Explainable Artificial Intelligence-Based Diseases Diagnosis From Unstructured Clinical Data and Decision Making Using Blockchain Technologies

Sumathi M.[1]*, S.P. Raja[2]

[1] School of Computing, SASTRA Deemed University, Thanjavur, Tamil Nadu (India)

[2] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, (India)

* Corresponding author: sumathi@it.sastra.edu

Received 30 September 2022 | Accepted 28 December 2022 | Published 6 February 2025

Abstract

In the digital era, health information is stored in digital form for easy maintenance, analysis and transfer. The proficiency of manual illness diagnosis and drug prediction in the medical field depends on the expertise availability, and experience of the specialists. In emergency and abnormal situation, the patient's life completely depends on expert's availability. Therefore, a different approach is needed to get around the difficulties in managing emergency cases. Artificial intelligence helps to take decisions in an accurate manner but does not provide the details of the decisions. The ability to treat emergency patients entirely depends on the particular hospitals. The clinical data includes numerical results, text prescriptions, scanned images, etc. Therefore, managing unstructured data with care is necessary for making clinical decisions. An explainable artificial intelligence-based disease diagnosis and blockchain-based decision-making system are presented in this work to address these challenges and improve patient care. A natural language processing system analyzes the unstructured data to identify different types of data and explainable AI diagnosis disease with justification and reason for the prediction. An ant colony optimization-based recommender system examines the predicted decision and identifies the specific drug for the disease. The disease decision and drug information are kept in a permissioned blockchain for confirmation. Decisions are validated by more than 50% of the experts present in the permissioned blockchain network, which consists of experts from various regions. As a result, the quickest and most accurate decisions possible are taken to handle emergency situations.

Keywords

Ant Colony Optimization, Blockchain, Clinical Care, CNNn-LSTM, Disease Diagnosis, Explainable AI, Unstructured Data.

DOI: 10.9781/ijimai.2025.02.003

I. Introduction

In the past, medical experts had examined minimal sized clinical data (CD) sets for disease predictions (DP) and drug recommendations. At present, automation devices produce an enormous amount of CD as a result of technological development. To analyse these huge amounts of data by an expert is not an easy task and requires more time to predict the diseases and drugs. It leads to error-prone in emergency handling cases. Currently, artificial intelligence (AI) predictions are favoured in CD analysis [1]. The capability of the training tool and model determine how accurate the current AI predictions are. But, the diagnosis (Dig) is unclear based on the predicted results of current AI, and it is challenging to manage the enormous volume of crucial data. As a result, the present AI has significant rates of false positives. Consequently, a new tool with enhanced functionality is needed. The vast majority of the CD is composed of tiny critical terms, and these tiny critical terms are crucial for disease Dig (DDig). Consequently, the work of significant phrase extraction is crucial in DDig [2].

Natural language processing (NLP) is a well-known method for gleaning tiny critical terms from a vast amount of structured, semi-structured, and unstructured CD. Hence, NLP is a useful tool for handling diverse data effectively. The CD includes text, numbers, scanned images, handwritten reports, and audio data. NLP extracts the tiny critical terms which have been required for decision-making (DM) from the diverse structured CD [3]. AI helps to analyse the large volume of digitalized medical information in a quicker way and produces the result without any explanation. This prediction is not an evident result for making a decision. Hence, the explainable AI (XAI) is preferable for making accurate results with evident. In an XAI, the results are discussed with justification and provide detailed results of the analysis. By using this result, the important features are analysed and better decisions are taken for a disease. Hence,

Please cite this article as:
Sumathi M., S.P. Raja. Explainable Artificial Intelligence-Based Diseases Diagnosis From Unstructured Clinical Data and Decision Making Using Blockchain Technologies, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 3, pp. 40-51, 2025, http://dx.doi.org/10.9781/ijimai.2025.02.003

XAI has been used in the proposed system. In healthcare domain, recommender system (RCS) has been used for improving clinical care and provides suggestions in different aspects like disease prediction, drug recommendation and medical information storage etc. This RCS helps patients and clinical care takers for taking accurate decisions in normal and emergency situations. To improve clinical care, the ant colony optimization (ACO) based RCS has been used in the proposed system. The ACO technique compares the historical disease and drug predication result with the new prediction results and produces the result as "recommended" or "not recommended".

Presently, blockchain (BC) technology is used in different domains for the sharing of information in an immutable and decentralized way. In a medical sector, sharing of information between experts in different geographical location is an essential task of handling the critical and emergency cases. The clinical information requires highly secure and immutable sharing. Thus, the BC technique is used in the proposed system. In a proposed system, the BC technology helps to exchange confidential data between AI and clinical professionals in a secured way without third parties. The property of BC is to provide decentralized and irreversible storage. Hence, the BC technique is used for transferring the predicted diseases and recommended drug information to a clinical expert in a confidential and immutable way [4]. The information stored in the block is validated by the clinical experts. If it is an accurate prediction, experts "approve to the prediction" otherwise "reject the prediction". Information about approved diseases and drug information is stored in BC. The information stored in BC cannot be changed in future. This immutable storage is useful for future DM [5].

