button code and the current time in milliseconds. Subsequently, the calculation is performed - the difference of 2 values (new current time value minus the old current time value). However, since the thread sleeps for 5 seconds the result will be the sum of all button hold times divided by the number of times the thread wakes up. The result therefore will be an arithmetic mean calculated according to (1). This result is stored in the database.

$$Score = \frac{\sum_{i=1}^{n}(new_i - old_i)}{n} \tag{1}$$

To capture the scroll-wheel activity we use the `nativeMouseWheelMoved` method. The principle is the same as in the previous methods - we create a thread that has a predefined sleep time of 5000 milliseconds (5s). If the method detects the scroll wheel movement it calls the thread method and passes a method parameter of positive or negative rotation (scroll up or down) to the thread. When the thread wakes up it will have several positive and negative scrolls. The following pair of values is considered as output:

- the total amount of scrolling – the sum of these numbers in absolute values divided by sleep time (in our case it is 5 seconds), e.g. positive rotation would be 15 and negative rotation -5. The result would look like this: $(abs\,(15) + abs\,(-5))\,/\,5 = 4$,

- the sum of counts – we add up the number of positive rotations

with the number of negative rotations, e.g. $15 + (-5) = 10$. The further this number is from zero, the more decisive it was for the user to choose to go up or down the screen. But the closer it is to zero, the user feels more indecisive (moved chaotically).

The speed of the cursor can be determined by the number of pixels per seconds that the cursor passes from point A to point B. We take advantage of methods:

- `nativeMouseMoved` – method, which outputs the current coordinates of the cursor when it moves,

- `nativeMouseDragged` – method, which outputs the current cursor coordinates, similar to `nativeMouseMoved`, while the cursor is dragged, not just moved. These situations occur when the user presses and holds the mouse button while moving the cursor.

The first movement of the mouse cursor sets the starting coordinates of the first point using the next thread method. The thread sleeps every 100 milliseconds (0.1 seconds), wakes up, sets the coordinates for the new point, calculates the distance according to (2) between them using the Pythagorean theorem and overwrites the coordinates of the first point with the new coordinates.

$$|AB| = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} \tag{2}$$

Distance calculations are added in the meantime to the local variable, and after the thread repeats this 50 times, the sum of all distances is divided by the data collection time (5 seconds). The calculation of the distance between of two points on the screen, $A\,[xA; yA]$, $B\,[xB; yB]$ which is represented by a 2D space, a plane, is determined as the size of the hypotenuse of the right triangle ACB (Fig. 1).
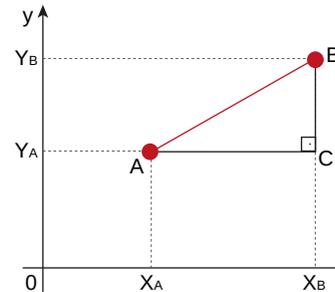
Fig. 1. The distance between of two points in a 2D plane.

B. Keyboard Events

Similarly to the raw mouse events data, raw keyboard events data have to be grouped by the type of action and unit of time.

To determine the frequency of keystrokes or the number of keystrokes per second we use the `nativeKeyPressed` method which calls a specific thread method and increases the value of a variable by one, similar to recording the mouse click frequency. The thread sleeps for 5 seconds, wakes up and divides the number of keystrokes by sleep time (5 seconds). As there is a much greater number of keys on a keyboard than there are buttons on a computer mouse, we've decided to divide the keys into three basic categories:

- modifier keys – Shift (character capitalisation), Ctrl and Alt (keyboard shortcuts), AltGr, Win (summoning Start menu and keyboard shortcuts in Windows), Fn (volume, brightness control, etc.),

- erasing keys – Delete (deletes character after the cursor) and Backspace (deletes the character before the cursor),

- all other keys – alphanumerical keys and any other key apart from modifiers and deletion keys.

We use two methods from the JNativeHook library to determine

key holding period (a time period which starts when a key is pressed and ends when it is released):

- `nativeKeyPressed` - at the moment a key is pressed it calls the thread method with the key code and current time in milliseconds as its method parameters.
- `nativeKeyReleased` – is executed when a pressed key is released and calls a method which passes the key code and current time in a millisecond as parameters to the thread method.

Since we can have multiple keys pressed at the same time, the key code parameter is very important for the calculation. When the key is released the codes are matched in the list, the timestamp of the keypress is found, subtracted from the key release timestamp and the result is converted from milliseconds to seconds. Partial results are obtained during thread sleep. When the thread is awakened the total result is obtained when the sum of all key holding period is divided by the number of all key holds. And since all keys are divided into three basic categories into modifying, erasing and all other keys, the hold time results are divided into these categories.

Pause between keystrokes is a period that starts when the last key is released and lasts until the next key is pressed. Fig. 2 shows the key holding period as well as the pause between keystrokes. To determine the pause period we used the thread we created, which calculates the frequency of keystrokes and the key holding period. After the key was released we checked whether the list of held keys is empty, and if so, we saved the value of time in milliseconds in a local variable. When a keystroke occurred we checked if it happened during the pause or whether the last key release timestamp was written in front of it. If this was true the difference between the two values in milliseconds was sent to the thread that counted the sum of all the pairs and the number of all the pairs. After the thread awoke the average value was calculated from the sum of the pauses and the number of them.
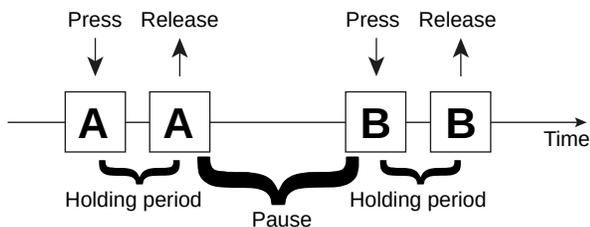

Fig. 2. A pause between keystrokes.

Summary description of all recorded events:

KPS_avg_ABC: the average of keystrokes per millisecond for the group of keys containing all the alphanumerical keys and all the other keys apart from modifiers and erasing keys,

KPS_avg_Ctrl: the average of keystrokes per millisecond for the group of modifier keys (Ctrl, Shift, Alt, etc.),

KPS_avg_DEL: the average of keystrokes per millisecond for the group of erasing keys,

DTK_avg_ABC: the average of key holding periods in milliseconds for the group of keys containing all the alphanumerical keys and all the other keys apart from modifiers and erasing keys,

DTK_avg_Ctrl: the average of key holding periods in milliseconds for the group of modifier keys (Ctrl, Shift, Alt, etc.),

DTK_avg_DEL: the average of key holding periods in milliseconds for the group of erasing keys,

pause_avg: the average of key holding periods in milliseconds for the group of modifier keys,

pause_max: maximal pause of all the keystroke pauses in milliseconds,

CPS_avg: the average of mouse clicks in milliseconds,

CPS_max: the maximal value of all the mouse clicks in milliseconds,

DTM_avg: the average of all the mouse holding periods in milliseconds,

DTM_max: maximum of all the mouse holding periods in milliseconds,

DM_avg: average count of all the pixels the mouse cursor passes on screen in 1000 milliseconds,

DM_max: maximal count of all the pixels the mouse cursor passes on screen in 1000 milliseconds

DiffChar: the difference in character count of two text strings,

SAM1: rate of valence (from -1 to 1), obtained by users in self-assessment,

SAM2: rate of arousal (from -1 to 1), obtained by users in self-assessment,

IV. Experiment – Preparation of a reference Database For the Classification of Emotional State Based on User Behavioural Characteristics

The experiment aimed to create a reference database of user behavioural characteristics for experimental classification of their emotional state.

Procedure for experimenting:

1. creation of a group to propose a reference database,
2. creation of quality measurement procedures (creation of user profile),
3. collection and processing users data,
4. comprehending data,
5. data analysis and interpretation of results,
6. description of used methods, descriptive statistics, multiple regression.

We asked 52 people (47 males and 5 females) of an average age of 21.46 years to undergo profiling by our Emotnizer application. The experiment was supervised by a psychologist. Not all participants met all the requirements, for example, they did not transcribe the entire text, and therefore we have excluded these erroneous data from the overall testing. The total number of people who met all requirements fell to 50 (45 men and 5 women) of an average age of 21.52 years. Most of the students study at the university in the field of applied informatics, therefore they can work with computers very well. Of the total number of people, 38% said they could typewrite. Also, 20% reported using a computer for less than 5 hours a week, 44% using a computer for 6 to 20 hours a week, and 36% using a computer for more than 20 hours a week. Up to 96% of people said they prefer to use the mouse wheel rather than the Home, End, etc. keys. as you scroll vertically across the screen. In this step we created the source data for the reference database (user profile). Based on this database it is possible to determine the user's behavioural characteristics, i.e. how they work with the text (read it, rewrite it) and compare it with the data obtained in the experiment.

To obtain the data, we used a trio of tests, which were prepared for us by a psychologist:

1. Questionnaire tasks to select options,
2. Rewriting the text,
3. Classification of the user's emotional state using self-assessment.

The choice questionnaires required the users to first enter their

username, age, gender, dominant hand (right or left), hours of computer use per week, etc. While users were solving this task, their work with the computer mouse was recorded.

The users' second task was to transcribe a predefined text (Fig. 3), which was supposed to evoke an emotion of anger or fear in the user. The text consisted of over 800 characters, including spaces. The users' task was to read the text and rewrite it correctly in the dialogue box. All events were measured and recorded when transcribing the text.
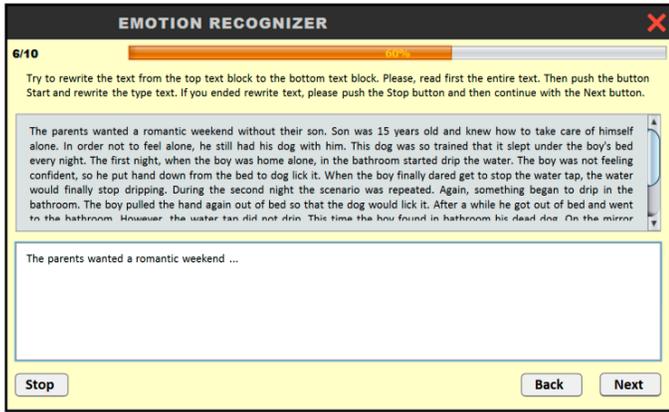

Fig. 3. Text transcription during user profile creation in Emotnizer.

Finally, we calculated the number of errors the users made when rewriting the text. The number of errors was measured using the matrix principle based on the Levenshtein distance between two strings (compared to the original text with the text rewritten by the user). The result was the number of characters that differed from the predefined text.

The users' third task was to evaluate their emotional state with the Self-Assessment Manikin, where they should determine the valence rate (SAM1) and arousal rate (SAM2) on a scale of 1 to 5 they felt during the transcription.

V. Results of Experiment

After the data collection phase we had the results from the test group. The following table (Table I) shows the descriptive characteristics recorded in the database. We have selected one of the participants for the demonstration. In the database, descriptive characteristics are stored in a single row for each user. For a clear view we chose to split the characteristics below one another.

TABLE I. Descriptive Characteristics of A single User

id	name	age	gender	hand	time
10	Vladimir	22	1	1	1

abreviations	mouse wheel	fingering	SAM1	SAM2	DiffChar
1	0	1	0	0	8

KPS_avg_ABC	KPS_avg_Ctrl	KPS_avg_DEL	KPS_avg_ABC	KPS_avg_Ctrl	KPS_avg_DEL
5662.07	372.73	355.56	91.01	166.05	84.72

Pause_avg	Pause_max	CPS_avg	CPS_max	DTM_avg	DTM_max
128.66	547.86	312.5	600	163.41	554.5

DM_avg	DM_max
89.76	301

From these descriptive characteristics we have created average descriptive statistics for 50 users and continued with multiple regressions. We could not use linear regression because of the independence of the parameters and their number. In multiple regression we focused on processing SAM1 (representing valence) and SAM2 (representing arousal) depending on all recorded events. Multiple regression calculation was performed using MS Excel (Data Analysis tool). We reduced the valence to two states: positive, negative (arousal was reduced similarly) from the original 5-point scale due to the neutral impact of the obtained data on the statistical evaluation.

According to Russel's model, the emotional state of happiness occurs when the user reaches a high value of valence and arousal. State of relaxation is defined by low arousal and high valence. The emotional state of anger is characterized by high arousal and low valence, while sadness is characterized by low arousal and high valence. In the reduction we evaluated happiness and state of relaxation as positive valence, state of anger and sadness as negative valence. Similarly, we have reduced the state of arousal.

When comparing the input parameters we got descriptive statistics for individual coefficients. The Table II shows the results of multiple regression for SAM1 and SAM2 (Table II).

TABLE II. Results of Multiple Regression for SAM1 and SAM2

	Coefficients for SAM1	Coefficients for SAM2
Intercept	-3.418	1.967
DiffChar	-0.008	-0.023
KPS_avg_ABC	0	0
KPS_avg_Ctrl	0.001	-0.001
KPS_avg_DEL	0.001	0
DTK_avg_ABC	0.01	-0.005
DTK_avg_Ctrl	-0.001	-0.001
DTK_avg_DEL	-0.001	0
pause_avg	0.005	-0.002
pause_max	0	0
CPS_avg	-0.001	0
CPS_max	0	0
DTM_avg	-0.002	-0.001
DTM_max	0	0
DM_avg	0	0.002
DM_max	0	0

These data were multiplied according to (3), where β are coefficients calculated by multiple regression and X are parameters from the keyboard and mouse. The results are predicated of SAM1 and SAM2.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_n X_{ni} \qquad (3)$$

To determine the overall success of the classification we applied the SAM1 and SAM2 predicates to the Emotnizer application we had designed. We verified the overall success of the classification under the guidance of a psychologist by providing another 50 users (average age of 21.15) a PDF file with instructions to use the application. In this file, we presented detailed instructions for creating a profile, and we also focused on recognizing emotions by this application. Users have been notified in advance to read this PDF instruction file first. After reading and understanding, they were tasked to click the Start button and start performing tasks. After pressing the Start button the program randomly chose 5 words from a set of 20 words. The users' task was to open a web browser and look up the first word generated. Users were tasked to open the first search result and copy at least 4 continuous sentences from there into any text editor. They were tasked to transcribe this copied text below itself. As a result, there should be two identical texts below each other in the text editor (the

reference database was created in the same way - users transcribed the predefined text into the dialogue box). In this case the choice of text was up to the user, the user could search for any text unlike the profile creation task. The same was true for the second word chosen by the application. For the last three words users had to think of a short (one sentence) definition that describes the word and the user had to write it in a text editor. After completing all tasks, users were tasked to click the Stop button.

Upon task completion, a program dialogue box appeared in which the users were shown a visual (using a graph) representation of their emotional recognition results by the program. SAM1 and SAM2 values on a scale of -1 to 1 were recalculated to the percentage of all emotion categories according to formula (4). As an example, we chose a formula to calculate the share for the emotion of happiness.

$$y = \left| \frac{(-1 - SAM1)}{2} \right| * \left| \frac{(-1 - SAM2)}{2} \right| \tag{4}$$

Finally, users were asked to objectively evaluate the success rate of detected emotions for all categories of emotions (happiness, anger, sadness and relaxation status) on a scale of 1 to 9 (9 is best) and submit the results (Fig. 4).
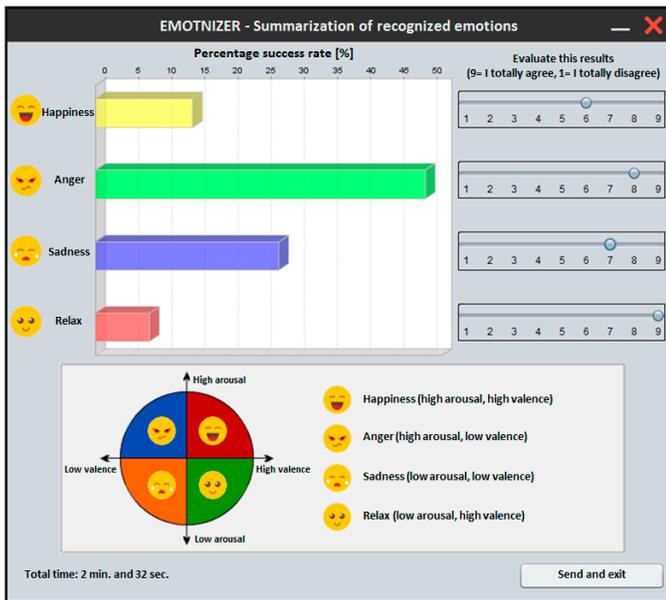

Fig. 4. Summary of recognised emotions.

Table III shows the descriptive characteristics we saw after testing. We have selected one of the participants for the demonstration.

TABLE III. Descriptive Characteristics of A single User

Predicting of happiness (%)	Predicting of anger (%)	Predicting of mourning (%)	Prediction of relaxation (%)
51.97	30.82	6.41	10.81
Evaluation of happiness	**Evaluation of anger**	**Evaluation of mourning**	**Evaluation of relaxation**
9	7	8	9

From the predicted emotions and their ratings we calculated the success rate percentage of each emotion for each user. We have recalculated partial success rates as percentage equivalents of the respondents' accuracy of prediction, represented by (5).

$$partial success = (evaluation - 1) * 100/8 \tag{5}$$

By summing all the partial success rates of the emotional state rating we gained the overall success rate for that user. By summing up the overall success rate of individual users divided by the number of users we obtained an overall average success rate of Emotnizer classification of 82.31% (Table IV).

VI. Discussion and Conclusion

Recognizing the emotional state of a user utilizing their behavioural characteristics in particular, while working with a mouse and a keyboard, is a relatively new and not sufficiently explored topic. Every person has a different typing speed and not all people are used to work with computers. Therefore, when creating Emotnizer application we had to work with the so-called user profile. Using the user profile we obtained user behavioural characteristics, and based on multiple regression and re-calculated SAM1 (valence) and SAM2 (arousal) predicates. We designed a classification algorithm for emotional states classification. Emotnizer is currently able to recognize 4 types of emotional states (from Russell's complex circular model): happiness, anger, sadness, and a state of relaxation.

A classifier is generally considered successful if it has at least 80% average success rate of the classification. According to the valence and arousal predicates, Emotnizer is currently able to classify emotional states with a success rate of 82.31%. Epp, Lee and Pentel also achieved similar values in their research [27]–[29]. In his research, Epp only recorded key holding periods and latency, not focusing on how respondents work with a computer mouse. In the resulting classification he found that the classifiers used by him could better identify anger and excitement (84%) but conversely had difficulty correctly determining self-esteem, hesitation, nervousness, relaxation, sadness or fatigue (77.4%). Based on the analysis of data from Table IV (average values for each emotional state) we see that classifiers were able to determine anger with a probability of 85.25%, state of relaxation with a probability of 85.00%, sadness of 80.25% and happiness of 78.75%. Thus, we achieved very similar classification values to Epp. Pentel also addressed the state of confusion in his classification. His C4.5 machine learning algorithm model for data processing predicted confusion with an accuracy of 84.47%, in his second experiment, using logistic regression, Support Vector Machine, Random Forest and C4.5 he reached the success rate of 93%. However, according to Rosenberg and Ekman [30], confusion, frustration or shame cannot be regarded as emotion but rather a manifestation (effect) of a particular emotional state.

We have now solved 4 basic emotions. These emotions are found in both Ekman's classification and Russell's circumplex model of affect. The reason why we have developed a classification algorithm for 4 basic emotions is that the larger the number of emotions, the lower the overall success of the classification algorithm. Epp, Lee and Pentel also had this problem. We want to expand our solution in the future. We assume that it is possible to increase the success of the classification and at the same time recognize (and classify) a greater number of emotional states. For example, with the help of a sufficiently long text it would be possible to observe the course of individual emotions in a certain time interval. Emotions would again be classified on the basis of predicates. The role of the psychologist would be to tag (evaluate) individual words and sentences from the text from the point of view of sentiment, and thus we would be able to clearly identify the degree of success in classifying individual emotional states.

Currently, our next goal is to integrate Emotnizer into the learning process. The whole application was developed in order to determine why students of applied informatics are unsuccessful in solving certain programming tasks. We want to continue to improve the application and use it to determine how students behave when programming:

TABLE IV. Results of the Emotional State Classification

id	Time [s]	Prediction of Emotnizer [%]				User's evaluation of Emotnizer [%]				
		Happiness	Anger	Sadness	Relax	Happiness	Anger	Sadness	Relax	Average
1	577	51.97	30.82	6.41	10.81	100	75	88	100	91
2	602	40.32	26.18	13.19	20.31	63	100	88	63	78
3	507	31.68	44.31	14.00	10.01	88	75	100	100	91
4	665	0.00	63.11	36.89	0.00	50	75	100	100	81
5	737	17.80	54.12	21.13	6.95	63	100	75	88	81
6	620	32.91	47.01	11.81	8.27	100	63	75	75	78
7	435	35.14	15.81	15.22	33.84	75	88	75	88	81
8	514	16.12	57.98	20.27	5.64	88	88	75	100	88
9	444	37.10	36.69	13.03	13.18	63	88	75	63	72
10	498	23.64	54.11	15.49	6.76	88	100	75	75	84
11	255	46.43	12.98	8.87	31.72	88	100	88	75	88
12	343	21.69	59.67	13.67	4.97	88	63	100	100	88
13	189	27.88	58.66	9.13	4.34	88	75	100	88	88
14	286	40.06	23.49	13.47	22.97	100	63	88	100	88
15	359	14.60	57.95	21.92	5.52	50	88	13	100	63
16	557	44.77	24.47	10.87	19.89	38	100	100	13	63
17	437	42.86	18.21	11.61	27.32	100	100	100	100	100
18	563	17.96	41.34	28.38	12.33	100	100	100	100	100
19	1135	22.61	0.00	0.00	77.39	50	100	50	88	72
20	571	28.63	37.46	19.22	14.69	100	63	50	100	78
21	1226	24.54	0.00	0.00	75.46	63	100	100	38	75
22	591	46.74	0.00	0.00	53.26	100	100	100	100	100
23	345	15.80	46.89	27.90	9.41	88	100	100	100	97
24	375	17.06	3.36	13.10	66.48	50	100	38	100	72
25	138	55.63	17.76	6.44	20.17	0	100	100	50	63
26	352	39.12	41.30	10.06	9.53	88	100	13	50	63
27	445	45.75	22.90	10.46	20.89	100	50	100	75	81
28	626	28.24	38.48	19.19	14.09	88	63	13	100	66
29	299	26.89	43.06	18.49	11.55	63	63	75	88	72
30	475	24.43	24.69	25.58	25.31	100	100	100	100	100
31	514	22.07	36.99	25.64	15.30	75	50	88	100	78
32	338	51.65	26.15	7.46	14.74	88	50	88	75	75
33	368	36.49	41.84	11.58	10.10	50	50	88	88	69
34	386	30.70	24.11	19.88	25.31	100	100	50	75	81
35	461	5.77	14.08	56.84	23.31	100	100	100	100	100
36	382	35.16	34.74	14.96	15.14	100	100	100	100	100
37	61	35.41	0.00	0.00	64.59	88	100	100	88	94
38	95	42.22	0.00	0.00	57.78	63	100	100	100	91
39	86	14.79	0.00	0.00	85.21	88	100	100	100	97
40	97	20.38	0.00	0.00	79.62	63	100	100	88	88
41	83	34.19	0.00	0.00	65.81	75	100	100	100	94
42	152	50.10	0.00	0.00	49.90	63	100	38	75	69
43	91	52.31	0.00	0.00	47.69	50	100	100	75	81
44	68	28.08	0.00	0.00	71.92	75	63	100	75	78
45	73	36.86	0.00	0.00	63.14	100	100	100	100	100
46	513	33.57	43.52	12.93	9.98	75	100	50	50	69
47	433	18.76	59.16	16.77	5.32	100	88	38	75	75
48	579	26.07	59.92	9.76	4.25	100	100	100	100	100
49	233	24.09	46.14	19.55	10.21	88	25	63	75	63
50	369	32.79	35.25	16.56	15.41	88	63	63	100	78

1. They must read the task assignment first (induction of emotion).
2. They propose a solution, they start programming (change of emotional state).
3. The application will detect programming errors (number of corrections in the code) and monitor the emotional state of students (whether, for example, frustration does not affect their programming skills and the end result).
4. Based on the results of behavioral characteristics together with a psychologist, we will propose a suitable solution to a possible problem.

References

[1] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human Emotion Recognition: Review of Sensors and Methods," *Sensors (Switzerland)*, vol. 20, no. 3, p. 592, Jan. 2020, doi: 10.3390/s20030592.

[2] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson, "Facial expression analysis with AFFDEX and FACET: A validation study," *Behavior Research Methods*, vol. 50, no. 4, pp. 1446–1460, Aug. 2018, doi: 10.3758/s13428-017-0996-1.

[3] M. Magdin, T. Sulka, J. Tomanová, and M. Vozár, "Voice Analysis Using PRAAT Software and Classification of User Emotional State," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 6, p. 33, 2019, doi: 10.9781/ijimai.2019.03.004.

[4] L. Abramson, I. Marom, R. Petranker, and H. Aviezer, "Is fear in your head? A comparison of instructed and real-life expressions of emotion in the face and body," *Emotion*, vol. 17, no. 3, pp. 557–565, Apr. 2017, doi: 10.1037/emo0000252.

[5] M. Magdin, L. Benko, and Š. Koprda, "A case study of facial emotion classification using affdex," *Sensors (Switzerland)*, vol. 19, no. 9, May 2019, doi: 10.3390/s19092140.

[6] D. H. Hockenbury and S. E. Hockenbury, *Discovering psychology*. Worth Publishers, 2007.

[7] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, Feb. 1971, doi: 10.1037/h0030377.

[8] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environ. Psychol. Nonverbal Behav.*, vol. 1, no. 1, pp. 56–75, Sep. 1976, doi: 10.1007/BF01115465.

[9] K. Bahreini, R. Nadolski, and W. Westera, "FILTWAM - A framework for online affective computing in serious games," in *Procedia Computer Science*, 2012, vol. 15, pp. 45–52, doi: 10.1016/j.procs.2012.10.057.

[10] G. M. PS Sreeja, "Emotion models: a review," *International Journal of Control Theory and Application*, vol. 10, no. 8, pp. 651–657, 2017.

[11] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.

[12] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 3, pp. 715–734, Jul. 2005, doi: 10.1017/S0954579405050340.

[13] M. Khari, A. K. Garg, R. Gonzalez-Crespo, and E. Verdú, "Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 7, p. 22, 2019, doi: 10.9781/ijimai.2019.09.002.

[14] T. Kanade, "Picture processing system by computer complex and recognition of human faces," 1974.

[15] D. Ghimire and J. Lee, "A robust face detection method based on skin color and edges," *Journal of Information Processing Systems*, vol. 9, no. 1,

pp. 141–156, 2013, doi: 10.3745/JIPS.2013.9.1.141.

[16] S. H. Yoo, S. K. Oh, and W. Pedrycz, "Optimized face recognition algorithm using radial basis function neural networks and its practical applications," *Neural Networks*, vol. 69, pp. 111–125, Sep. 2015, doi: 10.1016/j.neunet.2015.05.001.

[17] L. Surace, M. Patacchiola, E. B. Sönmez, W. Spataro, and A. Cangelosi, "Emotion recognition in the wild using deep neural networks and Bayesian classifiers," in *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, vol. 2017-January, pp. 593–597, doi: 10.1145/3136755.3143015.

[18] K. Eroğlu, T. Kayıkçıoğlu, and O. Osman, "Effect of brightness of visual stimuli on EEG signals," *Behavioural Brain Research*, vol. 382, p. 112486, Mar. 2020, doi: 10.1016/j.bbr.2020.112486.

[19] A. Meikleham and R. Hugo, "Understanding informal feedback to improve online course design," *European Journal of Engineering Education*, vol. 45, no. 1, pp. 4–21, Jan. 2020, doi: 10.1080/03043797.2018.1563051.

[20] L. Li, X. Zhu, Y. Hao, S. Wang, X. Gao, and Q. Huang, "A hierarchical CNN-RNN approach for visual emotion classification," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 15, no. 3s, pp. 1–17, Dec. 2019, doi: 10.1145/3359753.

[21] N. Y. Oktavia, A. D. Wibawa, E. S. Pane, and M. H. Purnomo, "Human Emotion Classification Based on EEG Signals Using Naïve Bayes Method," in *Proceedings - 2019 International Seminar on Application for Technology of Information and Communication: Industry 4.0: Retrospect, Prospect, and Challenges, iSemantic 2019*, 2019, pp. 319–324, doi: 10.1109/ISEMANTIC.2019.8884224.

[22] D. Ayata, Y. Yaslan, M. K.-I. U.-J. Of, and U. 2017, "Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods," *Istanbul University – Journal of Electrical and Electronics Engineering*, vol. 17, no. 1, p. 3129, 2017.

[23] M. Dedeoglu, J. Zhang, and R. Liang, "Emotion classification based on audiovisual information fusion using deep learning," in *IEEE International Conference on Data Mining Workshops, ICDMW*, 2019, vol. 2019-November, pp. 131–134, doi: 10.1109/ICDMW.2019.00029.

[24] A. A. Akinrinmade, E. Adetiba, J. A. Badejo, and A. A. Atayero, "Creation of a Nigerian Voice Corpus for Indigenous Speaker Recognition," *Journal of Physics: Conference Series*, vol. 1378, no. 3, 2019, doi: 10.1088/1742-6596/1378/3/032011.

[25] S. D. Morgan, "Categorical and dimensional ratings of emotional speech: Behavioral findings from the Morgan emotional speech set," *Journal of Speech, Language and Hearing Research*, vol. 62, no. 11, pp. 4015–4029, Nov. 2019, doi: 10.1044/2019_JSLHR-S-19-0144.

[26] S. Gong, H. Mao, Y. Wang, and A. Xu, "Machine learning in human-computer nonverbal communication," in *NeuroManagement and Intelligent Computing Method on Multimodal Interaction, AICMI 2019*, 2019, pp. 1–7, doi: 10.1145/3357160.3357670.

[27] C. Epp, M. Lippold, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in *Conference on Human Factors in Computing Systems - Proceedings*, 2011, pp. 715–724, doi: 10.1145/1978942.1979046.

[28] P.-M. Lee, W.-H. Tsui, and T.-C. Hsiao, "The Influence of Emotion on Keyboard Typing: An Experimental Study Using Auditory Stimuli," *PLoS One*, vol. 10, no. 6, p. e0129056, Jun. 2015, doi: 10.1371/journal.pone.0129056.

[29] A. Pentel, "Employing think-aloud protocol to connect user emotions and mouse movements," in *IISA 2015 - 6th International Conference on Information, Intelligence, Systems and Applications*, 2016, doi: 10.1109/IISA.2015.7387970.

[30] E. L. Rosenberg and P. Ekman, "Emotion: Methods of study.," in *Encyclopedia of Psychology, Vol. 3.*, American Psychological Association, 2000, pp. 171–175.

D. Držík

He is student at the Department of Informatics. He deals with programming in Java, Python and C. He specialization is using and implementation various algorithm for face detection, acquisition of user behavioral characteristics and classification emotion.

J. Reichel

He was born in Nova Bana, Slovakia, in 1988. He received the M.S. degree in computer science teaching from Faculty of Natural Sciences, Constantine the Philosopher University in Nitra, Slovak Republic, in 2012. In 2016, he received the Ph.D. degree in theory of teaching mathematics at Constantine the Philosopher University in Nitra. From 2016 to present, he is an Assistant Professor with the Department of Informatics, Constantine the Philosopher University, Nitra, Slovakia. His research interests are in the field of Educational Data Mining.

S. Koprda

He has been an employee of the Department of Informatics since 2005. Since 2015 he has been working as an Associate Professor at the Department of Informatics. He has many years of experience in the field of electrical engineering with a focus on digital and control technology. He is an expert in working with microcontrollers, specifically Arduino. In field of microcontrollers he specializes to programming microcontrollers for face recognition, fingerprint reader using the Atmel Studio programming environment and Arduino IDE.

M. Magdin

He works as a professor assistant at the Department of Informatics. He deals with the theory of teaching informatics subjects, mainly implementation interactivity elements in e-learning courses, face detection and emotion recognition using a webcam, speech recognition. He participates in the projects aimed at the usage of new competencies in teaching and also in the projects dealing with learning in virtual environment using e-learning courses.

Function Analysis of Industrial Robot under Cubic Polynomial Interpolation in Animation Simulation Environment

You Li[1], Juan Wang[1], Yiming Ji[2]*

[1] School of Arts Design, Wuchang University of Technology, Wuhan (China)
[2] School of Journalism, Zhongnan University of Economics and Law, Wuhan (China)

Abstract

In order to study the effect of cubic polynomial interpolation in the trajectory planning of polishing robot manipulator, firstly, the articular robot operating arm is taken as the research object, and the overall system of polishing robot operating arm with 7 degrees of freedom is constructed. Then through the transformation of space motion and pose coordinate system, Denavit-Hartenberg (D-H) Matrix is introduced to describe the coordinate direction and parameters of the adjacent connecting rod of the polishing robot, and the kinematic model of the robot is built, and the coordinate direction and parameters of its adjacent link are described. A multi-body Dynamic simulation software, Automatic Dynamic Analysis of Mechanical Systems (ADAMS), is used to analyze the kinematic simulation of the robot operating arm system. Finally, the trajectory of the robot manipulator is planned based on the cubic polynomial difference method, and the simulation is verified by Matrix Laboratory (MATLAB). Through calculation, it is found that the kinematic model of polishing robot operating arm constructed in this study is in line with the reality; ADAMS software is used to generate curves of the rotation angles of different joint axes and the displacement of end parts of the polishing robot operating arm changing with time. After obtaining relevant parameters, they are put into the kinematic equation constructed in this study, and the calculated position coordinates are consistent with the detection results; moreover, the polishing robot constructed in this study can realize the functions of deburring, polishing, trimming, and turning table. MATLAB software is used to generate the simulation of the movement trajectory of the polishing robot operating arm, which can show the change curve of angle and angular velocity. The difference between the angle at which the polishing robot reaches the polishing position, the change curve of angular velocity, and the time spent before and after the path optimization is compared. It is found that after path optimization based on cubic polynomial, the change curve of the polishing robot's angle and angular velocity is smoother, and the time is shortened by 17.21s. It indicates that the cubic polynomial interpolation method can realize the trajectory planning of the polishing robot operating arm, moreover, the optimized polishing robot has a continuous and smooth trajectory, which can improve the working efficiency of the robot.

Keywords

I. Introduction

AT present, China's manufacturing industry is developing rapidly, and customers have higher and higher requirements on the surface quality of castings. Labor costs and environmental awareness are also increasing, which greatly limits the development of China's foundry industry. In the process of different casting production, sand treatment, molding, core making, and other processes have gradually achieved the goal of automatic production and intelligent management, but the post-processing technology of casting is still in the stage of manual operation [1], [2]. Facing the high demand of enterprises on the surface quality of castings, the research on casting polishing technology and related equipment has become a big trend. The appearance of flash and burr on the surface of the casting can not only affect the appearance of the casting, but also affect the performance of the casting and even the service life of the casting [3], [4]. Therefore, polishing the surface of the casting is a very important step in the manufacturing industry.

With the continuous progress of industrial technology, enterprises have been gradually moving towards the direction of intelligence and automation. The castings polished by the robot have the advantages of better quality, higher efficiency, strong flexibility, and strong adaptability [5]. Therefore, robot polishing automatic production line and casting polishing have gradually become the best choice for enterprises in the manufacturing field. Moreover, the polishing of sports robot can reduce the risk of injury to human body, ensure the uniformity of polishing quality, improve the efficiency of operation,

* Corresponding author.
E-mail address: jym20200123@163.com

and reduce the cost [6], [7]. In recent years, experts at home and abroad have explored robot polishing equipment and its automatic polishing technology, but relatively few studies have been conducted on the polishing of large castings [8]. Polishing large castings requires a high load on the robot, while for castings with complex surface shapes, manual polishing is still required [9]. The research prospect of polishing robot is broad, and the intelligent processing system of robot is of great significance to improve the working efficiency of polishing. Therefore, it is very important to shorten the running time of the robot to improve the efficiency. Some studies have shown that polynomial interpolation can carry out the optimal time trajectory planning [10], [11]. Among them, cubic polynomial interpolation is widely used. Zubko et al. calculated the accelerated light scattering based on cubic polynomial functions and found that the interpolation method could accurately calculate the light scattering characteristics [12]. Zhang et al. showed that cubic polynomial interpolation could present a better linear relation than quadratic polynomials [13]. Moreover, a large number of studies have proved that the application of cubic polynomial interpolation in motion trajectory planning under different conditions and constraints can effectively improve the running speed [14]. Dincer et al. combined cubic polynomial interpolation with Bézier curves for the trajectory planning of 2-dof parallel structure, and found that using cubic polynomial interpolation could obtain smoother trajectory curves [15].

However, at present, the research on the path planning of seven-degree-of-freedom polishing robot by cubic polynomial interpolation is still relatively few. Therefore, firstly, the mechanism types of common industrial robot operating arms are discussed, and the polishing robot operating arm system is designed with articular robot operating arms as the research object. Through the transformation of the spatial coordinate system, the kinematic model of the polishing robot manipulator is constructed. Then, virtual simulation software is used to verify the kinematic model of the polishing robot operating arm. Finally, based on cubic polynomial interpolation, the trajectory of the polishing robot manipulator is planned. The results of this study are intended to lay a foundation for optimizing the movement trajectory of the robot's operating arm and improving its polishing efficiency.

II. Methodology

A. Mechanism Type of industrial Robot Manipulator

Common industrial robot operating arm structures include rectangular coordinate type, cylindrical coordinate type, polar coordinate type, and joint coordinate type, etc. [16]. Among them, the rectangular robot operating nose has the simplest structure, with only three axes in the cartesian coordinate system: the X axis, the Y axis (the horizontal motion axis), and the Z axis (the vertical motion axis). Articular robots operate arms that are similar to human arms, with structures that mimic the human upper and lower arms, shoulder, elbow, and wrist joints. These manipulators have 5 to 6 degrees of freedom and can touch all the points in the workspace, and their movement trajectory while working can form most of the spherical motion. The operational performance of common structural industrial robot operating arms is compared, and the results are shown in Table I. Among industrial robot operating arms with different structure types, articular robot operating arms are more compact in structure and manoeuvrable in working space. Therefore, it requires less joint torque, and its own energy consumption is low, and the joint sealing and dustproof effect is excellent. Therefore, joint coordinate robot arm is finally selected as the object of this study.

TABLE I. Performance Comparison of Operating Arms of Industrial Robots With Different Structures

Structures	Operational complexity	Workspace size	Flexibility	Their size
Rectangular coordinate	Easy	Small	Bad	Big
Cylindrical coordinate	Relatively easy	Relatively big	Normal	Relatively big
Polar coordinates	complex	Relatively small	Good	Relatively small
Joint coordinate	Relatively complex	Big	Very good	Very small

B. The Overall Design of Polishing Robot Operating Arm System

There are two main methods for robot polishing and casting, which are respectively used for medium and large castings, medium and small castings. The former is the polishing tool held by the robot arm for polishing processing, while the latter is the casting held by the robot arm close to the polishing tool for polishing processing [17]. The automobile cylinder castings are taken as the polishing object, and the corresponding polishing robot structure is designed to improve the polishing efficiency of automobile manufacturing industry and reduce the cost in the production process [18]. Therefore, the robot structure with strong flexibility and simple structure should be selected. In the end, the way of polishing the casting with the polishing tool held by the robot arm to polish the automobile cylinder block is chosen. The height of each stand in the car cylinder block convex table is different, and the position distribution is not uniform. Therefore, the polishing robot arm designed in this study must have such characteristics as high motion accuracy, low return clearance, low rotational inertia, and non-variability. Finally, the specific performance indicators are determined as follows: The grinding speed is 6000rpm/min. The grinding force is 110N. The quality roughness Ra of the cylinder surface after polishing should be equal to and less than 3.0. The arms of the polishing robot extend between 600 and 1000mm.

The operating arm of the polishing robot is mainly composed of the base, shoulder joint, big arm, elbow joint, forearm, and wrist joint. Moreover, the articular structure connected in series between the operating arms can best imitate the motion mode of human arms. The manipulator has 7 rotation axes, and the corresponding manipulator has 7 degrees of freedom. In the process of different coaxial rotation motions, it will receive loads in both radial and axial directions. Therefore, the connection mode of bearing between connecting rod and shaft adopts double-row angular contact ball. The specific structure of each joint of the manipulator is as follows: the shoulder joint is located at the top of the whole connecting rod drive chain and has 3 degrees of freedom (namely 3 rotating axes). The angles of different rotating axes are respectively 180°(-90°~+90°), 180°(-90°~+90°) and 210°(-60°~+150°); the elbow joint is located between the upper arm and the lower arm of the robot arm, and has 2 degrees of freedom (i.e. 2 joint axes). The angles of different joint axes are 240°(-180°~+60°) and 180°(-90°~+90°) respectively; The wrist joint has 2 degrees of freedom (i.e. 2 joint axes), and the angles of different joint axes are all 180°(-90°~+90°). The structure of the polishing robot arm designed in this study is briefly shown in Fig. 1.

In order to better meet the operating requirements of the polishing robot operating arm [19], AC servo motor is selected for this study. In addition, other joint modules include harmonic gear reducer, encoder, brake, transmission gear, and support frame. According to the actual volume of the cylinder block to be polished, the outer cylinder working space size of 350*350*350 mm^2 is finally selected in this study, and the cylinder with the length and height of the bottom surface of 500 mm and 350 mm are selected as the target working space.
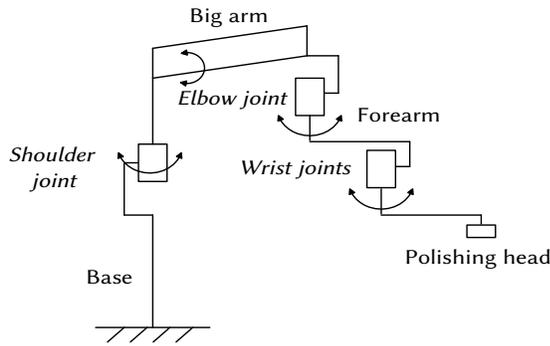
Fig. 1. Schematic diagram o polishing robot arm.

C. The Transformation of the Spatial Coordinate System of the Operating Arm of the Polishing Robot

Robot time - space motion is mainly divided into: firstly, forward kinematics, which refers to the relationship and parameters among the structure types and lengths of the robot's connecting rod, base and end-effector, etc, and finally, the pose problem of end-effector relative to the base coordinates is solved; secondly, inverse kinematics, it is necessary to use algebraic operation and other methods to obtain the relationship between each linkage structure, structural parameters, end-effector pose, and other information, and finally solve the variable values of each joint. In this study, the homogeneous transformation matrix method [20] is used to describe the spatial motion of the robot and the transformation relation of the linkage coordinate system, and the matrix calculation is used to solve the kinematic analysis problem of the robot manipulator.

In the spatial coordinate system, any point on the target object can be used to describe its pose. To determine the pose state of the target object on the relatively fixed coordinate system by using the 3*1 position matrix and 3*3 attitude matrix, the moving point coordinate system is {H}, and the fixed coordinate system is {O}. Then the 3*1 position matrix can be expressed as follows.

$$P_{0H} = \begin{Bmatrix} P_x \\ P_y \\ P_z \end{Bmatrix} = \begin{Bmatrix} x_{0H} \\ y_{0H} \\ z_{0H} \end{Bmatrix} \tag{1}$$

The cosine value of the included angle between {H} and {O}x, y, and z axes is used to describe the third-order attitude matrix.

$$R_{0H} = \begin{pmatrix} \cos(x_0, x_H) & \cos(x_0, y_H) & \cos(x_0, z_H) \\ \cos(y_0, x_H) & \cos(y_0, y_H) & \cos(y_0, z_H) \\ \cos(z_0, x_H) & \cos(z_0, y_H) & \cos(z_0, z_H) \end{pmatrix} \tag{2}$$

And the initial coordinate system of any point P in space is {i}, and the changed coordinate system is {j}. In the general transformation of coordinate equation, P_{ij} is the translation change matrix and R_{ij} is the rotation transformation matrix. Then homogeneous coordinate transformation is introduced to describe the transformation process between coordinate systems, and finally the matrix M_{ij} is obtained.

$$M_{ij} = \begin{pmatrix} 1 & 0 & 0 & P_x \\ 0 & 1 & 0 & P_y \\ 0 & 0 & 1 & P_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} n_x & o_z & a_z & 0 \\ n_y & o_y & a_y & 0 \\ n_z & o_z & a_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

$$M_{ij} = Trans(P_x, P_y, P_z)Rot(k, \theta) \tag{4}$$

Trans (P_x, P_y, P_z) is the homogeneous coordinate transformation matrix after translation; Rot (k, θ) is the homogeneous coordinate transformation matrix after rotation; k is the rotation axis vector; θ is the rotation angle; n, o, and a are the cosines of the 3 directions from

the coordinate system {j} to the coordinate system {i}.

D. Construction of the Kinematic Model of Polishing Robot Operating Arm

To construct the kinematic equation of the robot manipulator, the corresponding coordinate system should be set on each revolute joint. The corresponding homogeneous coordinate system equation is obtained from the relative position of each coordinate system, and then the kinematics equation of the robot manipulator is obtained [21]. In this study, D-H method is introduced to describe the coordinate direction and parameters of the adjacent connecting rod of the robot manipulator. The schematic diagram of connecting rod parameters of robot operating arm is shown in Fig. 2. The d_j, θ_j, a_j, and α_j in Fig. 2 are four parameters of D-H.
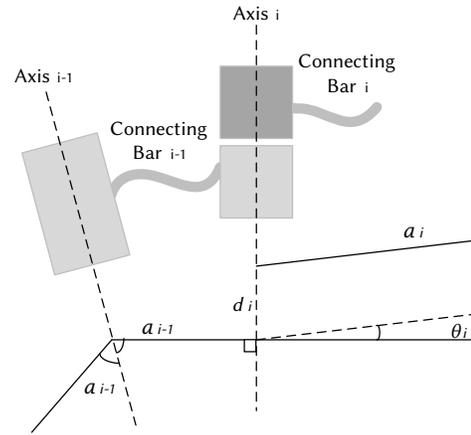
Fig. 2. Schematic diagram of parameters of adjacent connecting rod of robot operating arm.

Then in the d-h coordinate system, the pose matrix between adjacent links is as follows.

$$M_{ij} = Trans(0,0,d_j)Rot(z_{j-1}, \theta_j)Trans(a_j, 0,0)Rot(x_j, \alpha_j) \tag{5}$$

In Eq. (5), d_j is linkage migration; z_{j-1} is the axis; θ_j is the joint angle; a_j is the length of connecting rod; α_j is the torsion Angle of the connecting rod.

According to the homogeneous coordinate transformation method, the pose of the end-effector of the manipulator arm is solved.

$$M_{0n} = M_{0,1}M_{1,2}M_{2,3} \cdots M_{j-1,j} \cdots M_{n-1,n} \tag{6}$$

D-H parameters are obtained according to the linkage coordinate system, as shown in Table II.

TABLE II. Robot Manipulator Connecting Rod D-H Parameter Table

Connecting rod number j	$\theta_j(°)$	$\alpha_j(°)$	a_j(mm)	d_j(mm)
1	θ_1	0	0	0
2	θ_2	90	0	0
3	θ_3	-90	0	0
4	θ_4	0	a_3	d_4
5	θ_5	-90	0	d_5
6	θ_6	90	0	0
7	θ_7	90	0	0

a_3=450mm, d_4=400mm, d_5=-350mm.

Then, the 7 degrees of freedom is used to polish the structure characteristics of robot operating arm. Using joint variable collapse method, the 7-degree-of-freedom manipulator is transformed into a

typical 6-degree-of-freedom manipulator. The inverse kinematics of manipulator is solved by algebraic method. The kinematic equation of the robot manipulator is as follows.

$$
{}_{\tau}^{0}T = \begin{bmatrix} r_{11} & r_{12} & r_{13} & P_x \\ r_{21} & r_{22} & r_{23} & P_y \\ r_{31} & r_{32} & r_{33} & P_z \\ 0 & 0 & 0 & 1 \end{bmatrix} =
$$
$$
{}_{1}^{0}T(\theta_1){}_{2}^{1}T(\theta_2){}_{3}^{2}T(\theta_3){}_{4}^{3}T(\theta_4){}_{6}^{4}T(\theta_6){}_{7}^{6}T(\theta_7) \tag{7}
$$

After a series of solutions, it is concluded that when θ_5 is 0, the robot manipulator is in a singular configuration, that is, the joint axis 5 and 7 are in a straight line. Therefore, when calculating θ_7, the value of θ_5 can be referred to.

E. Virtual Simulation Analysis of Polishing Robot Operating Arm

In this study, ADAMS is used to analyze the kinematic simulation of the polishing robot operating arm system [22]. ADAMS is a popular virtual simulation software, and its simulation analysis steps are similar to those of CAE software [23]. The specific simulation process is shown in Fig. 3, which is mainly divided into 6 steps. Step 1: preparation before modeling. The main work is the choice of path, coordinate system, unit system, and the setting of gravity acceleration. Step 2: construction of the model. The main purpose of this step is to simplify the model. Under the premise of not affecting the visual effect, only the motion-related components should be retained, which can effectively reduce the difficulty of the model and improve the accuracy of model simulation. Step 3: constrains definition. Step 4: drive application. The Time function, STEP function, and AKISPL function can be used to define the driver. Step 5: verification of the model. Before the simulation, the model needs to be tested, and the model is verified successfully if there is no under constraint or over constraint. Step 6: simulation running. The Kinematic type of interactive simulation is chosen. The simulation time is set to 25s and the simulation times to 350.
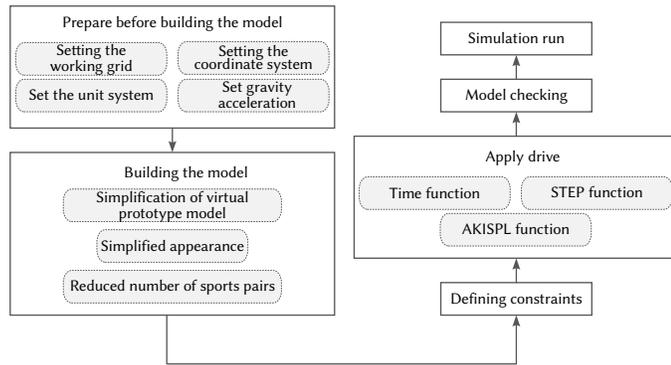

Fig. 3. Simulation process of polishing robot operating arm based on ADAMS software.

When the driver is applied, the AKISPL function needs to interpolate the data using the Akima fitting method. When the manipulator is conducting forward kinematic analysis, it is necessary to define the pose of each joint and then obtain the pose of the terminal part. Therefore, in the forward kinematics simulation of the robot operating arm, it is necessary to increase the driving force for all the 7 joint axes, so that the connecting rod outside the terminal can carry out known regular motion. Taking joint shaft 1 for example, the rotary drive diagram is chosen in the software Motions, and the rotation pair at the left and right turning joints of the shoulder is selected to create the rotation drive. Functions are input into Function(time): STEP(time,0,0d,3,90d)+ STEP(time,3,0d,6,-90d)+ STEP(time,9,0d,12,-90d)+ STEP(time,12,0d,15,90d), and the type is set to Displacement. The same method is used to apply the drive to the remaining joint shafts.

F. Trajectory Planning of Polishing Robot Manipulator Based on Cubic Polynomial Interpolation

Trajectory planning is mainly to translate the command input by the operator of the robot arm into a process of describing the specific movement trajectory. Its purpose is mainly to determine the movement trajectory, movement time, speed, and other parameters of the target joint [24]. Trajectory planning of joint space mainly uses different functions to describe trajectories, and the commonly used methods are polynomial interpolation and linear interpolation based on parabolic transition [25]. In order to meet the accuracy of robot arm motion trajectory, the cubic polynomial interpolation is adopted to plan robot arm motion trajectory. Assuming that the smooth trajectory function of the joint angles at the starting and ending positions in joint space is q(t), then the trajectory function q(t) of different joints should meet the constraints of position and velocity. In order to ensure the continuity of joint motion speed, the starting and ending joint speeds are set to 0. Then the trajectory function before and after can be expressed as follows.

$$
\begin{cases} q(0) = q_0 \\ q(t_f) = q_f \\ \overline{q}(0) = 0 \\ \overline{q}(t_f) = 0 \end{cases} \tag{8}
$$

In Eq. (8), q_0 is the initial joint angle; q_f is the joint angle of termination. t_f is the termination time.

The cubic polynomial of the constraint conditions of the boundary is determined through Eq. (8).

$$
q(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \tag{9}
$$

According to Eq. (9), the equations of joint velocity and acceleration are obtained.

$$
\begin{cases} \overline{q}(t) = a_1 + 2a_2 t + 3a_3 t^2 \\ \overline{\overline{q}}(t) = 2a_2 + 6a_3 t \end{cases} \tag{10}
$$

According to Eq. (9) and Eq. (10), different linear equations containing coefficients a_0, a_1, a_2, and a_3 are obtained.

$$
\begin{cases} q_0 = a_0 \\ q_f = a_0 + a_1 t_f + a_2 t_f^2 + a_3 t_f^3 \\ 0 = a_1 \\ 0 = a_1 + 2a_2 t_f + 3a_3 t_f^2 \end{cases} \tag{11}
$$

By solving Eq. (11), Eq. (12) is obtained as follows.

$$
\begin{cases} a_0 = q_0 \\ a_1 = 0 \\ a_2 = \frac{3}{t_f^2}(q_f - q_0) \\ a_3 = -\frac{2}{t_f^3}(q_f - q_0) \end{cases} \tag{12}
$$

The boundary constraint condition is set as a unique cubic polynomial, and the corresponding constraint condition is substituted into Eq. (8) to obtain the corresponding linear equation. In order to ensure the continuity and stability of the motion when the critical starting and ending velocities are 0, the cubic polynomial function can be expressed as follows.

$$
q(t) = q_0 + \frac{3}{t_f^2}(q_f - q_0)t^2 - \frac{2}{t_f^3}(q_f - q_0)t^3 \tag{13}
$$

Then the calculation equations of joint angular velocity and angular

acceleration are as follows.

$$\begin{cases} \overline{q}(t) = \frac{6}{t_f^2}(q_f - q_0)t - \frac{6}{t_f^3}(q_f - q_0)t^2 \\ \overline{\overline{q}}(t) = \frac{6}{t_f^2}(q_f - q_0) - \frac{12}{t_f^3}(q_f - q_0)t \end{cases} \tag{14}$$

According to Eq. (14), the function curve of joint angular velocity is a parabola, while the function curve of angular acceleration is a straight line. If the angular acceleration follows the principle of constant change, it will cause the vibration of the mechanical arm, thus increasing the wear rate of the joint parts. The robot arm doesn't stay in motion at all path points, so the information of intermediate points should be considered. The path is divided into several sections, and the adjacent path points are defined as the initial point and the termination point, then the cubic polynomial function can be used for interpolation processing, and various paths are connected smoothly. Therefore, the joint velocity is not equal to 0 just for different segments. Therefore, the constraint conditions in Eq. (8) are adjusted accordingly.

$$\begin{cases} \overline{q}(0) = \overline{q}_0 \\ \overline{q}(t_f) = \overline{q}_f \end{cases} \tag{15}$$

In this case, the cubic polynomial correlation equation is as follows.

$$\begin{cases} q_0 = a_0 \\ q_f = a_0 + a_1 t_f + a_2 t_f^2 + a_3 t_f^3 \\ \overline{q}_0 = a_1 \\ \overline{q}_f = a_1 + 2a_2 t_f + 3a_3 t_f^2 \end{cases} \tag{16}$$

The coefficients in Eq. (16) are further solved, and the results are shown in the following equation.

$$\begin{cases} a_0 = q_0 \\ a_1 = \overline{q}_0 \\ a_2 = \frac{3}{t_f^2}(q_f - q_0) - \frac{2}{t_f}\overline{q}_0 - \frac{1}{t_f}\overline{q}_f \\ a_3 = \frac{1}{t_f}(\overline{q}_0 + \overline{q}_f) - \frac{2}{t_f^3}(q_f - q_0) \end{cases} \tag{17}$$

Finally, Eq. (17) can be used to describe the trajectory of any position and velocity in the initial and final points of the path.

Then, MATLAB software is used for the simulation analysis of the motion trajectory planning of the polishing robot operating arm. The whole running time is set as 5s. The starting path is X0, the middle point X1 is passed through at 2.5s, and then the end point X2 is reached after 2.5s. At the start and end of the manipulator, the joint velocities are 0. First, the inverse kinematics equation is used to calculate the angles of each joint at different path points, and then interpolation is performed every 20ms. Finally, using cubic polynomial interpolation method to plan the motion trajectory, the curve of joint axis angle with time is obtained.

III. Results and Discussion

A. Verification of Kinematic Solution of Operating Arm of Polishing Robot

Since the forward kinematic solution method of the robot manipulator is relatively simple, while the reverse kinematic solution operation is complicated and there are many methods, the inverse kinematic solution method proposed has been verified by an example [26]. MATLAB software is used for verification. The corresponding solver is written first, and then the inverse solution of the end of the operating arm of the robot is obtained. Finally, the different solutions obtained are put into the forward kinematic equation to determine whether the equation is applicable. The coordinates of a target point are randomly selected as (-150, 550, -450), and the attitude of the end piece when the robot manipulates the arm to reach the target point is $\left(\frac{\pi}{7}, \frac{\pi}{5}, \frac{\pi}{6}\right)$. After the calculation of MATLAB software, the inverse solutions of each joint of the robot operating arm are obtained, and the results are shown in Table III. The results of 8 inverse kinematic solutions with different joint angles show the trend of arcs, and the forward kinematic equation is substituted to solve the problem.

TABLE III. Results of inverse Kinematic Solution of Robot Manipulator Under Target Pose

Grouping	$\theta_1(°)$	$\theta_2(°)$	$\theta_3(°)$	$\theta_4(°)$	$\theta_5(°)$	$\theta_6(°)$	$\theta_7(°)$
1	0.752	1.012	-0.230	1.021	-0.563	0.186	-1.689
2	0.752	1.022	1.235	-1.021	-0.676	0.770	-1.689
3	-1.221	1.233	1.089	1.021	0.913	0.182	0.105
4	-1.221	-1.233	-0.579	-1.021	0.587	0.771	0.105
5	3.879	-1.018	2.920	1.021	-0.563	0.182	-1.689
6	3.879	-1.018	4.379	-1.021	-0.676	0.770	-1.689
7	1.903	-1.231	4.241	1.021	0.913	0.186	0.105
8	1.903	1.231	2.560	-1.021	0.560	0.771	0.105

B. Polishing Robot Operating Arm Simulation Results

Fig. 4 shows the animation simulation model of polishing robot with different performance built by ADAMS software. It can be concluded that the polishing robot needs to have the performance of deburring work built on non-smooth surface, polishing work built on medium and large scale, trimming work of complex components, and return to the working origin.
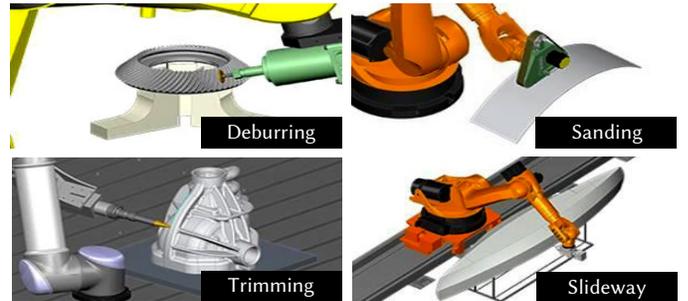

Fig. 4. The animation simulation model of polishing robot with different performance.

ADAMS software is used to process the curve graph of the output processing data of the module, so as to generate the function curve of the operating arm of the polishing robot with different coaxial changes over time to observe the specific situation of the changes of each parameter with time more intuitively. The curves of the angles of different joint axes and the displacement of end parts over time are shown in Fig. 5. It can be concluded from Fig. 5A that the change trend of the rotation Angle of axis 4 and axis 5 is basically the same, while the change trend of the rotation Angle of axis 6 and axis 7 is basically the same. As can be concluded from Fig. 5B, the initial displacement of the manipulator in the X-axis direction (End piece1) is about -450mm, and the position moves to the positive maximum at about 15s (about 600mm), and to the negative maximum at about 21s (about -500mm). The initial displacement on the Y-axis direction (End piece2) is about -350mm, and the position reaches the positive maximum at about 14s (about 810mm), and the position reaches the negative maximum at about 21s (about -500mm). The initial displacement on the z-axis direction (End piece3) is about 400mm, and the position reaches the maximum forward direction (about 550mm) at around 16s, and the position reaches the maximum negative direction (about -810mm) at

about 12s. As shown in Fig. 5, the movement trajectory curves of each joint axis and the end parts of the polishing robot constructed in this study present a continuous and smooth change trend, without any sudden change, which indicates that when the robot manipulators move, the joints and connecting rods present a stable and non-violent movement trend [27]. This also verifies the reliability of the mechanical structure of the polishing robot designed in this study. Finally, all can return to the starting position. The displacement of the end piece is calculated as (550.8mm, 400.5mm, -80.1mm), and the joint angles are θ_1=0°, θ_2=-90°, θ_3=-180°, θ_4=0°, θ_5=0°, θ_6=0°, θ_7=-90°, respectively. It is put into the kinematic equation constructed in this study, and the position coordinates obtained after calculation are consistent with the detection results, which reflects that ADAMS software is used for the simulation analysis of the polishing robot operating arm, with high accuracy and reliability [28].


(a)

(b)
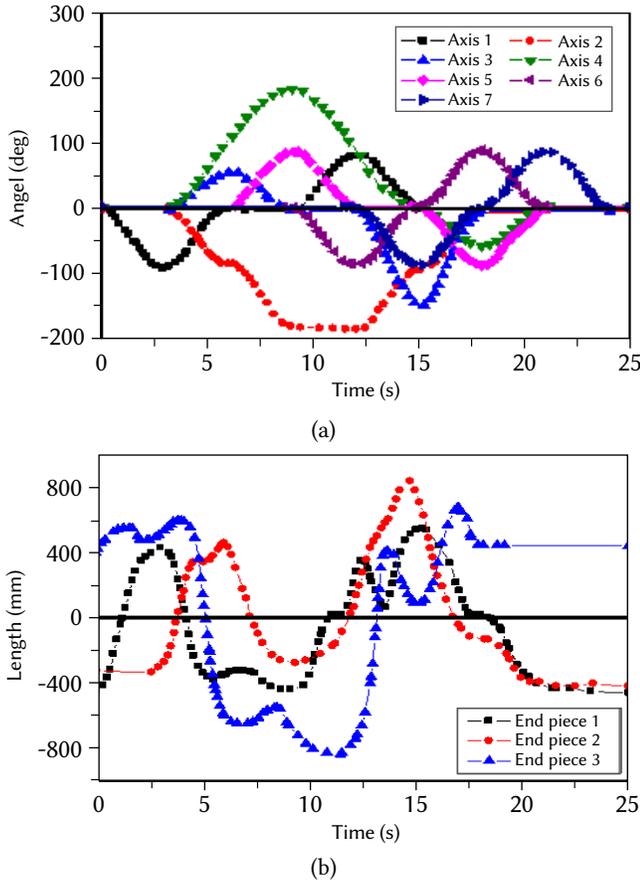
Fig. 5. The curve diagram of the rotation angle of the joint axis of the robot manipulator arm and the displacement of the terminal part with time based on ADAMS software. Note: (a) is the curve diagram of the rotation angle of joint axis changing with time; (b) is the curve graph of displacement of the terminal piece changing with time.

C. Simulation of Trajectory Planning of the Robot Operating Arm

MATLAB is the software often used for trajectory analysis simulation, which can be directly generated through programming data curve. The Robotic toolbox can be used to simulate track planning, and then output the relevant data curve. And for some special difference function, VC++ and other programming can be used to automatically generate the trajectory of the simulation program [29], [30]. MATLAB software is used to generate the movement trajectory of the polishing robot operating arm. The starting point X1 coordinate is set as (17.32, -100.00, 99.09), the intermediate point X2 coordinate is (92.17, -100.00, 79.37), and the termination point X3 coordinate is (54.88, -100.00,

-100.00). The inverse kinematics equation is used to calculate the angles of each joint of the robot arm at the three path points, and the results are shown in Table IV.

TABLE IV. Joint Angles At Different Path Points

Joints	Starting point X_1	Middle point X_2	End point X_3
1(°)	-6.45	-10.73	-6.33
2(°)	-5.76	-0.58	21.19
3(°)	-19.20	-13.41	2.73
4(°)	29.33	13.85	-6.61
5(°)	-30.00	-50.00	-80.00
6(°)	-11.56	-0.75	21.50
7(°)	36.08	60.70	85.61

After the command program is input into MARLAB software, the angle curves and angular velocity curves of each joint axis are calculated and obtained. The results are shown in Fig. 6. From the simulation curve diagram, it can be intuitively observed of the changes in the angle and angular velocity of each joint axis in the whole process from the starting point of path planning to the middle point and finally to the end point. It can be concluded from Fig. 6A that the rotation angle of joint pumping 1 changes gently, while the rotation angle of joint axis 2, 3, and 7 increases gradually, and the rotation angle of joint axis 4, 5, and 6 decreases gradually. As shown in Fig. 6B, taking the time to reach the intermediate point as the node (2.5s), the angular velocity of different joint axes at this point shows obvious changes, but the angular velocity of both the starting and ending positions is 0.


(a)

(b)
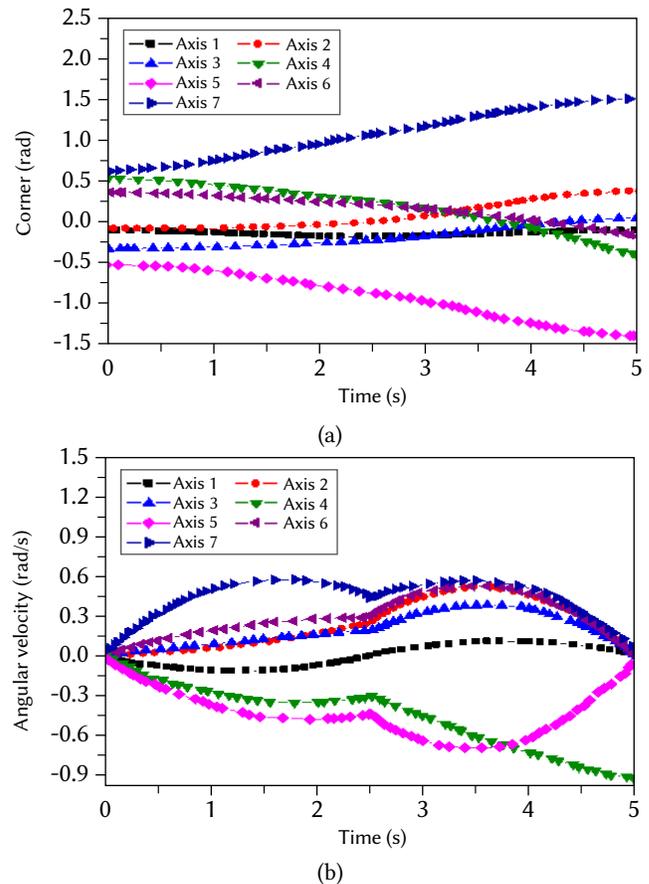
Fig. 6. The path planning simulation results of the polishing robot operating arm. Note: A is the angle curve of each joint axis of the robot operating arm; B is the angular velocity curve of each joint axis of the robot operating arm.

Taking joint axis 1 as an example, the change trends of angular

displacement and angular velocity before and after optimization are compared, and the results are shown in Fig. 7. It can be concluded from Fig. 7A and Fig. 7B that under the premise of satisfying each constraint condition, the changes of angular displacement and angular velocity of the optimized joint axis 1 are continuous smooth curves. Both the angular displacement and angular velocity curves of the joint axis 1 before optimization show a certain degree of mutation, and the time to reach the maximum displacement and angular velocity of the joint axis 1 before optimization is longer than that after optimization. Based on cubic polynomial interpolation, Walambe et al. realized the planning of vehicle prototype motion trajectory, and obtained the continuous, smooth, and optimized path trajectory, which is basically consistent with the results of this study [31].


(a)

(b)
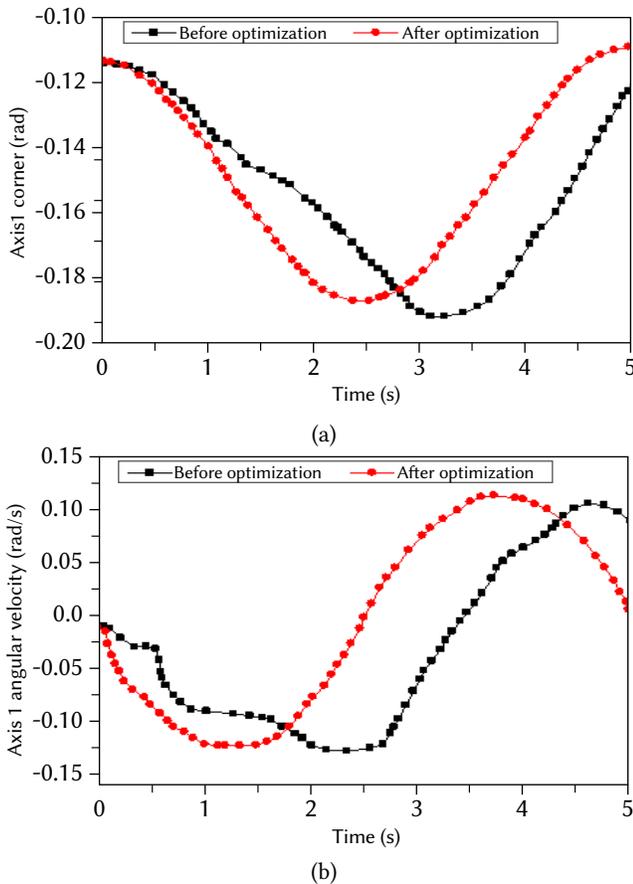
Fig. 7. Trajectory curves of joint axis 1 before and after optimization. Note: A is the angle curve of joint axis 1 before and after optimization; B is the angular velocity curves of joint axis 1 before and after optimization.

MATLAB software is used to simulate the movement trajectory of the polishing robot operating arm reaching the polishing target point before and after optimization, and the difference of operation time before and after optimization is compared, and the results are shown in Table V. The time for the optimized polishing robot to reach the polishing position required by the motion trajectory is 45.14s, while the time after the optimization is only 27.93s, which is shortened by about 17.21s. Moreover, the polishing position required by the trajectory after optimization is consistent with that before optimization, and there is no deviation in the position. This indicates that the trajectory planning of polishing robot based on cubic polynomial interpolation can better improve the motion speed and work efficiency of polishing robot operating arm [32]-[33].

TABLE V. Comparison of the Time to Reach the Polishing Position Before and After Optimization

Time (s)	t_1	t_2	t_3	t_4	t_5	Total
Before optimization	7.41	2.38	15.17	4.99	15.19	45.14
After optimization	4.22	3.19	7.29	3.95	9.28	27.93

IV. Conclusion

In order to explore the effect of cubic polynomial interpolation in polishing robot path optimization, in the study, firstly, a 7-degree-of-freedom articular polishing robot operating arm system is constructed. The d-h method is introduced to construct the kinematic model of the manipulator system. The kinematic model of the robot manipulator is simulated and analyzed by ADAMS software. It is found that the kinematic model of the robot operating arm is accurate. Then, the trajectory of the robot is planned based on cubic polynomial interpolation. Finally, MATLAB software is used for simulation and verification, and it is found that it could accurately show the displacement and angular velocity changes of the robot end parts. Moreover, the path planning of polishing robot is carried out using cubic polynomial interpolation method, which can effectively reduce the time to reach the polishing position compared with the unoptimized method. However, only simulation software is analyzed to carry out simulation tests, and subsequent application analysis is needed to verify the reliability of the polishing robot operating arm system proposed in this study. To sum up, the results of this study can provide theoretical basis for improving the working efficiency of polishing robot.

References

[1] L. Song, W.H. Liu, Y.M. Li, and F.H. Xin. Humidity-resistant inorganic binder for sand core making in foundry practice, China Foundry, vol. 16, no.4, pp. 267-271, 2019, doi:10.1007/s41230-019-8169-8.

[2] Y. Watanabe, T. Sugiura, H. Sato, and H. Tsuge. Fabrication of Al-Based Composites by Centrifugal Mixed-Powder Method and Their Application for Grinding Wheels, Journal of Materials Engineering and Performance, pp. 1-12, 2019, doi:10.1007/s11665-019-03917-3.

[3] Shaofei Wu. A Traffic Motion Object Extraction Algorithm, International Journal of Bifurcation and Chaos, vol. 25, no. 14, Article Number 1540039, 2015, doi:10.1142/S0218127415400398.

[4] G. Pradeep, S. K. Kumar, G. Mathivanan, and H. Praveen. Design and Fabrication of Punch Tool Arrangement to Remove Excess Base Fillings from Aluminium Die Casting, Engineering Reports, vol. 2, no. 1, pp.30-34, 2019.

[5] T. Zhang. Weighted hand-eye calibration algorithm for robot grinding, Journal of Mechanical Engineering, vol. 54, no. 17, pp. 142, 2018.

[6] K. Subrin, S. Garnier, T. Bressac, and B. Furet. Digital chain development for sanding application with a kinematically redundant robotic system, Procedia CIRP, vol. 82, pp. 515-520, 2019, doi:10.1016/j.procir.2019.04.063.

[7] H. Lyu, Y. Liu, J. Y. Guo, H. M. Zhang, and Z. X. Li. Tool-path generation for industrial robotic surface-based application, Advances in Manufacturing, vol. 7, no. 1, pp. 64-72, 2019, doi:10.1007/s40436-018-00246-x.

[8] W. X. Ng, H. K. Chan, W.K. Teo, and I.M. Chen. Capturing the tacit knowledge of the skilled operator to program tool paths and tool orientations for robot belt grinding, International Journal of Advanced Manufacturing Technology, vol. 91, no. 5-8, pp. 1599-1618, 2017, doi:10.1007/s00170-016-9813-8.

[9] D. Mahdy, R. Reda, N. Hamdi, and I.S.M. Khalil. Ultrasound-guided minimally invasive grinding for clearing blood clots: promises and challenges, IEEE Instrumentation and Measurement Magazine, vol. 21, no. 2, pp. 10-14, 2018.

[10] H. J. Lee, M. T. Soe, S. H. Chauhdary, S. Rhee, and M. S. Park. A data aggregation scheme for boundary detection and tracking of continuous objects in WSN, Intelligent Automation & Soft Computing, vol. 23, no. 1, pp. 135-147, 2017, doi:10.1080/10798587.2016.1183922.

[11] V. Andrean, P. Chang, and K. Lian. A review and new problems discovery

of four simple decentralized maximum power point tracking algorithms—Perturb and observe, incremental conductance, golden section search, and Newton's quadratic interpolation, Energies, vol. 11, no. 11, pp. 2966, 2018, doi:10.3390/en11112966.

[12] E. Zubko, G. Videen, J. A. Arnold, B. MacCall, A. J. Weinberger, and Y. Shkuratov. Interpolating light-scattering properties of irregularly shaped, absorbing particles, Optics letters, vol. 43, no. 17, pp. 4308-4311, 2018, doi:10.1364/OL.43.004308.

[13] B. Zhang, and Z. Feng. The Construction Method of Interpolation Polynomial by Fractal Interpolation, International Journal of Nonlinear Science, vol. 25, no. 1, pp. 48-52, 2018.

[14] Shaofei Wu, Mingqing Wang, Yuntao Zou. Research on internet information mining based on agent algorithm, Future Generation Computer Systems, vol. 86, pp. 598-602,2018, doi:10.1016/j.future.2018.04.040.

[15] Ü. Dinçer, and M. Çevik. Improved trajectory planning of an industrial parallel mechanism by a composite polynomial consisting of Bézier curves and cubic polynomials, Mechanism and Machine Theory, vol. 132, pp. 248-263, 2019, doi:10.1016/j.mechmachtheory.2018.11.009.

[16] L. Li, Y. Zhang, M. Ripperger, J. Nicho, and A. Fumagalli. Autonomous object pick-and-sort procedure for industrial robotics application, International Journal of Semantic Computing, vol. 13, no. 2, pp. 161-183, 2019, doi:10.1142/S1793351X19400075.

[17] J. R. Bourne, E.R. Pardyjak, and K.K. Leang. Coordinated bayesian-based bioinspired plume source term estimation and source seeking for mobile robots, IEEE Transactions on Robotics, vol. 35, no. 4, pp. 967-986, 2019, doi: 10.1109/TRO.2019.2912520.

[18] S. H. Kim, and J. H. Park. The Effect of Occupation-Based Bilateral Upper Extremity Training in a Medical Setting for Stroke Patients: A Single-Blinded, Pilot Randomized Controlled Trial, Journal of Stroke and Cerebrovascular Diseases, vol. 28, no. 12, pp. 104335, 2019.

[19] R. Hou, L. Wang, Q. Gao, Y. Hou, and C. Wang. Indirect adaptive fuzzy wavelet neural network with self-recurrent consequent part for AC servo system, ISA transactions, vol. 70, pp. 298-307, 2017, doi:10.1016/j.isatra.2017.04.010.

[20] H. Yan, Y. Cao, and J. Yang. Statistical tolerance analysis based on good point set and homogeneous transform matrix, Procedia Cirp, vol. 43, pp. 178-183,2016, doi:10.1016/j.procir.2016.02.042.

[21] R. R. Serrezuela, A. F. C. Chavarro, M. A. T. Cardozo, A. L. Toquica, and L. F. O. Martinez. Kinematic modelling of a robotic arm manipulator using Matlab, ARPN Journal of Engineering and Applied Sciences, vol. 12, no. 7, pp. 2037-2045, 2017.

[22] I. Pavlenko, J. Trojanowska, O. Gusak, V. Ivanov, J. Pitel, and V. Pavlenko. Estimation of the reliability of automatic axial-balancing devices for multistage centrifugal pumps, Periodica Polytechnica Mechanical Engineering, vol. 63, no. 1, pp. 52-56, 2019.

[23] X. H. Li, H. B. Yu, M. Z. Yuan, Y. Yin, and H. Pan. Study on the linear dynamic model of shield TBM cutterhead driving system, IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society, pp. 3864-3871, 2011.

[24] Z. Liu, J. Xu, Q. Cheng, Y. Zhao, and C. Yang. Trajectory planning with minimum synthesis error for industrial robots using screw theory, International Journal of Precision Engineering and Manufacturing, vol. 19, no. 2, pp. 183-193, 2018.

[25] S. Wu, M. Wang, Y. Zou. Bidirectional cognitive computing method supported by cloud technology, Cognitive Systems Research, vol. 52, pp. 615-621, 2018, doi:10.1016/j.cogsys.2019.03.015.

[26] D. Di Vito, C. Natale, and G. Antonelli. A comparison of damped least squares algorithms for inverse kinematics of robot manipulators, IFAC-PapersOnLine, vol. 50, no. 1, pp. 6869-6874, 2017, doi:10.1016/j.ifacol.2017.08.1209.

[27] S. Pellegrinelli, A. Orlandini, N. Pedrocchi, A. Umbrico, and T. Tolio. Motion planning and scheduling for human and industrial-robot collaboration, CIRP Annals, vol. 66, no. 1, pp. 1-4, 2017, doi:10.1016/j.cirp.2017.04.095.

[28] S. Wu, Nonlinear information data mining based on time series for fractional differential operators, Chaos, vol. 29, pp. 013114, 2019, doi:10.1063/1.5085430.

[29] L. Ji, H. Wang, T. Zheng, and X. Qi. Motion trajectory of human arms based on the dual quaternion with motion tracker, Multimedia Tools and Applications, vol. 76, no. 2, pp. 1681-1701, 2017, doi:10.1007/s11042-015-3099-y.

[30] J. Xu, H. Yang, M. Liu, J. Tian, and B. Liu. Research on winding trajectory planning for elbow pipe based on industrial robot, The International Journal of Advanced Manufacturing Technology, vol. 93, no. 1-4, pp. 537-545, 2017, doi:10.1007/s00170-017-0568-7.

[31] R. Walambe, N. Agarwal, S. Kale, and V. Joshi. Optimal trajectory generation for car-type mobile robot using spline interpolation, IFAC-PapersOnLine, vol. 49, no. 1, pp. 601-606, 2016, doi:10.1016/j.ifacol.2016.03.121.

[32] H. M. Alwan, and Z. H. Rashid. Motion Control of Three Links Robot Manipulator (Open Chain) with Spherical Wrist, Al-Khwarizmi Engineering Journal, vol. 15, no. 2, pp. 13-23, 2019, doi:10.22153/kej.2019.12.001.

[33] K. Leng, Y. Bi, L. Jing, H. Fu, I. Van Nieuwenhuyse. Research on agricultural supply chain system with double chain architecture based on blockchain technology, Future Generation Computer System, vol. 86, pp. 641-649, 2018, doi:10.1016/j.future.2018.04.061.

Li You

Li You Master of Fine Art currently a lecturer of Art and Design School in Wuchang Institute of Technology. Her research interests include digital media art, comprehensive material painting and animation modeling research. Dozens of her academic papers and paints have been published in Chinese Journals as well as international Journals. Email:120120962@wut.edu.cn

Juan Wang

Juan Wang was born in Zhongxiang, Hubei, P.R. China, in 1987. She received the Master degree from Hubei University of Technology, P.R. China. Now, she works in School of Arts Design, Wuchang University of Technology. Her research interests include animation design and digital illustration design. Her papers and works have won awards in provincial competitions for many times, and she has published many core journals and works. E-mail:120100054@wut.edu.cn

Yiming Ji

Yiming Ji now works as a lecturer in School of journalism at Zhongnan university of economics and law, Master of fine art. His research interests include digital media art, computer graphics, animation design and cartoon creation. His works have participated in international exhibitions, national exhibitions and provincial exhibitions for many times. Email: jym20200123@163.com

Deep Learning-based Side Channel Attack on HMAC SM3

Xin Jin[1], Yong Xiao[1], Shiqi Li[2]*, Suying Wang[2]

[1] CSG Electric Power Research Institute, Guangzhou (China)
[2] Open Security Research, Inc., Shenzhen (China)

Abstract

SM3 is a Chinese hash standard. HMAC SM3 uses a secret key to encrypt the input text and gives an output as the HMAC of the input text. If the key is recovered, adversaries can easily forge a valid HMAC. We can choose different methods, such as traditional side channel analysis, template attack-based side channel analysis to recover the secret key. Deep Learning has recently been introduced as a new alternative to perform Side-Channel analysis. In this paper, we try to recover the secret key with deep learning-based side channel analysis. We should train the network recursively for different parameters by using the same dataset and attack the target dataset with the trained network to recover different parameters. The experiment results show that the secret key can be recovered with deep learning-based side channel analysis. This work demonstrates the interests of this new method and show that this attack can be performed in practice.

Keywords

I. Introduction

SIDE CHANNEL analysis is a powerful technique that helps an adversary recover sensitive information without damaging the target. Target device leaks information (e.g. power consumption [1], Electromagnetic emanation [2], temperature [3], acoustic [4], etc.) that is related to sensitive data during calculation [5]. An adversary can make use of the leakage to recover sensitive data. In order to decrease the leakage, several countermeasures such as masking and hiding are used in the cryptographic implementation. However, even with countermeasures, adversaries can come up with more powerful methods to recover the sensitive information.

Following the current trend in the side channel analysis research area, recent works have demonstrated that deep learning algorithms were very efficient to conduct security evaluations of embedded systems and had many advantages compared to the other methods.

Nowadays, machine learning becomes a popular topic in many areas. It is usually divided into three classes: supervised learning, unsupervised learning and semi-supervised learning. In most situation, supervised learning is mainly used. There are many kinds of structures which are used in machine learning, such as Support Vector Machine (SVM) and Random Forest, etc. Deep learning is a kind of machine learning. It extracts features by several non-linear layers. Deep learning becomes popular since AlexNet [6] is proposed in 2012. Then, more and more complex network structures are proposed, such as VGGNet [7], GoogLeNet [8] and ResNet [9], etc. These networks work well in many areas, e.g. image recognition area, face recognition area and so on.

In recent years, it appears that deep learning techniques are applied in side channel analysis research area [10], [11]. Comparing to traditional side channel method, deep learning-based side channel analysis performs better. Deep learning-based side channel analysis performs better especially when the implementation has mask or jitter [12]. Without alignment or pre-processing, neural network can recover sensitive information as well, which is much more convenient than the traditional side channel method. Many researches are done in recent years. In 2013, Martinasek. et al., play an attack on an AES implementation working on PIC16F84A with only one hidden layer [13]. Another work [14] compares different kind of machine learning method on DPA contest V2 dataset. A research [15] proposed a CNN based side channel analysis and claim that this method is robust to trace misalignment. Different structures of network are applied on the dataset and come up with a CNN-best structure for the dataset. Picek et al. compare different methods for the class imbalance situation on DL-SCA [16]. Last, a research [17] come up with correlation-based loss function.

In this paper, we use a deep learning-based side channel analysis technique to analyze HMAC SM3 algorithm implementation. The structure of this paper is as follows: In Section II, we introduce the SM3 algorithm, HMAC SM3, basic idea of CNN as well as the attack path. This section will help readers have a basic understanding of the algorithm and the attack method. The attacks on real traces are demonstrated in Section III. In this section, the target, the structure of the network and the attack are illustrated. In the end, conclusion and future work are presented in Section IV.

II. Background

A. SM3 Algorithm

SM3 is the Chinese hash standard [18]. The structure of the

* Corresponding author.

E-mail address: shiqi.li@osr-tech.com

algorithm is shown in Fig. 1. The input data of the function is padded such that it can be split into N blocks of 512 bits. Each block will be treated in a same procedure: the former block calculates a new IV for the latter block through function f(), and the output of block N is the hash result of the algorithm.
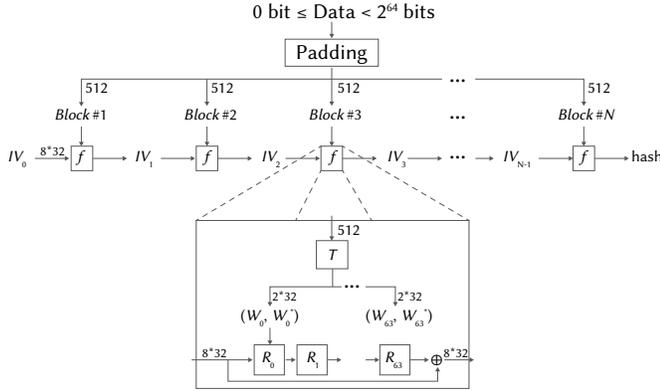

Fig. 1. Structure of SM3 algorithm.

The structure of function f() is show in Fig 2. The function T() convert the 512-bit input into 64 32-bit word pairs. Each pair (Wi, Wi*) are used during round Ri, and the result of each round is used as input of the next round. When the 64th round is completed, a final transformation is applied by adding the input of the first round and the output of the last round together as the output of the function f().
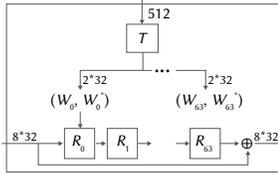

Fig. 2. Structure of the f function.

In order to explain the detail of each loop, we define the loop by function: $IV_i = f(IV_{i-1}, Block_i)$

The first constant IV_0 is:

$$IV_{0,0} = 0x7380166F \tag{1}$$

$$IV_{0,1} = 0x4914B2B9 \tag{2}$$

$$IV_{0,2} = 0x172442D7 \tag{3}$$

$$IV_{0,3} = 0xDA8A0600 \tag{4}$$

$$IV_{0,4} = 0xA96F30BC \tag{5}$$

$$IV_{0,5} = 0x163138AA \tag{6}$$

$$IV_{0,6} = 0xE38DEE4D \tag{7}$$

$$IV_{0,7} = 0xB0FB0E4E \tag{8}$$

The detail of each loop is as follows:

First, initialize the 8 32-bit local variables named a to h:

$$a_0 = IV_{i-1, \cdots 0} \tag{9}$$

$$b_0 = IV_{i-1, \cdots 1} \tag{10}$$

$$c_0 = IV_{i-1, \cdots 2} \tag{11}$$

$$d_0 = IV_{i-1, \cdots 3} \tag{12}$$

$$e_0 = IV_{i-1, \cdots 4} \tag{13}$$

$$f_0 = IV_{i-1, \cdots 5} \tag{14}$$

$$g_0 = IV_{i-1, \cdots 6} \tag{15}$$

$$h_0 = IV_{i-1, \cdots 7} \tag{16}$$

For each round R_j, $j \in [0,63]$, we compute:

$$SS1_j = ((a_j <<< 12) + e_j + (T_j <<< j)) <<< 7 \tag{17}$$

$$SS2_j = SS1_j \oplus (a_j <<< 12) \tag{18}$$

$$TT1_j = FF_j (a_j, b_j, c_j) + d_j + SS2_j + W_j^* \tag{19}$$

$$TT2_j = GG_j (e_j, f_j, g_j) + h_j + SS1_j + W_j \tag{20}$$

$$a_{j+1} = TT1_j \tag{21}$$

$$b_{j+1} = a_j \tag{22}$$

$$c_{j+1} = b_j <<< 9 \tag{23}$$

$$d_{j+1} = c_j \tag{24}$$

$$e_{j+1} = P_0 (TT2_j) \tag{25}$$

$$f_{j+1} = e_j \tag{26}$$

$$g_{j+1} = f_j <<< 19 \tag{27}$$

$$h_{j+1} = g_j \tag{28}$$

where all additions are done modulo 2^{32} <<< n means left rotation of n bits, constants T_j is:

$$T_j = \begin{cases} 0x79CC4519, & 0 \le j \le 15 \\ 0x7A879D8A, & 15 < j < 64 \end{cases} \tag{29}$$

Function FF_j is:

$$FF_j(X,Y,Z) = \begin{cases} X \oplus Y \oplus Z, & 0 \le j \le 15 \\ (X \wedge Y) \vee (X \wedge Z) \vee (Y \wedge Z) & 15 < j < 64 \end{cases} \tag{30}$$

Function GG_j is:

$$GG_j(X,Y,Z) = \begin{cases} X \oplus Y \oplus Z, & 0 \le j \le 15 \\ (X \wedge Y) \vee (\neg X \wedge Z) & 15 < j < 64 \end{cases} \tag{31}$$

Function P_k is:

$$P_k(X) = \begin{cases} X \oplus (X \lll 9) \oplus (X \lll 17), & k = 0 \\ X \oplus (X \lll 15) \oplus (X \lll 23), & k = 1 \end{cases} \tag{32}$$

Input plaintext of each block $Plain_{Block}$ is split into 32-bit words $Plain_{Block} = \{PB_0, PB_1, ..., PB_{15}\}$. Then the parameter W_j is computed as:

$$W_j = \begin{cases} PB_j, & 0 \le j \le 15 \\ P_1(W_{j-16} \oplus W_{j-9} \oplus (W_{j-3} \lll 15)) \oplus (W_{j-13} \lll 7) \oplus W_{j-6}, & 15 < j < 68 \end{cases} \tag{33}$$

And the parameter W_j^* is computed as:

$$W_j^* = W_j \oplus W_{j+4} \tag{34}$$

The function f() is finished by 32-bit XOR with the initial state:

$$IV_{i,0} = IV_{i-1,0} \oplus a_{64} \tag{35}$$

$$IV_{i,1} = IV_{i-1,1} \oplus b_{64} \tag{36}$$

$$IV_{i,2} = IV_{i-1,2} \oplus c_{64} \tag{37}$$

$$IV_{i,3} = IV_{i-1,3} \oplus d_{64} \tag{38}$$

$$IV_{i,4} = IV_{i-1,4} \oplus e_{64} \tag{39}$$

$$IV_{i,5} = IV_{i-1,5} \oplus f_{64} \tag{40}$$

$$IV_{i,6} = IV_{i-1,6} \oplus g_{64} \tag{41}$$

$$IV_{i,7} = IV_{i-1,7} \oplus h_{64} \tag{42}$$

B. SM3 Based HMAC

The HMAC stands for keyed- Hash Message Authentication Code and is a NIST standard which can be found in [19]. Fig. 3 presents the process of HMAC SM3.
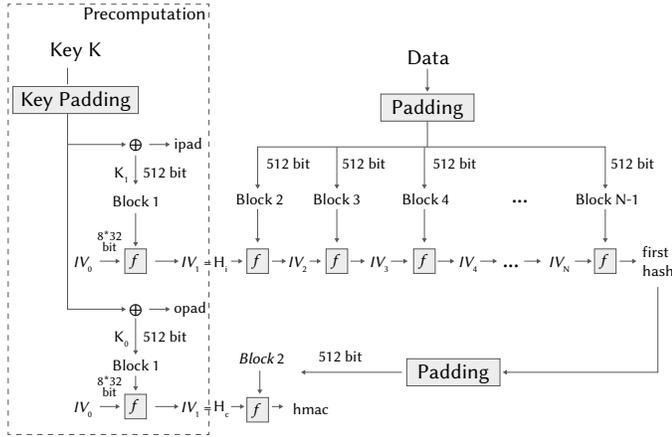

Fig. 3. Structure of HMAC SM3.