This article's remaining portion is organized as follows: In section 2, the methods now in use and the advantages and disadvantages of AI techniques have been discussed. In section 3, along with the essential equations and architecture, the proposed NLP data classification, XAI disease Dig, ACO drug prediction and BC based DM framework principles and algorithms are described. The experimental setup, dataset gathering, and suggested methodology for experimental outcomes in various parametric aspects are all covered in section 4. The proposed system security strength is covered in section 5 along with a proof of concept. In section 6, the proposed system finishes with future improvements.

II. Literature Review

A. Term Extraction Using NLP

Jignesh R.Parikh et al. have employed NLP to find rare diseases. Correlating genetic variation data with in-depth phenotypic data is necessary for effective genetic Dig. The manually extracted terms and the extracted NLP terms have been associated with preparing a rank list of the terms. Then, the higher ranked terms have been filtered out for analysis [6]. Julia Lve et al. had concentrated on the most challenging of de-identifying and analyzing the free text in CD. The generated text from a subsequent NLP text categorization was used for conducting the extrinsic evaluation. The important terms of the CD were selected, and fictional data was created for accessing the test data [7]. Essam H. Houssein et al. had combined NLP and AI to extract the information from the CD. The biomedical NLP helps for direct DM and has been used for unlocking the hidden CD content. Additionally, analyzing CD and extracting previously undiscovered CD is possible in biomedical NLP [8]. Comito et al. have been discussed laboratory test results, patients' basic information, CD, and social media data for their heterogeneous health data integration. In order to forecast the patient's future health information, a neural network model was deployed. The similar technique has been used to predict the similar diseases. Based on similarities, the combined supervised and unsupervised techniques with NLP were utilized to forecast the diseases [9].

Sungho Sim et al, had discussed how existing IoT and AI techniques have been worked together to Dig and monitor diseases in an emergency case. This work specifically discussed the rapid spread of infectious diseases. To determine the severity of the disorders, thorough analysis is necessary for disease prediction [10]. Haohui Lu et al. have discussed regular patient monitoring is an expensive and impractical chore. As a result, the bipartite graph has been used for framing the patient network among the patients who were all Dig with the same condition. To estimate the disease risk, network analysis and machine learning (ML) algorithms were integrated. Random forest (RF) offers higher prediction accuracy than other ML methods [11]. Flora Amato et al. collect the CD from the patients using IoT smart devices, however evaluating this data is a challenging task because data is collected from numerous devices. This problem was solved by extracting structured data (SD) from the CD and processing it semantically. The data was stored on a central server to facilitate various categorization and DM techniques. NLP were used to extract the information from CD's and data from smart devices. The feature extraction (FE) and reduction reduces the data size and enables us to handle large-sized data [12].

B. Disease Diagnosis Using AI

The objective of the Methods section is to describe materials and methods in a detailed way so that a knowledgeable reader could repeat the experiment. Possible sub-sections could be: *Participants, Materials, Tasks* and *Design and Analysis.* Notice that not all these sections are always applicable.

Eden et al. had been analyzed, the diabetes is the root cause of a number of medical problems, including heart attack, stroke, renal and heart failure, and coma. Diabetes is brought on by genetics, environmental circumstances, immune system attacks and insulin breakdown, immune system responses to an infection, among other things. The AI prediction is useful for forecasting the disease severity accurately. Applications of AI in drug therapy include, improving drug treatment procedures and foreseeing drug-drug interactions [13]. Elham Khodabandehloo et al. were discussed the early DDig to lessen the impact of disease severity. The AI-enabled smart house alerts practitioners for providing better care for elderly. In the present system, the forecast had been made without any justification. The practitioner recommended the rule-based prediction system for delivering better results. The rule-based approach is unable to handle a variety of users and abnormal events. Thus, the author has suggested flexible AI to identify the initial symptoms and provide a fine-grained justification for forecasts. This adaptable, data-driven XAI can be adjusted to various people and circumstances. A collaborative process has been used for anomaly identification. The author's concentrated on fixed parameters, quick analyses, and DM on specific criteria. The multi-criteria and long-term study has not been covered in this work [14].