The process of HMAC SM3 is as follows:

Derive key pair (K_i, K_o) from the key K.

Calculate the first hash with K_i and input data: first hash = $H(K_i \mid T)$

Calculate HMAC with K_o and first hash: HMAC = $H(K_o \mid$ first hash), where H() is the SM3 hash function.

C. Side Channel Analysis

Side Channel Analysis (SCA) is first proposed by Kocher et al. in 1996[22]. It is a technique to retrieve the secret information of an algorithm by monitoring the physical information of a device (such as power, heating, time consuming, electromagnetic signals, etc., as shown in Fig. 4. The reason that SCA can recover secret is that the physical signal of a cryptographic device demonstrates correlation with the internal statement.

It is much easier to recover information from side channel signals than directly breaking the core implementation. There are several kinds of SCA, e.g. simple power analysis, correlation power analysis, template attack, etc. Simple power analysis [23] is an easy way to recover secret information. By observing the side channel signals, the attacker can find the difference and recover the sensitive information according to the differences. Correlation power analysis (CPA) [24] needs much more traces. When using CPA to recover sensitive information, we need to guess the secret key to calculate a certain mid-value. Since different traces correspond to different plaintext, we can have a set of mid-value for every guesses. By computing the correlation between mid-values and the side channel signals, we can figure out the correct guess. Template attack (TA) is another kind of passive attack. It has two stages: first, template building, second, template matching. Deep learning based SCA is similar to TA. We will discuss TA and deep learning based SCA in the following section.

D. Deep Learning Based Side Channel Analysis

Template attack [20] is a traditional method of side channel analysis. During the attack, we should take a reference set of trace from a reference device first in the learning phase. For this set, we
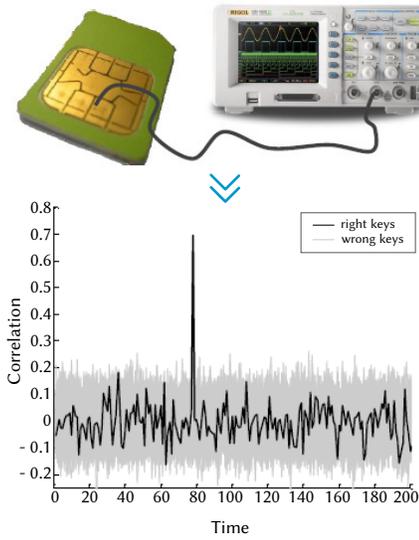
Fig. 4. Side Channel Leakage.

know the key, the plaintext and all the details of every trace. We can set up templates for each mid-value using the reference set. For attacking phase, we can use the templates in the learning phase to attack the target trace set to recover the mid-value such that the secret key can be recovered as well.

Deep learning-based side channel analysis is similar to the traditional template at-tack, which has two phases: a learning phase and an attacking phase. The whole procedure is shown in Fig. 5.
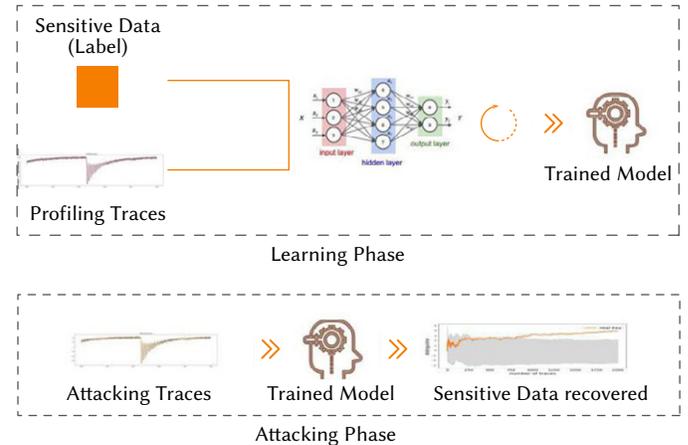
Fig. 5 .The procedure of deep learning-based side channel analysis.

In the learning phase, a trace set is collected from the reference device. For each trace, we add a label that is related to the sensitive data (e.g. key). The neural network will be trained with the traces and labels. Parameters in the neural network is updated. The goal of the learning phase is to try to make the output (prediction) of the neural network be closer to the true label. After the learning phase, the parameters updated in the network will be saved.

The network saved at the end of learning phase can be applied to the attacking phase. In the attacking phase, we have a trace set with only traces but we do not know the label. With each trace, the network can give us a prediction of the label using the parameters saved in the learning phase. We can recover the sensitive data (e.g. key) according to the predictions.

E. Neural Network Architecture

In this paper, we will primarily focus on Convolutional Neural Network. We will introduce the basic idea of the Convolutional Neural Network.

Convolutional Neural Network (CNN) [21] is a popular network in deep learning do-main. Usually, CNN is consisting of three kinds of layers: convolutional layers, pooling layers, as well as fully connected layers. The convolutional layers work as pattern detectors by convolution operations. Kernels in Convolutional layers can be updated by backpropagation. In this way, we can get different kernels using the same set of input. These kernels can detect different kinds of edges to track different features. Usually one kernel corresponds to one feature and a convolutional layer contains several kernels. In the meantime, since the kernels doing convolution by moving through the whole dataset, same pattern in different positions can be detected by the same kernel. We should notice that the kernels in the convolutional layer always have a small size compare to the input data, which reduce the computation complexity of the neural network.

Pooling layers always come after convolutional layers, which reduce the size of the inputs. We can choose average pooling to reduce the size by local averaging or max pooling to reduce the size by picking up the max value in a certain area. By averaging or max pooling, the pooling layer extract features of the input and reduce the size. This operation makes the CNN more robust to the shift and deformation of the input. In addition, it can reduce the possibility of overfitting since it reduces the size of the input data.

Fully connected layer usually comes at the end of the neural network. Each neural in fully connected layers are connect to every input.

We can choose the number of convolutional layers, pooling layers and fully connected layers arbitrarily. With more layers, the neural network can learn more com-plex features. However, with more layers, the network will be easier to get over-fit. Thus, the structure should be chosen carefully according to the input data.

The detailed architecture of MLP_{best}, $MLP_{monobit}$, $MLP_{multi-label}$, CNN_{best}, $CNN_{monobit}$ and $CNN_{multi-label}$ is shown in Fig. 6. The mark "FC-200" means a fully connected layer of 200 neurons, "conv11-64, Relu" means 64 convolutional kernels of size 11 using Relu activation function, and "average pooling, 2 by 2" means an average pooling layer, whose pooling window size is 2 and the stride is 2.

For a deep insight into the differences of identity model and our multi-label model, the architecture of the output layers of CNN_{best} and $CNN_{multi-label}$ (the same for MLP_{best} and $MLP_{multi-label}$) is depicted in Fig. 7, the output layer of CNN_{best} has 256 output neurons with softmax activation function while the output layer of $CNN_{multi-label}$ has 8 neurons with sigmoid activation function. Correspondingly, CNN_{best} uses cross-entropy as the loss function and $CNN_{multi-label}$ utilizes the binary cross-entropy since there are 8 binary labels.

F. Attack Path

Since SM3 algorithm has no secret key, we cannot attack SM3 directly. We can only attack HMAC SM3 to recover the key K used in the HMAC process. In order to recover the key K, we should recover the key pair (K_i, K_o) first. Thus, we should recover the first hash IV: Hi and the second hash IV: Ho to recover the key pair.

We use Ho and first hash result to calculate the HMAC. To recover Ho, we should know the first hash result first. In order to get the first hash result, we should recover the first hash $IV(H_i)$ first. Recovering Hi and Ho can use the same process. If H_i is recovered, Ho can be easily recovered as well. In this paper, we only consider recovering H_i and our target is to recover a_0, b_0, c_0, d_0, e_0, f_0, g_0 and h_0 related to H_i.

MLP_{best}	$MLP_{monobit}$	$MLP_{multi-label}$	CNN_{best}	$CNN_{monobit}$	$CNN_{multi-label}$
6 weights layers			8 weights layers		
352,456 params	301,201 params	302,608 params	66,652,544 params	65,607,809 params	65,636,488 params
input					
FC-200, Relu			conv11-64, Relu		
			average pooling, 2 by 2		
FC-200, Relu			conv11-128, Relu		
			average pooling, 2 by 2		
FC-200, Relu			conv11-256, Relu		
			average pooling, 2 by 2		
FC-200, Relu			conv11-512, Relu		
			average pooling, 2 by 2		
FC-200, Relu FC-200, Relu			conv11-512, Relu		
			average pooling, 2 by 2		
			FC-4096, Relu		
			FC-4096, Relu		
FC-256 softmax	FC-1, sigmoid	FC-8, sigmoid	FC-256, sigmoid	FC-1, sigmoid	FC-8, sigmoid

Fig. 6. Details of NN architecture.
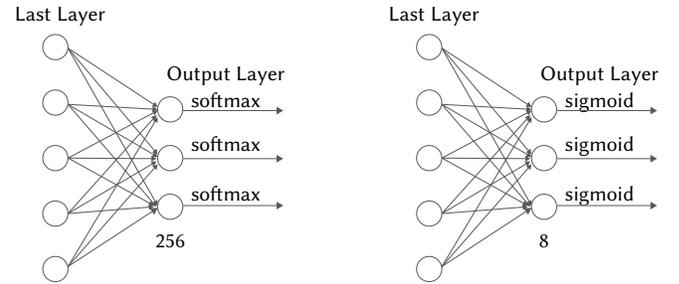
Fig. 7. The difference of output layers between CNN_{best} and $CNN_{multi-label}$. Left: the output layer of CNN_{best}. Right: the output layer of $CNN_{multi-label}$.

In order to illustrate the attack path more clearly, we denote part of the $TT1_j$ and $TT2_j$ computation as $\delta_{1,j}$ and $\delta_{2,j}$ respectively:

$$\delta_{1,j} = FF_j(a_j, b_j, c_j) + d_j + SS2_j \tag{43}$$

$$\delta_{2,j} = FF_j(e_j, f_j, g_j) + h_j + SS1_j \tag{44}$$

We can first recover $\delta_{1,0}$ and $\delta_{2,0}$ according to Equation (3) and Equation (4) respectively when j is equal to 0. With $\delta_{1,0}$ and $\delta_{2,0}$ known, $TT1_0$ and $TT2_0$ can be easily calculated since W_0^* and W_0 are known. Then, we can recover a_0 by targeting at $TT1_0 \oplus a_0$, recover b_0 by targeting at $TT1_0 \oplus a_0 \oplus (b_0 <<<9)$. c_0 can be recovered through targeting at the computation $c_0 + FF_0(TT1_0, a_0, (b_0 <<<9) + SS2_1 + W_1^*$. After $\delta_{1,0}$, a_0, b_0 and c_0 are recovered, we can simply recover d_0 by computing $d_0 = \delta_{1,0} - (a_0 \oplus b_0 \oplus c_0) - SS2_0$. Similarly, we can recover e_0, f_0, g_0 and h_0 with $TT2_0$ and W_1. Thus, the IV H_i can be recovered.

III. ATTACK ON REAL TRACES

A. Experiment Setup and Data Set

The testing target is a software HMAC SM3 running on a 32-bit microprocessor Infineon TC1782. The experiment setup consists of a high-performance Digital Storage Oscilloscope (DSO), high-precision XYZ stage and near-field high-bandwidth EM probe, as shown in Fig. 8.

EM traces are acquired when the HMAC SM3 is running. A single measurement contains 50,000 points, representing the computation of first hash.

Fig. 8. Experiment equipment and devices.

We collect two set of traces: Set A: 200,000 traces with input data and the first hash IV Hi varying; Set B: 50,000 traces with variance input data and fixed first hash IV Hi. Set A is used in the learning phase, while Set B is used for the attack phase. In the training phase, 180,000 traces of Set A are used as training set while the rest 20,000 traces are used as validation set to choose the best network parameter.

B. Neural Network Structure

Fig. 9 shows the structure of the network. We only use one convolutional layer with kernel size 3 and 32 convolutional filters. For the pooling layer, we use set both the pooling size and the stride to 2. The first fully connected layer has 1024 neuros while the second has 512 neuros. The input layer contains 5000 neuros while the output layer contains 9 neuros, which stands for Hamming Weight 0 to Hamming Weight 9.
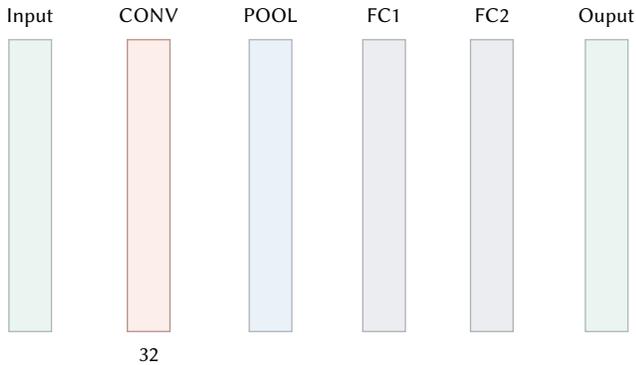
Fig. 9. Structure of the Neural Network.

The network contains 82,450,569 parameters in total, as shown in Fig. 10.

C. Experimental Result

We try to recover $\delta_{1,0}$ first. Instead of recover all 32-bits of $\delta_{1,0}$, we recover $\delta_{1,0}$ byte by byte. With learning rate 0.0001, batch size 200, we trained each model 10 epochs using Set A. The training result is shown in Fig. 11. the blue line corresponds to the training set while the orange line corresponds to the validation set. We can find that for every byte, the loss increases and the accuracy decrease in the validation set after several epochs. Thus, we save the network with best performance instead of the network obtained when training is finished.

```
Layer (type)              Output Shape          Param #
=====================================================
Input (InputLayer)        (None, 5000, 1)       0
-----------------------------------------------------
block1_conv (Conv1D)      (None, 5000, 32)      128
-----------------------------------------------------
block1_pool (AveragePoolingl (None, 2500, 32)   0
-----------------------------------------------------
flatten (Flatten)         (None, 80000)         128
-----------------------------------------------------
fc1 (Dense)               (None, 1024)          81921024
-----------------------------------------------------
fc2 (Dense)               (None, 512)           524800
-----------------------------------------------------
predictions (Dense)       (None, 9)             4617
=====================================================
Total params: 82, 450, 569
Trainable params: 82, 450, 569
Non-trainable params: 0
-----------------------------------------------------
```

Fig. 10. Parameters and Structure of the Neural Network.
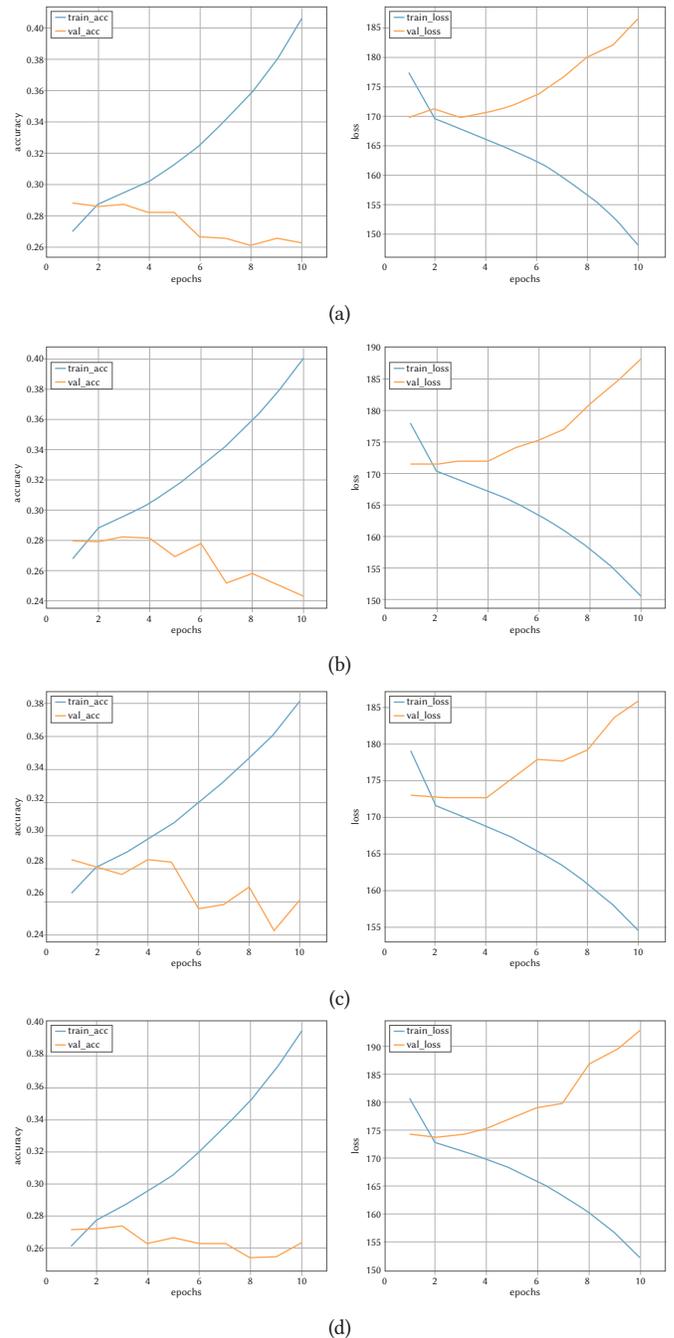
Fig. 11. Training result of $\delta_{1,0}$: (a) Byte3 (b) Byte 2 (c) Byte 1 (d) Byte 0.
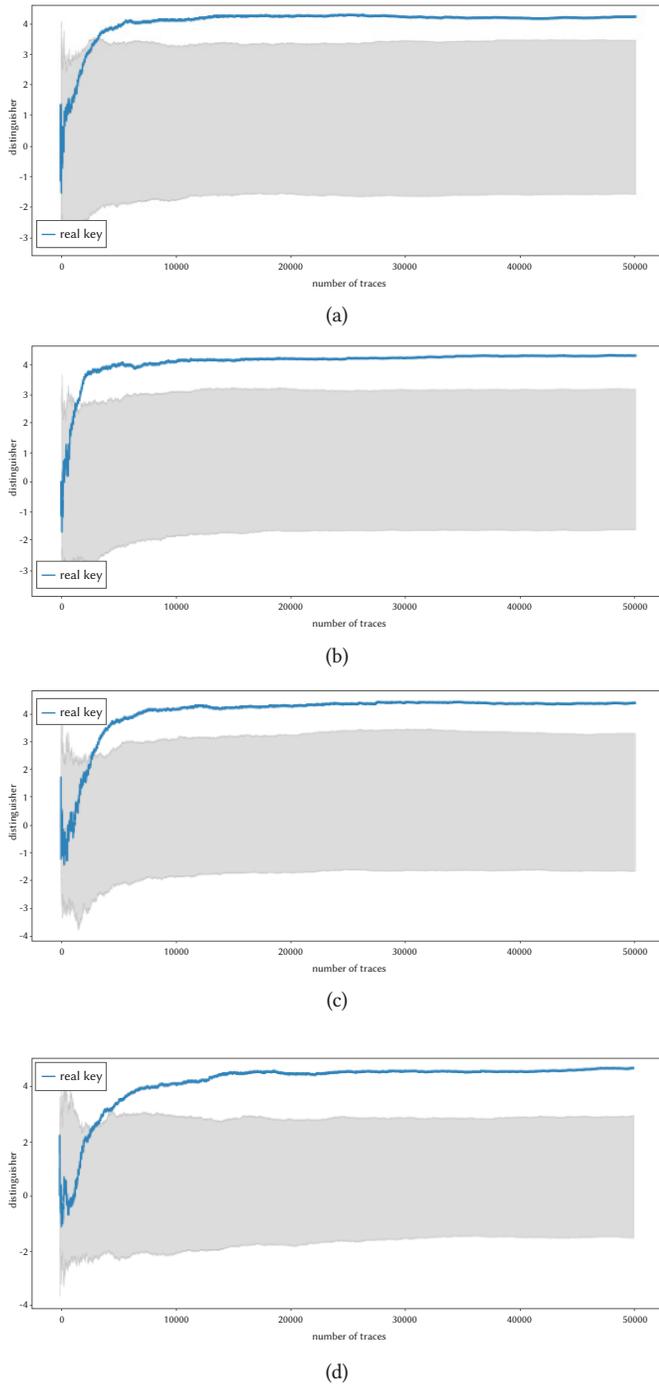
Fig. 12. Attack result of $\delta_{1,0}$: (a) Byte 3 (b) Byte 2 (c) Byte 1 (d) Byte 0.

The attack results on Set B are shown in Fig. 12. The line in blue indicates the expected value of different bytes. We can find that all bytes in $\delta_{1,0}$ can be recovered with only several thousands of traces in Set B.

With $\delta_{1,0}$ recovered, we can calculate $TT1_0$ for every trace according to the corresponding W_0^*. The EM traces leaks information related to a_0 when calculating $a_0 \oplus TT1_0$. The training result is shown in Fig. 13.

We recover a_0 byte by byte as well. The result is shown in Fig. 14. The line in blue indicate the expected value of different a_0 bytes. We can find from the result that we need to use almost all traces in Set B to recover all the four bytes of a_0. Unlike the result of $\delta_{1,0}$, the correct candidates of a_0 are not very distinguishable from other candidates. The result of $\delta_{1,0}$ seems more distinguishable than that of a_0, the reason

may be that the process that leaks information about a_0 is a XOR operation while the leakage about $\delta_{1,0}$ is an ADD operation.
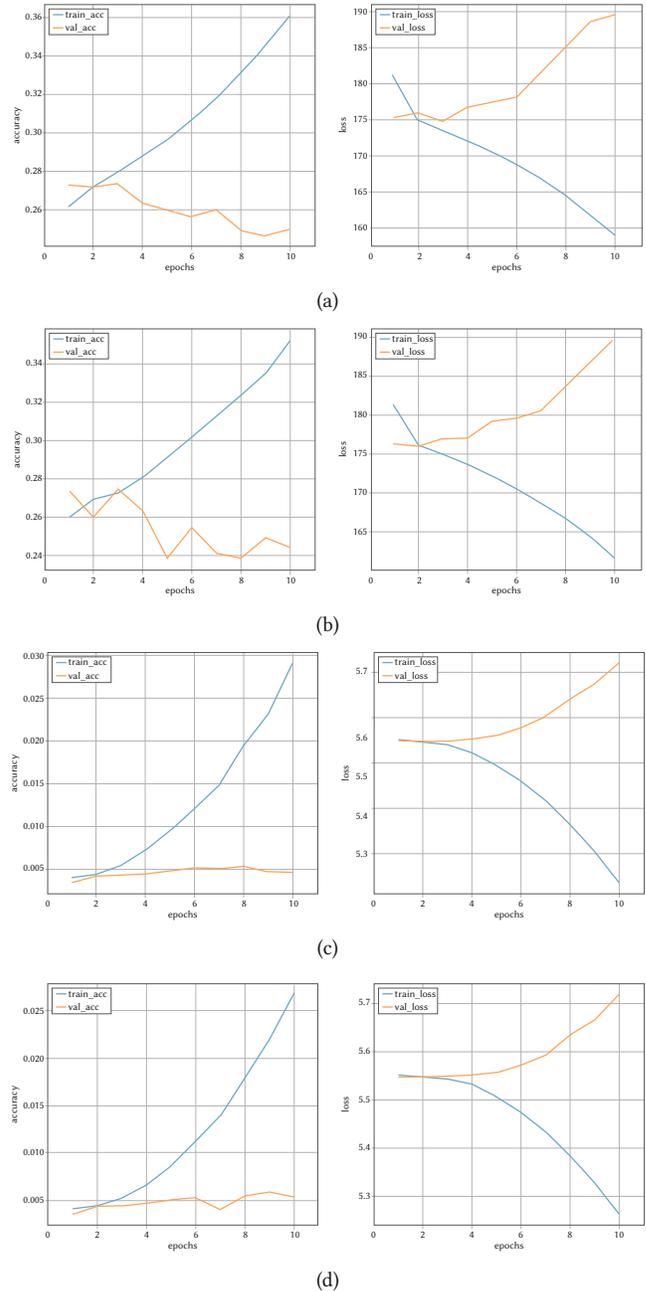

Fig. 13. Training result of a_0: (a) Byte 3 (b) Byte 2 (c) Byte 1 (d) Byte 0.

(a)

(b)

(c)

(d)
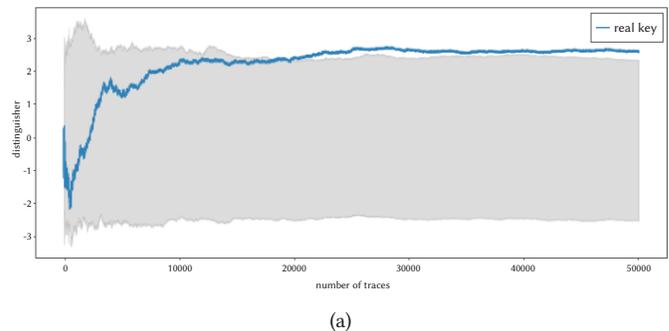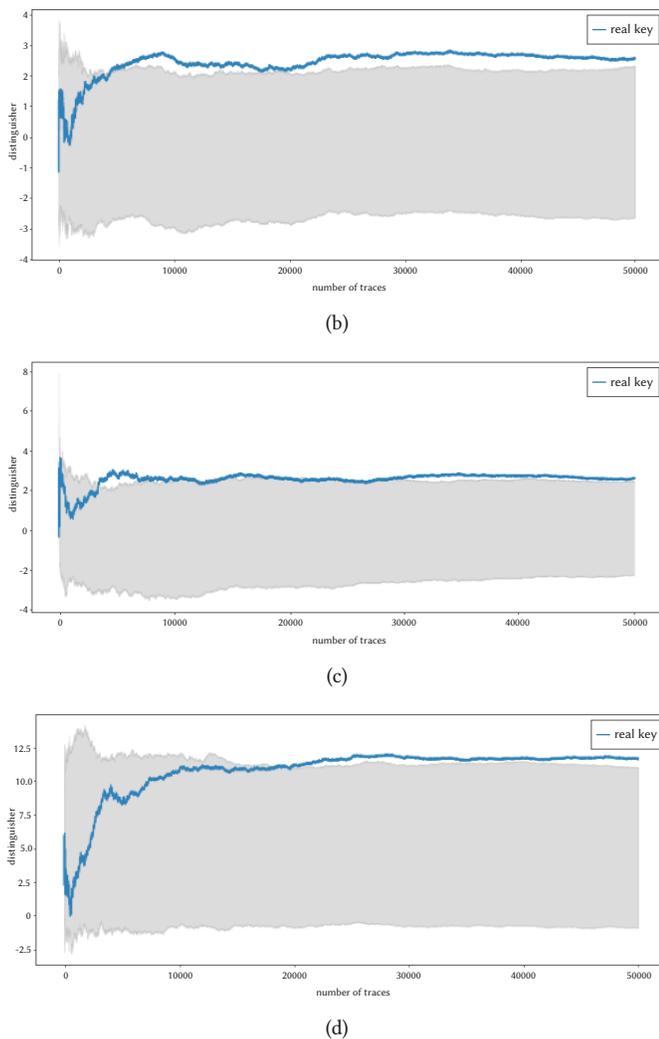
Fig. 14. Attack result of a_0: (a) Byte 3 (b) Byte 2 (c) Byte 1 (d) Byte 0.

We can repeat the attack on all the other parameters using the same set of data by changing the target label to recover b_0, c_0, $\delta_{2,0}$, e_0, f_0 and g_0. Then, d_0 and h_0 can be calculated simply.

IV. Conclusion and Future Work

In this paper, we demonstrate a Deep Learning-based Side Channel Attack on HMAC SM3 algorithm. In order to recover the key used in HMAC SM3, the attacker should recover two IVs: Hi and Ho. In this paper, we only focus on recovering Hi since the method of recovering the two IVs are the same. We try to recover $\delta_{1,0}$, $\delta_{2,0}$, a_0, b_0, c_0, d_0, e_0, f_0, g_0 and h_0 to recover H_i. The experiment result shows that we can recover the IV with 50,000 traces. In addition, we can find that when we focus on an add operation, the attack result is much better than focusing on a XOR operation. Thus, we need more traces to recover parameters when focusing on XOR operations. Although the correct candidate for XOR operation is not quite distinguishable from other candidates, we can recover the correct candidate. This situation may be solved if more traces are added to the attacking set.

In this paper, we focus on a software implementation of HMAC SM3 without any countermeasures. In future work, we can try several different HMAC implementations: (a) hardware implementation without countermeasures; (b) software implementation with some countermeasures; (c) hardware implementation with countermeasures. By doing experiments on different implementations, we can check whether deep learning works well on both unprotected and protected situations. In addition, we can try to figure out the difference of the structure of the network when attacking a hardware implementation and a software implementation.

References

[1] Kocher P, Jaffe J, Jun B. "Differential power analysis, advances in cryptology"-CRYPTO'99, Proc.19th Annual International Cryptology Conf, pp. 388-397, 1999.

[2] Gandolfi K, Mourtel C, Olivier F. "Electromagnetic analysis: Concrete results", International workshop on cryptographic hardware and embedded systems. Springer, Berlin, Heidelberg, pp. 251-261, 2001.

[3] Brouchier J, Kean T, Marsh C, et al. "Temperature attacks." IEEE Security & Privacy, vol. 7, no. 2, pp. 79-82, 2009.

[4] Thompson, J.N. "Insect Diversity and the Trophic Structure of Communities". In: Ecological Entomology. New York. pp. 165-178, 1994.

[5] Robyns, Pieter, Peter Quax, and Wim Lamotte. "Improving CEMA using correlation optimization.", 2018.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems, 2012.

[7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv: pp. 1409-1556, 2014.

[8] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[9] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[10] Bartkewitz T, Lemke-Rust K. "Efficient template attacks based on probabilistic multi-class support vector machines", International Conference on Smart Card Research and Advanced Applications. Springer, Berlin, Heidelberg, pp. 263-276, 2015.

[11] Heuser A, Zohner M. "Intelligent machine homicide", International Workshop on Constructive Side-Channel Analysis and Secure Design. Springer, Berlin, Heidelberg, pp. 249-264, 2012.

[12] Prouff, Emmanuel, et al. "Study of Deep Learning Techniques for Side-Channel Analysis and Introduction to ASCAD Database." IACR Cryptology ePrint Archive, vol. 53, 2018.

[13] Martinasek, Zdenek, and Vaclav Zeman. "Innovative method of the power analysis." Radio engineering, vol. 22.2, pp. 586-594, 2013.

[14] Maghrebi, Houssem, Thibault Portigliatti, a nd Emmanuel Prouff. "Breaking cryptographic implementations using deep learning techniques." International Conference on Security, Privacy, and Applied Cryptography Engineering. Springer, Cham, 2016.

[15] Cagli, Eleonora, Cécile Dumas, and Emmanuel Prouff. "Convolutional neural networks with data augmentation against jitter-based countermeasures." International Conference on Cryptographic Hardware and Embedded Systems. Springer, Cham, 2017.

[16] Picek, Stjepan, et al. "The curse of class imbalance and conflicting metrics with machine learning for side-channel evaluations." IACR Transactions on Cryptographic Hardware and Embedded Systems, 2018.

[17] Robyns, Pieter, Peter Quax, and Wim Lamotte. "Improving CEMA using correlation optimization." IACR Transactions on Cryptographic Hardware and Embedded Systems, vol. 2019, no. 1, pp. 1-24, 2018. Doi: 10.13154/tches.v2019.i1.1-24

[18] China's Office of Security Commercial Code Administration: Specification of SM3 Cryptographic Hash Function, http://www.oscca.gov.cn/UpFile/20101222141857786. Pdf, 2010.

[19] Turner, James M. "The keyed-hash message authentication code (HMAC)." Federal Information Processing Standards Publication, vol.

198-1, 2008.

[20] Chari S, Rao J R, Rohatgi P. "Template attacks", International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, pp. 13-28, 2002.

[21] O'Shea K, Nash R. "An introduction to convolutional neural networks", arXiv preprint arXiv:1511.08458, 2015.

[22] Kocher P C. "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems", Annual International Cryptology Conference. Springer, Berlin, Heidelberg, pp. 104-113, 1996.

[23] Mayer-Sommer R. "Smartly analyzing the simplicity and the power of simple power analysis on smartcards", International Workshop on Cryptographic Hardware and Embedded Systems. Springer, Berlin, Heidelberg, pp. 78-92, 2002.

[24] Brier E, Clavier C, Olivier F. "Correlation power analysis with a leakage model", International workshop on cryptographic hardware and embedded systems. Springer, Berlin, Heidelberg, pp. 16-29, 2004.

Suying Wang

Suying Wang (1991-), She received her B.S. degree in Zhejiang University in 2014, and M.S. degree in Royal institute of Technology in 2016. She currently works at Open Security Research, Inc. Her main research interests include Side Channel Analysis and Deep Learning-based Side Channel Analysis. Email: suying.wang@osr-tech.com

Yong Xiao

Yong Xiao was born in Jingzhou, Hunan, China, in 1978. He received the Ph.D. degree from Wuhan University. He is currently a Senior Engineer. His research interests include intelligent electric power technology and measurement automation technology.

Shiqi Li

Shiqi Li (1985-), He received the B.S. and M.S. degrees in Electrical Engineering from Katholieke Universiteit Leuven, Belgium in 2007 and 2009 respectively. His main research interests include Side Channel Analysis, Fault Injection and Cache timing analysis on TEE systems. He currently works as a senior security researcher in the Security Laboratory from Open Security Research, Inc. in Shenzhen, China. Email: shiqi.li@osr-tech.com

Xin Jin
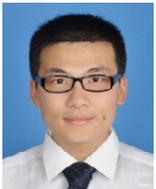
Xin Jin, He received the Master degree from the North China Electric Power University (NCEPU), Bao Ding, China, in 2011. He is currently a senior engineer of the China Southern Power Grid Research Institute and serves as the director of the Intelligent Measurement and New Technology Laboratory. His main research areas are power line carrier communication and wireless communication technology. Email: jinxin1@csg.cn

A Feature Extraction Method Based on Feature Fusion and its Application in the Text-Driven Failure Diagnosis Field

Shenghan Zhou, Bang Chen, Yue Zhang, HouXiang Liu, Yiyong Xiao*, Xing Pan

School of Reliability and Systems Engineering, Beihang University, Beijing 100191 (China)

uniR
LA UNIVERSIDAD
EN INTERNET

Abstract

As a basic task in NLP (Natural Language Processing), feature extraction directly determines the quality of text clustering and text classification. However, the commonly used TF-IDF (Term Frequency & Inverse Document Frequency) and LDA (Latent Dirichlet Allocation) text feature extraction methods have shortcomings in not considering the text's context and blindness to the topic of the corpus. This study builds a feature extraction algorithm and application scenarios in the field of failure diagnosis. A text-driven failure diagnosis model is designed to classify and automatically judge which failure mode the failure described in the text belongs to once a failure-description text is entered. To verify the effectiveness of the proposed feature extraction algorithm and failure diagnosis model, a long-term accumulated failure description text of an aircraft maintenance and support system was used as a subject to conduct an empirical study. The final experimental results also show that the proposed feature extraction method can effectively improve the effect of clustering, and the proposed failure diagnosis model achieves high accuracies and low false alarm rates.

Keywords

I. Introduction

IN the information age, data is generated all the time, especially text data. Because of its convenience, flexibility and universality, the quantity of text data is growing exponentially in, for example, applications such as twitter, online articles and shopping reviews. These words usually contain a lot of useful information, but also a lot of interference information. So, the text data needs to be processed to mine the useful information.

For structured data, scholars usually write corresponding programs directly to mine information with the help of computers. However, for unstructured data such as text, the computer cannot recognize it directly. Therefore, it is necessary for the computer to first understand the text before mining the text for information. The recognition process involves converting text symbols into numeric symbols that can be recognized by computers. Scholars have tried many methods to resolve issues around recognition, but the original rule-based methods obviously do not solve this problem very well [1]. However, inspired by neural network, many scholars put forward some effective new technologies [2], and text feature extraction is one of them.

Since computer cannot directly recognize unstructured data, the text needs to be transformed into a structured format for processing. At present, the VSM (Vector Space Model) [3] is widely used to make the transformation; but a document may contain thousands of words, which easily leads to dimension explosion and expensive calculations.

* Corresponding author.
E-mail address: xiaoyiyong@buaa.edu.cn

Therefore, feature extraction is usually used for further dimension reduction. Feature extraction can find the most representative text features that use low-dimensional feature vectors to represent text data based on the original text [4]. TF-IDF [5] and LAD [6], are two classical models for text feature extraction, and are easy to operate. However, TF-IDF only extracts keywords without considering the context. Although LDA considers the context, it tends to lose features and be ambiguous. Inspired by Zhao [7] et al. 's work of fusing χ^2 statistics-based feature and LDA semantic-based feature to improve the performance of a feature extraction, this paper proposes a text feature extraction method based on feature fusion, which combines the TF-IDF and LDA methods.

As NPL technology advanced, it was used in the field of text-driven failure diagnosis. Dnyanesh [8] et al. proposed a novel ontology-based text mining methodology to construct the D-matrices by automatically mining the unstructured repair verbatim data collected during failure diagnosis and used it to do failure pattern recognition. Rodrigues [9] et al. used text mining and neural networks to identify and classify aircraft failure patterns. However, most of the current text-based failure diagnosis models are supervised, which means these models apply to only text data with labels. However, in fact, the failure-description text is often unlabeled because of high labor costs. Therefore, based on the proposed feature extraction method, this paper developed a text-driven unsupervised failure diagnosis model. The extracted feature vectors are clustered to obtain the pseudolabel data, and then the pseudolabel data is used to train the classifier for feature diagnosis. This failure diagnosis model can classify and automatically judge which failure mode the failure described in the text belongs to once a failure-description text is entered.

To verify the effectiveness of the proposed feature extraction algorithm and failure diagnosis model, a long-term accumulated failure-description text of an aircraft maintenance and support system was used as a subject to conduct an empirical study.

The main contribution of this paper is to propose a more effective feature extraction method, by fusing TF-IDF and LDA, two typical feature extraction methods, and apply it to the field of failure diagnosis, by establishing an intelligent text-driven failure diagnosis model with the help of machine learning methods such as clustering and classification. The remaining of this paper is organized as follows: Section II presents an overview of text feature extraction and feature fusion. Section III is the introduction of the basic principle of the proposed feature extraction algorithm. Section IV presents the proposed text-driven failure diagnosis model. Section V is the experiment and the discussion of the experimental results. Section VI covers conclusions and discussion.

II. Related Work

A. Text Feature Extraction

As the basic work of text processing, text feature extraction has always been a hot research topic in NLP. So far, most of the existing feature extraction methods are based on the bag-of-words model [10], the topic model [11] and the word embedding model [12] - [13].

The bag-of-words model [10] adopts one-hot encoding to generate word vectors. Each word vector's dimension is equal to the size of the word vocabulary. In this vector, only one dimension's value is 1 and the rest are 0. Obviously, this kind of vector composed of 0 and 1, cannot represent a word accurately, because different words have a different importance to the text. Scholars usually use the TF-IDF method to assign weights to one-hot vectors. TF-IDF [5] is a widely used weighting technique, and plays an important role in the field of information retrieval. A TF-IDF is easy to carry out and usually performs well in short-text dataset, but performs badly in a long-text dataset or a class imbalance dataset.

A topic model [11] is a kind of topic generation models and a three-layer Bayesian probability model. The topic model's core idea is that a document selects a topic according to a certain probability, and a topic also selects a word according to a certain probability. LDA [6], as a classical topic model, is widely used in the task of text classification and text clustering. Compared with TF-IDF, LDA considers the context and performs much better in long-text dataset. However, LDA, as an unsupervised model, tends to lose features, is ambiguous in the process of feature extraction and performs badly in short-text datasets.

A word embedding model is designed to solve the dimension explosion problem of the bag-of-words model when processing long-text dataset. Through neural networks [13], word cooccurrence matrices, probabilistic models and other methods, a word embedding model maps the bag-of-words model's one-hot vector to a continuous vector space with a much lower dimension to enable a dimension reduction [12] - [13]. Word2vec is the most widely used word embedding model framework; it includes two word-vector-generation models, Skip-Gram and CBOW (Continuous Bag-of-Words Model). Skip-Gram and CBOW are three-layer neural networks with different inputs. Skip-Gram inputs the current word to predict the surrounding words, while CBOW inputs the surrounding words to predict the current word. Obviously, the two models take the context into account. However, the two models usually perform badly in short-text datasets [14].

B. Feature Fusion

Feature fusion originates from data fusions that were originally conducted in the military. In recent years, with the rapid development of AI, data fusion has been widely applied in intelligent medical [15], intelligent industry [16], intelligent transportation [17] and so on. Data fusion is a framework, which contains fusion modes and tools. Data fusion mainly uses different fusion modes and tools to combine different data sources, which may generate improved new data for certain application scenarios [18]. Whether the data after a fusion is effective or not mainly depends on the application scenario. In most cases, data fusion can effectively enhance the authenticity and availability of data [19], which is why data fusion is needed.

Feature fusion, as a technology of data fusion; uses given feature sets to generate new fusion features [19], and is very suitable for classification tasks. Liu [20] et al. fused two groups of feature vectors into a unit vector, and extracted features from high-dimensional vector space. They proposed a serial feature fusion algorithm and applied this algorithm to do face recognition. Their experiments showed that this algorithm could reach an accuracy rate of 98.5% with only 25 features. Yang [19] et al. proposed a new serial feature fusion algorithm for unstructured data, and tested it on the CENPARMI handwritten digital library, the NUST603 handwritten Chinese character library and the ORL face image library. The experimental results showed that their algorithm effectively improved the classification accuracy. Sun [21] et al. also proposed a new feature fusion algorithm based on CCA (Canonical Correlation Analysis), which performed well in small sample dataset with high dimensions. They first extracted two groups of feature vectors with the same pattern, then established correlation criterion functions between them, and finally extracted their representative features to form effective recognition vectors.

III. Text Feature Extraction Method Based on Feature Fusion

Text feature extraction is a process of text vectorization. As the first step of text processing, it directly determines the effects of the follow-up processes. However, text feature extraction is a very complex problem, because it involves conversion of abstract character symbols into concrete number symbols under the premise of maintaining the meaning of the original text. TF-IDF and LDA are two commonly used feature extraction methods proposed by scholars. The TF-IDF method is easy to execute and usually performs well in short-text datasets, but ignores the context. LDA considers the context and performs well in long-text dataset, but easily leads to blindness. Therefore, this paper proposes a text feature extraction method based on feature fusion, which combines the TF-IDF and LDA methods and is named TI-LDA.

A. TF-IDF Feature Extraction Method

The TF-IDF feature extraction method is actually the TF-IDF weighting of one-hot vector generated by a bag-of-words model. A One-hot vector whose dimension is equal to the size of the word vocabulary is discrete, and only one dimension's value is 1 and the rest are 0. The one-hot representation of a sentence can be obtained by adding the one-hot vector of all the words in the sentence. However, obviously, not every word is equally important to a sentence, so using TF-IDF to weight each word is necessary. The core idea of TF-IDF is the importance of a word is positively correlated with the frequency of its occurrence in a given text, and negatively correlated with the frequency of its occurrence in all texts of the corpus. The TF-IDF is composed of TF (Term Frequency) and IDF (Inverse Document Frequency). TF refers to the frequency of a word's occurrence in a given text, and its calculation formula is:

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

(1)

where $n_{i,j}$ is the number of times that word t_i appears in text d_j and

the denominator is the total number of times that all word appears in text d_j.

IDF is used to measure the frequency of a word's occurrence in all texts of the corpus, and its calculation formula is:

$$IDF_i = log_2 \frac{|D|}{|\{j : t_i \in d_j\}| + 1}$$ (2)

where $|D|$ is the total number of texts in the corpus and $|\{j : t_i \in d_j\}|$ is the number of texts containing t_i in the corpus.

The calculation formula of TF-IDF is:

$$TF_{i,j} IDF_i = TF_{i,j} \times IDF_i$$ (3)

The normalized formula of TF-IDF is:

$$TF_{i,j} IDF_i' = \frac{TF_{i,j} IDF_i}{\sqrt{\sum_{j=1}^{|d_j|} (TF_{i,j} IDF_i)^2}}$$ (4)

The TF-IDF algorithm is simple in principle, easy to operate and efficient to calculate, which make it suitable for short-text mining. However, it ignores the context to carry out vectorization and easily causes dimension explosion when dealing with long text.

B. LDA Feature Extraction Method

LDA is a statistical topic model which represents the topics of each document in the form of a probability distribution. An LDA believes that a document consists of several topics, and each topic consists of several words. When generating a document d with K topics, the probability of a word w being selected is:

$$p(w|d) = \sum_K p(w|t_k) \times p(t_k|d)$$ (5)

where K and k are the total number of topics and indexes of the group of topics and t_k stands for the topic k. For example, there are three topics: animals, actions and names. Different words are distributed under each topic. "cat", "dog" and "pig" belong to the topic of animals. "sitting", "running" and "standing" belong to the topic of actions. "Tony", "Jack" and "Lucy" belong to the topic of names. Suppose a sentence that says the dog is sitting need to be generated. The first step is to select the topics under the condition of the target semantics. The second step is to choose the words under the condition of the selected topics. In the end, the sentence "the dog is sitting" can be generated.

Therefore, the distribution of words in a given document can be obtained as Fig. 1.
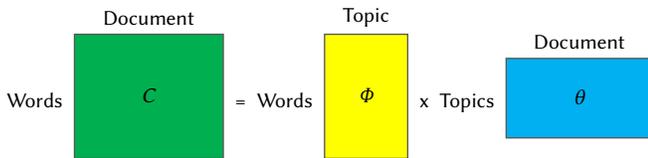

Fig. 1. LDA Model Diagram.

The process of generating a document by the LDA model can be summarized as detailed below:

- Step 1: Sample from the Dirichlet distribution α to generate the topic distribution θ_i of document i.;
- Step 2: Sample from the Multinomial distribution θ_i to obtain the topic $z_{i,j}$ of word j in document i.;
- Step 3: Sample from the Dirichlet distribution β to generate the topic distribution $\phi_{z_{i,j}}$ of topic $z_{i,j}$;
- Step 4: Sample from the Multinomial distribution $\phi_{z_{i,j}}$ to obtain word $\omega_{i,j}$.

Based on the above process, the following joint distribution can be obtained:

$$p(\omega_i, z_i, \theta_i, \Phi | \alpha, \beta) =$$
$$\prod_{j=1}^{N} p(\theta_i | \alpha) p(z_{i,j} | \theta_i) p(\Phi | \beta) p\left(\omega_{i,j} \middle| \theta_{z_{i,j}}\right)$$ (6)

By integrating θ_i and Φ, and summing z_i, the maximum likelihood estimation of the word distribution can be obtained as follows:

$$p(\omega_i | \alpha, \beta) = \int_{\theta_i} \int_{\Phi} \sum_{z_i} p(\omega_i, z_i, \theta_i, \Phi | \alpha, \beta)$$ (7)

Finally, parameters in an LDA model can be obtained by a Gibbs sampling [22].

The LDA model can effectively extract the semantic information of the text, in consideration of the context. However, in the selection of topic words, LDA has a certain blindness, that easily causes ambiguity and feature loss.

C. TI-LDA Feature Extraction Method

The feature fusion algorithm has been widely used in the field of AI for applications such as target tracking, pattern recognition and image understanding. In the field of pattern recognition, Jian Yang [19] et al. proposed two fusion strategies, parallel feature fusion and serial feature fusion. They also verified the robustness and practicability of the two fusion strategies through experimentation.

Suppose A and B are two different feature spaces of sample space Ω, meanwhile suppose $\alpha \in R^n$ and $\beta \in R^n$ are a feature vector of A and B, respectively. Then, the parallel feature fusion can be expressed as:

$$\gamma = \alpha + i\beta$$ (8)

where i is the imaginary component and $\gamma \in R^{\max(n,m)}$ is a feature vector of the new feature spaces. In the parallel feature fusion process, feature vectors α and β may have different dimensions; the feature vector with the lower dimension needs to be supplemented with 0 before fusion. Take $\alpha = (a_1, a_2, a_3)^T$ and $\beta = (b_1, b_2)^T$ for example. First, add 0 to supplement vector β to create a three-dimensional feature vector $(b_1, b_2, 0)^T$, then carry out the feature fusion according to equation (8), and the final result is $\gamma = (a_1 + ib_1, a_2 + ib_2, a_3 + i0)^T$.

Serial feature fusion can be expressed as:

$$\gamma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$ (9)

Compared with parallel feature fusion, serial feature fusion does not need to consider the vectors with different dimension. Also take $\alpha = (a_1, a_2, a_3)^T$ and $\beta = (b_1, b_2)^T$ for example; fuse directly and the final result is $\gamma = (a_1, a_2, a_3, b_1, b_2)^T$.

For a given text sample space Ω, suppose A and B is the feature vector space based on TF-IDF and LDA, meanwhile $\alpha_i \in R^n$ is a feature vector in A and $\beta_i \in R^m$ is a feature vector in B. Adopting the parallel feature fusion strategy, according to equations (8), the sample space Ω can be represented as:

$$\Omega = (\gamma_1, \gamma_2, \dots, \gamma_j)^T = \begin{vmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{j1} & \gamma_{j2} & \cdots & \gamma_{jk} \end{vmatrix} =$$
$$\begin{vmatrix} \alpha_{11} + i\beta_{11} & \alpha_{12} + \beta_{12} & \cdots & \alpha_{1k} + \beta_{1k} \\ \alpha_{21} + i\beta_{21} & \alpha_{22} + i\beta_{22} & \cdots & \alpha_{2k} + i\beta_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{j1} + i\beta_{j1} & \alpha_{j2} + i\beta_{j2} & \cdots & \alpha_{jk} + i\beta_{jk} \end{vmatrix}$$ (10)

While adopting the serial feature fusion strategy, according to equations (9), the sample space Ω can be represented as:

$$\Omega = (\gamma_1, \gamma_2, ..., \gamma_j)^T = \begin{vmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{j1} & \gamma_{j2} & \cdots & \gamma_{jk} \end{vmatrix} =$$

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{j1} & \alpha_{j2} & \cdots & \alpha_{jk} \\ \beta_{j1} & \beta_{j2} & \cdots & \beta_{jk} \end{vmatrix}$$

(11)

IV. Text-Driven Failure Diagnosis Model Based on TI-LDA

To make full use of failure-description text and understand the role the TI-LDA text feature extraction method plays in the field of fault diagnosis, this paper researched TI-LDA in the failure diagnosis field and designed a text-driven failure diagnosis model, that is suitable for small data samples and the main framework is as shown in Fig. 2. Before any further processing, the text data needs to be preprocessed. Specifically, for English text data, stop words need to be removed, while for Chinese text, word segmentation is also needed because there is no distinct identifier for separation. In addition, there are differences in word granularity, part of speech, polyphonic characters and so on between Chinese NLP and English NLP. Although the model is mainly for Chinese text, there is usually a mixed use of Chinese and English for failure-description text. For this situation, the model treats English words in the text as special Chinese characters. After the preprocessing, feature extraction is done to obtain text vectors, the obtained feature vectors are processed by CFSFDP (Clustering by Fast Search and Find of Density Peaks) clustering to mark the pseudolabels for failure text. The obtained pseudolabel data cannot be directly put into the classifier for training, because the class imbalance problem often exists in the failure text, which will affect the performance of the classifier. Therefore, this paper adopts the SMOTE (Synthetic Minority Oversampling Technique) oversampling method to balance the pseudolabel data. Finally, the balanced data is put into the SVM classifier to train the failure diagnosis model.
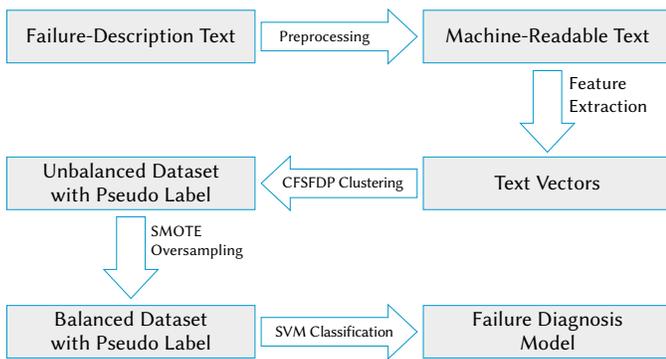

Fig. 2. Flow Chart of Aircraft Failure Diagnosis Model.

A. Text Preprocessing

As mentioned above, different preprocessing strategies should be adopted for English text and Chinese text. The English text should be processed by removing stop words directly, while the Chinese text should be segmented first and then the stop words should be removed. At present, the common Chinese word segmentation methods are mainly based on dictionaries, statistics and rules, and a dictionary-based method is the most effective and widely used. Common dictionary-based word segmentation systems include Jieba, the CAS (Chinese Academy of Sciences) segmentation system, Smallseg and Snailseg. Their functions are compared in Table I. It can be seen from Table I that Jieba is more powerful and suitable for the text data used in this article, so this paper adopted Jieba for word segmentation.

TABLE I. Comparison of Dictionary-Based Word Segmentation Systems

Segmentation system	Custom dictionary	Part-of-speech tagging	Keyword extraction
Jieba	√	√	√
CAS	√	√	×
Smallseg	√	×	×
Snailseg	×	×	×

There are often a large number of stop words in the text, such as emotional particles and punctuation marks, that have no contribution to the semantic expression. If text is directly used for subsequent processing, these stop words will inevitably cause too high a dimension for the text vector, increase the calculation cost, and interfere with text clustering. Therefore, these stop words need to be removed with the help of a stop words list. A simpler approach is to directly use the stop word list established by professional organizations, such as the NLTK (Natural Language Toolkit) English stop word list and the Baidu Chinese stop word list. Although this method is simple, it does not achieve the best effect. If you want to achieve the best effect, you need to establish a special stop table in accordance with the specific situation of the text data.

B. Feature Extraction

The text feature extraction function adopts the TI-LDA method proposed in this paper. In terms of the selection of fusion strategy, this paper selects the serial pattern feature fusion. By comparing equations (10) and (11), it can be found that the parallel feature fusion will continue to reduce the dimension. However, before the fusion of text features, the preprocessing and feature extraction will have already reduced the dimension of the text data. Obviously, more features will be lost if we use the parallel feature fusion. Therefore, this paper uses the serial feature fusion to do feature fusion.

C. Text Clustering Based on CFSFDP

Text clustering is a key element in the failure diagnosis model, and its main function is to mark the pseudolabels for the failure text. Therefore, choosing a clustering method suitable for the text data is very important. The current clustering algorithms can be broadly divided into partition-based methods [23], hierarchical-based methods [24] - [25], density-based methods [26], grid-based methods [27] and model-based methods. Because the failure-description text studied in this paper is typical of small data sampls, this paper uses CFSFDP to do the clustering. CFSFDP is a clustering method for small data sampls published in Science by Rodriguez [28] et al. Compared with the typical partition-based K-Means [29] method, CFSFDP not only handles clusters with aspherical shapes but also automatically determines the number of clusters. Compared with the typical density-based DBSCN (Density-Based Spatial Clustering of Applications with Noise) method, CFSFDP doesn't need to iterate repeatedly to determine the density threshold.

CFSFDP assumes that the center of the cluster is surrounded by some points with low local density, and these points are far away from other points with high local density. Therefore, the clustering centers can be obtained by calculating the nearest distance, and the remaining points can be divided into their categories according to their order of density. Suppose p_i and p_j are two different points of discrete data point

set D = $\{p_1, p_2, ..., p_n\}$, and define p_i's local density ρ_i as the number of points in the circle with p_i as the center and d_c as the radius. Then, ρ_i can be calculated by the following formulas:

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \tag{12}$$

where function:

$$\chi(x) = \begin{cases} 1, x < 0 \\ 0, x \geq 0 \end{cases} \tag{13}$$

Here, d_{ij} is the distance between p_i and p_j, and d_c is the cutoff distance, that needs to be determined manually.

Define the set of points with a higher density than p_i as $I_S^i = \{k \in I_S : \rho_k > \rho_i\}$, define the distance δ_i to be:

$$\delta_i = \begin{cases} \min_{j \in I_S^i} d_{ij} & I_S^i \neq \emptyset \\ \max_{j \in I_S} d_{ij} & I_S^i = \emptyset \end{cases} \tag{14}$$

When the data point p_i has the largest local density, $I_S^i = \emptyset$ and δ_i represents the maximum distance between p_i and p_j in the data set I_S, otherwise δ_i represents the minimum distance between p_i and p_j in the data set I_S^i.

To comprehensively measure the local density p_i and distance δ_i, another variable γ_i needs to be introduced. In addition, the calculation criteria of γ_i is:

$$\gamma_i = \rho_i \delta_i \tag{15}$$

The clustering centers can be selected according to the value of γ_i, because the clustering centers usually have a larger value of γ_i. If all values of γ_i are arranged in descending order and plotted on a two-dimensional plane, it can be found that the values of a in the nonclustering central interval are relatively smooth. If we arrange all γ_i in descending order and plot them in coordinates, you can see the value of γ_i is generally small and changes stably in the interval of non-nonclustering centers, while the value of γ_i is generally large in the interval of clustering centers and there is an obvious jump of γ_i's value near the critical point. Therefore, the number of clustering centers and classes can be determined based on the above characteristics.

D. Oversampling Based on SMOTE

In most of the failure monitoring data, including the failure-description text, there exists a class imbalance problem. The major class usually has more samples than the minor class. In fact, the major class occurs frequently but usually does less harm, while the minor class occurs occasionally but does great harm. This kind of unbalanced class data is a great challenge to classification. If the unbalanced data is directly used for classification, the minor-class samples will be submerged in the major class samples, which often results in high false alarm rates for the major-class and high missing alarm rates for the minor-class [30].

At present, there are mainly two ways of data balancing processing of oversampling. One is directly copying the samples of the minor class, the other is artificially generating the samples of the minor class according to the minor class's characteristics. The former is easy to do but easily causes overfitting, while the latter is more complex but difficult to overfit. This paper adopts the SMOTE [31] oversampling method, which is based on the latter, because the sample size of the data used in this paper is small.

Based on the above considerations, the SMOTE [31] algorithm, a widely used and relatively mature oversampling method, was adopted in this paper. The basic function of SMOTE is to manually add the minor-class samples to the new sample set by analyzing the characteristics of the minor-class. The SMOTE processes to solve a class imbalanced problem are as follows:

Define a sample set $X = \{x_i | i = 1,2,3, ..., m\}$ of a minor class. For any point in X, calculate the Euclidean distance between this point and all remaining points to obtain the k nearest points. Here, this paper assumes the multiplier of oversampling as n, that is randomly selecting n points in the k nearest points to generate set $Y = \{\hat{x}_j | j = 1,2,3, ..., n\}$. By a random linear interpolation, add the new sample to X, which is shown in the following formula:

$$x_{new}(i,j) = \{x_i + rand(0,1)(\hat{x}_j - x_i) | i = 1,2,3, ..., m ; j = 1,2,3, ..., n\} \tag{16}$$

where $rand(0,1)$ is a random number between 0 and 1. The above formula can generate m samples of the minor class to achieve the purpose of balancing the data set.

E. Classification Based on SVM

The pseudo-label data after balanced processing, needs to be put into the classifier for training. To do the selection of the classifier, the SVM classifier is selected in this paper. SVM, as a supervised learning method suitable for data with small sample sizes, was first proposed by Vapnik et al. [32] Due to its characteristics of easy operation and high robustness, SVM has been widely used in the field of feature diagnosis, and this paper also uses SVM to do classification. SVM is mainly based on statistics. SVM first maps the input data from the low-dimensional space to the high-dimensional space to make the problem linearly separable, then finds an optimal hyperplane in the high-dimensional space to divide the data. Therefore, the selection of the optimal hyperplane directly determines the classification effect.
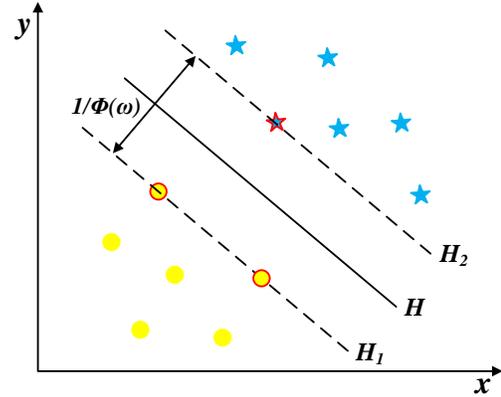
Fig. 3. Linear Binary Classification Diagram.

Because the linear binary classification of SVM is the basis and prototype of SVM. This process first consider the linear binary classification problem, as shown in Fig. 3. Define the sample points (x_i, y_i), $i = 1, 2, ..., s$, $x_i \in R^m$, $y_i \in \{1, -1\}$. Based on the above conditions, a classification hyperplane H is constructed:

$$wx + b = 0 \tag{17}$$

H can divide the above sample points into two classes, and the formula is:

$$\begin{cases} wx_i + b \geq 0 & y_i = 1 \\ wx_i + b \leq 0 & y_i = -1 \end{cases} \tag{18}$$

H can separate two different classes of samples, but the goal of SVM is to find the optimal hyperplane, that is, the maximum distance between the two classes of samples. Therefore, the objective function is:

$$min\Phi(w) = ||w||^2/2 \tag{19}$$

The constraint is:

$$y(wx_i + b) - 1 \geq 0 \quad i = 1, 2, \ldots, l \tag{20}$$

To find the optimal solution of equation (19), the Lagrange multiplier is introduced, and equation (19) can be linearized into

$$\min_a W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j < x_i \bullet x_j > \tag{21}$$

where $a_i \geq 0$ and $i = 1, 2, \ldots, l$.

The constraint is updated to:

$$\sum_{i=1}^{l} y_i a_i = 0, a_i \geq 0 \tag{22}$$

When $a_i \geq 0$, these sample points are referred to as support vectors. The optimal classification discriminant function is:

$$f(x) = sgn(\sum y_i a_i (x \bullet x_i) + b) \tag{23}$$

So far, the linear binary classification problem is solved. In addition, the nonlinear binary classification problem can be converted to a linear binary classification problem by kernel functions. The objective function of the nonlinear binary classification problem is:

$$\min_a W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j K < x_i \bullet x_j > \tag{24}$$

The optimal classification discriminant function of the nonlinear binary classification problem is:

$$f(x) = sgn(\sum y_i a_i K(x \bullet x_i) + b) \tag{25}$$

Here, $K(x \bullet x_i)$ is the kernel. The kernel function needs to be selected artificially, and the Gauss kernel function is used in this paper:

$$K(||x_i - x_c||) = exp(\frac{-||x_i - x_c||^2}{2\sigma^2}) \tag{26}$$

where x_c is the center of the kernel function, and σ is the width parameter of the kernel function.

So far, the binary classification problem has been solved. For the multiclassification problem, this paper builds an SVM multiclassification framework based on a binary tree, which is shown in Fig. 4.
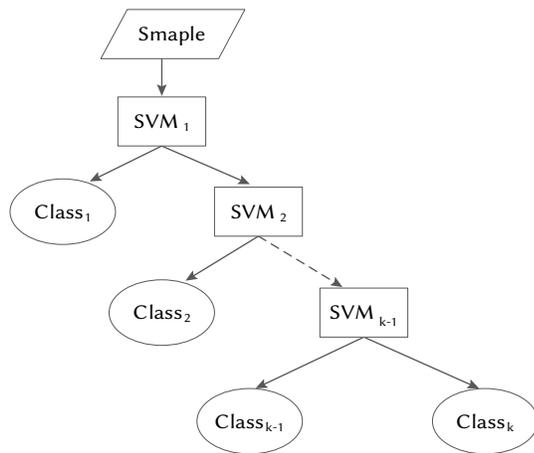
Fig. 4. The SVM Multiclassification Framework.

V. Experiments and Result Analysis

To verify the effectiveness of the proposed TI-LDA feature extraction algorithm and the text-driven failure diagnosis model, this paper used the failure-description Chinese text accumulated and recorded by an aircraft maintenance and support system to design verification experiments and analyzes the experimental results in detail. After eliminating the repeated and missing data, 1683 effective failure-description texts are obtained. In addition, some failure-description texts are shown in Table II.

TABLE II. Examples of Aircraft Failure-Description Text

Text number	Content
1	电机不工作，离合器坏 Motor didn't work, and clutch was damaged
2	渗油，密封圈损坏 Oil leakage, and seal ring was damaged
3	漏油致起落架放不下，密封圈损坏 Oil leakage caused undercarriage unable to be put down, and seal ring was damaged
…	…
1682	四台发动机尾喷口堵盖过大 The block cover of four engines' jet nozzle was too large
1683	信号灯故障 Signal lamp had a breakdown

To obtain better effectiveness for removing the stop words, this paper designed a special stop words list according to the characteristics of the corpus and the existing stop word list. Through the analysis of the text used in this paper, it is easy to find that two-word stop words are the most common, followed by professional characters, letters, and three-word stop words. Therefore, based on the above characteristics and the commonly used stop words lists, this paper designed a special stop word list to removing stop words. Some of the stop words are shown in Table III.

TABLE III. Stop Words List

Number	Stop word	Number	Stop word
1	———	8	]
2	》），	9	?
3	）÷（１ -	10	。
4	"，	11	竟然 (Actually)
5	℃	12	看 (Look)
6	[⑤]］	13	快 (Almost)
7	[	…	…

A. Effectiveness Verification Experiment of TI-LDA

To verify the effectiveness of TI-LDA, this paper first used TF-IDF, LDA and TI-LDA to extract the feature vectors of the failure-description texts respectively, then clustered the three sets of feature vectors using CFSFDP, and finally evaluated the effectiveness of the feature extraction by comparing the effects of clustering.

This paper used TF-IDF, LDA and TI-LDA to extract the features of the preprocessed text, and the normalized feature vectors obtained are shown in Table IV, Table V and Table VI.

TABLE IV. Normalized Feature Vectors of TF-IDF

Number	Dimension					
	1	2	3	…	2310	2311
1	0.662	0.350	0.000	…	0.000	0.000
2	0.000	0.350	0.270	…	0.000	0.000
3	0.662	0.000	0.270	…	0.000	0.000
…	…	…	…	…	…	…
1682	0.000	0.000	0.000	…	0.781	0.000
1683	0.000	0.000	0.000	…	0.000	0.599

TABLE V. Normalized Feature Vectors of LDA

Number	Dimension					
	1	2	3	...	2310	2311
1	0.726	0.056	0.056	...	0.000	0.000
2	0.042	0.042	0.043	...	0.000	0.000
3	0.000	0.971	0.000	...	0.000	0.000
...	...	...	...	...	...	...
1682	0.913	0.019	0.420	...	0.000	0.000
1683	0.913	0.018	0.288	...	0.000	0.000

TABLE VI. Normalized Feature Vectors of TI-LDA

Number	Dimension					
	1	2	3	...	2310	2311
1	0.662	0.350	0.000	...	0.000	0.000
	0.726	0.056	0.056	...	0.000	0.000
2	0.000	0.350	0.270	...	0.000	0.000
	0.042	0.042	0.043	...	0.000	0.000
3	0.662	0.000	0.270	...	0.000	0.000
	0.000	0.971	0.000	...	0.000	0.000
...	...	...	...	...	...	...
	...	...	...	...	...	...
1682	0.000	0.000	0.000	...	0.781	0.000
	0.913	0.019	0.420	...	0.000	0.000
1683	0.000	0.000	0.000	...	0.000	0.599
	0.913	0.018	0.288	...	0.000	0.000

TABLE VII. Relative Distance Values of TF-LDA

Number	Distance	Number	Distance
R_{11}	0.000	R_{598749}	2.796
R_{12}	1.883	R_{598750}	2.117
R_{13}	1.883	R_{598751}	2.362
R_{14}	1.883	R_{598752}	3.555
R_{15}	1.866	R_{598753}	3.709
...	...	...	...
R_{598743}	2.383	$R_{16831678}$	1.270
R_{598744}	3.593	$R_{16831679}$	1.247
R_{598746}	2.462	$R_{16831681}$	1.558
R_{598747}	2.283	$R_{16831682}$	1.983
R_{598748}	2.192	$R_{16831683}$	0.000

TABLE VIII. Relative Distance Values of LDA

Number	Distance	Number	Distance
R_{11}	0.000	R_{598749}	0.488
R_{12}	1.161	R_{598750}	0.393
R_{13}	1.161	R_{598751}	0.348
R_{14}	1.161	R_{598752}	1.176
R_{15}	1.067	R_{598753}	1.094
...	...	...	...
R_{598744}	1.128	$R_{16831679}$	0.956
R_{598745}	1.094	$R_{16831680}$	0.723
R_{598746}	0.553	$R_{16831681}$	0.219
R_{598747}	0.925	$R_{16831682}$	0.983
R_{598748}	2.209	$R_{16831683}$	0.000

For the three sets of vectors, this paper used the CFSFDP method for text clustering. According to the principle of CFSFDP, the relative distance values between the vectors was first calculated, and the results are shown in Table VII, Table VIII and Table IX. Then, we calculated the values of γ_i and drew them in descending order on a two-dimensional plane, as shown in Fig. 5, Fig. 6, and Fig. 7. Finally, we determined the number of clustering centers and classes based on the numerical variation diagram of γ_i. It's easy to see from Fig. 5, Fig. 6 and Fig. 7 that although the methods of feature extraction are different, the number of classes are the same.

TABLE IX. Relative Distance Values of TI-LDA

| Number	Distance	Number	Distance
R_{11}	0.000	R_{598749}	4.087
R_{12}	2.637	R_{598750}	2.727
R_{13}	2.637	R_{598751}	3.598
R_{14}	2.637	R_{598752}	4.864
R_{15}	2.620	R_{598753}	3.709
...	...	...	...
R_{598744}	3.587	$R_{16831679}$	2.009
R_{598745}	3.698	$R_{16831680}$	1.738
R_{598746}	3.536	$R_{16831681}$	2.474
R_{598747}	2.292	$R_{16831682}$	2.968
R_{598748}	2.209	$R_{16831683}$	0.000
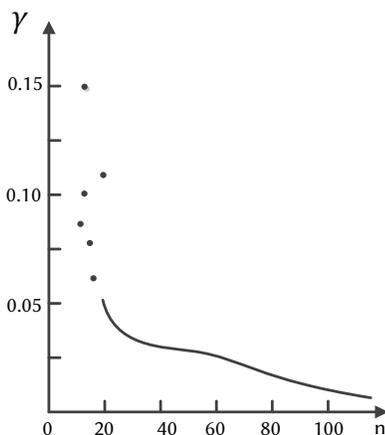
Fig. 5. γ_i Value Variation Diagram of TF-IDF.
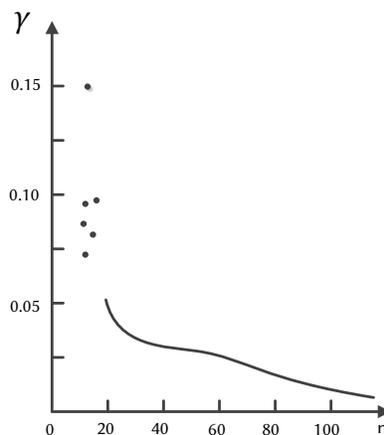
Fig. 6. γ_i Value Variation Diagram of LDA.
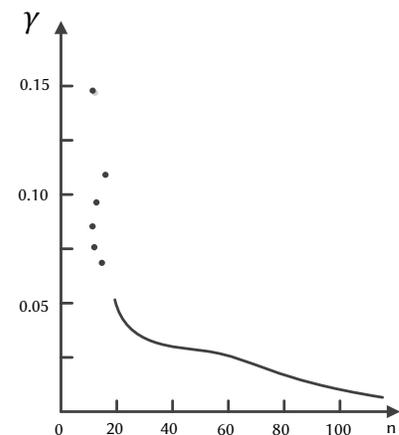
Fig. 7. γ_i Value Variation Diagram of TI-LDA.

To evaluate the clustering effect, this paper looks at two aspects, the intraclass compactness and the interclass separability, and uses the average intraclass compactness ($\overline{CP}$) and the average interclass separability ($\overline{SP}$) indicators to do the evaluation. The smaller the $\overline{CP}$'s value, the higher the compactness of the entire data set; the larger the $\overline{SP}$'s value, the higher the separability of the entire dataset is. To reflect the clustering effect on the entire dataset, this paper defines a comprehensive evaluation indicator $\overline{CS} = \overline{SP}/\overline{CP}$. It's easy to observe that the larger $\overline{CS}$'s value, the better the clustering comprehensive effect. The specific results are shown in Table X.

TABLE X. Comparison TABLE of Clustering Indicators

Indicator	Method		
	TF-IDF	LDA	TI-LDA
$\overline{CP}$	0.0014	0.0003	0.0002
$\overline{SP}$	1.6085	0.8628	1.0214
$\overline{CS}$	1148.9	2876.0	5107.0

In terms of intraclass compactness, based on Table X, because TI-LDA's $\overline{CP}$ value is the smallest, TI-LDA has the highest intraclass compactness and makes obvious improvements compared with TF-IDF. In terms of interclass separability, TF-IDF performs best because of the higher dimension of the feature vectors; it is followed in performance by TI-LDA and LDA. For overall performances, TI-LDA gains the highest marks and is far ahead of TF-IDF and LDA. Altogether, the TI-LDA method proposed in this paper, effectively improves the clustering effect.

B. Effectiveness Verification Experiment of Text-Driven Failure Diagnosis Model

Because the TI-LDA method proposed in this paper is better than TF-IDF and LDA based on Table X, the subsequent processing of the confirmatory experiments is based on the clustering results of TI-LDA. Through CFSFDP text clustering, the text data in this paper was divided into six completely different failure types. The first failure type is a transmitter failure; the second is a signal failure, which mainly is a signal problem of different monitors; the third is the failure of the aircraft's flight parameter indicators; the fourth is a generator failure, which is mainly caused by a generator overload and a signal failure; the fifth is engine failure; and the last is the failure caused by mechanical fatigue. The details are shown in Table XI.

Fig. 8. Sample Number of Each Failure Type Under Original Data.

Fig. 9. Sample Number of Each Failure Type Under Oversampling Data.

In this paper, a number of samples of each failure type were calculated, as shown in Fig. 8. It can be seen from Fig. 8 that the text data presents an obvious class imbalance, with a large quantity gap between the major class and the minor class. In addition, the second type of failure and the sixth type of failure accounts for more than 90% of the failures, while the remaining four types accounts for a relatively small percent. In this paper, SMOTE is mainly used for data equalization to solve the class-unbalanced problem and to sample a number of each failure type after oversampling, as shown in Fig. 9.

TABLE XI. Clustering Results of TI-LDA

| Class number	Clustering center	Text content of clustering center	Failure class
1	210	起飞后不能加高压，发射机故障 High voltage cannot be added after take-off and transmitter was damaged	Transmitter
2	365	电台不能发射，收不到信号，射频的K14a、b开关片地线断，23MZ的本振的L3电感器未并联上，16HZ频率调偏 The radio cannot transmit and the signal cannot be received. The ground wire of the switch pieces K14a and K14b of rf was broken. The L3 inductor of the 23MZ local vibration was not connected, and the frequency was deviated	Signal
3	773	校"I"状态记忆灯亮，地速指示极小（0021）且不动，低频分机故障 Status indicator light of "I" was on. The reading of the ground speed indicator was minimal (0021) and fixed. The low frequency extension broken down	Indicator
4	1416	1发起动发电机起动超负荷信号灯亮，减速器轴承漏光 The starting generator of the no. 1 engine was overloaded and the signal light was on. The reducer bearing was lightly leaking	Generator
5	1509	发动机停车后余油管大量漏油 Lots of oil leaked from the residual oil pipe after the engine stopped	Engine
6	1679	在"陆"位置加不上高压 High pressure cannot be applied at position "Land"	Mechanical fatigue

Comparing Fig. 8 and Fig. 9, the class imbalanced problem has been significantly improved and there is no significant difference in the number of samples between different classes after oversampling.

This paper used the original data and the oversampling data to train the SVM classifiers. By comparing the classification effect of the two classifiers, the effectiveness of the proposed text-driven failure diagnosis model was verified. Usually, the classification effect of classifier is mainly judged by the classification accuracy. However, for the text data with a class imbalance problem in this paper, the classification accuracy is not comprehensive. Therefore, this paper decided to use the confusion matrix from *FNR* (False Negative Rate), *FPR* (False Positive Rate), *Acc* (Accuracy), *Recall* and F_1 to evaluate the classification effect. The specific results are shown in Table XII.

TABLE XII. Comparison TABLE of Evaluation Indicators of SVM Classifier

Indicator	Original data	Oversampling data
FNR	0.149	0.021
FPR	0.015	0.008
Acc	0.982	0.989
Recall	0.851	0.979
F_1	0.721	0.968

From Table XII, it can be shown that the classifier trained by the original data is very close to the classifier trained by the oversampling data in *Acc*, and both have high accuracy, which further reflects the validity and feasibility of the proposed failure diagnosis model from a data perspective. However, for other classification indicators, the failure diagnosis model trained by the oversampling data has obvious improvements in *Recall* and F_1, which reflects that there exist false high accuracies of the classifier trained by the original data. According to the characteristics of the data in this paper, the major class usually has many more samples than the minor class. Therefore, classifiers tend to group minor-class samples into the major class, which often results in high false alarm rates of the major class and high missing alarm rates of the minor class. This can also be seen from the *FNR* and *FPR*. The *FNR* and *FPR* of the classifier trained by the oversampling data are much smaller, so there are fewer mistakes in the classification task, which effectively avoids the high false alarm rates.

VI. Conclusion and Discussion

In a variation from the traditional methods of failure diagnoses based on structured data, this paper proposes a text-driven failure diagnosis model by using NLP technology, which fills in a gap for research on failure diagnoses based on unstructured data, especially text data.

To resolve the shortcomings of traditional TF-IDF and LDA text feature extraction methods, this paper proposes TI-LDA, a new text feature extraction method, based on serial feature fusion, and uses the CFSFDP clustering method to verify the effectiveness of TI-LDA. The final experimental results show that the feature vectors extracted by TI-LDA can effectively improve intraclass compactness and interclass separability, compared with the methods using TF-IDF and LDA alone.

In this paper, the TI-LDA method is applied to the field of failure diagnosis, and a text-driven failure diagnosis model based on TI-LDA is created by combining the machine learning methods such as CFSFDP clustering, SMOTE oversampling and SVM classification. This failure diagnosis model can classify and automatically judge which failure mode the failure described in the text belongs to once a failure-description text is entered. Through an effectiveness verification experiment, it was found that the failure diagnosis model proposed in this paper has a high accuracy, and effectively solves the problem of high false accuracies and high false alarm rate caused by the class imbalance problem.

It's worth mentioning that this text-driven failure diagnosis model is unsupervised, which means it does not need any label data, so it has better portability and lower labor costs. This model has very broad application prospects, especially in the failure diagnosis field for large and complex equipment such as for aircraft, high-speed rail, and even for the medical field.

However, the model can still be improved. For example, the TI-LDA text feature extraction method fuses according to a ratio of 1:1, but this ratio may be not the best fusion ratio. So, our next work will study how to find the best fusion ratio to obtain the best fusion effect. In addition, in order to simplify the study, this paper does not consider the interference of abnormal or unreal text data. Future work will focus on ways to identify abnormal text data so as improve the accuracy of the model.

References

[1] Cambria, E. & White, B., "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]", Computational Intelligence Magazine IEEE, vol. 9, no. 2, pp. 48-57, 2014.

[2] Tom, Y., Devamanyu, H., Soujanya, P. & Erik, C., "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]", IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55-75, 2018.

[3] Salton, G., Wong, A. & Yang, C., "A vector space model for automatic indexing", Communications of the ACM, vol. 18, no. 11, pp. 613-620, 1975.

[4] Liang, H., Sun, X., Sun, Y. & Gao, Y., "Text feature extraction based on deep learning: a review", EURASIP journal on wireless communications and networking, vol. 2017, no. 1, pp. 1-12, 2017.

[5] Sparck Jones, K., "A statistical interpretation of term specificity and its application in retrieval", Journal of documentation, vol. 28, no. 1, pp. 11-21, 1972.

[6] Blei, D. M., Ng, A. Y. & Jordan, M. I., "Latent dirichlet allocation", Journal of machine Learning research, vol. 3, no. Jan, pp. 993-1022, 2003.

[7] Wang, F., Xu, T., Tang, T., Zhou, M. & Wang, H., "Bilevel feature extraction-based text mining for fault diagnosis of railway systems", IEEE transactions on intelligent transportation systems, vol. 18, no. 1, pp. 49-58, 2016.

[8] Rajpathak, D. G. & Singh, S., "An Ontology-Based Text Mining Method to Develop D-Matrix From Unstructured Text", IEEE Transactions on Systems Man & Cybernetics Systems, vol. 44, no. 7, pp. 966-977, 2014.

[9] Rodrigues, R. E. R. S., Balestrassi, P. P., Paiva, A. P., Garcia-Diaz, A. & Pontes, F. J., "Aircraft interior failure pattern recognition utilizing text mining and neural networks", Journal of Intelligent Information Systems, vol. 38, no. 3, pp. 741-766, 2012.

[10] Harris, Z. S., "Distributional structure", Word, vol. 10, no. 2-3, pp. 146-162, 1954.

[11] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R., "Indexing by latent semantic analysis", Journal of the American society for information science, vol. 41, no. 6, pp. 391-407, 1990.

[12] Roweis, S. T. & Saul, L. K., "Nonlinear dimensionality reduction by locally linear embedding", science, vol. 290, no. 5500, pp. 2323-2326, 2000.

[13] Bengio, Y., Ducharme, R. E. J., Vincent, P. & Jauvin, C., "A neural probabilistic language model", Journal of machine learning research, vol. 3, no. Feb, pp. 1137-1155, 2003.

[14] Ming, T., Lei, Z., Xianchun, Z. & Others, "Document vector representation based on Word2Vec", Computer Science, vol. 43, no. 6, pp. 214-217, 2016.

[15] Adali, T. U. L., Levin-Schwartz, Y. & Calhoun, V. D., "Multimodal data fusion using source separation: Application to medical imaging", Proceedings of the IEEE, vol. 103, no. 9, pp. 1494-1506, 2015.

[16] Jing, L., Wang, T., Zhao, M. & Wang, P., "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox", Sensors, vol. 17, no. 2, pp. 414, 2017.

[17] Wang, S., Deng, Z. & Yin, G., "An accurate GPS-IMU/DR data fusion method for driverless car based on a set of predictive models and grid constraints", Sensors, vol. 16, no. 3, pp. 280, 2016.

[18] Khaleghi, B., Khamis, A., Karray, F. O. & Razavi, S. N., "Multisensor data fusion: A review of the state-of-the-art", Information fusion, vol. 14, no. 1, pp. 28-44, 2013.

[19] Yang, J., Yang, J., Zhang, D. & Lu, J., "Feature fusion: parallel strategy vs. serial strategy", Pattern recognition, vol. 36, no. 6, pp. 1369-1381, 2003.

[20] Liu, C. & Wechsler, H., "A shape-and texture-based enhanced Fisher classifier for face recognition", IEEE transactions on image processing, vol. 10, no. 4, pp. 598-608, 2001.

[21] Sun, Q., Zeng, S., Liu, Y., Heng, P. & Xia, D., "A new method of feature fusion and its application in image recognition", Pattern Recognition, vol. 38, no. 12, pp. 2437-2448, 2005.

[22] Geman, S. & Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE Transactions on pattern analysis and machine intelligence, no. 6, pp. 721-741, 1984.

[23] Visser, E., Nijhuis, E. H., Buitelaar, J. K. & Zwiers, M. P., "Partition-based mass clustering of tractography streamlines", NeuroImage, vol. 54, no. 1, pp. 303-312, 2011.

[24] Bouguettaya, A., Yu, Q., Liu, X., Zhou, X. & Song, A., "Efficient agglomerative hierarchical clustering", Expert Systems with Applications, vol. 42, no. 5, pp. 2785-2797, 2015.

[25] Tang, X. & Zhu, P., "Hierarchical clustering problems and analysis of fuzzy proximity relation on granular space", IEEE Transactions on Fuzzy Systems, vol. 21, no. 5, pp. 814-824, 2012.

[26] Lu, J. & Zhu, Q., "An effective algorithm based on density clustering framework", Ieee Access, vol. 5,4991-5000, 2017.

[27] Tu, L. & Chen, Y., "Stream data clustering based on grid density and attraction", ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 3, no. 3, pp. 1-27, 2009.

[28] Rodriguez, A. & Laio, A., "Clustering by fast search and find of density peaks", Science, vol. 344, no. 6191, pp. 1492-1496, 2014.

[29] Jain, A. K., "Data clustering: 50 years beyond K-means", Pattern recognition letters, vol. 31, no. 8, pp. 651-666, 2010.

[30] Japkowicz, N. & Stephen, S., "The class imbalance problem: A systematic study", Intelligent data analysis, vol. 6, no. 5, pp. 429-449, 2002.

[31] Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P., "SMOTE: synthetic minority over-sampling technique", Journal of artificial intelligence research, vol. 16,321-357, 2002.

[32] Cortes, C. & Vapnik, V., "Support-vector networks", Machine learning, vol. 20, no. 3, pp. 273-297, 1995.

Shenghan ZHOU

Shenghan ZHOU, Doctor of management science, Associate Professor, graduated from the Beihang University in 2009, working in Beihang University. His research interests include system safety, data mining, and risk management.

Bang CHEN

Bang CHEN was born in Xinyang, Henan, China in 1995. He received the B.S. degree in safety engineering from Northeastern University Shenyang, Liaoning, China, in 2019. He is currently pursuing the M.S. degree in safety science and engineering at Beihang University, Beijing China. His research areas are data mining and nature language processing.

Yue ZHANG

Yue ZHANG was born in Shanxi Province, China. 1997. She received the B.S. degree in engineering management from Nanjing Agricultural University, Nanjing, China, in 2019. She is currently pursuing the M.S. degree in safety science and engineering at Beihang University, Beijing, China. Her research areas are data mining, factory layout and path optimization.

HouXiang LIU

HouXiang LIU was born in Anhui Province, China. 1997. He received the B.S. degree in safety engineering from Northeastern University, Shenyang, China in 2019. He is currently pursuing the M.S. degree in industrial engineering at Beihang University, Beijing, China. His research areas are data analysis and traffic optimization.

Yiyong XIAO

Yiyong XIAO, Doctor of System Engineering, Associate Professor, graduated from the Beihang University in 2003. Working in Beihang University. His research interests include economic affordability, data mining, network security optimization and algorithm research.

Xing PAN

Xing PAN graduated the doctor degree in management science from the Beihang University, in 2005, where he is currently a Senior Engineer. His research interests include systems engineering, data mining, and risk assessment.

Deraining and Desnowing Algorithm on Adaptive Tolerance and Dual-tree Complex Wavelet Fusion

Jingfeng Zang*, Ningxue Xu, Rui Liu, Yuhuan Shi

Changchun University of Science and Technology, College of Electronic and Information Engineering, No. 7089 Weixing Road, Changchun, 130022 (China)

uniR

LA UNIVERSIDAD
EN INTERNET

Abstract

Severe weather conditions such as rain and snow often reduce the visual perception quality of the video image system, the traditional methods of deraining and desnowing usually rarely consider adaptive parameters. In order to enhance the effect of video deraining and desnowing, this paper proposes a video deraining and desnowing algorithm based on adaptive tolerance and dual-tree complex wavelet. This algorithm can be widely used in security surveillance, military defense, biological monitoring, remote sensing and other fields. First, this paper introduces the main work of the adaptive tolerance method for the video of dynamic scenes. Second, the algorithm of dual-tree complex wavelet fusion is analyzed and introduced. Using principal component analysis fusion rules to process low-frequency sub-bands, the fusion rule of local energy matching is used to process the high-frequency sub-bands. Finally, this paper used various rain and snow videos to verify the validity and superiority of image reconstruction. Experimental results show that the algorithm has achieved good results in improving the image clarity and restoring the image details obscured by raindrops and snows.

Keywords

I. Introduction

THE video image captured by the outdoor video image system in severe weather (rain and snow) is severely degraded. Generally, the color and contrast of captured images and videos will drop sharply, which greatly affects the normal operation of outdoor video image systems, such as security surveillance systems, video recognition, intelligent perception, environmental safety, and intelligent transportation. Because there is a sequence of feature information in the video image, the research on the video image rain and snow removal algorithm is the main research direction in this field. Video-based image rain and snow removal algorithms are roughly divided into several types of algorithms based on the time domain, space domain, and color characteristics. Among them, the algorithm based on the image space domain started earlier and the research is more in-depth.