Thomas et al. discussed about the four main factors that affect the accuracy rate of AI models. The data source had been considered as the primary factor. The CD gathered from various sources may be out-of-date, biased, incorrect, and lacking. As a result, false Dig had been made. However, both people and information sources are important to take accurate decision. The bias of AI Dig is the second factor. The training data or human classification may have skewed the AI Dig. Hence, the properties of the dataset and categorized by humans should be examined before moving forward with the Dig model's finalization. AI Dig analysis is considered as the third factor. The performance of the locked AI model depends on the first learning set's initial value. It results in errors, over- Dig or under Dig for certain patients. Therefore, before making a forecast, it is necessary to assess the performance of an AI. The way that Dig work is organized and divided can also have an impact on the accuracy of the Dig. Therefore, prior to Dig the legal obligation for labor must be established [15].

Norma Latif Fitriyani et al. had been discussed the CD with normal, abnormal and outlier data. The outlier data must be taken out of the dataset before it can be used to make predictions. The outlier data had been found and eliminated by DBSCAN method. The unbalanced training dataset was balanced by SMOTE-ENN, and the prediction model was learned and created using XGBoost ML algorithm. To verify the effectiveness of the AI model, the gathered patient data has been compared with statistical significance [16]. Romany Fouad Mansour et al. had been discussed an automated prediction that integrates the IoT and AI in order to make a decision quickly. The Crow Search Optimization algorithm-Cascaded Long Short Term Memory (CSO-CLSTM) model was used for DDig. The weight and bias settings of the CLSTM technique were adjusted using the CSO approach. The outliers have been eliminated by an isolated forest technique. The accuracy of the CSO-LSTM approach was Dig diabetes and heart disease 96.16% and 97.26%, respectively [17]. Vijendra Singh et al. had been discussed the ML based hybrid classification technique for DDig. The support vector machine (SVM) and dimensionality reduction methods were combined to identify the chronic kidney disease. Compared to existing techniques, this combination provided higher accuracy rate [18].

C. Recommendation System-Based Drug Prediction

Dinakaran et al. had discussed the spectral deep feature classification-based drug RCS. The drug ratio is determined by feature selection and mutual molecular weights. The characteristics are using the candidate selection technique, and the molecular weight has been determined by the intra-class activation function. Then, the suggested drug level had been determined using the active adaptive recurrent neural network. More than 94% of the predicted accuracy has been produced by the RCS [19]. Farman Ali et al. were proposed an ontology-based RCS for IoT-based healthcare. The Protégé web ontology language had been used to predict the patients' dietary habits, medical history, and medication needs. Wearable health monitoring devices were used to monitor the data on the patient's body. Then, the RCS had been automated by fuzzy logic and semantic web rule languages. Without any preparation, this procedure performs straight analysis of the data. As a result, the RCS's accuracy rate suffers [20]. Farman Ali et al. were proposed the wearable sensors and social networking data based health monitoring system. Bidirectional long-short-term memory had been utilized to predict abnormal conditions, including high blood pressure, diabetes, mental health issues, and medication adverse effects. It offered a greater accuracy rate in the health monitoring system and handled diverse data [21].

Luis Fernando Granda Morales et al. had discussed the clustering and collaborative filtering based drug RCS for diabetes. The dimensionality reduction and clustering had been performed to predict the drug for diabetes. The collect user profile was compared to other patient profile for finding the similar characteristics. At last, the RCS group the patients based on the identified values. Then the quality of the recommendation and predictions were compared to the benchmark data for evaluating the accuracy rate [22]. Qing et al. had been created the unified framework with RCS and knowledge graph. The basic low-dimensional entities were learned from knowledge graph and integrated this result in neural factorization machine for improving accuracy result. Compared to existing RCS, the integrated RCS provided higher prediction accuracy [23].

D. Blockchain-Based Clinical Data Storage

Nowadays BC technology used in different sectors like agriculture, medical, manufacturing etc. Guofeng et al. had used the BC technique in agriculture to store the information in a secured and fine-grained accessible manner. Both horizontal and vertical partitions are applied to partition the data and stored in a block instead of actual data. This partition based technique provides higher confidentiality of sensitive information. Additionally, the cipher policy attribute based encryption and data encryption standard techniques were used for providing data confidentiality to block information [24]. Tetiana Hovorushchenko et al. had discussed the method for blockchain based medical data storage. The header node contains the date, time, version, metadata, digital signature, encrypted data and hash code of the previous block etc. By the same way, the next blocks were created and added to the existing blocks. This technique provides integrity to user data [25].

Suruchi et al. had reviewed the BC technology in healthcare applications. By using a BC network, patients CD from various organizations can be stored in a single network, which makes it easier for patients, doctors and other caretakers to make accurate decisions. The BC method offers patients CD protection and immutability [26]. Roberto Cerchione et al. were proposed the BC-based CD storage. To provide higher security, the CD had been kept in the permissioned BC. By using an information processing principle, the CD can be verified and transferred to storage. This procedure reduces processing and storage expenses, medical errors, and raises service standards. As a result, the BC approach contributes to enhancing patient care in a safe manner [27]. Gaofan Lin et al. had stored CD and offered fine-grained data sharing amongst caregivers using the BC approach. To guarantee privacy for sensitive information, the encrypted data had been kept on the block. Caretakers gave secure fine-grained access through the deployment of the ciphertext searching. As a result, patient data had been kept confidential and accurate [28].