For example, Hase [1] and others use the size and speed of moving objects to propose a new method that can remove snow noise from the continuous images recorded by the camera. Garg [2]-[3] and others conducted a comprehensive analysis of the visual impact of the rain imaging system for the first time, and proposed to analyze the motion and optical characteristics of rain and snow, calculated the dynamics and optical models of rain and snow to identify and process the frame difference method of rain and snow. Camera parameters were used to reduce the visibility of rain during image acquisition, but the removal effect of this algorithm is not ideal in the case of heavy rain and snow

scenes and when there are high-speed moving objects in degraded images. Zhang et al. [4] proposed a k-means clustering method for all the same coordinates pixels along the video timeline, which can detect and clear away raindrops, this method is proposed based on the combination of the time and chroma characteristics of raindrops in the video. The clustering method needs to be applied to the entire video sequence, and this method cannot be used online. Then through the improved frame difference method or clustering method they achieve image space to remove rain (snow). Brewer et al. [5] adopted a multi-step method to eliminate rainwater and similar effects in video clips, a new method of raindrop pixel replacement is constructed by detecting raindrops in five consecutive frames. Zhao Xudong et al. [6] proposed an improved histogram model for detecting and removing rainwater in video, using the brightness distribution range of raindrops to determine whether pixels are covered by raindrops. Barnum et al. [7]-[8] combined the fringe model and rain and snow statistical characteristics to create an overall effect model of dynamic weather in frequency space, suitable for video detection and rain and snow removal with scenes and camera motion, but the model construction process is more complicated. Bossu et al. [9] proposed to use the classic Gaussian mixture model to segregate the foreground model from the image sequence for detecting rain and snow. Based on the choice rules of luminosity and size, the geometric moment method is used to estimate the histogram of the rain and snow stripe direction, and the expectation-maximization algorithm is used to achieve the purpose of removing rainwater in the video image. Kang et al. [10]-[11] proposed for the first time that a single image can be denoised by a sparse coding algorithm and morphological component analysis model. Then, they used the time information of consecutive frames and reused the dictionary learned in the previous frame of the video, while

* Corresponding author.

E-mail address: zjf0319830@sohu.com

holding video time consistency, frame-by-frame image deraining, the rain removal based on the video is an extension of the rain removal method of a single image, but this method is only applicable to static scenes. Yu [12] proposed a video defogging technique based on adaptive tolerance, and combined it with the guiding filter to apply video defogging. Abdel-Hakim [13] used inter-frame information to recover low-rank components between adjacent frames to extract rain-free distortion signals, and used low-rank recovery methods to remove raindrops and snowflakes from the video frames. Kim et al. [14] proposed a new algorithm for removing rain and snow stripes through low-rank matrix video sequences, and extended this algorithm to three-dimensional stereo video cancellation, but when processing dynamic video images, the snow removal effect is not ideal. Ren et al. [15] proposed a video denoising and drainage model based on matrix decomposition, which solved the problem that the rain and snow removal method could not cope with heavy rain, blizzard and dynamic scenes. Zheng et al. [16] proposed a method of through multi-guide filter in a single image to remove raindrops and snowflakes, the low-frequency sub-band obtained by decomposition is a non-raindrop and snowflake image, and the high-frequency sub-band is input to the guide filter as an input image. Wang et al. [17] proposed an algorithm that uses wavelet fusion technology to achieve video rain removal, combined with wavelet multi-layer decomposition technology to determine the particular layer of rain noise. L. Dou et al. [20] improved the traditional dual-tree complex wavelet image restoration method and proposed a new algorithm. The image is decomposed into two parts of low-frequency and high-frequency using the complex wavelet transform, the high-frequency components of the image are repaired by the total change model, and the improved curvature-driven diffusion is used to repair the low-frequency components. J. Zang et al. [31] improved on the traditional dual-tree complex wavelet and proposed an improved method to eliminate raindrops, this method uses the fusion rule of principal component analysis for the low-frequency sub-bands obtained by the complex wavelet decomposition, and uses local energy matching for the high-frequency sub-bands fusion rules.

On this basis, aiming at the problem of distortion in bright areas of fusion images under more complicated rain and snow weather backgrounds, this paper proposes a method to eliminate video rain and snow by combining adaptive tolerance and dual-tree complex wavelet. The adjustment mechanism based on adaptive tolerance can not only solve the problem of distortion in bright areas, but also not cause transmittance. The error compensation is obtained, so that a higher quality fusion image can be restored. At the same time, the video image is decomposed by dual-tree complex wavelet, and different sub-images are formulated with different fusion rules to improve the quality of the video image under the complex weather background. The performance of this method has been improved and is better than the methods already proposed.

II. Video Enhancement Based on Adaptive Tolerance

A. Video Enhancement of Dynamic Scene

For dynamic scene videos which are caused by camera motion, each frame of the video is enhanced as a single image in our method, which requires the processing rate of each video frame to be fast enough. Firstly, to estimate the rough transmission amount of each frame by the dark prior channel, the guide filter is used to refine the rough transmission amount quickly. Then it is needed to be judged whether a large bright region exists in the intercepted video frame. If [12] it exists, the transmission requires to be corrected with the help of the adaptive tolerance, if not, the transmission can be directly used to recover the scene radiance. The flowchart of the video enhancement based on adaptive tolerance is shown in Fig. 1.
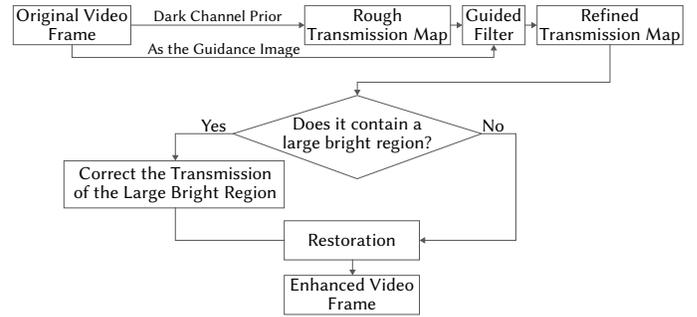
Fig. 1. Flowchart of video enhancement based on adaptive tolerance.

B. Dark Channel Prior and Rough Transmission Map

The [18] dark channel prior can be understood as a statistic for outdoor fog-free images. Most of the local areas in outdoor fog-free images will contain some very dark pixels, which exist in at least one color channel. The min operation is taken to both the local patch and three color channels, and it is obtained:

$$\min_{k \in \Omega(l)}(I^C(k)) = \tilde{t}(l) \min_{k \in \Omega(l)}(J^C(k)) + (1 - \tilde{t}(l))A^C \qquad (1)$$

l and k respectively refer to image pixels and local area pixels, t(l) is the transmittance, and A is the atmospheric light value, Ω (l) is a local area window centered on l. The purpose of this method is to restore the fog-free image J from the fog-free image I. Then the dark channel of the source image is obtained by:

$$J^{dark}(l) = \min_C(\min_{k \in \Omega(l)}(J^C(k))) \rightarrow 0 \qquad (2)$$

the rough transmittance can be estimated as follows:

$$\tilde{t}(l) = 1 - \omega \min_C(\min_{k \in \Omega(l)}(\frac{I^C(k)}{A^C})) \qquad (3)$$

Select the pixel with the highest fog density in the dark channel (the first 0.1% of the brightest), and then select the pixel with the highest intensity among the pixels corresponding to this point in the source fog image as A. Fig. 3 shows the rough transmission map estimated from Fig. 2.

Fig. 2. Original video frame.   Fig. 3. Rough transmission map.

C. The Guided Filter and Refined Transmission Map

If the rough transmission is directly used to recover the enhanced image, halo artifacts will appear at the edges with discontinuous depth. Therefore, a guided filter is introduced to refine the rough transmittance. The refined transmittance is defined as:

$$t(y) = a_l^T I(k) + b_l \quad \forall k \in \omega_l \qquad (4)$$

In the local window ω_p, there is a linear relationship between $I(k)$ and $t(k)$, and the coefficient a_l and b_l is defined as:

$$a_l = (\sum_l + \varepsilon U)^{-1}(\frac{1}{|\omega|}\sum_{k \in \omega_l} I(k) \tilde{t}(k) - \mu_l \bar{t}(l)) \qquad (5)$$

$$b_l = \bar{t}(l) - a_l^T \mu_l \qquad (6)$$

Refer to (5)-(6), Equation (4) indicates that refined transmission can obtain the refined transmittance for each frame. Refining the transmission by the guided filter is not only efficient, but can also produce good results comparable with the results of soft matting. The result is shown in Fig. 4.

Fig. 4. Refined transmission map.　　Fig. 5. Restored video frame, U_{max}=50.

D. The Image Restoration Based on the Adaptive Tolerance

The [12] inaccurate transmission estimated can be corrected according to the similarity degree between the source image I and the atmospheric light A. If a region of the source image satisfies that the value of $|I-A|$ is smaller than or equal to the tolerance, the transmission of the region considered to be a bright region will be recalculated. The region whose value of $|I-A|$ is bigger than the tolerance keeps its transmission unchanged, because it is considered to the region which meets the dark channel prior. Tolerance value can be determined adaptively according to the histogram of the transmission of the original video frame, which meets the requirement of automaticity in video enhancement, and the color distortion can be avoided effectively in image restoration.

Thus, if the original video frame contains a large bright region, it is needed to search out the first maximum point (U_{max}, P_{max}) on the left side of the histogram curve of the refined transmission map. Equation (7) can be used to correct transmission by substituting U_{max}. According to the value of $|I-A|$ in each color channel, the three corrected transmission maps $q^c(l)$ can be separately used to recover the three color channels of the enhanced image. When U_{max} was set to be '50', the restored video frame is shown in Fig. 5.

$$q^C(l) = min(max(\frac{U_{max}}{|I^C(l) - A^C|}, 1) \cdot max(t(l), t_0), 1)$$
$$c \in \{r, g, b\} \tag{7}$$

U_{max} should be set to be '0', if the original video frame does not contain a large bright region, and the transmission will not be corrected.

III. Image Fusion Based on Dual-Tree Complex Wavelet

A. Dual-tree Complex Wavelet

The dual-tree complex wavelet transform adopts a binary tree structure, and the signal is decomposed and reconstructed by two independent real number filter banks. The complex wavelet transform is realized by a real wavelet transform. The problem that the complex wavelet transform cannot be completely reconstructed is solved.

Among [21] them, the one-dimensional complex wavelet function is represented by Equation (8):

$$\psi(\alpha) = \psi_h(\alpha) + i\psi_g(\alpha) \tag{8}$$

the two-dimensional complex wavelet function is expressed by Equation (9):

$$\psi(\alpha, \beta) = \psi(\alpha)\psi(\beta) = [\psi_h(\alpha) + i\psi_g(\alpha)][\psi_h(\beta) + i\psi_g(\beta)] \tag{9}$$

Use $R[\psi(\alpha, \beta)]$ to represent the real part and $I[\psi(\alpha, \beta)]$ to represent the imaginary part:

$$R[\psi(\alpha, \beta)] = \psi_h(\alpha)\psi_h(\beta) - \psi_g(\alpha)\psi_g(\beta) \tag{10}$$

$$I[\psi(\alpha, \beta)] = \psi_g(\alpha)\psi_h(\beta) + \psi_h(\alpha)\psi_g(\beta) \tag{11}$$

The discrete wavelet transform can only get three sub-bands in different directions for each layer of decomposition. However [20], each layer of dual-tree complex wavelet transform can be decomposed to obtain 6 sub-bands, as shown in Fig. 6. This method improves the accuracy of image decomposition and reconstruction, and can retain more detailed information about the image.
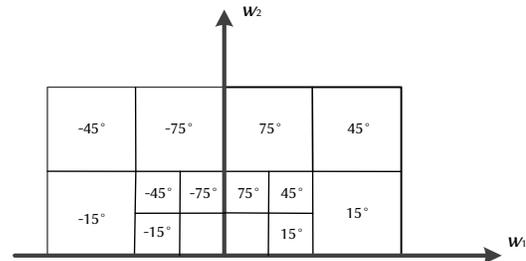
Fig. 6. Direction Sub-images Distribution.

B. Image Fusion

The procedure of fusing together several images in the same scene collected by different sensors to obtain an image with complementary information is called image fusion. Fig. 7 shows the main process of image fusion. First, the input image is preprocessed, mainly the process of noise reduction and correction, then the processed image is matched with features, and finally, image fusion is performed. The most basic and common image preprocessing technique is image enhancement. Image [22] enhancement is mainly for specific processing of images, obtaining new images with better visual quality. Image matching is a process of aligning multiple images obtained at different times, different viewing angles or different sensor devices. Through the matching algorithm, the similarity and consistency of the image are found in textures, features and structures, then similar images are obtained. Image [23] fusion is widely used in security surveillance, military defense, biological monitoring, remote sensing and other fields.
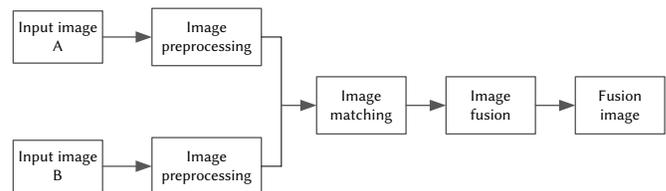
Fig. 7. Image Fusion flow chart.

IV. The Proposed Video Deraining and Desnowing Algorithm

The deraining and desnowing algorithm on dual-tree complex wavelet fusion and adaptive tolerance proposed in this paper is mainly divided into three modules to complete: improved image preprocessing based on adaptive tolerance, optimized image feature matching based on SIFT and dual-tree complex Image fusion of wavelet. Fig. 8 shows the overall flow chart of the fusion algorithm.
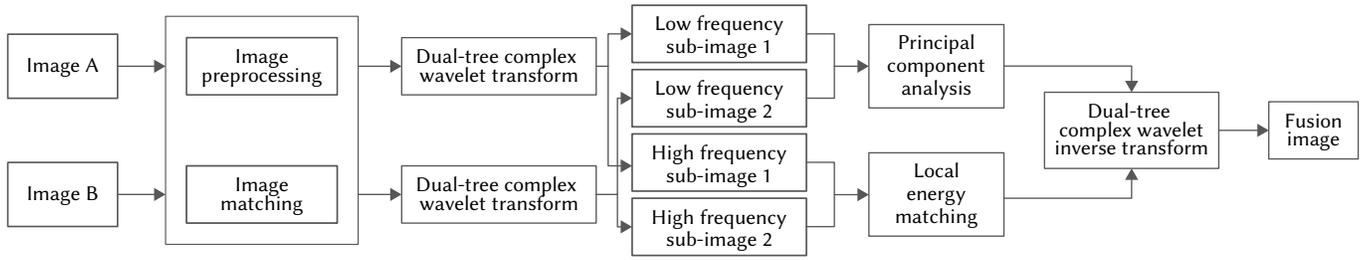
Fig. 8. The proposed algorithm flow chart.

A. Image Preprocessing Based on Adaptive Tolerance

The image acquired by the sensor usually weakens the contrast, which makes the edge and contour information in the image appear inconspicuous to a certain extent, and the imaging effect becomes blurred. In order to solve the problem of bright area distortion in traditional image preprocessing algorithms, this article [24] proposes an improved image preprocessing algorithm based on adaptive tolerance in the chapter 'Video enhancement based on adaptive tolerance'. It can adaptively determine the tolerance parameters of the bright area according to the image, correct the transmittance value of the bright area, and effectively reduce the error compensation of the transmittance, maximize the reduction of image distortion after deraining and desnowing, and improve the quality of video images quality after deraining and desnowing.

B. Image Matching

The SIFT algorithm is the mainstream algorithm for image feature matching, which has far and wide application on the seeking of key points, remaining the immutability and stability of brightness, translation, rotation, and scale, in different scale-spaces. To calculate the Euclidean distance between the two sets of feature points, the method of using feature points to extract feature descriptors is applied first, and then the match between feature descriptors is apparently figured out. The smaller the Euclidean distance, the higher the similarity. When the Euclidean distance is less than a certain threshold, the match is considered as a success. The SIFT algorithm of MLESAC is outstandingly proposed in [25] and turned into a basis of the improvement in [20]. The approach involves three steps:

The first step is to detect and extract the feature points of the preprocessed image and the reference image.

The second is a rough match. The Euclidean distance of the key point feature vector is used for rough matching to generate unmatched feature points.

The last step is exact matching. For the feature point pairs matched roughly in the second procedure, the algorithm [25] applies to statistical estimation. After excluding the mismatching feature points, exact matching of the feature points happens with eventual success.

C. Image Fusion Based on Dual-tree Complex Wavelet

The [26] fusion image obtained based on the discrete wavelet transform fusion algorithm usually has artifact effects on the edges. To [27] solve the above problems, this paper improves the traditional dual-tree complex wavelet, different from wavelet transform, dual-tree complex wavelet transform has good direction selectivity and reduces the translation sensitivity of wavelet transform. The [20] fusion process is divided into the following three modules:

(1) Decompose the input image

The dual-tree complex wavelet transform first decomposes in the row direction and then decomposes in the column direction to obtain two low-frequency sub-bands denoted as $I_A^N(p,q)$ and $I_B^N(p,q)$, and six high-frequency sub-bands $\{I_A^{l,k}(p,q) \setminus I_B^{l,k}(p,q)\}$ in different directions.

(2) Sub-image fusion

The low-frequency images after wavelet decomposition display the overall characteristics of the image and contains some significant details (such as edges and contours). The selection of which fusion rule directly effects the consequence of image fusion. In order to obtain a clearer image fusion result, this paper adopts the weighting criterion and uses principal component analysis to fuse low-frequency images. The [27] fusion process is as follows:

$$I_{LF}^N(p,q) = w_A \times I_A^N(p,q) + w_B \times I_B^N(p,q) \tag{12}$$

$I_A^N(p,q)$ [29] and $I_B^N(p,q)$ respectively represent the low-frequency coefficients of the source image at the baseband (p, q) position, $(x,y)^T$ are their corresponding feature vectors, then the weighting coefficient is defined as:

$$w_A = \frac{x}{x+y}, w_B = \frac{y}{x+y} \tag{13}$$

The [17] area covered by raindrops or snowflakes in the video image is obviously brighter, and the energy of this pixel will be higher than that of the blocked background pixel. This paper uses local energy to determine the local features of high-frequency images. The [29] local energy is defined as:

$$E_A^{l,k}(p,q) = \sum_{m=-1}^{1}\sum_{n=-1}^{1} R(m+2,n+2)[I_A^{l,k}(p,q)]^2 \tag{14}$$

The partial area matching degree is similar to the covariance of two partial windows, which represents the similarity. The partial energy matching degree is defined as:

$$M_{A,B}(p,q) = \frac{2\sum_{m=-1}^{1}\sum_{n=-1}^{1} R(m+2,n+2)I_A^{l,k}(p,q)I_B^{l,k}(p,q)}{E_A^{l,k}(p,q) \times E_B^{l,k}(p,q)} \tag{15}$$

Given a matching degree threshold η, its value range is generally [0.5, 1]. Because the pixel partial energy in the area covered by raindrops and snowflakes is higher than that of the background pixels, when $M_{A,B}(p,q) < \eta$, high-frequency images are used with lower energy, and the fused high-frequency images is defined as:

$$I_{HF}^{l,k}(p,q) = \begin{cases} I_A^{l,k}(p,q) & E_A^{l,k} \geq E_B^{l,k}(p,q) \\ I_B^{l,k}(p,q) & others \end{cases} \tag{16}$$

When $M_{A,B}(p,q) \geq \eta$, weights are assigned according to energy. The [26] local weight coefficient is as follows:

$$w_\alpha(p,q) = \frac{1}{2} + \frac{1}{2}\left(\frac{1-M_{A,B}(p,q)}{1-\eta}\right) \tag{17}$$

$$w_\beta(p,q) = 1 - w_\alpha(p,q) \tag{18}$$

(3) Reconstructed image

After the fusion is completed, the high-frequency and low-frequency images are subjected to the inverse transformation of the dual-tree complex wavelet, and the final clear fusion image is obtained by reconstruction.

Fig. 9. Frames of dynamic video containing large bright regions: (a) the 8th frame (b) the 144th frame (c) the 184th frame.

Fig. 10. Enhanced frames of dynamic video without adaptive tolerance: (a) result of the 8th frame (b) result of the 144th frame (c) result of the 184th frame.

Fig. 11. Enhanced frames of dynamic video using adaptive tolerance: (a) result of the 8th frame (b) result of the 144th frame (c) result of the 184th frame.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Results of Image Preprocessing Algorithm Based on Adaptive Tolerance

The dynamic video is processed in the experiments. These videos contain large bright regions. From the video frame of the dynamic background in Fig. 9, it can be seen that the camera of the video is moving, the aircraft is moving, and the sky and the aircraft belong to a large area of bright areas. In Fig. 10, it can be seen that if these bright areas are not corrected, there will be obvious color distortion in the areas. With the video enhancement based on adaptive tolerance, the corrected results are greatly improved in Fig. 11, and the distortion can be effectively avoided and the image can be clearer. The [12] processing rate of static video is greatly increased and it can reach 25 fps.

B. Experimental Effects of Image Matching

The traditional SIFT feature matching effect is shown in Fig. 12. The improved SIFT feature matching effect proposed in this paper is shown in Fig. 13. The comparison shows that the improved matching method proposed in this paper obtains more correct matching points and higher accuracy.

Fig. 12 The traditional SIFT feature matching effect.

Fig. 13 The improved SIFT feature matching effect.

C. Experimental Results of Image Fusion

The [20] first test video frame is an outdoor rainy image. In order to make the result more convincing, the deraining effect of this algorithm is compared and verified with the Dou algorithm. The results are shown in Fig. 14.

The [20] second test video frame is the outdoor rainy image, the desnowing effect of this algorithm is compared and verified with the Dou algorithm. The results are shown in Fig. 15.

It can be seen from the subjective evaluation of Fig. 14(c) and Fig. 15(c) that the effect by the fusion algorithm proposed in this paper is better. It can well remove raindrops or snowflakes from the video image, enhance the image detail information covered by raindrops or snowflakes, and better retain the image edge information.

The experimental results are evaluated from the four objective indicators of the representative evaluation indicators: average gradient, standard deviation, spatial frequency, and entropy, and the fusion results are quantitatively analyzed. The results are shown in Table I and Table II. From the perspective of the objective evaluation indicators, this method is superior to other fusion methods under the traditional fusion rules, and it can effectively retain image details.
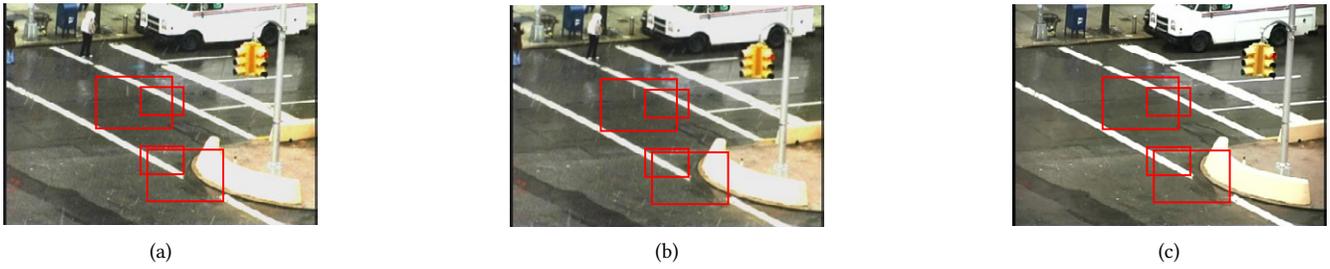
Fig. 14 The results of the first test video frame. (a)The source image. (b)Dou 's algorithm. (c) Improved algorithm.
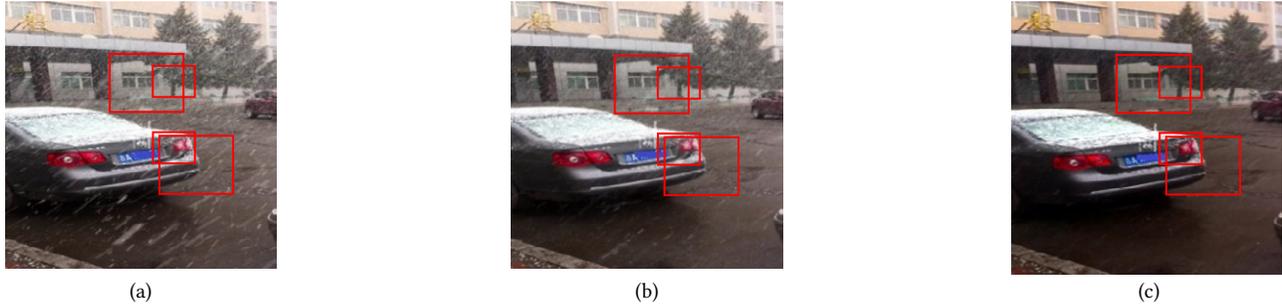

Fig. 15. The results of the second test video frame: (a) The source image (b) Dou's algorithm (c) Improved algorithm.

TABLE I. The First Test Video Frame

index algorithm	Average Gradient	Standard Deviation	Spatial Frequency	Entropy
Dou's algorithm	20.7218	51.2670	40.1258	5.1536
The proposed algorithm	25.0816	60.3450	45.8465	6.7690

TABLE II. The Second Test Video Frame

index algorithm	Average Gradient	Standard Deviation	Spatial Frequency	Entropy
Dou's algorithm	21.5478	51.3213	40.3448	5.2014
The proposed algorithm	25.3647	60.4478	45.9657	6.9985

VI. Conclusion

This paper proposes a new algorithm combining dual-tree complex wavelet fusion and adaptive tolerance. The algorithm adopts the fusion rule based on double-tree complex wavelet analysis to improve the quality of the video image under the complex weather background. The adjustment mechanism based on adaptive tolerance can not only solve the problem of distortion in the bright area, but also not make the transmittance be miscompensated, so as to restore a higher quality fused image. The experimental results show that the overall visual effect of the image obtained by the fusion algorithm is better, and it can better recover the details of the image obscured by raindrops or snowflakes.

Acknowledgment

References

[1] H. Hase, K. Miyake and M. Yoneda, "Real-time snowfall noise elimination," *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, Kobe, vol. 2, pp. 406-409, 1999, doi:10.1109/ICIP.1999.822927.

[2] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, Washington, DC, USA, pp. I-I, 2004, doi:10.1109 / CVPR.2004.1315077.

[3] K. Garg and S. K. Nayar, "Vision and Rain," *International Journal of Computer Vision* 75, pp. 3–27, 2007, doi:10.1007/s11263-006-0028-6.

[4] X. Zhang, H. Li, Y. Qi, W. K. Leow and T. K. Ng, "Rain Removal in Video by Combining Temporal and Chromatic Properties," *2006 IEEE International Conference on Multimedia and Expo*, Toronto, pp. 461-464, 2006, doi:10.1109/ICME.2006.262572.

[5] N. Brewer, N. Liu, "Using the shape characteristics of rain to identify and remove rain from video," *Lecture Notes in Computer Science*, vol. 5342, pp. 451-458, 2008, doi:10.1007/978-3-540-89689-0_49.

[6] X. Zhao, P. Liu, J. Liu, X. Tang, "The application of histogram on rain detection in video," *Proceedings of the 11th Joint Conference on Information Science*, pp. 382-387, 2008, doi:10.2991/jcis.2008.65.

[7] P. Barnum, T. Kanade, & S. Narasimhan, "Spatio-temporal frequency analysis for removing rain and snow from videos," *The First International Workshop on Photometric Analysis for Computer Vision*, 1-17 (2008).

[8] P. Barnum, S. Narasimhan, & T. Kanade, "Analysis of Rain and Snow in Frequency Space," *International Journal of Computer Vision* 86, pp. 256, 2010, doi:10.1007/s11263-008-0200-2.

[9] J. Bossu, N. Hautière, & J. Tarel, "Rain or Snow Detection in Image Sequences Through Use of a Histogram of Orientation of Streaks," *International Journal of Computer Vision* 93, pp. 348–367, 2011, doi: 10.1007/s11263-011-0421-7.

[10] L. Kang, C. Lin and Y. Fu, "Automatic Single-Image-Based Rain Streaks Removal via Image Decomposition," *in IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1742-1755, 2012, doi:10.1109/ TIP.2011.2179057.

[11] L. Kang, C. Lin and Y. Lin, "Self-learning-based rain streak removal for image/video," *2012 IEEE International Symposium on Circuits and Systems (ISCAS)*, Seoul, pp. 1871-1874, 2012, doi:10.1109/ISCAS.2012.6271635.

[12] L. Yu, Y. Piao, X. Yan, "Video Defogging Based on Adaptive Tolerance," *TELKOMNIKA: Indonesian Journal of Electrical Engineering*, vol. 10, pp. 1644-1654, 2012, doi:10.11591/telkomnika.v10i7.1556.

[13] A. E. Abdel-Hakim, "A Novel Approach for Rain Removal from Videos Using Low-Rank Recovery," *2014 5th International Conference on*

Intelligent Systems, Modelling and Simulation, Langkawi, pp. 351-356, 2014, doi:10.1109 / ISMS.2014.161.

[14] J. Kim, J. Sim and C. Kim, "Video Deraining and Desnowing Using Temporal Correlation and Low-Rank Matrix Completion," *in IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2658-2670, 2015, doi:10.1109/TIP.2015.2428933.

[15] W. Ren, J. Tian, Z. Han, A. Chan and Y. Tang, "Video Desnowing and Deraining Based on Matrix Decomposition," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 2838-2847, 2017, doi:10.1109/CVPR.2017.303.

[16] X. Zheng, Y. Liao, W. Guo, X. Fu, X. Ding, "Single-Image-Based rain and snow removal using multi-guided filter," *20th International Conference on Neural Information Processing (ICONIP). Springer Berlin Heidelberg*, vol. 8228, pp. 258-265, 2013, doi:10.1007/978-3-642-42051-1_33.

[17] Z. Wang, "Method of Removing Rain (Snow) from Video Images Based on Wavelet Fusion," *Journal of BeiHua University (Natural Science)*, vol. 19, no. 1, pp. 136-140, 2018.

[18] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713-724, June 2003 doi:10.1109/TPAMI.2003.1201821.

[19] K. He, J. Sun, X. Tang, "Guided image filtering," *European Conference on Computer Vision 2010, Part I, Lecture Notes in Computer Science*, vol. 6311, pp. 1-14, doi:10.1007/978-3-642-15549-9_1.

[20] L. Dou, D. Xu, J. Li, "Image Inpainting Based on Dual-tree Complex Wavelet Transform," *Computer Science*, vol. 44, no. 6, pp. 180-191, 2017.

[21] J. Du, S. Chen, "Adaptive PCNN Image Fusion Algorithm Based on Double Tree Complex Wavelet Transform," *Infrared Technology*, vol. 40, no. 10, pp. 1002-1007, 2018.

[22] W. Yang, Q. Chai, L. Wang, "Multi-focus image fusion method based on double-tree complex wavelet transform," *Electro-Optic Technology Application*, vol. 24, no. 3, pp. 59-62, 2009.

[23] F. Yang, L. Zheng, "New feature extraction algorithm based on SIFT," *Applied Science and Technology*, vol. 1, no. 1, pp. 1-5, 2018.

[24] Y. Xiang, H. Qin, J. Li et al., "Multi-focus image fusion using a guided-filter-based difference image," *Applied Optics*, vol. 55, no. 9, pp. 2230-2239, 2016, doi:10.1364/AO.55.002230.

[25] J. Li, Y. Yang, "An Improved MLESAC Algorithm for Estimating Fundamental Matrix," *Computer Engineering*, vol. 38, no. 19, pp. 215-217, 2012.

[26] M. R. Metwalli, A. H. Nasr, O. S. Farag Allah and S. El-Rabaie, "Image fusion based on principal component analysis and high-pass filter," *2009 International Conference on Computer Engineering & Systems*, Cairo, pp. 63-70, 2009, doi:10.1109/ICCES.2009.5383308.

[27] S. Lu, L. Zou, X. Shen, W. Wu, W. Zhang, "Multi-spectral remote sensing image enhancement method based on PCA and IHS transformations," *Journal of Zhejiang University-SCIENCE A*, vol. 12, pp. 453–460, 2011, doi:10.1631/jzus.A1000282.

[28] X. Wang, Y. Zhou, "Infrared and visible fusion in undecimated dual-tree complex wavelet domain," *Computer Engineering and Design*, vol. 38, no.3, pp. 730-734, 2017.

[29] F. Wu, H. Li, "Image Fusion Algorithm Based on NSCT and PCA," *Aeronautical computing technique*, vol. 45, no. 3, pp. 48-50, 2015, doi:10.12968/prtu.2015.45.48.

[30] J. Zang, G. Ren, J. Dong, Y. Piao, S. Jim, "Removal of rain video based on temporal intensity and chromatic constraint of raindrops," *Evolutionary Intelligence*, vol. 12, pp. 349–355, 2019, doi:10.1007/s12065-018-0185-x.

[31] J. Zang, G. Ren, Y. An, Y. Piao, "Removal of rain from video based on dual-tree complex wavelet fusion," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 1, pp. 105-113, 2020, doi:10.3233/JIFS-179385.

Jingfeng Zang

Jingfeng Zang, associate professor of Changchun University of Science and Technology, received a bachelor's degree from Changchun University of Science and Technology in 2000, a master's degree from Changchun University of Science and Technology in 2004, and a doctorate from Changchun University of Science and Technology in 2010. From November 2013 to September 2014, he was a visiting scholar at the Norwegian University of Science and Technology, mainly engaged in the development and application of heterogeneous multicore embedded systems. In recent years, he has published more than 10 academic papers on video and image processing, including 6 papers in EI, 3 core papers, 1 utility model patent, and won 1 third prize for scientific and technological progress in Jilin Province. As a backbone, he participated in 1 National Natural Science Foundation project, 1 Ministry of Science and Technology project, 4 Jilin Science and Technology Department project, Changchun City Science and Technology Department project, Jilin Province Education Department project, etc., mainly responsible for hardware system design and The hardware transplantation of software algorithms has accumulated rich scientific research experience, and made great breakthroughs in video image processing related theories and key technologies. At present, he is mainly engaged in the research of video image processing technology, video surveillance, intelligent information processing and the transformation of technological achievements of joint enterprises.

Ningxue Xu

Ningxue Xu, graduated from China in 2019 with a bachelor's degree in electronic information engineering from Anyang Institute of Technology. Currently studying for a master's degree in electronic science and technology in the School of Electronic Information Engineering of Changchun University of Science and Technology. Mainly research sensing and signal processing technology, convolutional neural network technology, video security monitoring technology, image registration, fusion, matching technology, etc. Good at researching multi-focus image fusion technology based on deep learning.

Rui Liu

Rui Liu, graduated from the School of Electronic Information Engineering, Changchun University of Science and Technology in 2019 with a bachelor's degree in engineering, Jilin, China. Currently studying for a master's degree in the School of Electronic Information Engineering, Changchun University of Science and Technology, majoring in electronic science and technology. The main research directions are sensing and signal processing technology, video image processing technology, video security monitoring technology, and deep learning. Good at researching face state detection and evaluation based on deep learning.

Yuhuan Shi

Yuhuan Shi, graduated from Jilin, China in 2019 with a bachelor's degree in electronic information engineering from Beihua University. He is currently a master's student at Changchun University of Science and Technology. The major is electronic science and technology. Current research interests include image processing or video processing based on deep learning, surveillance video analysis, and video content understanding. The specific research direction is pedestrian detection and pedestrian re-identification based on deep learning. Good at researching pedestrian re-identification technology and pedestrian detection technology based on deep learning.

The Construction Site Management of Concrete Prefabricated Buildings by ISM-ANP Network Structure Model and BIM under Big Data Text Mining

Guiming Xu*

Changzhou Vocational Institute of Engineering, Architectural Engineering Institute, Wujin District, Changzhou City 213164 (China)

unir

LA UNIVERSIDAD
EN INTERNET

Abstract

In the construction industry, prefabricated buildings have developed rapidly in recent years due to their various excellent properties. To expand the application of big data text mining and Building Information Model (BIM) in prefabricated building construction, with concrete as a form of expression, the construction management of concrete prefabricated buildings is discussed. Based on the Interpretative Structural Model (ISM) and Analytic Network Process (ANP), the importance of the safety factors on the construction sites of concrete prefabricated buildings are assessed. Based on BIM, an optimized construction management platform for concrete prefabricated buildings is built, whose realization effects are characterized. The results show that prefabricated buildings have developed rapidly from 2017 to 2019. Compared with traditional buildings, they can significantly reduce the waste of resources and energy, and the savings of water resource utilization can reach 80%. Among the various safety impact elements, construction management has the greatest impact on construction safety, and the corresponding weight value is 0.3653. The corresponding weight of construction personnel is 0.2835, the corresponding weight of construction objects is 0.1629, the corresponding weight of construction technology is 0.1436, and the corresponding weight of construction environment is 0.0448. This building construction management platform is able to control the construction progress in real-time and avoid the occurrence of construction safety accidents. The final layout of the construction site shows a good effect, and the deviation between the actual construction schedule and the expected construction schedule is small, which is of great significance for the smooth development of concrete prefabricated buildings. This is a catalyst for the future development of concrete prefabricated buildings and the application of big data technology.

I. Introduction

THE traditional construction industry mainly uses the way of on-site pouring construction. Such constructions require a lot of manpower costs. Meanwhile, during the process of construction, construction wastes, environmental pollutions, and other problems will be produced inevitably [1], [2], while prefabricated buildings have solved these problems perfectly [3]. The so-called "prefabricated buildings" specifically refer to the prefabricated components at the factory or project construction site. With the aid of some connection paths, the scattered components are finally connected as an integrity, which is a means for building construction. Fewer consumables, fast construction, and low strength are the unique properties of prefabricated buildings [4]. Since their invention, the forms of prefabricated buildings have evolved from wood and light steel in the early times and concrete structures in current times. For prefabricated buildings, the completion of the construction goals is achieved

through industrialization [5]. Due to the excellent characteristics of prefabricated buildings, China has provided strong support for the development of this industry. In recent years, concrete prefabricated buildings have developed rapidly due to their ease of installation, high efficiency, and the ability to achieve mass production.

Compared with traditional building construction, the proportion of construction personnel required for concrete prefabricated buildings is greatly reduced. Most of the components are prefabricated in the workshop and have a high degree of mechanization; the workload of the entire construction process is reduced; there are higher requirements for the planning of the construction site; at the same time, there are also higher requirements for the details, such as the accuracy of the prefabricated components; moreover, the entire construction process has a higher difficulty compared with the traditional construction [6]; therefore, it has become the main development direction for the transformation and upgrading of traditional buildings. But at the same time, the construction process and the production of prefabricated parts require precise technical support. This building model puts forward more stringent requirements for safety management. Therefore, the safety management of prefabricated building construction must be

* Corresponding author.

E-mail address: 8000000295@czie.edu.cn

paid attention to. Traditional management methods are more limited in the treatment of accidental hazards. It is currently difficult to meet the management requirements of prefabricated buildings. The rise of big data text mining and Building Information Modeling (BIM) technology provides a good direction for solving this problem. However, the effect of BIM technology in the application process is not very satisfactory. Concrete prefabricated buildings have developed rapidly in recent years. This construction model can effectively reduce the risk of accidents. At the same time, the mechanized construction method has brought convenience to construction management. However, the actual scenarios require the cooperation of multiple people and multiple devices, as well as higher requirements on construction techniques. On the contrary, it can easily lead to safety accidents. Therefore, in the construction process of concrete prefabricated buildings, it is very important to study and discuss the safety management.

Therefore, to apply big data technology to expand the development of concrete prefabricated buildings in the construction industry and achieve effective management and regulation of construction sites, the Interpretative Structural Model (ISM) and Analytic Network Process (ANP) are combined. On this basis, BIM technology is introduced to build a construction management platform for concrete prefabricated buildings. It is hoped to provide some data references for the combination of modern analysis technologies and prefabricated buildings.

II. Literature Review

A. Research Progress of Prefabricated Buildings

At present, scholars from various countries have explored the prefabricated buildings. Aiming at the limitations of prefabricated buildings, Zhao et al. realized the real-time collection and analysis of component information through the combination of Radio Frequency Identification (RFID) and Building Information Model (BIM) technologies [7]. Hwang et al. discussed the important factors affecting the volume structure of prefabricated components in prefabricated buildings and proposed a feasible strategy through the application of BIM [8]. From the perspective of technology promotion and green production, Wu et al. analyzed the key factors affecting prefabricated buildings by taking the Index of Relative Importance (IRI) as an evaluation indicator; they found that the technology lock-in, incentive policies, and standardization are the important influencing factors [9]. Shi et al. proposed a bolt end-plate connection method and a new construction method, tested the seismic performance of the new prefabricated steel frame based on prefabricated buildings, and found that the proposed method could make the test piece have the good horizontal bearing capacity and elongation performance [10]. Tan et al. explored the obstacles in the application of BIM technology in prefabricated buildings. Finally, by introducing the Interpretative Structural Model (ISM), a corresponding strategy was developed, which provided technical and management solutions for the application of BIM in prefabricated buildings [11].

B. Research Progress of Concrete Prefabricated Buildings

Thanks to the excellent performance of concrete, the concrete prefabricated buildings have developed rapidly in recent years. Sebaibi and Boutouil used the ground blast furnace slag and ultra-fine silicate cement to improve the energy efficiency of the concrete prefabricated building construction by eliminating heat treatment and reducing the environmental impact of concrete [12]. Tsangouri et al. evaluated the design feasibility and structural integrity of prefabricated building construction components, including concrete cantilever slab bearing steel, through incremental bending tests, which provide test and data support for the development of concrete prefabricated building construction [13]. Ren et al. used a batch of concrete specimens as

samples to study the effect of passenger and freight trains with different wheel weights and different speeds on the mechanical properties of prefabricated concrete [14].

In summary, there are many discussions on the application of BIM technology in prefabricated buildings. However, at the same time, there are limitations in the application of BIM technology in prefabricated buildings, while explorations that combine BIM technology with ISM are rare. Among the discussions on concrete prefabricated buildings, those on construction site management are few.

III. Method

A. A Brief Introduction to Concrete Prefabricated Buildings

The construction of concrete prefabricated buildings requires coordination of various factors to ensure its smooth progress. For the management of construction site of concrete prefabricated buildings, the construction schedule, construction quality, construction safety, and construction costs should be considered as the management and control objectives for the construction site management of concrete prefabricated buildings. Here, a construction site management platform is built by considering the above factors that affect the construction process. Examples of prefabricated components used in the construction of concrete prefabricated buildings are shown in Fig. 1. For prefabricated buildings, from 2017 to 2019, in the three major types of areas for the development of prefabricated buildings: key areas, active promotion areas, and encouraging areas, the data on the new construction area of prefabricated buildings and the differences between prefabricated and traditional buildings are analyzed.

Fig. 1. Examples of prefabricated components used in the construction of concrete prefabricated buildings.

B. Construction Safety Factors Based on ISM-ANP Network Structure Model

ISM is able to decompose complex systems, thereby analyzing the relations among various constituent elements of each subsystem obtained by the decomposition. The relations among the objects to be analyzed are expressed by directed graph and matrix theory; at present, ISM has been widely used in solving complex economic problems [15], [16]. Directed connection graph, adjacency matrix, and accessibility matrix are the three major elements that constitute ISM [17]. In the adjacency matrix, the system is defined as:

$$S = (S_1, S_2, \cdots, S_n) \tag{1}$$

Where: n represents the number of elements in the system, S_1 represents the first component in the system, the rest may be deduced by analogy.

In further, the adjacency matrix is defined as:

$$A = \left(a_{ij}\right)_{n \times n} \qquad (2)$$

$$a_{ij} = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \qquad (3)$$

Where: A represents the adjacency matrix, and a_{ij} represents the vector in the i-th row and j-th column of the matrix. If a_{ij} is 1, element S_i in the system will have an effect on element S_j; if a_{ij} is 0, element S_i in the system has no effect on element S_j.

The accessibility matrix is defined as:

$$(A + I) \neq (A + I)^2 \neq \cdots \neq (A + I)^k \neq (A + I)^{k+1} = M \qquad (4)$$

Where: k corresponds to the number of nodes, $A + I$ represents the matrix, and M represents the accessibility matrix.

ANP is obtained by improving the Analytic Hierarchy Process (AHP). In contrast, ANP considers the interaction between factor groups at the same level and the various elements within the factor group. Its specific implementation process mainly includes the construction of structural models, the construction and calculation of the judgment matrix and super matrix [18], [19]. For example, the unweighted super matrix is constructed as:

$$W_{ic} = \begin{bmatrix} w_{i1}^{(c_1)} & w_{i1}^{(c_2)} & \cdots & w_{i1}^{(c_n)} \\ w_{i2}^{(c_1)} & w_{i2}^{(c_2)} & \cdots & w_{i2}^{(c_n)} \\ \vdots & \vdots & \vdots & \vdots \\ w_{in}^{(c_1)} & w_{in}^{(c_2)} & \cdots & w_{in}^{(c_n)} \end{bmatrix} \qquad (5)$$

Where: $w_{i1}^{(c_1)}$ represents the vector of the corresponding position in the unweighted super matrix, and the rest may be deduced by analogy.

The weighted super matrix is defined as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \qquad (6)$$

Where: a_{11} represents the vector located in the i-th row and j-th column of the weighting matrix, and the rest may be deduced by analogy.

Furthermore, the calculation of the weighted super matrix is defined as:

$$\overline{W} = \overline{W_{ic}} = a_{ij} W_{ic} (i = 1,2,\cdots,n; c = 1,2,\cdots,n) \qquad (7)$$

Where: $\overline{W}$ represents the weighted super matrix, i and c are factors.

Furthermore, when the weighted super matrix satisfies the following equation, the limit super matrix is obtained so that the importance weight of each component is reflected.

$$W^\infty = \lim_{t \to \infty} W^t \qquad (8)$$

The implementation steps of ISM and ANP in the application are shown in Fig. 2. The implementation process of ISM mainly includes the setting of key issues, the selection of influencing factors, the analysis of influencing factors, and the analysis of the correlation between various factors. Among them, the analysis of correlation is the core. This link mainly includes the construction of adjacency and reachability matrix, region hierarchical division, and relationship generation. The realization process of ANP is an analysis of the relationship between system goals, decision criteria, and various factors.

On this basis, ISM is utilized to analyze the correlation of factors affecting the safety of construction sites of concrete prefabricated buildings. ANP is utilized to analyze the importance of factors affecting the safety of construction sites of concrete prefabricated buildings. The analysis of the correlation of construction site management factors of concrete prefabricated buildings based on ISM is shown in Fig. 3.

As shown in Fig. 3, the factors mainly include the construction safety organization, construction safety supervision, construction safety protection, and construction technical disclosure management. These components of the construction site safety management influence each other; only through coordination and cooperation can the safety management of the concrete prefabricated buildings be guaranteed. Therefore, according to the ANP network structure model, based on the correlation analysis of construction safety management, under the big data text mining technology, the construction personnel, construction technology, construction objects, construction environment, and construction management are fully considered. Furthermore, based on the ISM-ANP network structure model, the network structure relations of the construction safety management influencing factors of concrete prefabricated buildings are shown in Fig. 4.

C. Construction of An Optimized Construction Management Platform Based on BIM

BIM applies parameterized models to complete the integration of information in relevant projects, and finally, realizes the update and circulation of the entire project cycle. In this process, the application of BIM also provides a platform for collaborative work [20], [21]. Therefore, BIM technology is critical in promoting the development
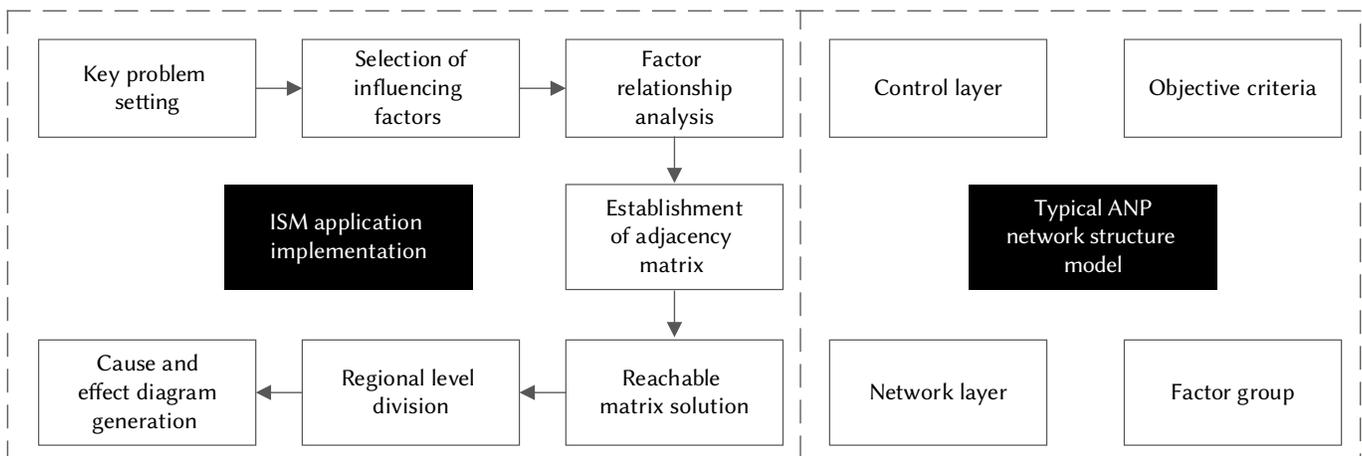

Fig. 2. Implementation of ISM and ANP.
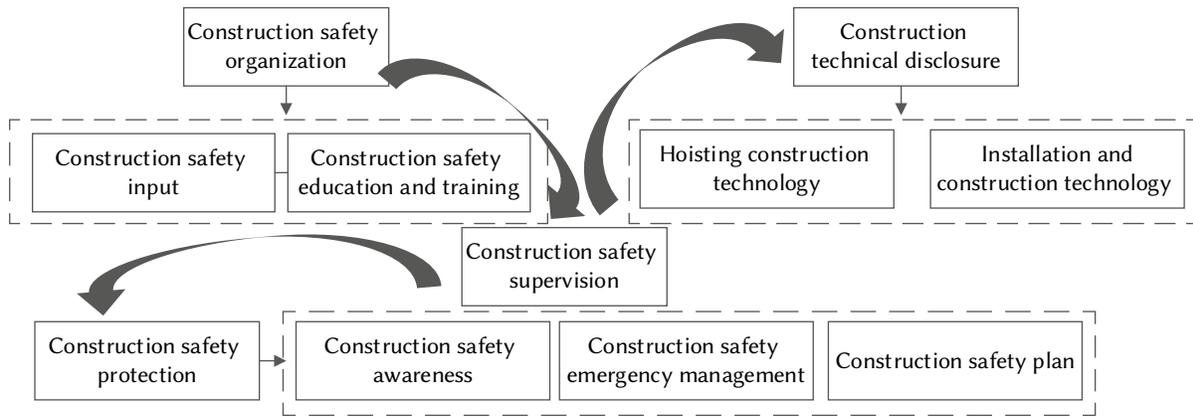
Fig. 3. Correlation between management factors of the construction site.
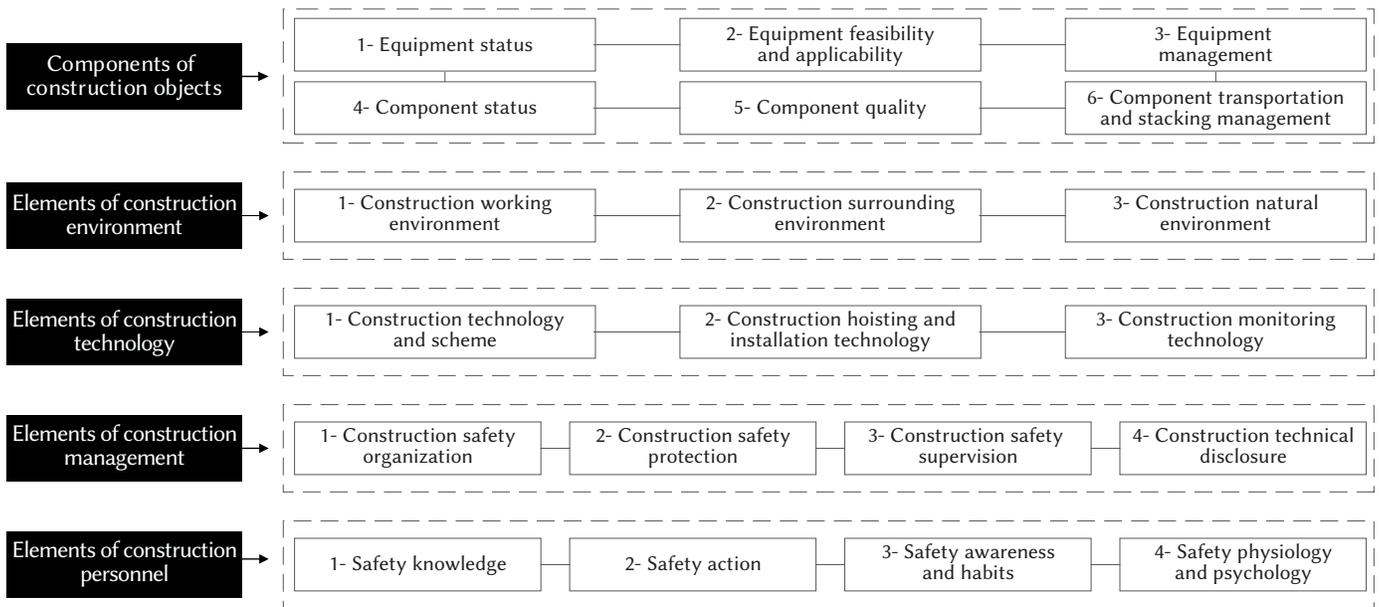
Fig. 4. Network structure relations of construction safety management influencing factors of concrete prefabricated buildings.
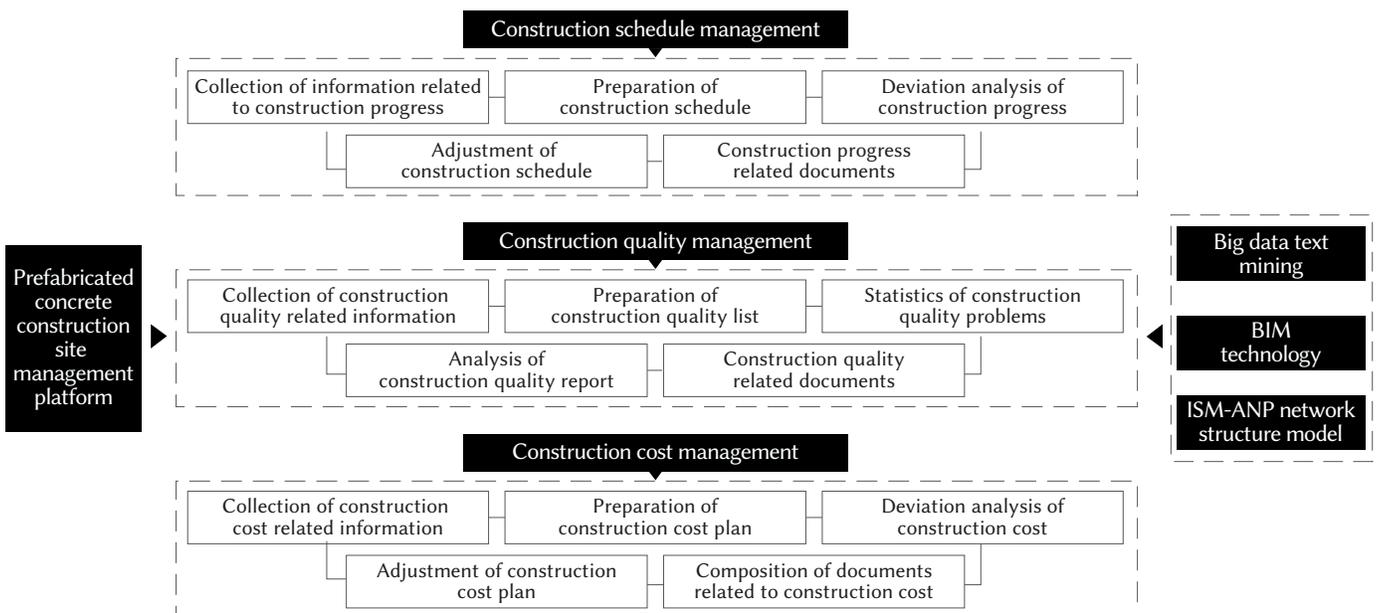
Fig. 5. Construction site management platform for concrete prefabricated buildings.

and upgrading of the construction industry. This is also true in the construction of concrete prefabricated buildings. Currently, BIM technology has been well applied in the construction industry. From the perspective of architecture, BIM technology has excellent performances such as visualization, collaboration, simulation, optimization, and drawing [22], [23]. At the same time, incorporating BIM technology into prefabricated buildings has many obvious advantages. From the perspective of the participating contractor, the application of BIM technology facilitates communication during project implementation, and the visualization function of BIM will be well played, which is beneficial to the successful completion of construction projects. From the perspective of the designer, the application of BIM technology can cooperate with professional technicians to identify conflicts, thereby improving efficiency and reducing errors, making control over details more intuitive. From the perspective of the producer, the fine modeling of BIM technology can meet the requirements for prefabricated components, which will significantly improve the pass rate in the production process; thus, the production will be completed accurately. During the production of prefabricated components, the applications of big data-based text mining and Radio Frequency Identification (RFID) chips are also very helpful for improving the efficiency of production management. From the perspective of the project construction party, the application of BIM technology can realize the regulation of construction site management, promote the improvement of assembly accuracy and quality, and avoid the occurrence of errors in the construction process. In summary, it is not difficult to find that BIM technology has high applicability in the construction of prefabricated buildings, which is also applicable in the construction site management of concrete prefabricated buildings.

Based on this, the BIM technology is introduced; combined with the ISM-ANP construction site safety management factors of concrete prefabricated buildings, an optimized construction site collaborative management platform for concrete prefabricated buildings is built. Specifically, the idea of big data text mining is introduced. In big data text mining, it mainly focuses on the research of structured data, such as data based on relationship and data warehouse. From the perspective of text mining, its main purpose is to extract effective information from the relevant information pattern recognition and retrieval. In short, hidden knowledge can be found from batch data through text mining. Based on the above analysis, the construction of collaborative management platform of concrete prefabricated buildings by ISM-ANP is discussed, whose core concept is consistent with big data text mining, such as seeking effective information and mining hidden information mechanism. This is the reason why the idea of big data text mining is introduced.

The construction schedule, construction quality, construction costs, construction safety, and other influencing factors are comprehensively considered, as well as the construction design, component production, prefabricated component transportation, and key parties of the construction. The constructed construction site management platform for concrete prefabricated buildings based on big data text mining and BIM technology is shown in Fig. 5.

IV. Results and Discussion

A. Development Analysis of Prefabricated Buildings

The distribution and changes in construction areas of prefabricated buildings during 2017-2019, as well as the difference between prefabricated buildings and traditional buildings in terms of resource and energy utilization, are shown in Fig. 6 (a) and Fig. 6 (b).

(a)

(b)
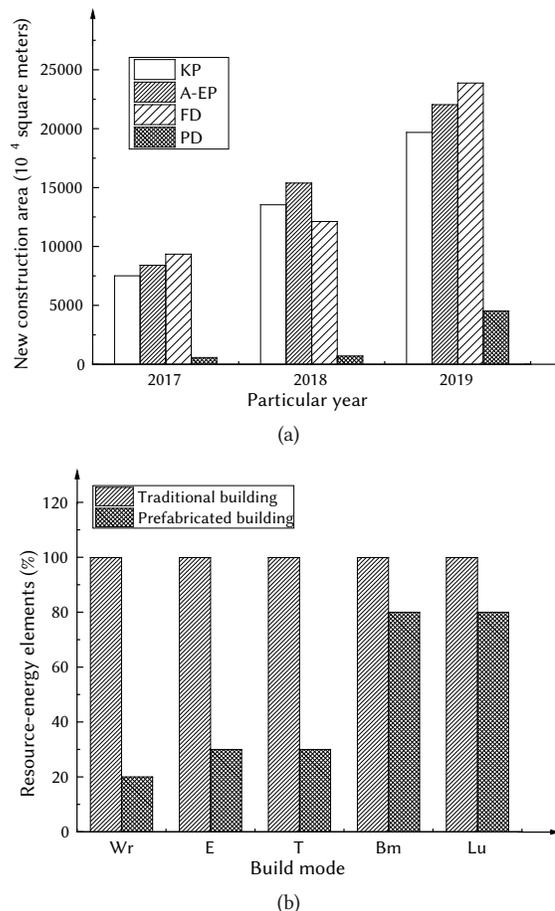
Fig. 6. Development analysis of prefabricated buildings: (a) Changes in the new construction areas of prefabricated buildings in the past three years, where KP indicates the key promotion area, A-EP indicates the active and encouraging promotion area, FD indicates full decoration, and PD indicates prefabricated decoration; (b) Comparison between traditional buildings and prefabricated buildings, where Wr represents water resources, E represents energy, T represents time, Bm represents building materials, and Lu represents land use.

As shown in the changes in the new construction areas of prefabricated buildings in the past three years, whether in the key promotion areas, active promotion areas, or encouraging promotion areas of prefabricated buildings, the new construction areas of prefabricated buildings show an annual increase trend. Besides, the area of newly-started prefabricated buildings in the key promotion areas has the most significant growth. In terms of the energy resource utilization, compared with traditional buildings, no matter in the utilization of water resources, energy, construction time, building materials, or construction land use, the energy resource consumption of prefabricated buildings is significantly lower than that of traditional buildings. Especially, in water resource utilization, energy utilization, and construction time, the differences between the two building construction modes are particularly significant. Specifically, the utilization of water resources can be saved by 80%, and energy utilization and construction time can be saved by 70%.

The above analysis shows that the proportion of prefabricated buildings currently used in the construction industry is developing in an increasing direction. At the same time, compared with traditional buildings, prefabricated buildings show many excellent performances in the utilization of resources and energy, which significantly reduce the consumption and waste of resources and energy. In addition, the applications of various modern mechanical methods in the construction process make prefabricated buildings also have significant advantages

in terms of improving construction efficiency. With the continuous development and deepening of big data mining technology, as well as the implementation of modern mechanical automation construction methods, it is believed that the construction of prefabricated buildings will have greater development and application potentials in the future. In the construction industry, the application of concrete is extremely extensive; thus, the construction of concrete prefabricated buildings will inevitably develop into the mainstream of prefabricated buildings in the future.

B. The Importance Analysis of Construction Safety Factors of Concrete Prefabricated Buildings

Based on the composition of the network structure of the construction safety management influencing factors of concrete prefabricated buildings, the analysis results of the importance of each construction safety factor are shown in Fig. 7.
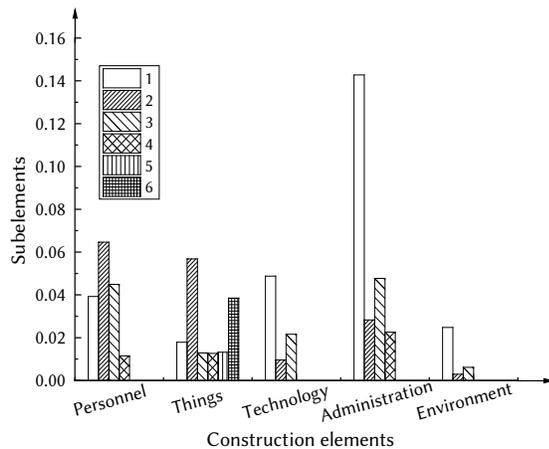
Fig. 7. Analysis results of the importance of factors affecting construction safety.

As shown in Fig. 7, the average weight corresponding to the construction management factor is 0.3653, which is the most important among all the construction safety influencing factors, followed by the construction personnel factor, whose corresponding average weight value is 0.2835, the construction objects factor, whose corresponding average weight value is 0.1629, and finally, the construction technology factor and the construction environment factor, whose corresponding average weight values are 0.1436 and 0.0448, respectively.

The above analysis shows that, among all the construction safety influencing factors of concrete prefabricated buildings, construction management is the most important one since it reflects the concerns from senior management personnel and persons responsible for construction safety, as well as the amount of human and material resources invested in the construction process, safety management implementation measures, and organizational arrangements. If insufficient attention is paid to this link, incorrect decision-making will occur. No matter what advanced construction techniques and methods are used and how professional the technical personnel is, the construction works will be vain. However, from another perspective, the smooth development and realization of the entire construction process of concrete prefabricated buildings require coordination and cooperation between various links. Any problem in any link will have some negative impacts on the final result. Therefore, it is necessary to deal with the coordinated development between construction management, personnel, objects used, and environment. At the construction safety level, management is the priority. For the overall construction process, the site management of the construction of concrete prefabricated buildings is also very critical. This also lays a

good foundation for the application of collaborative BIM technology in the construction of concrete prefabricated buildings.

C. Realization Effect of the Optimized Construction Management Platform of Concrete Prefabricated Buildings

Based on the optimized construction management platform of concrete prefabricated buildings, the results of the comparison between the actual construction progress and the expected construction progress, and the changes in the flow of people at safety exits in a certain area at the construction safety level, are shown in Fig. 8 (a) and Fig. 8 (b).

(a)

(b)
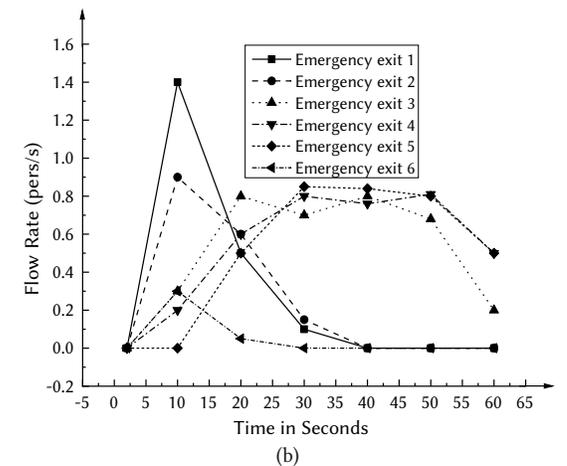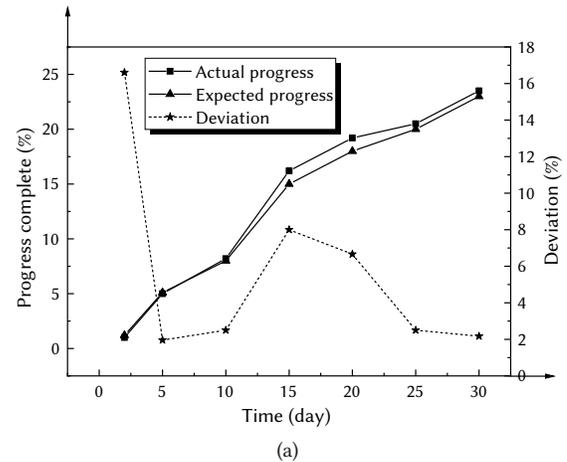
Fig. 8. Implementation of an optimized construction management platform for concrete prefabricated buildings: (a) Comparison of actual and expected construction progress; (b) Changes in the number of people at safe exits.

By analyzing the data changes in Fig. 8, it is found that the optimized construction management platform for concrete prefabricated buildings is more accurate in controlling the construction progress. Meanwhile, the deviation between the actual construction progress and the expected construction progress is not very huge. The construction management platform based on BIM technology has the characteristics of visualization; thus, it is able to display the construction progress in real-time; in turn, it analyzes the deviation between the actual construction schedule and the expected construction schedule. By finding the deviation, the construction schedule is adjusted. Through the control of construction safety management elements of concrete prefabricated buildings based on the ISM-ANP network structure model, it is possible to effectively control the persons entering and exiting the safety exits of the construction site to avoid congestion and safety accidents.

By analyzing the data changes in Fig. 8, it is found that the optimized construction management platform for concrete prefabricated buildings is more accurate in controlling the construction progress. Meanwhile, the deviation between the actual construction progress and the expected construction progress is not very huge. The construction management platform based on BIM technology has the characteristics of visualization; thus, it is able to display the construction progress in real-time; in turn, it analyzes the deviation between the actual construction schedule and the expected construction schedule. By finding the deviation, the construction schedule is adjusted. This is similar to the effective information acquisition emphasized in the big data text mining idea, which shows that the construction of collaborative management platform and big data text mining has achieved the preliminary integration. Through the control of construction safety management elements of concrete prefabricated buildings based on the ISM-ANP network structure model, it is possible to effectively control the persons entering and exiting the safety exits of the construction site to avoid congestion and safety accidents.

Based on the optimized construction management platform of concrete prefabricated buildings, the realization effects of the prefabricated buildings in terms of transportation, prefabricated component production, and the construction site management of concrete prefabricated buildings are shown in Fig. 9.
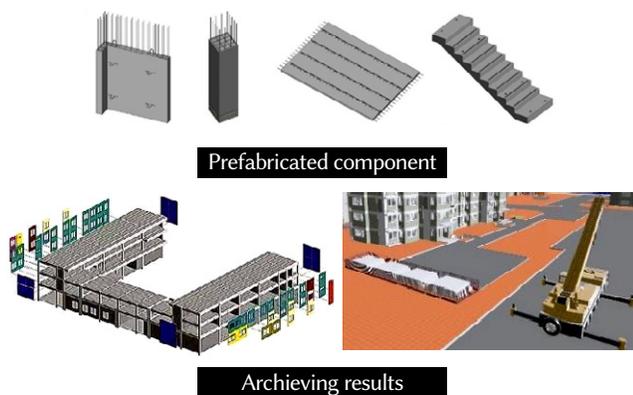

Fig. 9. Actual implementation results of construction site management of concrete prefabricated buildings.

By using the ISM-ANP network structure model to evaluate the safety elements of concrete prefabricated buildings, in addition to the construction of an optimized management platform for construction sites of prefabricated buildings based on big data text mining and BIM technology, the construction site management of the prefabricated buildings shows excellent results. Based on the BIM technology and the ISM-ANP network structure model, the participating organizations or personnel can coordinate to arrange the layout of the construction site, thereby reducing the major changes during the construction process. The practical application value and specific application of BIM technology in prefabricated buildings are briefly described from the aspects of design, manufacturing, and installation. Regarding the application effect of BIM technology in the construction of concrete prefabricated buildings, the conclusions obtained here are basically the same as the conclusions drawn by Liao [24]. However, due to the introduction of big data mining technology, the implementation of construction site management here not only optimizes the quality of concrete prefabricated buildings but also combines the advantages of modern data technology, which has stronger autonomous operation and controllability. Meantime, it can capture the effective information in real time, further optimize the management quality, and improve the efficiency.

V. Conclusion

The innovative introduction of big data text mining and BIM technologies has helped in building the construction site management platform for concrete prefabricated buildings. It is found that construction management is the most critical influencing factor in site safety. Big data mining and BIM technologies have great application potential in site management for the construction of concrete prefabricated buildings. The constructed management platform has achieved good results. However, in the construction of the ISM-APN network structure model, the sub-elements under each component have not been analyzed in detail. At the same time, due to the influence of multiple factors such as experimental conditions, the construction personnel, construction technology, construction objects, construction environment, and other influencing factors that affect the management of the construction site are not analyzed. In addition, the application scope of BIM technology in prefabricated buildings is not very comprehensive. In the future, the sub-factors of each influencing factor will be discussed, and more in-depth analysis and discussion on the application of BIM technology will be conducted from the perspectives of site equipment management and safety protection management.

References

[1] T. Wang, G. Q. Zhou, D. Y. Chao, and L. J. Yin, "Influence of hydration heat on stochastic thermal regime of frozen soil foundation considering spatial variability of thermal parameters," *Applied Thermal Engineering*, vol. 142, pp. 1-9, 2018, doi: 10.1016/j.applthermaleng.2018.06.069.

[2] Y. W. Lee, J. Seo, S. Kim, Y. J. Kang, and D. Won, "Cyclic Lateral Performance Evaluation of Precast Double-Skinned Composite Tubular Columns," *International Journal of Steel Structures*, vol. 18, no. 1, pp. 97-113, 2018, doi: 10.1007/s13296-018-0308-3.

[3] G. Tumminia, F. Guarino, S. Longo, M. Ferraro, M. Cellura, and V. Antonucci, "Life cycle energy performances and environmental impacts of a prefabricated building module," *Renewable and Sustainable Energy Reviews*, vol. 92, pp. 272-283, 2018, doi: 10.1016/j.rser.2018.04.059.

[4] Q. Shen, Y. Yu, J. Hou, Q. Wang, L. Zhang, and X. Meng, "The Testing Research on Prefabricated Building Indoor Thermal Environment of Earthquake Disaster Region," *Procedia Engineering*, vol. 205, pp. 453-460, 2017, doi: 10.1016/j.proeng.2017.10.406.

[5] R. S. Xin, H. L. An, S. Ren, J. T. Yao, and J. Pan, "Behaviors of Continuous Cooling Transformation and Microstructure Evolution of a High Strength Weathering Prefabricated Building Steel," *Solid State Phenomena*, vol. 279, pp. 21-25, 2018, doi: 10.4028/www.scientific.net/SSP.279.21.

[6] Z. Dong, G. Wu, X. L. Zhao, H. Zhu, and X. Shao, "Behaviors of hybrid beams composed of seawater sea-sand concrete (SWSSC) and a prefabricated UHPC shell reinforced with FRP bars," *Construction and Building Materials*, vol. 213, no. 20, pp. 32-42, 2019, doi: 10.1016/j.conbuildmat.2019.04.059.

[7] L. Zhao, Z. Liu, and J. Mbachu, "Development of Intelligent Prefabs Using IoT Technology to Improve the Performance of Prefabricated Construction Projects," *Sensors*, vol. 19, no. 19, pp. 4131, 2019, doi: 10.3390/s19194131.

[8] B. G. Hwang, M. Shan, and K. Y. Looi, "Key constraints and mitigation strategies for prefabricated prefinished volumetric construction," *Journal of Cleaner Production*, vol. 183, no. 10, pp. 183-193, 2018, doi: 10.1016/j.jclepro.2018.02.136.

[9] G. Wu, R. Yang, L. Li, X. Bi, B. Liu, and S. Li, "Factors influencing the application of prefabricated construction in China: From perspectives of technology promotion and cleaner production," *Journal of Cleaner Production*, vol. 219, no. 10, pp. 753-762, 2019, doi: 10.1016/j.jclepro.2019.02.110.

[10] G. Shi, H. Yin, and F. Hu, "Experimental study on seismic behavior of full-scale fully prefabricated steel frame: Global response and composite action," *Engineering Structures*, vol. 169, no. 15, pp. 256-275, 2018, doi: 10.1016/j.engstruct.2018.05.052.

[11] T. Tan, K. Chen, F. Xue, and W. Lu, "Barriers to Building Information

Modeling (BIM) implementation in China's prefabricated construction: An interpretive structural modeling (ISM) approach," *Journal of Cleaner Production*, vol. 219, no. 10, pp. 949-959, 2019, doi: 10.1016/j.jclepro.2019.02.141.

[12] N. Sebaibi, and M. Boutouil, "Reducing energy consumption of prefabricated building elements and lowering the environmental impact of concrete," *Engineering Structures*, vol. 213, pp. 110594, 2020, doi: 10.1016/j.engstruct.2020.110594.

[13] E. Tsangouri, O. Remy, F. Boulpaep, S. Verbruggen, G. Livitsanos, and D. G. Aggelis, "Structural health assessment of prefabricated concrete elements using Acoustic Emission: Towards an optimized damage sensing tool," *Construction and Building Materials*, vol. 206, no. 10, pp. 261-269, 2019, doi: 10.1016/j.conbuildmat.2019.02.035.

[14] J. Ren, S. Deng, and K. Wei, "Mechanical property deterioration of the prefabricated concrete slab in mixed passenger and freight railway tracks," *Construction and Building Materials*, vol. 208, no. 30, pp. 622-637, 2019, doi: 10.1016/j.conbuildmat.2019.03.039.

[15] G. Li, D. Huang, C. Sun, and Y. Li, "Developing interpretive structural modeling based on factor analysis for the water-energy-food nexus conundrum," *Science of the Total Environment*, vol. 651, pp. 309-322, 2019, doi: 10.1016/j.scitotenv.2018.09.188.

[16] Y. Li, B. Sankaranarayanan, D. T. Kumar, and A. Diabat, "Risks assessment in thermal power plants using ISM methodology," *Annals of Operations Research*, vol. 279, no. 1-2, pp. 89-113, 2019, doi: 10.1007/s10479-018-3121-7.

[17] S. G. Azevedo, T. Sequeira, M. Santos, and L. Mendes, "Biomass-related sustainability: A review of the literature and interpretive structural modeling," *Energy*, vol. 171, no. 15, pp. 1107-1125, 2019, doi: 10.1016/j.energy.2019.01.068.

[18] Z. Y. Sun, J. L. Zhou, and L. F. Gan, "Safety assessment in oil drilling work system based on empirical study and Analytic Network Process," *Safety Science*, vol. 105, pp. 86-97, 2018, doi: 10.1016/j.ssci.2018.02.004.

[19] K. J. Wang, Y. D. Lestari, and V. N. B. Tran, "Location Selection of High-tech Manufacturing Firms by a Fuzzy Analytic Network Process: A Case Study of Taiwan High-tech Industry," *International Journal of Fuzzy Systems*, vol. 19, no. 5, pp. 1-25, 2017, doi: 10.1007/s40815-016-0264-z.

[20] A. Francisco, H. Truong, A. Khosrowpour, J. E. Taylor, and N. Mohammadi, "Occupant perceptions of building information model-based energy visualizations in eco-feedback systems," *Applied Energy*, vol. 221, pp. 220-228, 2018, doi: 10.1016/j.apenergy.2018.03.132.

[21] Z. P. Song, G. L. Shi, J. B. Wang, et al., "Research on management and application of tunnel engineering based on BIM technology," *Journal of Civil Engineering and Management*, vol. 25, no. 8, pp. 785-797, 2019, doi: 10.3846/jcem.2019.11056.

[22] M. Marzouk, and A. Othman, "Modeling the performance of sustainable sanitation systems using building information modeling," *Journal of Cleaner Production*, vol. 141, no. 10, pp. 1400-1410, 2017, doi: 10.1016/j.jclepro.2016.09.226.

[23] J. Yuan, X. Li, X. Xiahou, N. Tymvios, Z. Zhou, and Q. Li, "Accident prevention through design (PtD): Integration of building information modeling and PtD knowledge base," *Automation in construction*, vol. 102, pp. 86-104, 2019, doi: 10.1016/j.autcon.2019.02.015.

[24] Z. Liao, "Application of BIMTechnologyin Design and Construction of Assembled Buildings," *Paper Asia*, vol. 34, no. 4, pp. 47-50, 2018.

Xu Guiming

Xu Guiming was born in August 1967 in Jiangyan District, Taizhou City, Jiangsu Province, China. In 1989, he obtained a bachelor's degree from Chongqing University, China, and in 2004, he obtained a master's degree in engineering from Hohai University, China. Now, he is a teacher at Changzhou Engineering Vocational and Technical College in China. His research interests include building information modeling , electromechanical engineering installation and computer engineering drawing, etc.

TD²SecIoT: Temporal, Data-Driven and Dynamic Network Layer Based Security Architecture for Industrial IoT

Dawit Dejene[1], Basant Tiwari[1], Vivek Tiwari[2]*

[1] Department of Computer Science, Institute of Technology, Hawassa University (Ethiopia)
[2] Computer Science & Engineering, Dr. S P Mukherjee IIIT-NR, Naya Raipur, Chhattisgarh (India)

UNIR
LA UNIVERSIDAD
EN INTERNET

Abstract

The Internet of Things (IoT) is an emerging technology, which comprises wireless smart sensors and actuators. Nowadays, IoT is implemented in different areas such as Smart Homes, Smart Cities, Smart Industries, Military, eHealth, and several real-world applications by connecting domain-specific sensors. Designing a security model for these applications is challenging for researchers since attacks (for example, zero-day) are increasing tremendously. Several security methods have been developed to ensure the CIA (Confidentiality, Integrity, and Availability) for Industrial IoT (IIoT). Though these methods have shown promising results, there are still some security issues that are open. Thus, the security and authentication of IoT based applications become quite significant. In this paper, we propose TD²SecIoT (Temporal, Data-Driven and Dynamic Network Layer Based Security Architecture for Industrial IoT), which incorporates Elliptic Curve Cryptography (ECC) and Nth-degree Truncated Polynomial Ring Units (NTRU) methods to ensure confidentiality and integrity. The proposed method has been evaluated against different attacks and performance measures (quantitative and qualitative) using the Cooja network simulator with Contiki-OS. The TD²SecIoT has shown a higher security level with reduced computational cost and time.

Keywords

Cooja, Cryptography, Elliptic Curve Cryptography (ECC), IoT, NTRU, Security, CIA, Dynamic Network Layer.

I. Introduction

THE Internet of Things (IoT) is emerging as a novel thought for future technological advancement, which allows interaction between the physical and virtual world. IoT devices play the role of internet devices and provide mountainous improvement in Information and Communication Technology (ICT) applications [1]. It is predicted that total IoT devices will be 24 billion worldwide in the coming year, which will be connected through the internet [1]. IoT provides exclusive identity to objects that are accessible from the network for tracking down the status and location of the devices [2]. Later on, there will be numerous sorts of things conceivably interfacing with the internet such as heterogeneous industrial machines, self-driving cars, robots, smart supermarkets, mobile phones, shoes, plants, watches, etc. [3], [4].

IoT is applicable in various areas of fields such as Smart Homes, Smart Cities, Smart Industries, Military, eHealth, and several real-world applications by connecting domain-specific sensors [4] and more specifically known as Industrial IoT [2],[4]. IoT also provides tracking, monitoring, and controlling of devices and their services. This influences the minimum human interactions with the attached physical articles to work accordingly. IoT uses Wireless Sensor Network (WSN) infrastructure and technologies such as Radio Frequency ID Devices, IR sensors, laser scanner, Global Positioning System (GPS), etc. [4],[6]. The collected data through the IoT system is real-time data. IoT enables devices (things) to interact and coordinate with each other to perform domain-specific tasks that result in the less human intervention [7].

The Industrial IoT (IIoT) is composed of wireless sensors to control several processes remotely that are constantly connected with physical objects embedded in industrial devices. These devices are vulnerable for different security attacks like Distributed Denial of Services (DDoS), Man in the Middle attack (MiTM) targeted on power grids, hacking of industrial control systems, etc. [8]. These attacks are increasing and have a significant impact on the progress of industries. The ubiquitous nature of IIoT expands the scope of risks, especially when intersecting with hackers or illegitimate users. For example, in healthcare IIoT, riding insulin pumps or pacemakers is a striking concern. Similarly, in the food supply chain, it is a big concern that any agricultural drone surveying a farmland strike may come under the hacker control , etc.[6]. Thus, creation, movement, and processing of goods and services of industries need to be secured.

It is estimated that by 2023 the industrial IoT may have a 14.2 trillion growth to the global economy [9]. Hence, ensuring the CIA (Confidentiality, Integrity, and Availability) of IoT devices is unquestionable. This security is related to information that is either

* Corresponding author.

E-mail address: viveknitbpl@gmail.com

stored or in transit. Therefore, it is vital to have a systematic study and research on designing and improving the security problems of IoT [10]. The standard IoT systems are composed of four main layers known as Information Generation (Perception), Information Transmission (Network), Information Management (Processing), and Integrated Application layer. Each layer is vulnerable for attacks associates with it [11]. An attacker can send a malicious code that can affect the integrity of the data. The Network layer is composed of routers, computers with wireless or wired networks with greater vulnerability for attacks. This includes hacker intrusion and network content security, Sybil, sinkhole, sleep deprivation, code injection, and man-in-the-middle attacks commonly pertain to this layer. The Application layer is a layer where smart environments are achieved. This layer ensures the CIA of data in real-time. Code injection, DoS, spear-phishing, and sniffing are attacks associated with the application layer [7].

Nowadays, actuators in IIoT service infrastructure are also added into the production floor that must be authenticated before the transmission of sensed data [12]. This is a safeguard for data that should be initiated from authentic devices rather than a fraudulent source. Thus, it is essential to ensure the IIoT devices and underlying data to provide data integrity and confidentiality. The solution must protect firmware (program enabled machines) against malicious activities to take less computational time as well as battery power. This work ensures a comprehensive package of security solutions, including confidentiality and integrity requirements, with rigid security components for IIoT devices. There is no universal agreed-upon common security mechanism available to implement in the device [13]. In this view, the manufacturer applies the techniques in devices which is available with them. This limitation leads to immature standards to protect IIoT devices. Moreover, this leads to little or no encryption for protecting sensed data either at the storage or in transit. This opens the way for security threats that include various attacks. There is some industrial process that is more vulnerable such that Human/Machine interface, industrial control system, Machine logic controller and industrial SCADA system, etc.

The securing IoT system is complex as well as challenging since IoT applications are distributed and heterogeneous. In addition, the centralized cloud services cannot cover attack detection, due to resource restrictions, distribution, scalability, latency, and mobility of devices, etc. [14]. Thus, securing IoT requires to be design strong encryption and authentication method and protocol with cryptographic techniques [7].

Proposed work deals with security augmented IoT architecture for ensuring data integrity and authentication in IoT industrial applications. This integrates ECDSA and NTRU cryptosystems using a co-design method, which overcomes the shortcomings of other traditional algorithms. The proposed model employs a dynamic key cryptosystem which uses temporality and dynamicity techniques to ensure confidentiality and integrity. The combination is needed to achieve the stringent performance in IIoT infrastructure.

This paper discussed the basic concepts of IoT and security problems. Section II reviews related works and literature on IoT security. Section III discusses and illustrates the proposed TD²SecIoT architecture in detail. Section IV deals with simulation setup and configurations. Section V presents experimental results and comparative analysis of the proposed work with existing research work based on key generation, authentication, and encryption time and power (energy) usage. At the same time, Section VI deals with security analysis of the proposed work. Section VII concludes the work by discussing the major findings, future works, and recommendations.

II. Related Work

Several applications of cryptosystems are proposed in recent years for ensuring security in resource-constrained IoT devices and applications [15]. Most of the current methods developed to secure IoT are based on symmetric and asymmetric cryptographic algorithms. Some of the researchers also employ these algorithms in a hybrid fashion [16].

S. B. Sadkhan, and Zainab Hamza [16] introduced various cryptosystems for IoT environment. They reported that RSA is the slowest to decrypt and takes the utmost time to encrypt as compared to other methods like ECC, and also demands more memory space. Blowfish takes less time among all for both encryption and decryption and demands the least memory. Blowfish is effective on software but depends on the platform on which it executes. AES and DES require optimal memory space as compared with RSA. Both AES and Blowfish have high unpredictability in the output information and are less vulnerable to attacks. D. Rivera *et al.* [17] proposed a security protocol that employs AES and SHA-3 for IoT smart toy platform. The author focused on confidentiality, integrity, and authenticity for the data generated by the smart toys. Their proposed platform involved different modules of the system and also have concluded that denial of service attacks on the smart IoT toys might be possible. V. C. Ashan [18] modified Welch Gong (WG-7) stream cipher method, which employs an involution function block to increase the security of data. It is a cryptosystem that is widely used in private networks to secure the data from several attacks, including denial of service. The author reported that this cryptosystem is insecure against algebraic types of attacks.

L. Ding *et al.* [19] proposed a modified WG-8 algorithm that is a variant of a known WG stream cipher method. This method has a relatively simple structure and uses 80 bits of secrete key and 80 bits of initial vectors (IV) as an input. The author figured out that these keys are generated in linear and nonlinear relationships. Hence, it is easily identified by the attacker in less than three minutes and vulnerable to key attacks. T. Suzaki *et al.* [20] stated that TWINE is another lightweight block cipher that supports 80 and 128-bit keys. This algorithm is platform-dependent, so it is efficient on the unification of encryption and decryption on software. The authors analyzed the security of the method against impossible deferential and saturation attacks, and they found that its security level should be improved in the future. William Diehl et al. [21] reported that TWINE has the highest Throughput-to-Area (TP/A) ratio as compared with the other stream ciphers. L. Singh and M. Singh [22] stated Elliptic Curve Cryptography (ECC) is a public-key cryptography that uses the curve equation $y^2 = x^3 + ax + b$. This is known as Weierstrass equation, with $4a^3 + 27b^2 \neq 0$, where a, b are constants. ECC can be implemented in various tasks of public-key cryptography like encryption, authentication, digital signature, and key agreement.

T. Daisy *et al.* [23] stated and defined the ECC equation degree and functionality. They conclude that higher-order encryption and decryption is possible provided that the equation should satisfy the standards proposed by NIST [24]. A. Bansal [25] proposed a security system based on the integration of ECC with the cloud. The author used ECC only to encrypt some bits of the data block in order to reduce computational cost on the client-side, which increases resource sharing in the cloud. The author reported that this security method is more effective than RSA with regard to computational cost. S. Singh. *et al.* [26] carried out a comparison between ECC and RSA after a thorough experimental analysis. They concluded that ECC is superior over the RSA method in terms of the ciphertext sizes, digital signature signing, and verification time. This makes ECC ideal for replacing the RSA method. S. Shankar *et al.* [27] proposed ECC based secure key distribution model for secure medical data transmission and resolving

replay attacks by using timestamp values. P. Dhillon and S. Kalra [28] stated it is 10,000 times more difficult to break ECC as compared to an equivalent 2048-bit RSA without degrading the system performance. H. Hasan *et al.* [29] proposed BROSMAP protocol based on ECC for securing IoT. The authors stated that ECC is four times faster than RSA because it uses a smaller key size. M. Ahmed. *et al.* [30] proposed a payload based mutual authentication method that used the Advanced Encryption Standard (AES) with 128 bits key. P. Patil *et al.* [31] reported AES utilizes more computational resources. This indicates that a lightweight model is necessary to have a secure IoT environment.

M. Bunder *et al.* [32] reported that RSA is the first public-key cryptosystem. D. John *et al.* [33] proposed a modified RSA algorithm called binary RSA encryption for securing data in transit in wireless networks. A. Mektoubi *et al.* [34] proposed a hybrid algorithm that combined RSA and ECC methods. RSA is used to encrypt/decrypt the data and the ECC for digital signature, which exploits MQTT (Message Queuing Telemetry Transport) protocol. J. Wang *et al.* [35] discussed that the security of ECC is mainly based on the intractability of the Elliptic Curve Discrete Logarithm Problem (ECDLP). The author also stated that it is easy to calculate another point on the elliptic curve with point-multiplication defined on the finite field. Thus, it may not be appropriate to use for data integrity.

S. Singh *et al.* [36] and R. Tso and Y. Jheng [37] proposed the literature of the Nth-degree Truncated Polynomial Ring Units (NTRU) algorithm that is based on the difficulty of finding particularly small vectors in lattices and degree N with coefficients in the domain of $Z[X]/(X^N-1)$. The authors stated that it is the first system that does not depend on complicated mathematical problems. R. Chen and D. Peng [38] proposed an NTRU based authentication scheme that is used in the handover process of the wireless network. The authors stated that the conventional methods are unsafe with regard to quantum attacks, whereas the NTRU is a lightweight public-key cryptography algorithm based on lattice theory, which is powerful against quantum adversaries. Authors concluded that NTRU has many distinct benefits, such as less memory and computing requirements, higher encryption, and decryption. Y. Gao, X. Yin, and C. Hao [39] have conducted a security analysis to prove that NTRU is secure under the assumption of the NTRU lattice Approximate Closest Vector Problem. Further, R. Hassan and M.G Nadia. [40] described the NTRU algorithm that is based on the mathematical technique called polynomial algebra. R. Jha and A. K. Saini [41] conducted a comparison analysis and concluded that the NTRU algorithm is better than RSA and ECC. The authors concluded that NTRU security does not rely on the integer factorization problem or discrete logarithm problem. They mentioned that NTRU requires O(n log n) time, whereas the RSA, ECC, etc. requires O(n³) time. This makes NTRU ideal for constrained devices such as IoT sensors nodes and sensor-embedded devices.

III. Proposed Work

Related work depicts that privacy and confidentiality should be considered as a fundamental security aspect of IoT and its applications. Cryptographic techniques are used to secure sensor's data and data in transmission. Traditional cryptographic algorithms such as AES, DES, Blowfish, 3DES provided admirable security, but they demand higher computational cost. ECC and NTRU, on the other hand, are ideal for constrained devices such as IoT sensors and sensor-embedded devices. Several security methods have been developed to ensure the confidentiality and integrity of the data for IoT applications. There are some issues with them:

- Most of the methods employ complex mathematical calculations and are based on prime factorization. Moreover, these methods require higher computational resources, time, and even are

susceptible to IoT related attacks. For instance, zero-day attacks and several variants of existing attacks are the main issues of the applications.

- It is quite important to consider the sensors power, memory, etc. since it is employed at resource-constrained devices and in real-time operations. Hence, security models should demand low power, memory, and execution time without compromising security.

The proposed work TD²SecIoT makes it different from state-of-the-arts on the following points:

- Mutual authentication is made between two communicating parties without assistance from a certification authority, except the first time when the system is setup. So, the use of the Key Distribution Centre (KDC) is reduced.

- Two motes can generate keys between themselves by sensed data values.

- TD²SecIoT demands a smaller key size with less key generation, authentication, encryption/decryption time, as compared with other existing works. Hence, demand less computation time, energy requirement, and memory.

- It also prevents numerous attacks such as Reply attack, Man in the Middle (MiTM) attack, chosen cipher attack, Quantum attack, and Lattice-Based attack, which makes it efficient for securing resource-constrained IoT sensors and actuators.

Basically, IoT architecture has four layers that are used to manage IoT services. Such conventional architecture lacks security aspects [1], [2], [7]. The proposed TD²SecIoT offers to integrate a new layer called "Security Layer" in between the Perception layer and Network layer, as depicted in Fig. 1.
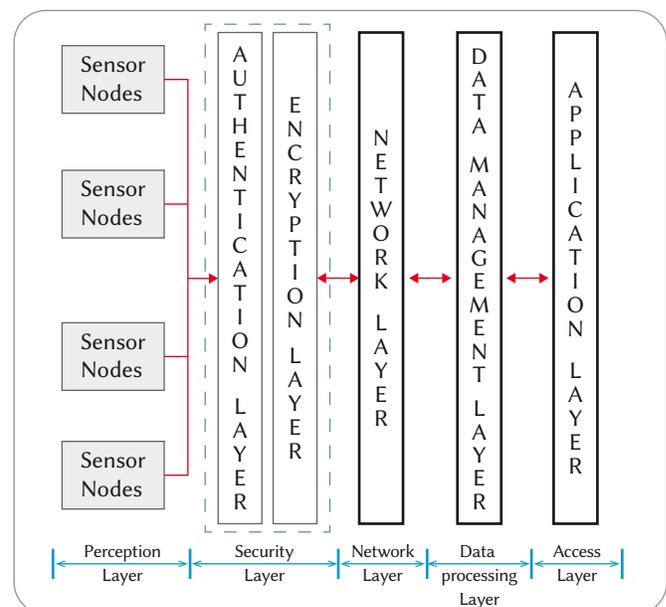
Fig. 1. The proposed architecture.

The security layer is further divided into two sub-layers as it is illustrated in Fig. 1. These layers constitute the Authentication layer; a layer that is responsible for the mutual authentication between the source and receiver nodes, and the Encryption layer; which is responsible for ensuring data integrity and confidentiality during transit. Each sensor node will have both the capabilities since they are embedded in every IoT devices that are deployed in smart environments or any industrial application. In this work, industrial IoT devices are considered for simulating our novel method TD²SecIoT.

The proposed security augmented architecture, which consists of authentication and encryption sub-layer, is discussed in the following sub-sections.

A. Authentication Sub-layer

The TD²SecIoT proposes mutual authentication as one of its tasks, which establishes an authentic two-way communication between the source and receiver node, as depicted in Fig. 2. Once a new node is connected to the network, the third party, called Key Distribution Centre (KDC) is responsible for assigning key pair that will become a certificate for that node. These keys are used during the re-certification at the time of network reset or restart. In this regard, we proposed an ECC based mutual authentication for authentic communication between the source and receiver node.
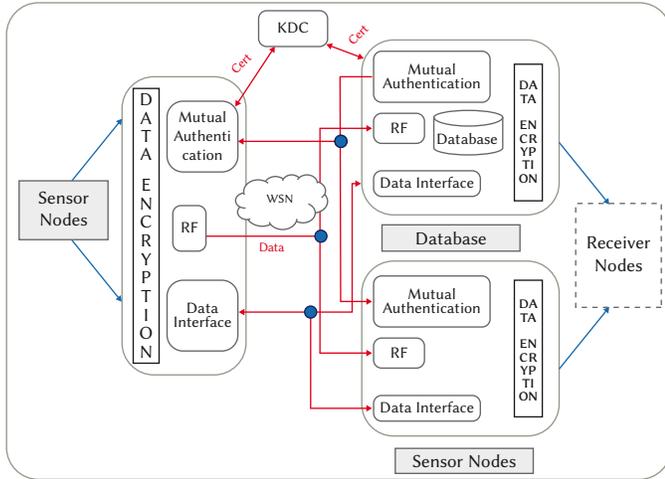
Fig. 2. The proposed Mutual Authentication Process.

1. Mutual Authentication Setup

The mutual authentication process begins by taking a key pair (certificate) from KDC as the network is reset/restarted, or any new node joins the communication. Source node and receiver node both send their public keys Q_{src} and Q_{rcv} to KDC to enable mutual authentication. The KDC uses its private key to generate the certificates and calls the ECDSA method to generate certificates and transfer certificates (r_{src}, s_{src}) to the source node. Similarly, (r_{rcv}, s_{rcv}) are transferred to receiver node with its own public key. The certificate consists of a pair of integers (r_{src}, s_{src}) for the source node and (r_{rcv}, s_{rcv}) for the receiver node. Here, r_{rcv} and s_{rcv} are the x coordinates of the distinct elliptic curve points. These certificates and a public key of KDC are used at the time of mutual authentication process when they send a certificate to each other. They use Q_{ca} (public key of KDC) in the verification phase.

2. Mutual Authentication Between Nodes

As source and receiver nodes are ready to communicate, they need to calculate the hash value using key pairs (obtained from the KDC), and further exchange and verify to each other. If the hash is matched, both devices become authentic and start communication for data sharing. The brief steps of the process are:

The mutual authentication process starts with exchanging public keys Q_{src} and Q_{rcv}, between Source (SRC) and Receiver (RCV) nodes. Further, they generate the secret key by multiplying their own private key to other's public key. The secret key is used to send the hash in encrypted form. Now, the RCV node will send the certificate/hash H_{rcv} (hash of the receiver) to SRC in encrypted form. The SRC then decrypts H_{rcv} and obtains the certificate of RCV node.

Furthermore, the SRC encrypts its certificate/hash and sends H_{src} to RCV, which further decrypts it and obtains the SRC's certificate. Then, RCV checks received hash, and if found identical, then SRC and RCV authenticate to each other and are ready to start communication (IoT data transaction). Otherwise, it rejects the H_{rcv} (send the negative acknowledgment to SRC). It may restart the mutual authentication process if required.

3. Mutual Re-authentication

Mutual Re-Authentication between nodes will happen only when a new node is trying to join the network for the first time or when some failure occurs (network failure, system failure, etc.). This will force each node in the network to re-authenticate each other by exploiting the same authentication procedures, as discussed in subsection A-2.

B. Encryption Sub-layer

This section deals with the integrity of data being transferred among nodes in the IIoT system, which is another task of TD²SecIoT. The data is gotten encrypted at the source node using NTRU approach through a dynamic key, which is generated using the previously stored data in the database, as shown in Fig. 3.
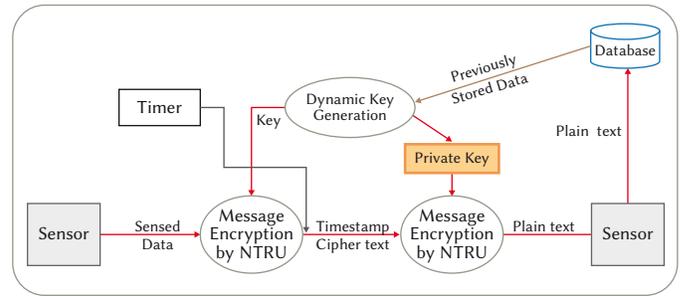

Fig. 3. Encryption Sub-layer.

Another feature of this sublayer is to provide a timestamp to each packet that increases the security level of the proposed system. System time (via timer) is stamped on the ciphertext generated so that the receiver understands that data does not eavesdrop during transmission. This mechanism is useful in identifying replay attacks [27]. Furthermore, timestamped cipher text routes to the receiver node. The communication methods represented by arrows in the Fig. 3 shows the transfer of packets between the sensor and receiver node using radio communication. The receiver node, which is another IoT mote or a server, first delimits timestamped ciphertext to obtain the actual text message and then decrypts the ciphertext using NTRU algorithm. The private key is used for such encryption, which is generated during key generation. Further, decrypted plain text has to be forwarded to the next IIoT layer, where it gets to settle down in repository.

RF propagation time has been used in work for the timestamp calculation [42], which determines the maximum time required for delivery. This timestamp value is used to check a possible reply attack by comparing the timestamp and current time. The Cooja simulation platform [43] facilitates to calculate the shortest distance between the two most distant motes in the network (aka. Network Diameter). The propagation time calculation steps are summarized as follows:

RF propagation time = range / RF wave speed

Where:

    RF wave speed = 3x108 m/s

    RF propagation: maximum time for propagation

    Range = Network Diameter

Finally, the difference in delimited timestamp and current time must be less than or equal to RF propagation time.

C. Algorithmic Procedure For Encryption

Consider, a sender (source node) needs to send confidential data (encrypted by NTRU cryptosystem) to a receiver node. At this point, the source employs the receiver's credential that is a public key, to encrypt the data. The NTRU cryptosystem consists of three phases: System Setup, Key Generation, Encryption, and Decryption that is described in following subsections 1, 2, and 3 with pseudocode. At the end of this section, we also described the key-regeneration process in subsection 4.

1. System-Setup Procedure

Every new node needs to set up three integers (*N, p, q*) and four polynomial sets (L_f, L_g, L_r, L_m) from the ring R = Z[X]/(X^N - 1), where 'N', 'p' and 'q' are prime numbers, and 'q' is considerably larger than 'p' with gcd(*p, q*)=*1*. Usually, 'p' is set to 3 (ternary polynomials), and 'q' for efficiency reasons is generally set to an integer number power of two i.e., q =2^k. The prime condition of product f * f_p =1 (mod p) and f * f_q =1(mod q) must hold in the key generation process.

The following procedure has been used to select the next prime number from any randomly selected sensed data and to assign the value for N. The minimum value of N is set to 167. Shen et al. [44] suggest various security levels and found N= 167 give a better security level. Similarly, for standard, high, and highest security level, 251, 347, and 503 are suggested, respectively.

Input: integer n

Algorithm:

{Set n ← stored random value}

{Set x | x ← 1000}

{Set f | f ← 0}

{Set i | i ← n+1}

{Set j | j ← 2}

For each i loop i to x do

    For each j loop j to j <i do

        If i mod j=0 then

            Set {f ← 1} && break the loop;

        Else

            Return i

        End if

End for

  End for

{Set N ← i}

Output: N (A prime number)

Now, the initial parameter <N, p, q> are passed to the key generation and encryption/decryption operation.

2. Key Generation

This process performs the following procedure to generate keys (public and a private key) that are used for encryption and decryption process:

Input: A mote picks two random polynomials f ∈ Lf, g ∈ Lg

Output: The private key (f, fp) and mote's public key h.

Process:

1. Receiver mote computes:

    fq & fp satisfying fq * f ≡ 1 (mod q) and fp * f ≡ 1 (mod p).

2. Receiver mote computes:

    h = fq *g (mod q), this is a public key, and the private key is (f, fp).

---

3. Encryption and Decryption Operation

Source mote encrypts message m ∈ M with coefficients {-1, 0, 1} using NTRU encryption algorithm. The key generation procedure is described below:

NTRU Encryption algorithm

Input: m ∈ M, h, interval [-p/2, p/2], r, <N, p, q>

1. Select random polynomial 'r' with small coefficient not restricted to the set {-1, 0, 1}

2. Compute

    c = pr * h + m (mod q) = r * h+ m (mod q)

3. Attach timestamp,   c= (c||t)

4. Transfer 'c' to receiver.

Output: Cipher Text 'c'

The receiver mote decrypts ciphertext 'c' using the decryption function of NTRU, the receiver's private key, and parameters established during the setup procedure. The key generation procedure is as follows:

NTRU Decryption algorithm

Input: c, r, interval [-*q/2, q/2*], f

1. Select random polynomial 'r'

2. Compute

    a = f * c (mod q)

    a = f * (pr * h + m) (mod q)

    a = f * (pr * fq * g + m) (mod q)

    a = *pr* * g + f * m (mod q)

3. Calculate

    b = a (mod p) = f * m (mod p)

    , hence *pr* * g (mod p) = 0 & f * f_p = 1

    b = pr * g + f * m (mod p)

    b = f * m (mod p)

    m = fp * b (mod p)

    m= m (mod p)

4. Transfer (m) to receiver or database.

Output: Plain Text 'm'

The aforesaid encryption and decryption procedure are shown diagrammatically in the Fig. 4.
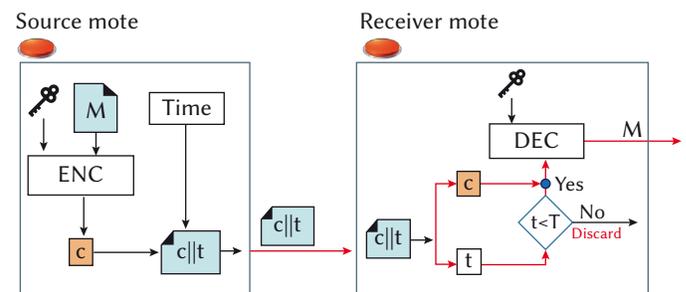

Fig. 4. Encryption and Decryption Process.

The database scheme uses specific sensed data values for a selection of a new key pair for NTRU encryption/decryption. Data exchange between connected nodes will continue with current parameters and keys until a fixed minimum or maximum period is over. Such a key pair is used for NTRU Encryption and Decryption during this interval.

4. Key Regeneration

The new key pair is needed to be generated based on two following factors:

(1) The minimum allotted period is over, and the maximum allotted period is not over. For example, it was decided that one set of keys will be used for a minimum of 7 hours and a maximum of 8 hours. Then a new key pair will be regenerated between this period from the previously stored data in the database server.

(2) Key regeneration gets initiated, when a constantly very small amount of data without significant change is collected or sensed for some continuous period of time. Such continuous similar values for a while (assume 10 minutes) indicates that most of the devices are stopped. It may happen due to employee shift change (since most of the industries are operating in this fashion).

IV. Simulation Setup & Configuration

Contiki-OS (a Linux based operating system) with a Cooja network simulator has been used for implementation [43]. A virtual machine of 64 bits Ubuntu operating system called Instant Contiki-2.7 is installed on a 64bits-x64 based processor. 8Gb RAM and Intel(R) Core (TM) i7-7500 CPU @ 2.90 GHz processor has been used. All the necessary packages and libraries of the Contiki-OS and Cooja simulator are also installed with an updated Java run time environment and open JDK-8 that supports the graphical representation of the Cooja simulator for the visualization of the radio traffic, packet flow, etc. A C-language code is written for encryption/decryption and mutual authentication and is compiled with the MSPGCC compiler (version 4.6.3).

Another useful tool, 'Power Tracker', is used, which is available in the menu Tools with the name "Mote Radio Duty Cycle" [45]. This tool allows us to measure the mean energy consumed by all the nodes and also measure the power consumption individual node wise.

Further, the network is built by inserting nodes in the Network window. We used Sky mote type because it supports 6LoWPAN, which is an advanced version of IPv6 [43]. It works with low-power radio frequency at the physical layer. We have implemented TD²SecIoT using IPv6, because it facilitates IoT devices communication over the Internet separately one at a time. A Unit Disk Graph Medium (UDGM) has been used as a radio frequency model, which takes the ideal transmission range disk in which motes are allowed to share the data [45].

Tmote sky mote has the ability of fast wake up from sleep, which takes less than 6µs, and it is equipped with a 16 bits RISC processor that demands less power during active and in sleep mode [46]. Tmote Sky is equipped with an integrated onboard antenna, with a 50m range indoors and 125m range outdoors, radio (Chipcon CC2420), which provides reliable wireless communication. In this implementation, we use the Sky mote with (8MHz MSP430) low power microcontroller, 10KB RAM, and 48KB flash memory.

Sky motes also provide 250Kbps, 2.4GHz, and IEEE 802.15.4. There are three types of integrated sensors in Tmote Sky, includes Humidity, Temperature, and Light sensors. Tmote Sky provides ST M25P80 40MHz serial code flash for external code and data storage that can hold 1024 kilobytes of data. Sky motes also provide 250Kbps, 2.4GHz, and IEEE 802.15.4. This flash is divided into 16 segments, each with 64kB in size. Initially, a border router and a server are included in the simulated network. Additionally, it has been included five more sensor nodes of the type "Sky Mote" in the simulation to collect the data. We have also chosen the program 'temp-sensor.c', which is part of the Contiki OS and compiled on the sky mote 6. This application is used to sensor the real-time temperature of the environment where the mote is placed.

A. Algorithmic Simulation

A C-language code is written for ECC and NTRU for the mutual authentication and further compiled on each mote in the network. We have deployed five motes as clients, one border router mote with compiled 'border-router.c' as a gateway that uses RPL for routing the packets, and one server mote as a web server (complied on sky motes).

The motes start transferring sensed data once the mutual authentication is successfully completed. This data is encrypted using NTRU cryptosystem, as described in section III. The simulation of the proposed work "TD²SecIoT" using the Cooja network simulator of the Contiki Operating System (OS) with multiple windows is shown in Fig. 5. The left-top corner is the "Network Window", where network topology or motes positioning are shown. It allows us to access every mote in our network. The next window is the ''Simulation Control window," which facilitates 'start', 'pause', make a 'step' forward, and 'reload' the simulation. The window on the top-right corner is used to take notes about the simulation called "Notes Window". The window in the middle is called "Mote output", where the outputs of serial ports of each node are displayed. The lower window is the "Timeline Window|" that shows the data packets delivered over time from the source sensor node to the receiver node. This window is used to investigate the time taken for authentication and encryption/decryption of sensed data. Furthermore, a program called 'tunslip6', home/user/Contiki-2.7/tools/tunslip.c, is used in order to connect the border router with Cooja.
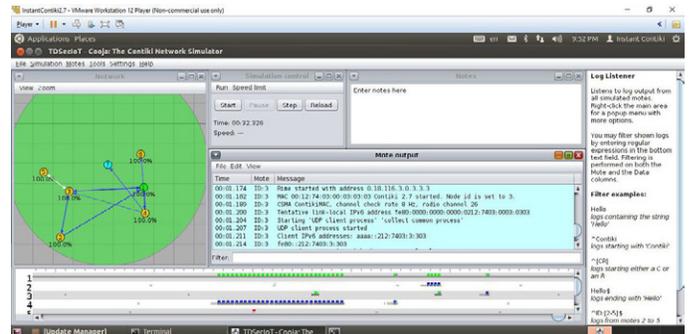
Fig. 5. TD²SecIoT simulation on Cooja.

Table I summarizes the configuration setup, protocols, key exchange times, IP version, and the preferred distance between the motes.

TABLE I. Summary of the Simulation Configuration

Parameters	Value
Motes	Sky
TX Ranges	10m
IP Version	IPV6
MAC Layer	IEEE 802.15.4
Network Layer	6LoWPAN
Radio Access	CSMA
Duty Cycle	Contiki MAC
Maximum Neighbors	5
Key Exchange life	28,700-28,900 sec. time

B. Database Implementation

Contiki has a built-in database management system called "Antelope"; a lightweight SQL-like database manager for resource-

constrained IoT devices, which enables a sensor or a mote to act as a database server [43]. It is a relational database management system that enables the dynamic creation of databases and complex data querying.

Additionally, a centralized database is designed and implemented using Contiki to store the sensed data (received from the sensor mote). This data consists of information about Sensor Nodes (node ID), Node Sequence (node_Seq), Sensed Temperature (node_TempC), and Sensed Light value of the node (node_Light) and it is represented using a relational database as depicted in Table II.

TABLE II. Relational Database Under Contiki

node_ID	node_Seq	node_TempC	node_Light
---	---	---	---

Antelope procedure (Contiki directory /home/user/ Contiki-2.7/examples/antelope/netdb/netdb.csc) has been used to store the sensed data from the sender node to sky server mote6. Furthermore, a python code 'Pydev' is written to get data from the sky mote server and further saves it into the MySQL database. Python needs an extension called 'pySerial' in order to communicate with the serial port of the sky server mote. In this regard, Python extension 'pyserial-3.1.1.tar.gz' has been installed. 'MySQL-python-1.2.4b4.tar.gz' file is utilized to connect the database to the MySQL library. The following commands in MySQL are written to create the database on sky mote6:

CREATE DATABASE mote6; ---(i)

USE mote6; ---(ii)

Create table BarelTempData (NODE_ID INT PRIMARY KEY AUTO_INCREMENT, NODE_SEQ INT (11) NOT NULL, NODE_TEMPC INT (4) NOT NULL, NODE_LIGHT INT (4) NOT NULL) -------------------(iii)

The command (i) creates a database called mote6, command (ii) make mote6 usable for database, and command (iii) creates a table named BarelTempData. This table has four fields Node_id, Node_sequence, Node_tamprature, and Node_light. Finally, we run the code "python serial-mysql.py" for transferring the sensed data and execute the query 'select * from BarelTempData' in MySQL to retrieve the data as shown in Fig. 13.

V. Experimental Results

Fig. 6 depicts the simulation of the mote output. The first four lines show the connection establishment between two motes. Further, the paired server mote waits for the authentication request from the client to process the mutual authentication between motes in the network.
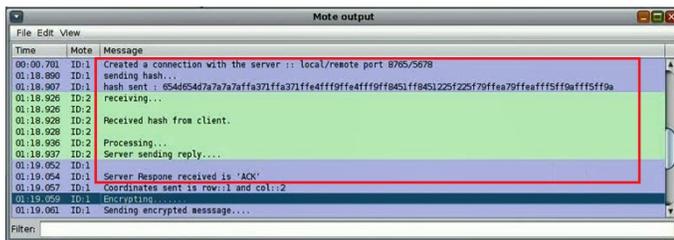
Fig. 6. Connection Establishment of the Motes.

This process takes very less time (in milliseconds) to establish the connection, as indicated in Fig. 6. The further authentication process is shown in Fig. 7.
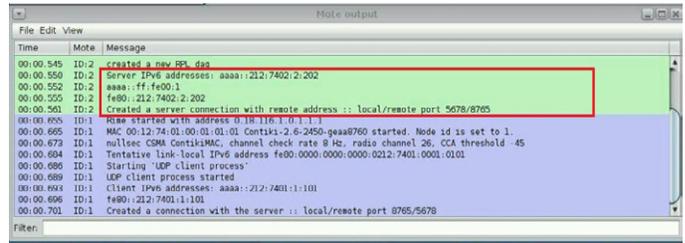
Fig. 7. Mutual Authentication Process.

Fig. 7 depicts the remaining authentication process and the time taken by devices (client, server motes etc.) to get authenticated to each other. Once the connection is established (as shown in Fig. 6), the client sends the authentication request to the server with hash. The server verifies the hash and sends the positive acknowledgment if it is matched. This indicates mutual authentication between motes is successfully established. Motes now share any sensed data by using the NTRU encryption and decryption method.

Fig. 8. Encryption Decryption Process.

Fig. 8 depicts the encryption and decryption of the proposed work and the time taken by the method to complete the processes. The encryption and decryption time of the plain text is 2 ms and 7 ms, respectively, as shown in Fig. 8. It has been observed that the proposed method is faster and more secure than the other methods [22], [47], [48], and [49].

The maximum stack usage in Cooja is the estimated RAM consumption. Figu. 9 depicts the stack usage of mote, where the Y-axis shows the stack usage of a mote in bits, and the X-axis shows the execution time to transfer the data. We have seen that the proposed method is in-line with the mote's memory specification, and the code can be emulated on the actual motes.
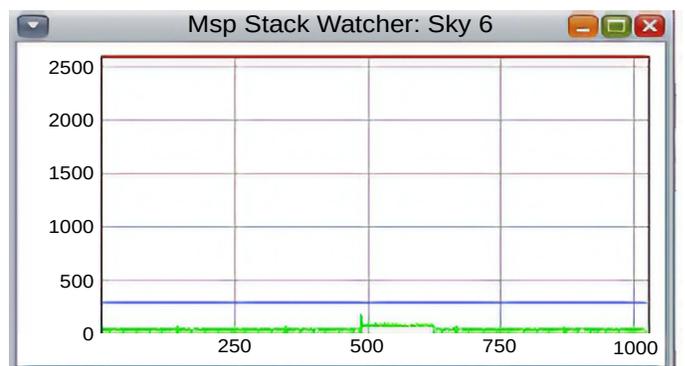
Fig. 9. Stack usage of motes.

Fig. 10 depicts the average power consumption (Milli-watt (mW)) of various states. For example, the energy needed for CPU in sensor motes, Transition energy, Listen energy, and LPM (Low Power Mode). Motes are indicating on the x-axis, and the four colors indicate the power consumption of the motes at different states. The following formula and procedure were employed in the presented research work
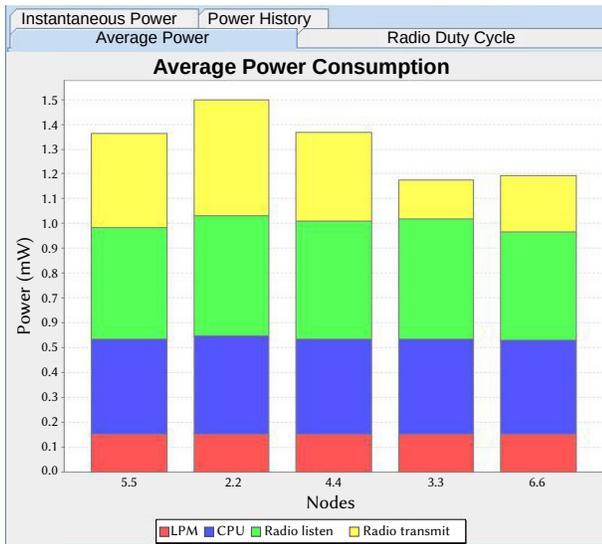
Fig. 10. Average Power Consumption of five sensor nodes.

to calculate the corresponding energy consumption for each one of the above states:

CPU Energy (CP) $= c*1.8tm$

LPM Energy (LP) $= 1*0.545tm$

Transmit Energy (LT) $= t*17.7tm$

Listen Energy (LR) $= r*20tm$

Total Energy $= CP+LP+LT+LR$, and

$tm = c+l$

Where:

tm: total time,

c: CPU time

l: is the time that the sensor was in Low Power Mode (LPM)

t: transmit time

r: Listen time

Based on the results obtained from the simulation of the proposed work (shown in Fig. 10), it is clear that the nodes demand less power to transmit (0.49mw) and receive data (0.35mw). Finally, the whole process in the proposed method takes a total of 1.41mw. This has a direct impact on motes life. The proposed method has reduced the power consumption (71.8%) as compared to the work presented in [48]. It is worth noticing that the proposed method does not compromise the security level.

Fig. 11 depicts the bar graph to show the number of packets received by the motes on the x-axis. The red bar at the y-axis indicates received packets, and the blue bars indicate the retransmitted or duplicate packets. In the simulation, we have seen that there is no packet duplicity (retransmission) to the mote, so the number of packets retransmitted is zero. UDP uses unreliable packet delivery, which results in packet loss due to electrical interference, congestion, or physical disconnection, etc. This indicates that the proposed method reduces congestion and physical disconnection. It is efficient for preventing the retransmission of the packets. Further, it has a great impact on the processing and lifetime of the motes with reliability.

The temperature captured by motes is shown in Fig. 12, where the x-axis indicates the time of sensing and y-axis shows the temperature range in Celsius. This data is stored temporarily in the mote configured with an antelope database. Further, it is sent to MySQL database server after connecting mote's serial port to the database as shown in Fig. 13.
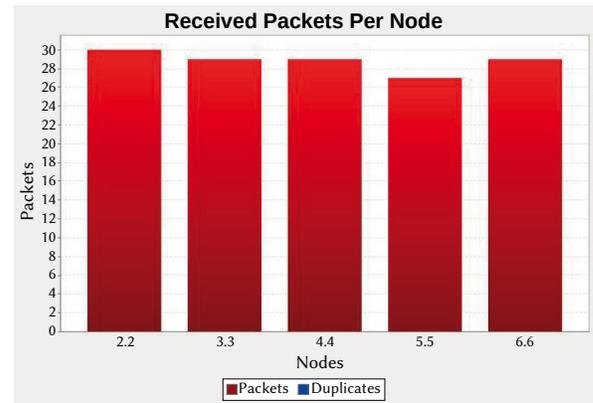

Fig. 11. Number of packets received by each mote.
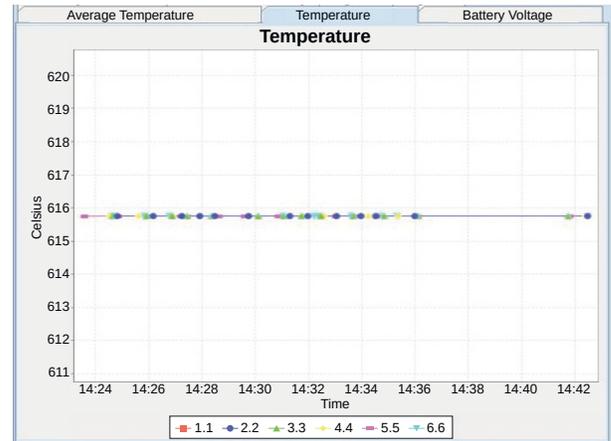
Fig. 12. Sensed temperature values by six motes.

```
+ - - - - - + - - - - - - + - - - - - - - + - - - - - - +
|node_ID | node_Seq |node_TempC |node_Light |
+ - - - - - + - - - - - - + - - - - - - - + - - - - - - +
|     1 |        0 |        616 |        155 |
|     2 |        1 |        616 |        176 |
|     3 |        2 |        616 |        166 |
|     4 |        3 |        616 |        151 |
|     5 |        4 |        616 |        137 |
```

Fig. 13. Antelope database output of temperature values.

In Fig. 12, the doted colors indicate different motes with data at a given time interval. We simulate it with temperature ranges for the operational barrel temperature zones in the first 14:24ms. It has been observed that motes have sensed the higher temperature in degrees Celsius, as it is recorded in Fig. 13. We used this data for generating the dynamic key that elevates the security level of the proposed method TD²SecIoT.

TABLE III. COMPARATIVE ANALYSIS OF TD²SECIOT AND STATE OF THE ART WORKS

Parameters	TD²SecIoT	[22]	[47]	[48]	[49]
Key generation(s)	0.195	0.195	0.20161	NA	0.77
Authentication(s)	0.0215	NA	NA	NA	0.22
Encryption (s)	0.002	0.093	0.41137	NA	NA
Decryption (s)	0.007	0.14	0.37679	NA	NA
Power(mw)	1.41	NA	NA	5.0	NA

NA: Not Available, s: time in second, mw: Milli-watt

Fig. 6, 7 and 8 show timing (in millisecond) at left most of the side, that is converted into seconds for the sake of comparison. This time is used to record key generation time, mutual authentication time, as well as encryption/decryption time. Table III represents the comparative analysis of proposed TD²SecIoT with state-of-the-art works [22], [47], [48], and [49].

The same result can also be visualized as depicted in Fig. 14. It is clearly observed that the proposed model (TD²SecIoT) performed better than other works.
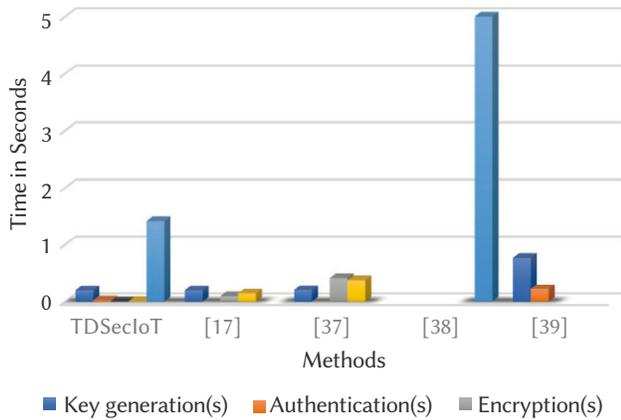

Fig. 14. Comparative Analysis of TD²SecIoT.

VI. Security Analysis & Discussion

The security analysis of the TD²SecIoT against IoT attacks has been analyzed as follows:

Replay Attack: It comes under the network attack where attackers try to delay valid data transmission repeatedly so that it seems fraudulent [50], [51], [52]. This process will take some time to modify or generate new data [53]. TD²SecIoT incorporates timestamp (attached to the ciphertext) to prevent replay attacks. It works because the replay causes a time greater than the maximum propagation time, so the receiver will discard the ciphertext without any further processes.

Man-in-The-Middle Attack: An attacker observes a session opening on a network. Once the authentication process is completed between two parties, it attacks the client computer to immobilize it through IP spoofing (claim the legitimate client) [50], [54]. This attack has been handled under TD²SecIoT, where the SRC sends the message containing its identity directly in an encrypted form. The man in the middle cannot decipher the message, because it does not have the private key of the receiver nor the sensed data values by which keys pairs are generated.

Security against Chosen Cipher Text Attacks: The proposed scheme is secure against chosen-ciphertext attacks [51], [55]. If the attacker anyhow gets the key from the previous ciphertext and tries to decipher the new ciphertext, it cannot be able to decrypt it because the proposed work uses a dynamic and data-driven key feature.

Key-Recovery Attack: It is an attack that attempts to recover the cryptographic key of an encryption scheme [50]. Since we have used a dynamic key that changes with the specified time and is generated using the previously-stored sensed data, it is difficult for the adversary to recover the keys.

Non-Repudiation: Non-repudiation is the assurance that someone cannot deny something [50], [53]. That is, the SRC mote cannot deny that ciphertext is not sent by it. The KDC or receiver mote himself

can verify that the SRC sends the ciphertext through the verification procedure.

Identity Attacks: It is also called 'spoofing.' An attack is trying to gain access to a node by masquerading as an authenticated one [50], [51], [55]. The proposed ECDSA authentication scheme makes it hard to masquerade. Attackers first need to contact KDC to get a certificate and key pair that is used at the time of mutual authentication. Thus, our scheme is more secure against these types of attacks.

Confidentiality: The proposed scheme employs ECC based process to generate key pairs for sharing an encrypted certificate issued by the KDC. It is further used for mutual authentication that ensures all nodes in the network are authenticated. Data transfer between the nodes is possible after encrypting the data, which makes network analysis more difficult for the adversary. Thus, the proposed scheme provides data confidentiality.

Integrity: It ensures that an unauthorized node does not alter information. The proposed scheme uses NTRU Encrypt for encryption and decryption processes to ensure integrity. If an unauthorized node has altered timestamped ciphertext from C to C', it can be revealed during the decryption process by inspecting the timestamp value. This alteration can also be checked at the time of the verification process. It might be denied by the receiver to accept, and hence integrity is ensured.

Several security methods have been proposed by other researchers to ensure confidentiality and integrity against various types of attacks in IoT applications. Some state-of-the-art works have been compared with proposed TD²SecIoT and presented in Table IV. TD²SecIoT offers stronger Confidentiality, Integrity, Unforgeability, and Non-repudiation than existing techniques [23], [27], [56]. The standard ECC and NTRU methods have been employed separately in existing works [23], [27], [56] and were not modified to synchronize with the process. TD²SecIoT process is strong and advanced from others in such a way that it is using NTRU for encryption and ECC for mutual authentication. The aforesaid methods have been modified (at the system setup procedure), which uses previously stored value to generate the key. The existing works [23], [27] do not ensure a replay attack, whereas the proposed work handles it through time stamping. The proposed work ensured security against key recovery attack that is not offered by other researchers [23], [27], [56]. TD²SecIoT offers a dynamic key concept that makes it difficult to recover the keys. The proposed scheme is also secured (through dynamic and data-driven key feature) against chosen-ciphertext attacks, whicht is missing in existing techniques [23], [27], [56].

TABLE IV. Qualitative Comparison of Proposed TD²SecIoT With State of the Art Works

Qualitative Parameters	TD²SecIoT	[23]	[27]	[56]
Confidentiality	Strong	Moderate	Moderate	Moderate
Integrity	Strong	Moderate	Moderate	Poor
Unforgeability	Strong	Moderate	Moderate	Moderate
Non-repudiation	Strong	Moderate	Poor	Moderate
Security against Reply Attack (Timestamping)	Yes	No	Yes	No
Security against Key-Recovery Attack	Yes	No	No	No
Security against Chosen Cipher Text Attacks	Yes	No	No	No
Dynamic key generation	Yes	No	No	No

VII. Conclusion

This paper proposes a security architecture (TD²SecIoT) for ensuring the security of IoT applications. The architecture performs the mutual authentication, encryption, and decryption of the data between motes with less computational cost and higher security levels by integrating public key cryptosystems ECC and NTRU based methods.

The presented system has been evaluated, analyzed, and proved that TD²SecIoT gives better performance as per the security matrices. It has been depicted from the simulation that the proposed method demands a smaller key size and takes less time to key generation, authentication, encryption/decryption, as compared with other existing works. TD²SecIoT ensures integrity and confidentiality with better security with less computational costs. It also prevents various attacks like a replay attack, a man in the middle attack, a chosen cipher attack, a Quantum attack, and a Lattice-Based attack. It makes TD²SecIoT efficient for securing resource-constrained IoT sensors and actuators.

The implementation was on Cooja network simulator and Contiki operating system. The following points have been recommended as future work:

1. Emulation of the code to the actual device.
2. Ensuring the availability issue to prevent against DoS attack.
3. Incorporate IDS/IPS by using an appropriate learning algorithm to secure IIoT.

References

[1] A. R. Sfar, Z. Chtourou, and Y. Challal, "A systemic and cognitive vision for IoT security: a case study of military live simulation and security challenges," Smart, Monit. Control. Cities, IEEE, 2017, pp. 17–19.

[2] Miettinen, M., Marchal, S., Hafeez, I., Asokan, N., Sadeghi, A.R. and Tarkoma, S., 2017, June. Iot sentinel: Automated device-type identification for security enforcement in iot. In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 2177-2184

[3] Frustaci, Mario, Pasquale Pace, Gianluca Aloi, and Giancarlo Fortino. "Evaluating critical security issues of the IoT world: Present and future challenges." IEEE Internet of things journal, vol. 5, no. 4, 2017, pp. 2483-2495.

[4] Zhai, Chuanying, Zhuo Zou, Qiang Chen, Lida Xu, Li-Rong Zheng, and Hannu Tenhunen. "Delay-aware and reliability-aware contention-free MF–TDMA protocol for automated RFID monitoring in industrial IoT." Journal of Industrial Information Integration, vol. 3, 2016, pp. 8-19.

[5] S. N. Swamy, D. Jadhav, and N. Kulkarni, "Security threats in the application layer in IOT applications. In International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), IEEE, 2017 , pp. 477-480.

[6] Cheng, Jiangfeng, Weihai Chen, Fei Tao, and Chun-Liang Lin. "Industrial IoT in 5G environment towards smart manufacturing." Journal of Industrial Information Integration, vol. 10, 2018, pp. 10-19.

[7] S. Vashi, J. Ram, J. Modi, S. Verma and C. Prakash, "Internet of Things (IoT): A vision, architectural elements, and security issues," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 492-496.

[8] Urquhart, Lachlan, and Derek McAuley. "Avoiding the internet of insecure industrial things." Computer law & security review, vol. 34, n. 3, 2018, pp. 450-466.

[9] L. Urquhart and D. McAuley, "Avoiding the Internet of Insecure Industrial Things," Comput. Law Secur. Rev. Elsevier, 2017, pp. 1–17.

[10] Yousefi, A. and Jameii, S.M., 2017, May. Improving the security of internet of things using encryption algorithms. In 2017 IEEE International Conference on IoT and Application (ICIOT), pp. 1-5.

[11] R. Marwat, "Internet of Things." [Online]. Available: https://www.slideshare.net/RehmatMarwat/introduction-to-internet-of-things-45172425. [Accessed: 07-Aug-2019].

[12] R. Elhabob, Y. Zhao, I. Sella, and H. Xiong, "An efficient certificateless

[13] Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. "The industrial internet of things (IIoT): An analysis framework", Computers in industry, Elsevier, 2018, Vol. 101, pp. 1-12.

[14] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," Futur. Gener. Comput. Syst. Elsevier, 2017, pp. 761-768.

[15] S. Vashi, J. Ram and J. Modi, "A Vision, Architectural Elements, and Security Issues" in International conference on I-SMAC, IEEE,2017, pp. 492-496.

[16] S. B. S.- Smieee, "Cryptosystems used in IoT- Current Status and Challenges," ICCIT, IEEE, 2017, pp. 58–62.

[17] D. Rivera, A. Garc, L. Mart, B. Alarcos, and G. Ana, "Secure Communications and Protected Data for a Internet of Things Smart Toy Platform," J. Internet of Things , IEEE, vol. 4662, 2019, pp. 1–11.

[18] V. C. Ashan, "Implementation of WG Stream Cipher with Involution Function," Procedia Technol., vol. 24, 2016, pp. 790–795.

[19] L. Ding, C. Jin, J. Guan, and Q. Wang, "Cryptanalysis of Lightweight WG-8 Stream Cipher," ICCIT, IEEE, vol. 9, n. 4, 2014, pp. 645–652.

[20] T. Suzaki, K. Minematsu, S. Morioka, and E. Kobayashi, "TWINE: A Lightweight Block Cipher," Lect. Notes Comput. Sci., vol. 7707, 2013, pp. 339–354.

[21] W. Diehl, F. Farahmand, P. Yalla, J. Kaps, K. Gaj, and C. Engineering, "Comparison of Hardware and Software Implementations of Selected Lightweight Block Ciphers," International Conference on Field Programmable Logic and Applications (FPL), IEEE, 2017, pp. 1-4.

[22] L. D. Singh and K. M. Singh, "Implementation of Text Encryption using Elliptic Curve Cryptography," Procedia Comput. Sci. Elsevier, vol. 54, n. 1, 2015, pp. 73–82.

[23] T. D. P. Bai, K. M. Raj, and S. A. Rabara, "Elliptic Curve Cryptography based Security Framework for Internet of Things (IoT) Enabled Smart Card," WCCCT, IEEE, 2016, pp. 1–4.

[24] F. Özdemir Sönmez, "A Conceptual Model for a Metric Based Framework for the Monitoring of Information Security Tasks' Efficiency" Procedia Comput. Sci. Elsevier, vol. 54, no. 1, 2019, pp. 181–188.

[25] A. Bansal, "Providing Security , Integrity and Authentication Using ECC Algorithm in cloud storage," International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2017, pp. 1-5.

[26] S. R. Singh, "Performance Evaluation of RSA and Elliptic Curve Cryptography," Trans. Inf. Secur, vol. 65537, IEEE, 2016, pp. 302–306.

[27] S. K. Shankar, A. S. Tomar, and G. K. Tak, "Secure Medical Data Transmission by using ECC with Mutual Authentication in WSNs," Eco-friendly Comput. Commun. Syst. vol. 70, Elsevier, 2015, pp. 455–461.

[28] P. D. K. and S. Kalra, "Elliptic Curve Cryptography for Real Time," Futur. Gen. Comp., IEEE, 2016 pp. 1–6.

[29] H. Hasan, T Salah, D Shehada., "Secure Lightweight ECC-Based Protocol for Multi- Agent IoT Systems," 13th Int. Conf. Wirel. Mob. Comput. IEEE, 2017, pp. 1-8.

[30] M. Ahmad, F. Khan, M. Alam, and M. Usman, "A payload-based mutual authentication scheme for Internet of Things," Futur. Gener. Comput. Syst. Elsevier, 2017, pp. 1028-1039.

[31] P. Patil, P. Narayankar, D. G. Narayan, and S. M. Meena, "A Comprehensive Evaluation of Cryptographic Algorithms: DES ,DES, 3DES, AES, RSA and Blowfish," Procedia - Procedia Comput. Sci. Elsevier, vol. 78, 2016, pp. 617–624.

[32] M. Bunder, A. Nitaj, W. Susilo, and J. Tonien, "A generalized attack on RSA type cryptosystems" Theor. Comput. Sci. vol. 704, Elsevier, 2017, pp. 74–81.

[33] D. John and L. Martin, "Binary RSA Encryption Algorithm,", In International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), IEEE, 2016, pp. 178–181.

[34] A. Mektoubi, H. L. Hassani, H. Belhadaoui, and Rifi, "New approach for securing communication over MQTT protocol A comparaison between RSA and Elliptic Curve," Third International Conference on Systems of Collaboration, IEEE, 2016, pp. 1-6.

[35] J. Wang, K Han, A Alexandridis, Z Zilic, Y Pang, W Wu., "A novel security scheme for Body Area Networks compatible with smart vehicles," Comput. Networks, Elsevier, vol. 143, 2018, pp. 74–81.

Note — at the top of the right column (continuation of [12]): public key cryptography with authorized equality test in IIoT". Journal of Ambient Intelligence and Humanized Computing, Springer, vol. 11, n. 3, 2020, pp. 1065-1083.

[36] S. Singh, S Padhye "Cryptanalysis of NTRU with n Public Keys," In Asia Security and Privacy (ISEASP), IEEE, 2017, pp. 1-6.

[37] R. Tso and Y.-S. Jheng, "Security Analysis of a NTRU-based Mutual Authentication Scheme," In 18th Asia-Pacific Network Operations and Management Symposium (APNOMS), IEEE, 2016, pp. 1-3.

[38] R. Chen and D. Peng, "A novel NTRU-based handover authentication scheme for wireless networks," Futur. gen. comp, vol. 7798, IEEE, 2017, pp. 1–4.

[39] Y. Gao, X. Yin, C. Hao, "Sequential Digital Multi-Signature Scheme Based on NTRU," Int. Conf. Comput. Commun. IEEE, 2016, pp. 236–240.

[40] H. R. Yassein and N. M. Al-saidi, "A Comparative Performance Analysis of NTRU and Its Variant Cryptosystems," In International Conf. on Current Research in Computer Science and Information Technology (ICCIT), IEEE, 2017, pp. 115-120.

[41] R. Jha and A. K. Saini, "A Comparative Analysis & Enhancement of NTRU Algorithm for Network Security and Performance Improvement," In International Conference on Communication Systems and Network Technologies, IEEE, 2011, pp. 80–84.

[42] A. Joshi, S. Dhongdi, K. R. Anupama, P. Nahar and Rishabh Sethunathan," Implementation of Protocol Stack for Three-DimensionalWireless Sensor Network",Procedia Comput. Sci. Elsevier, vol. 54, no. 1, 2016, pp. 193–202.

[43] I. Romdhani, "Cooja_Simulator_Manual." [Online]. Available:https://www.researchgate.net/publication/304572240_Cooja_Simulator_Manual. [Accessed: 07-Aug-2019].

[44] X. Shen, Z. Du, and R. Chen, "Research on NTRU algorithm for mobile java security". International Conference on Scalable Computing and Communications; Eighth International Conference on Embedded Computing, pp. 366-369, 2009, IEEE.

[45] M. J. Kaur and P. Maheshwari, "Building Smart Cities Applications using IoT and Cloud-based Architectures," In International Conference on Industrial Informatics and Computer Systems (CIICS), IEEE, 2016, pp. 1-5.

[46] Tmote Sky: Product Description. [Online] http://www.eecs.harvard.edu/~konrad/projects/shimmer/references/tmote-sky-datasheet.pdf. [Accessed: 11-Aug-2019].

[47] P. K. Panda, "A Hybrid Security Algorithm for RSA Cryptosystem," In International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, 2017, pp. 1-6.

[48] T. K. Goyal and V. Sahula, "Lightweight security algorithm for low power IoT devices," In Int. Conf. Adv. Comput. Commun. Informatics, IEEE, 2016, pp. 1725–1729.

[49] S. Athmani, A. Bilami, and D. E. Boubiche, "EDAK: An Efficient Dynamic Authentication and Key Management Mechanism for heterogeneous WSNs," In Future Generation Computer Systems, Vol. 92, Elsevier,2019, pp. 789-799.

[50] J. Bozzelli, "Temprature sensors," 2014. [Online]. Available: https://www.ptonline.com/columns/how-to-set-barrel-zone-temps. [Accessed: 25-Jun-2020].

[51] J. Sengupta, S. Ruj, and S. D. Bit "A Comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT", Journal of Network and Computer Applications, 149, p. 102481, 2020.

[52] Tiwari, Vivek, and Basant Tiwari. "A Data Driven Multi-Layer Framework of Pervasive Information Computing System for eHealthcare", International Journal of E-Health and Medical Communications (IJEHMC), vol. 10, n. 4, 2019, pp. 66-85.

[53] Choudhary, M., Tiwari, V., & Venkanna, U., "Iris anti-spoofing through score-level fusion of handcrafted and data-driven features", Applied Soft Computing, vol. 91, 106206, 2020 Elsevier. https://doi.org/10.1016/j.asoc.2020.106206,

[54] Yadav, Sonal, Vivek Tiwari, and Basant Tiwari. "Privacy preserving data mining with abridge time using vertical partition decision tree." In Proceedings of the ACM Symposium on Women in Research 2016, pp. 158-164.

[55] Nema, Aditi, Basant Tiwari, and Vivek Tiwari. "Improving accuracy for intrusion detection through layered approach using support vector machine with feature reduction." In Proceedings of the ACM Symposium on Women in Research 2016, pp. 26-31.

[56] Suo, Hui, Jiafu Wan, Caifeng Zou, and Jianqi Liu. "Security in the internet of things: a review." In 2012 IEEE international conference on computer science and electronics engineering, 2012, vol. 3, pp. 648-651.

Dr. Basant Tiwari

Dr. Basant Tiwari, is currently serving as Assistant Professor in Computer Science department at Hawassa University, Ethiopia. He has rich experience in teaching undergraduate and postgraduate classes. He has many international and national publications to his credit in conferences and Journals and edited Handbooks in Springer and IGI Global. He is a Senior Member of IEEE, Senior Member of ACM, CSI and IACSIT. Prof. Basant Tiwari has organized various National and International conferences and delivered invited talks and also chaired the technical sessions. He is reviewer of various reputed International journals and Books. He did his M. Tech. (CSE) from Rajiv Gandhi Technical University, Bhopal and Ph. D. from School of Electronics, Devi Ahilya University, Indore, India. His current area of research is Pervasive computing specially in Healthcare and IoT with Information and Network Security.

Dawit Dejene

Dawit Dejene earned his MSc. in computer science from Hawassa University, Ethiopia in 2018, and is currently a lecturer in department of computer science in the institute of technology. His research interests include network security, IoT security & analytics, Intelligent Security, Big Data Analytics and AI.

Vivek Tiwari

Professor in the Department of Computer Science and Engineering at DSPM IIIT-Naya Raipur, C. G. India. He received the B.Eng. degree (Comp. Sci. & Engg.) from the Rajiv Gandhi Technical University, Bhopal, in 2004, and the M.Tech. degree (Comp. Sci. & Engg.) from SATI, Vidisha (MP) in 2008. He has obtained Ph.D. degree from National Institute of Technology, Bhopal (MA-NIT), India in 2015. Dr. Tiwari is a recipient of Young Scientist Fellowship (MPYSC_2014_814) for the year 2014-2016 by the MPCST (Madhya Pradesh Council of Science & Technology), Govt. of M.P. His broad research interest areas include Data Mining, Data Warehousing, Pattern Recognition, Machine learning, and Predictive analytics.

Machine Learning Based Method for Estimating Energy Losses in Large-Scale Unbalanced Distribution Systems with Photovoltaics

Karar Mahmoud[1,3], Mohamed Abdel-Nasser[2,3*], Heba Kashef[3], Domenec Puig[2], Matti Lehtonen[1]

[1] Department of Electrical Engineering and Automation, Aalto University, Espoo FI-00076 (Finland)
[2] Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona 43007 (Spain)
[3] Department of Electrical Engineering, Aswan University, Aswan 81542 (Egypt)

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

In the recent years, the penetration of photovoltaics (PV) has obviously been increased in unbalanced power distribution systems. Driven by this trend, comprehensive simulation tools are required to accurately analyze large-scale distribution systems with a fast-computational speed. In this paper, we propose an efficient method for performing time-series simulations for unbalanced power distribution systems with PV. Unlike the existing iterative methods, the proposed method is based on machine learning. Specifically, we propose a fast, reliable and accurate method for determining energy losses in distribution systems with PV. The proposed method is applied to a large-scale unbalanced distribution system (the IEEE 906 Bus European LV Test Feeder) with PV grid-connected units. The method is validated using OpenDSS software. The results demonstrate the high accuracy and computational performance of the proposed method.

Keywords

I. Introduction

THE high increase in the demand for electricity has no longer been satisfied by the non-renewable energy sources [1]. The hybrid renewable energy system (e.g. solar and wind turbine) is one of the most employed renewable energy sources to fulfil the high energy demand, in addition to their friendly nature to the environment [2], [3]. The consumption of renewable energy sources has a positive impact on economic growth [4]. Estimation of losses in the distribution systems is affected by the fluctuated output power of renewable energy sources. The impact of photovoltaics (PV) fluctuation cannot be ignored due to its high capacity at the large power grids.

The impact of PV on the electrical distribution network can be analyzed by comparing the output of the system before and after connecting the PV source through different PV scenarios. The effect of PV on the losses and voltages is studied by using DIGSILENT power factory software [5]. The DIGSILENT power factory depends on making a balance between the load and the production of the PV system. When the generation of the PV is more than the load power, the excess power is migrated to the grid. On the other hand, when the load power is more than the PV generation, the excess power is fed by the grid.

In the literature, a day-ahead method for loss estimation depending on insufficient historical data mining is proposed in [6]. This method is based on creating a similar day matrix that is obtained using the statistical analysis of different weather conditions. The impact of the PV system in a low voltage network has been tested through three scenarios [7]. Based on this study, the penetration of the PV system should not exceed 50% with respect to the total load; otherwise, it will contribute to unbalancing voltage and high network losses. For achieving the lowest penetration of the PV, it should be allocated along with the feeders. The deviation in the maximum power point (MPP) in the grid-connected PV system occurrs due to the loss factors that are caused by the various variations in frequency-voltage, irradiance, DC load, and solar cell characteristics [8]. Indeed, the level of penetration of the PV system greatly affects the system losses. In [9], three tests have been applied to different IEEE systems (13, 30 and 69 bus systems) with four different simulation cases. For reducing the losses, the PV system is placed at the bus that is containing the peak value rather than the average value. Implementing different PVs in the systems can contribute to increased energy losses and voltage fluctuation. In [10], the annual energy losses with variant generators are computed, and the impact of different types of DGs at energy losses is analyzed.

Several methods are used for power flow calculations in distribution systems, e.g. Newton Raphson and Gauss-Seidel methods, which are used for non-linear loss calculations [11]-[13]. A fast and accurate method for loss calculation in balanced distribution systems is discussed in [14], which is based on the machine learning techniques. The model is constructed using the Regression Tree technique for various generations of load profiles. Another machine learning method for state estimation of the system using neural methods is discussed

---

in [15]. This method is applied to small-scale balanced distribution systems without renewable energy. In spite of the previous machine learning-based methods could overcome the computational burden of the iterative methods, they were applied to small systems.

Modern distribution systems require simulation algorithms for estimating energy losses with renewable energy sources, such as PV. In this paper, we propose a machine learning-based method for performing real-time simulations for unbalanced power distribution systems with PV. In our approach, the losses of the large scale systems have been calculated in a very short time and high accuracy using a neural network model. Unlike the existing iterative methods, the proposed method can deliver accurate results in a very short time. The proposed method is applied to a large-scale unbalanced distribution system (the IEEE 906 Bus European LV Test Feeder) with a PV grid-connected unit.

The rest of this paper is organized as follows. Section II explains the proposed methodology. Section III presents the results. Section IV concludes the paper and provides some lines of future work.

II. Proposed Method

A. Data Structure and Preparation

In the proposed method, machine learning algorithms are utilized to model the relationship between the input and its corresponding output. In our case, the inputs are active and reactive power profiles of all loads and PVs for the three-phases in the per-unit scale for different time instants.

The active and reactive power profiles for each phase are expressed by (1), and (2), respectively. The P_{ph} matrix involves the values of active power of all PV units and loads for all time instants $\{t_1 ... t_N\}$ in phase $ph \in \{A, B, C\}$ of the distribution system. While the Q_{ph} matrix involves the values of reactive power of all PV units and loads for all time instants $\{t_1 ... t_N\}$ in phase $ph \in \{A, B, C\}$ of the distribution system. NL and N_{PV} stand for the total numbers of loads and PV units in the distribution system, respectively.

On the other hand, the output of the machine learning algorithm is the power losses for all the branches of the distribution system. If we consider that the output is represented by a matrix PL_{ph} as expressed in (3), each element in this matrix $\left(L_{ph}^{t_m} \right)$ represents the total system losses at time instant t_m for ph phase of the distribution system.

Note that a power flow tool is required to construct the output matrix $PL_{ph\ based}$ on the input P_{ph} and Q_{ph} matrices. Indeed, there are available tools that can be used for loss calculations in distribution systems. The OpenDSS supports all frequency domains that relate to the smart grids with renewable energy systems. To validate our approach, we employed OpenDSS software [16] as a benchmark.

B. Constructing A Machine Learning Model

Indeed, there are many machine learning algorithms that can be utilized to model the energy losses in distribution systems, for example, regression trees, Gaussian processes, logistic regression, support vector machines, and XGboost. In this paper, we use neural networks because it is simple to use, the availability of the neural network tools with graphical user interface (GUI), allowing reproducibility of the studied cases.

In short, neural networks could deal with complex systems, and so they are widely used in data modelling and statistical analysis [17]. The training process of the network is performed by adjusting the weights and biases until reaching the minimum threshold. The techniques of learning neural networks are based on minimizing errors between the output and the desired target. There are different learning techniques such as feed-forward backpropagation (where the errors are directed back to the network input until achieving the network goal) and cascade forward backpropagation.

Here, we describe the architecture of the neural network model while highlighting the way to train the model. For this purpose, Fig. 1 shows the utilized neural network model for the IEEE 906 Bus European LV Test Feeder. The model includes one hidden layer (10 neurons and one output layer). Note that the dimension of the input is 907, which represents the loads, and the output is the total system losses. We build the model using the feed-forward neural networks with initial parameters. The learning algorithm for training the network is the Levenberg Marqudarable (LM) algorithm [18], [19] that has high efficiency. In the training phase, the actual losses are compared with the model output until the network stopping goal is accomplished (maximum number of 300 epochs, minimum 5% gradient, 1e-3 goal error). If the set goal is not met, the weights and biases are updated with the learning machine rate until the error is equal or less than the set goal. We use the MATLAB Neural Network Toolbox to construct the neural network model.
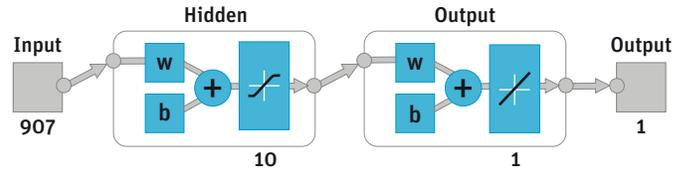

Fig. 1. The architecture of the neural network model (MATLAB 2014a, The MathWorks, Natick, 2014).

$$P_{ph} = \begin{bmatrix} \overbrace{\begin{matrix} P_{PV,1}^{t_1} & P_{PV,1}^{t_1} & \cdots & P_{PV,N_{PV}}^{t_1} \end{matrix}}^{PV\ Active\ Power\ Profiles} & \overbrace{\begin{matrix} P_{LD,1}^{t_1} & P_{LD,1}^{t_1} & \cdots & P_{LD,NL}^{t_1} \end{matrix}}^{Load\ Active\ Power\ Profiles} \\ P_{PV,1}^{t_2} & P_{PV,1}^{t_2} & \cdots & P_{PV,N_{PV}}^{t_2} & P_{LD,1}^{t_2} & P_{LD,1}^{t_2} & \cdots & P_{LD,NL}^{t_2} \\ & & & \vdots \\ P_{PV,1}^{t_N} & P_{PV,1}^{t_N} & \cdots & P_{PV,N_{PV}}^{t_N} & P_{LD,1}^{t_N} & P_{LD,1}^{t_N} & \cdots & P_{LD,NL}^{t_N} \end{bmatrix} \left. \right\} time\ instants \quad \forall\, ph \in \{A, B, C\} \quad (1)$$

$$Q_{ph} = \begin{bmatrix} \overbrace{\begin{matrix} Q_{PV,1}^{t_1} & Q_{PV,1}^{t_1} & \cdots & Q_{PV,N_{PV}}^{t_1} \end{matrix}}^{PV\ Ractive\ Power\ Profiles} & \overbrace{\begin{matrix} Q_{LD,1}^{t_1} & Q_{LD,1}^{t_1} & \cdots & Q_{LD,NL}^{t_1} \end{matrix}}^{Load\ Reactive\ PowerProfiles} \\ Q_{PV,1}^{t_2} & Q_{PV,1}^{t_2} & \cdots & Q_{PV,N_{PV}}^{t_2} & Q_{LD,1}^{t_2} & P_{LD,1}^{t_2} & \cdots & Q_{LD,NL}^{t_2} \\ & & & \vdots \\ Q_{PV,1}^{t_N} & Q_{PV,1}^{t_N} & \cdots & Q_{PV,N_{PV}}^{t_N} & Q_{LD,1}^{t_N} & Q_{LD,1}^{t_N} & \cdots & Q_{LD,NL}^{t_N} \end{bmatrix} \left. \right\} time\ instants \quad \forall\, ph \in \{A, B, C\} \quad (2)$$

$$PL_{ph} = \begin{bmatrix} L_{ph}^{t_1} & L_{ph}^{t_2} & \cdots & L_{ph}^{t_N} \end{bmatrix}^t \qquad\qquad\qquad \forall\, ph \in \{A, B, C\} \quad (3)$$

C. Solution Steps

Fig. 2 shows the steps for loss estimation using a neural network model in distribution systems with PV. The model is constructed using a synthetic dataset (one-month dataset of loads and PV generation). Specifically, we generate the possible scenarios of loads (43200 load points with 15 minutes time step). Another option is to utilize the reliable forecasting models for PV proposed in [20], [21] to generate the datasets. The corresponding power losses of these datastes are computed offline using OpenDSS. The input (load factors and PV power) and output data (power losses) of the OpenDSS are fed into the neural network to construct the model. Once the training process is completed, the model would be ready for solving the power losses for any input data rapidly and accurately without iterative processes employed in state-of-the-art methods. For testing the proposed model, we use a one-day dataset (1440 samples) at six different resolutions.
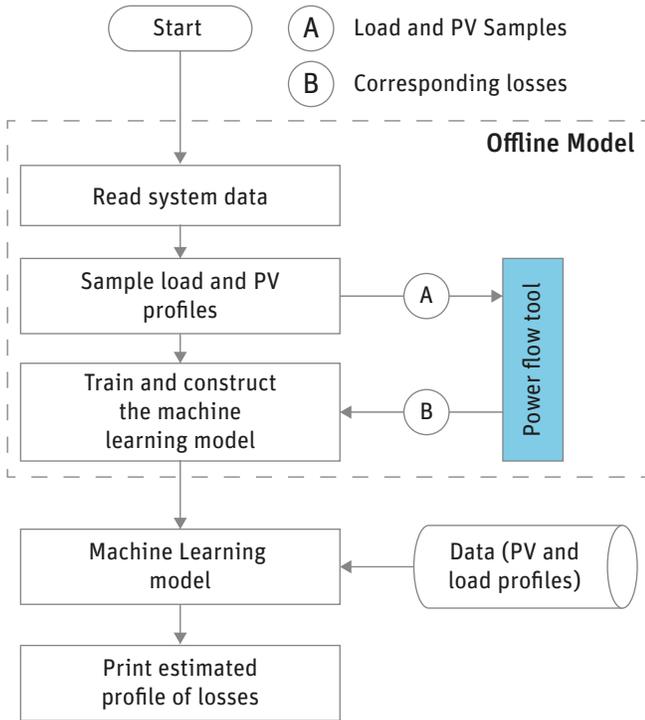

Fig. 2. Flowchart of the proposed method.

Specifically, the following steps are carried out for constructing the offline model and estimating the losses:

a) Read the data of the distribution system interconnected with PV.

b) Read the load and PV states for which the losses are required to be estimated.

c) Generate/read the synthetic dataset and use OpenDSS to calculate the corresponding power losses.

d) Train the neural network and save the model
  ◦ Initialize and create a network model.
  ◦ Train the created network with its parameters, where the input and the output are being compared until reaching the minimum value of error (5% threshold).

e) Call the model for estimating the losses.

f) The output of the network is simulated and compared to the OpenDSS output for assessing the accuracy of the proposed method.

g) Print the analytical and graphical results.

D. Evaluation Metrics

The efficacy of the proposed method is quantified by how close the estimated losses are to the exact ones calculated by OpenDSS. Here, different types of errors are computed:

• The mean square error (MSE) in which the average of the squared difference between the estimated and actual values of the power loss is measured by the following formula:

$$MSE_{ph} = \frac{1}{L}\sum_{t=t_1}^{t_N}(Y_{ph,t}^{dss} - Y_{ph,t}^{nn})^2 \quad \forall ph \in \{A, B, C\}$$
(4)

where $Y_{ph,t}^{dss}$ and $Y_{ph,t}^{nn}$ are the exact and estimated losses of phase *ph* at time instant *t*. *L* represents the number of time instants.

• The root means square error (RMSE), where the square root of the average of the squared values of the difference between the actual and the estimated values are calculated using the following formula:

$$RMSE_{ph} = \sqrt{\frac{1}{L}\sum_{t=t_1}^{t_m}(\left|X_{ph,t}^{dss}\right| - \left|X_{ph,t}^{nn}\right|)^2} \quad \forall ph \in \{A, B, C\}$$
(5)

• The mean absolute error (MAE) where the average difference between the two methods is calculated using the following formula:

$$MAE_{ph} = \frac{1}{L}\sum_{t=t_1}^{t_N}(|X_{ph,t}^{dss} - X_{ph,t}^{nn}|), \quad t = \{t_1,......t_m\}, ph \in \{A, B, C\}$$
(6)

• The mean absolute percentage error (MAPE) where the accuracy of the proposed method is expressed as a percentage defined by the following equation:

$$MAPE_{ph} = \frac{1}{L}\sum_{t=t_1}^{t_N}\left|\frac{X_{ph,t}^{dss} - X_{ph,t}^{nn}}{X_{ph,t}^{dss}}\right| \quad \forall ph \in \{A, B, C\}$$
(7)

• The sum of the squared error (SSE) which is the measure of the scale of variation between the two methods, given by the following equation:

$$SSE_{ph} = \sum_{i=1}^{n}\sum_{t=t_1}^{t_N}(X_{ph,t}^{dss} - X_{ph,t}^{nn})^2 \quad \forall ph \in \{A, B, C\}$$
(8)

• The relative error (RE) which is provided by the following equation:

$$RE_{ph} = \frac{1}{L}\sum_{t=t_1}^{t_N}\frac{X_{ph,t}^{dss} - X_{ph,t}^{nn}}{X_{ph,t}^{dss}} \quad \forall ph \in \{A, B, C\}$$
(9)

All of these errors are calculated for the proposed method to evaluate its accuracy rate compared to the exact one. We have utilized different error formulae (4)-(9) to test the accuracy of the proposed method sufficiently.

III. Results and Discussion

A. Test System and Dataset

Here, the performance of the proposed method is tested for estimating the energy losses on the MATLAB environment. The results are implemented at Intel® Core™ i5-5200U CPU @ 2.20GHz, 4.00 GB RAM, and 64-bit Operating System. The IEEE European low voltage test feeder [22] with 907 bus and 50 Hz frequency (Fig. 3) is used to validate the accuracy and the computational efficiency of the proposed method. Two PV units are connected at busses 639 and 906, and 55 single-phase loads with different daily load shapes (1400 loads points per minute step) are distributed along with the system.
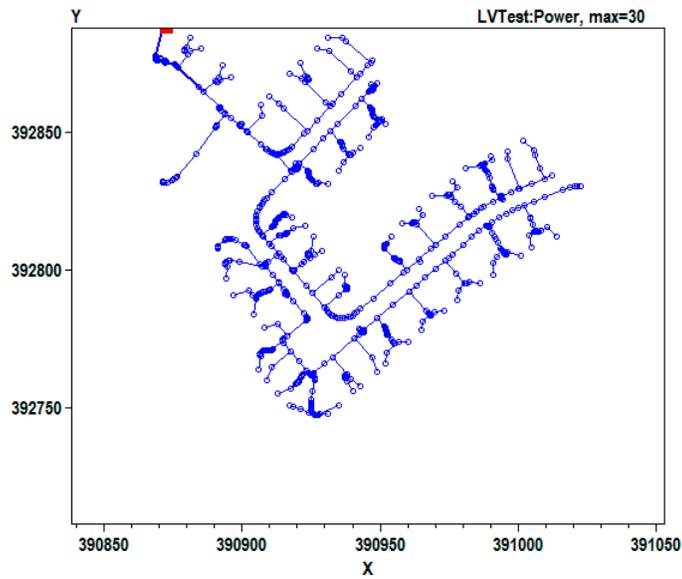
Fig. 3. The single-line diagram of low voltage (LV) test feeder [16].
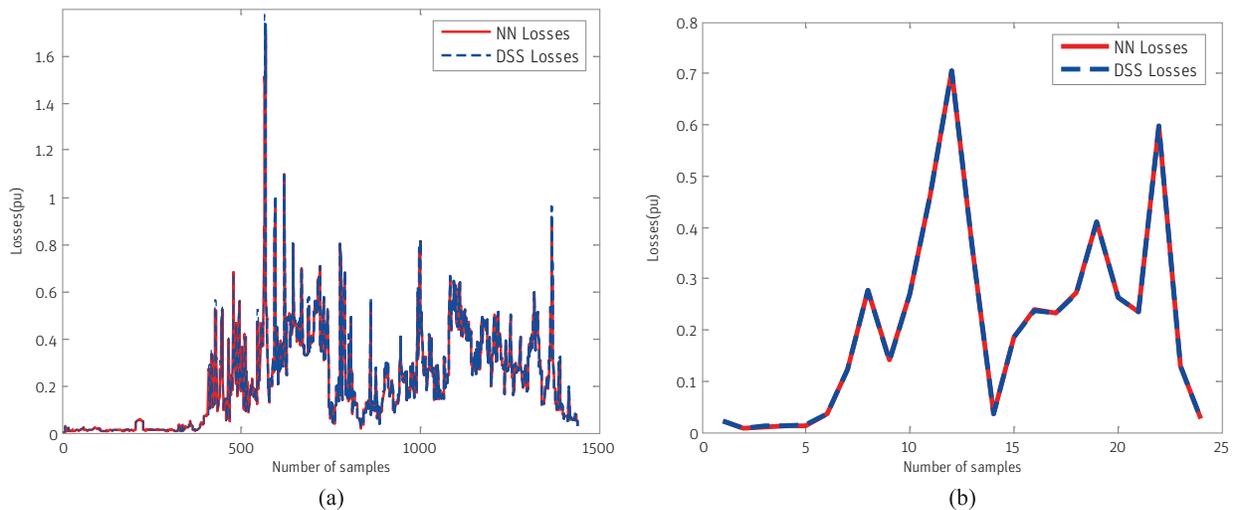
Fig. 4. The estimated losses at different time resolutions using the proposed model and OpenDSS. (a) 1min resolution and (b) 1hr resolution.

To construct the offline loss model, we used a dataset of load and PV generation profiles that contains 43200 samples. For analyzing the performance of the proposed method, we have performed the following experiments:

- The power loss is analyzed for six different time resolutions for phases A, B and C, separately. The estimated losses are compared to the exact power loss of OpenDSS software.
- For accuracy validation, MSE, RMSE, MAE, MAPE, and SSE are calculated for all the three-phases.
- To highlight the computational efficiency of the proposed method, the execution time of the proposed method for estimating losses is computed and compared to the execution time of the exact approach.

B. Performance Analysis

The performance of the proposed method is compared to the exact iterative time-series power flow approach (OpenDSS) for phases A, B, and C. Specifically, we estimate the power loss profile for a day with six different time resolutions (1min, 5min, 10min, 15min, 30min and 1hr). For example, the numbers of samples in datasets for the day per 1 min and 1 hr are 1440 samples (24*60) and 24, respectively.

Fig. 4 shows the estimated losses at 1min and 1hr resolutions using the proposed model and OpenDSS. It is evident that the estimated losses during the day for the two resolutions almost match those of the exact method. Another notice is the loss profile of 1m resolutions differs from the 1hr resolutions, higher fluctuations appear in 1m resolutions than 1hr resolutions. This means that the higher resolution of datasets can sufficiently represent the actual loss profiles in which the PV and load profiles have intermittent nature. However, the computational burden of the existing iterative methods will be increased when higher PV and load datasets are required to be analyzed. To solve this issue, the proposed method can accurately calculate the losses with large datasets (high resolution) in a very short time, thanks to the developed offline model.

Table I, Table II and Table III summarize the values of MSE, RMSE, MAE, MAPE, and SSE for phases A, B, and C, respectively at the six resolutions. It is obvious that the values of all errors are very small with respect to the exact solutions of the losses. Further, they are low at 1h resolution (coarse resolution) while they are high at 1m resolution (finest resolution). The lowest value of MSE appeared at 1h resolution for phase A (6.1190e-07) while the largest value appears at 1m resolution for phase B (3.2222e-04). The same trend is noticed for

MAE, MAPE, and SSE. For the three phases, the RE values are less than 0.05. Note that for this test system, the estimated results of phase A are more accurate than those of phase B and C, but this is not a general rule for distribution systems.

C. Computational Performance of the Proposed Method

For a further description of the contribution of the proposed machine learning-based method, the execution times for solving the losses during the day with six resolutions are computed for the proposed method and exact method (OpenDSS). Table IV shows the computational times of the two methods. The execution time required for the OpenDSS is very long compared to the proposed method. For 1hr resolution, the OpenDSS takes approximately 1.5 sec while the proposed method takes only 0.02 seconds for obtaining the results. The execution time of OpenDSS is greatly increased with the data resolution, for example, in the case of 1min resolution, the OpenDSS takes around 41 sec. However, our proposed method takes less than 0.04 sec.

D. Comparison

To demonstrate the performance of the NN model, we compare it with a support vector regression (SVR) model. Table V, Table VI and Table VII show the MSE, RMSE, MAE, MAPE, SSE and RE values of the SVR model for phases A, B, and C, respectively at the six resolutions. For phase A, the SVR model achieves MMSE of 0.0028 and 3.3389e-04 with 1m and 1h resolutions, receptively. With phase B, it gives MSE of 0.0031 and 3.8189e-04 with 1m and 1h resolution, respectively. In the case of phase C, the SVR model gives a MSE less than 7e-04 with all resolutions. For the three phases, the RE values are less than 0.23, which are much higher than those of the NN model (RE values < 0.05). In general, the comparison between the errors of the NN model shown in Tables I-III and the errors of the SVR model shown in Tables V-VII reveals that the NN model achieves prediction errors much lower than those of the SVR model. Therefore, it seems that the NN model is more suitable for this task.

TABLE I. MSE, RMSE, MAE, MAPE, AND SSE FOR PHASES A BY THE PROPOSED METHOD

Errors	1min	5 min	10min	15min	30min	1hr
MSE	1.1678e-04	5.3016e-05	2.0498e-05	5.5643e-07	4.8525e-07	6.1190e-07
RMSE	0.0108	0.0073	0.0045	7.4594e-04	6.9660e-04	7.8224e-04
MAE	0.2661	0.0579	0.0390	0.0033	0.0017	0.0017
MAPE	0.0305	0.0280	0.0243	0.0189	0.0180	0.0180
SSE	0.1682	0.0153	0.0030	5.3417e-05	2.3292e-05	1.4686e-05
RE	0.0262	0.0232	0.0234	0.0167	0.0166	0.0150

TABLE II. MSE, RMSE, MAE, MAPE, AND SSE FOR PHASES B BY THE PROPOSED METHOD

Errors	1min	5min	10min	15min	30min	1hr
MSE	3.2222e-04	2.0802e-04	1.9646e-04	1.7727e-06	1.5828e-06	1.2248e-06
RMSE	0.0179	0.0144	0.0140	0.0013	0.0013	0.0011
MAE	0.3525	0.1071	0.1071	0.0048	0.0023	0.0024
MAPE	0.0504	0.0449	0.0416	0.0309	0.0295	0.0280
SSE	0.4640	0.0599	0.0283	1.7018e-04	7.5975e-05	2.9395e-05
RE	0.0435	0.0395	0.0348	0.0289	0.0276	0.0258

TABLE III. MSE, RMSE, MAE, MAPE, AND SSE FOR PHASES C BY THE PROPOSED METHOD

Errors	1min	5min	10min	15min	30min	1hr
MSE	2.2134e-04	1.457e-04	1.0609e-04	1.7281e-06	1.0019e-06	8.8531e-07
RMSE	0.0149	0.0121	0.0103	0.0013	0.0010	9.4091e-04
MAE	0.1639	0.1041	0.1040	0.0054	0.0025	0.0019
MAPE	0.0482	0.0431	0.0388	0.0294	0.0262	0.0232
SSE	0.3187	0.0420	0.0153	1.6590e-04	4.8089e-05	2.1248e-05
RE	0.0421	0.0374	0.0338	0.0275	0.0243	0.0211

TABLE IV. COMPUTATIONAL PERFORMANCE OF THE PROPOSED METHOD (TIME IN SECONDS)

Method	1min	5min	10min	15min	30min	1hr
OpenDSS	41.490263	7.521413	3.939896	2.685175	1.570552	1.40819
NN phase A	0.037296	0.018610	0.017173	0.016421	0.014474	0.02584
NN phase B	0.034559	0.014244	0.013253	0.011219	0.012274	0.02462
NN phase C	0.039015	0.014491	0.013141	0.010945	0.021032	0.02186

TABLE V. MSE, RMSE, MAE, MAPE, and SSE for Phases A by SVR

Errors	1min	5min	10min	15min	30min	1hr
MSE	0.0028	0.0014	0.0017	6.3287e-04	8.3239e-04	3.3389e-04
RMSE	0.0533	0.0370	0.0415	0.0252	0.0289	0.0183
MAE	1.1800	0.4249	0.4249	0.1630	0.1626	0.0548
MAPE	0.2061	0.1841	0.1747	0.1528	0.1526	0.1326
SSE	4.0875	0.3936	0.2480	0.0608	0.0400	0.0080
RE	0.1173	0.0988	0.0801	0.0927	0.0932	0.0607

TABLE VI. MSE, RMSE, MAE, MAPE, and SSE for Phases B by SVR

Errors	1min	5min	10min	15min	30min	1hr
MSE	0.0031	0.0012	8.4249e-04	0.0010	4.1261e-04	3.8189e-04
RMSE	0.0555	0.0346	0.0290	0.0324	0.0203	0.0195
MAE	1.1540	0.2367	0.2064	0.2218	0.0800	0.0546
MAPE	0.3222	0.2917	0.2705	0.2671	0.2374	0.2053
SSE	4.4408	0.3455	0.1213	0.1008	0.0198	0.0092
RE	0.2206	0.1858	0.1603	0.1526	0.0892	0.0082

TABLE VII. MSE, RMSE, MAE, MAPE, and SSE for Phases C by SVR

Errors	1min	5min	10min	15min	30min	1hr
MSE	6.6348e-04	4.1257e-04	5.1903e-04	3.0905e-04	3.3014e-04	2.4974e-04
RMSE	0.0258	0.0203	0.0228	0.0176	0.0182	0.0158
MAE	0.5486	0.1541	0.1541	0.0636	0.0635	0.0355
MAPE	0.2707	0.2502	0.2382	0.2292	0.2235	0.2385
SSE	0.9554	0.1188	0.0747	0.0297	0.0158	0.0060
RE	0.2129	0.1986	0.1900	0.1885	0.1839	0.1884

IV. Conclusion

In this paper, an efficient method has been proposed for performing time-series simulations for unbalanced power distribution systems with PV. Unlike the related iterative methods, our proposed method is based on machine learning algorithms. The proposed method has been applied to the IEEE 906 Bus European LV Test Feeder with PV grid-connected units. The proposed method is validated using OpenDSS software. The test of the proposed method has been carried out at six different resolution times (1hr, 30min, 15min, 10min, 5min, and 1min). When comparing between the trained model and OpenDSS, the calculated results have a strong matching. The calculation time required by the OpenDSS for computing losses is too long compared to the proposed method, especially at the high resolution (i.e. 1 min). The experimental results show that the NN model outperforms the SVR model for time-series simulations. The results demonstrate the effectiveness of the proposed method.

The main goal of this work is to show that the application of machine learning is a promising approach to estimate energy loss in large-scale distribution systems. So, the contribution of this work is to demonstrate the applicability this approach while not trying to choose the most proper machine learning technique. Further, we believe that the results of an efficient machine learning technique can yield acceptable results for this application. The future work will be directed to consider diverse renewable energy sources, such as wind turbine generating systems.

References

[1] Said, S. M., Aly, M., Hartmann, B., Alharbi, A. G., & Ahmed, E. M. "SMES-Based Fuzzy Logic Approach for Enhancing the Reliability of Microgrids Equipped With PV Generators," IEEE Access, vol. 7, pp. 92059-92069, 2019.

[2] Ali, Abdelfatah, D. Raisz, K. Mahmoud and M. Lehtonen. "Optimal Placement and Sizing of Uncertain PVs Considering Stochastic Nature of PEVs," IEEE Transactions on Sustainable Energy, vol. 11, no. 3, pp. 1647-1656, July 2020, doi: 10.1109/TSTE.2019.2935349.

[3] R. A. Verzijlbergh, L. J. De Vries, and Z. Lukszo, "Renewable Energy Sources and Responsive Demand. Do We Need Congestion Management in the Distribution Grid?," IEEE Trans. Power Syst., vol. 29, no. 5, pp. 2119–2128, Sep. 2014.

[4] V. Khare, S. Nema, and P. Baredar, "Solar–wind hybrid renewable energy system: A review," Renew. Sustain. Energy Rev., vol. 58, pp. 23–33, May 2016.

[5] X. Yang, M. Xu, S. Xu, and X. Han, "Day-ahead forecasting of photovoltaic output power with similar cloud space fusion based on incomplete historical data mining," Appl. Energy, vol. 206, pp. 683–696, Nov. 2017.

[6] M. Bhattacharya, S. R. Paramati, I. Ozturk, and S. Bhattacharya, "The effect of renewable energy consumption on economic growth: Evidence from top 38 countries," Appl. Energy, vol. 162, pp. 733–741, Jan. 2016.

[7] T. Al Momani, A. Harb, and F. Amoura, "Impact of photovoltaic systems on voltage profile and power losses of distribution networks in Jordan," in 2017 8th International Renewable Energy Congress (IREC), 2017, pp. 1–6.

[8] N. I. Zolkifri, C. K. Gan, A. Khamis, K. A. Baharin, and M. Y. Lada, "Impacts of residential solar photovoltaic systems on voltage unbalance and network losses," in TENCON 2017-2017 IEEE Region 10 Conference, 2017, pp. 2150–2155.

[9] T.-F. Wu, C.-H. Chang, Y.-D. Chang, and K.-Y. Lee, "Power loss analysis of

grid connection photovoltaic systems," in 2009 International Conference on Power Electronics and Drive Systems (PEDS), 2009, pp. 326–331.

[10] S. Daud, A. Fazliana, A. Kadir, and C. K. Gan, S. Daud, A. F. A. Kadir and C. K. Gan, "The impacts of distributed Photovoltaic generation on power distribution networks losses," 2015 IEEE Student Conference on Research and Development (SCOReD), Kuala Lumpur, 2015, pp. 11-15, doi: 10.1109/SCORED.2015.7449305.

[11] V. H. M. Quezada, J. R. Abbad and T. G. S. Roman, "Assessment of energy distribution losses for increasing penetration of distributed generation," in IEEE Transactions on Power Systems, vol. 21, no. 2, pp. 533-540, May 2006, doi: 10.1109/TPWRS.2006.873115.

[12] H. Le Nguyen, "Newton-Raphson method in complex form [power system load flow analysis]," in IEEE Transactions on Power Systems, vol. 12, no. 3, pp. 1355-1359, Aug. 1997, doi: 10.1109/59.630481.

[13] H. Liu and T. Feng, "Study on the Convergence of Solving Linear Equations by Gauss-Seidel and Jacobi Method," in 2015 11th International Conference on Computational Intelligence and Security (CIS), 2015, pp. 100–103.

[14] K. Mahmoud and M. Abdel-Nasser, "Fast-yet-Accurate Energy Loss Assessment Approach for Analyzing/Sizing PV in Distribution Systems using Machine Learning," IEEE Transactions on Sustainable Energy, vol. 10, no. 3, pp. 1025-1033, July 2019, doi: 10.1109/TSTE.2018.2859036.

[15] M. Abdel-Nasser, K. Mahmoud, and H. Kashef, "A Novel Smart Grid State Estimation Method Based on Neural Networks," Int. J. Interact. Multimed. Artif. Intell., vol. 5, no. 1, p. 92, 2018.

[16] Electric Power Research Institute, OpenDSS, Distribution System Simulator [Online], Available: http://sourceforge.net/projects/electricdss/.

[17] Gurney K. An introduction to neural networks. CRC press, 1997.

[18] O. N. A. Al-allaf, "Face Recognition System Based on Different Artificial Neural Networks Models and Training Algorithms," International Journal of Advanced Computer Science and Applications, vol. 4, no. 6, pp. 40–47, 2013.

[19] M. I. A. Lourakis, "A Brief Description of the Levenberg-Marquardt Algorithm Implemened by levmar The Levenberg-Marquardt Algorithm," Foundation of Research and Technology, vol. 4, no. 1, pp. 4–9, 2005.

[20] Abdel-Nasser, M., Mahmoud, K. "Accurate photovoltaic power forecasting models using deep LSTM-RNN," Neural Comput & Applic vol. 31, pp. 2727–2740 (2019). https://doi.org/10.1007/s00521-017-3225-z

[21] Abdel-Nasser M, Mahmoud K, Lehtonen M. "Reliable Solar Irradiance Forecasting Approach Based on Choquet Integral and Deep LSTMs," IEEE Transactions on Industrial Informatics, 2020 doi: 10.1109/TII.2020.2996235.

[22] IEEE PES Distribution Systems Analysis Subcommittee Radial Test Feeders [Online], Available: http://ewh.ieee.org/soc/pes/dsacom/testfeeders.html

Karar Mahmoud

Karar Mahmoud received the B.S. and M.Sc. degrees in electrical engineering from Aswan University, Aswan, Egypt, in 2008 and 2012, respectively. In 2016, he received the Ph.D. degree from the Electric Power and Energy System Laboratory (EPESL), Graduate School of Engineering, Hiroshima University, Hiroshima, Japan. Since 2010, he has been with Aswan University where he is presently Assistant Professor. Currently, he is a Postdoctoral Researcher at the School of Electrical Engineering, Aalto University, Finland. His research interests include Power Systems, Renewable Energy Resources, Smart Grids, Distributed Generation, and Optimization.

Mohamed Abdel-Nasser

Mohamed Abdel-Nasser is currently a postdoc researcher at the University Rovira i Virgili (Spain) and Assistant Professor at the Electrical Engineering Department, Aswan University (Egypt). He received his PhD in Computer Engineering from Universitat Rovira i Virgili (Spain) in 2016. He has published +65 papers in international journals and conferences. His research interests include the application of machine learning and deep learning to several real-world problems, smart grid analysis, and time-series forecasting.

Heba Kashef

Heba Kashef received a B.Sc. degree in Electrical Engineering from the Faculty of Engineering, Aswan University, Aswan, Egypt, in 2012. She is currently working in her Msc studies the Faculty of Engineering, Aswan University, Aswan, Egypt. Her interests include machine learning, smart grid analysis, renewable energy sources.

Domenec Puig

Domenec Puig received the M.S. and Ph.D. degrees in computer science from the Polytechnic University of Catalonia, Barcelona, Spain, in 1992 and 2004, respectively. In 1992, he joined the Department of Computer Science and Mathematics, Rovira i Virgili University, Tarragona, Spain, where he is currently working as a Professor since 2017. He has been the Head of the Intelligent Robotics and Computer Vision Group, Rovira i Virgili University since 2006. His research interests include artificial intelligence and mobile robotics.

Matti Lehtonen

Matti Lehtonen received the M.Sc. and Licentiate degrees in electrical engineering from the School of Electrical Engineering, Aalto University (formerly Helsinki University of Technology), Espoo, Finland, in 1984 and 1989, respectively, and the D.Sc. degree from the Tampere University of Technology, Tampere, Finland, in 1992. Since 1987, he has been with VTT Energy, Espoo, and since 1999, he has been with the School of Electrical Engineering, Aalto University, where he is currently the Head of Power Systems and High Voltage Engineering Department. His main activities include earth fault problems, and harmonic related issues and applications of information technology in distribution automation and distribution energy management.