E. Summary

- NLP has been used for extracting critical CD from the diverse data in an efficient manner. The existing works concentrate on specific data type not to diverse data.
- XAI is used for DDig with a justification of the prediction. This feature helps to take accurate decision making than the AI techniques. Only limited features are analyzed in the existing works.
- The RCS-based drug prediction helps to predict the drug in a fast and accurate manner.
- The BC technique helps to store the medical data in a secure and immutable way. The inter-organization based network formation is not yet designed.

Based on the analysis provided above, the unstructured CD requires special care to DDig and drugs prediction. The main contribution of the proposed effort is outlined below in order to meet these needs:

III. Contribution of Proposed Work

- NLP is used for assessing the enormous amount of unstructured data (USD) gathered from many sources with varying levels of scarcity. The preferred phrases have been extracted from the structured, semi-structured, and unstructured CD by utilizing medical code generation (MCG), pattern matching (PM), keyword extraction (KE), and visual feature extraction (VFE) approaches.
- Similarity mapping is used for designing the highly efficient XAI training model (for providing explanation and understanding of DDig). By using single mapping process, the 100% similarity mapping is an impossible task. Hence, the multi-level training model is designed for reaching the nearly 100% prediction accuracy.
- The ACO RCS performed the depth-first search on the predicted diseases based on historical data for predicting the best drug for the disease. The ACO RCS provides higher clinical care to the patients on time.

- By using BC network, the reliable and secured communication channel has been created between AI and clinical specialists. The blocks in BC stores CD securely, and helps to validate prediction results while managing emergency situations. Based on accepted prediction, the clinical professionals construct and add the safe and immutable blocks in the BC network. This immutable storage will be beneficial for future prediction.

IV. Proposed Work

Currently, health data is gathered from a variety of sources, including IoT sensor devices, CD, narrative descriptions from clinical experts, patient basic information, and scanned images. It is a challenging task to manage this enormous data in its original form and making decisions on DDig. Consequently, the DDig and drug prediction leads to inaccurate DM. In order to address these problems, the enormous USD must be transformed into SD. Among the numerous data, only minimal size data is useful for making predictions about the nature of diseases, their severity, etc. To increase prediction accuracy and to provide better clinical care, it is necessary to identify and extract the necessary terms of the obtained data. Recently AI has been utilized in the healthcare industry to accurately predict diseases. Diverse AI technologies are being developed by researchers and AI specialists to increase the precision of DDig, promote drug development, provide more advanced patient monitoring, and cater to the needs of the person.

The AI tools help to reduce mistakes in manual DDig. Currently, AI models are predicting diseases without clear justification (no clear justification for why the condition is predicted). Consequently, it is challenging to grasp DDig. Likewise, the current AI techniques analyze the diseases using similar values, and it is hard to map with 100% similarity. As a result, these technologies aim to increase the accuracy of DDig and similarity mapping. The dissimilar values are not analyzed or not taken into account for further processing. Poor prediction in DDig is leading to unclear and different results. To increase the accuracy of the prediction, to design a new AI tool for identifying the diseases with a clear understanding is required.

In order to offer better clinical care, a clinical expert must get the drug RCS based on the DDig. The RCS suggests a drug based on an analysis of previous drug information for the same disease. Depending on the type of RCS, the RCS accuracy rate varies. The previous drug suggestion success rate must be examined and projected in order to establish the current drug RCS and increase RCS accuracy. Currently, a third-party services (CSP) using cloud storage site to store and maintain the predicted CD. This CSP leads to security breaches, including erroneous RCS and adversary (Ã) access or avoiding authorized user access. A novel storage approach with security, integrity, and availability is required to ensure data privacy and security of CD. In Fig. 1, the suggested work has been displayed.

The proposed method uses cosine similarity algorithms, PM, and sentence matching (SM), word embedding (WE), KE, and MCG to extract the necessary terms from the collected data CD. Following that, XAI-based multi-level Convolution Neural Networks (CNNn) and local interpretable model-agnostic explanations (LIME) algorithms are used to match the retrieved phrases. To map all the extracted terms to the trained model at multi-level, an integrated model has been created. By this mapping, nearly 100% similarity and accuracy of predictions is achieved. The ACO has been used to identify drugs after DDig in order to improve clinical care for patients. Finally, the BC has been used for reliable and secure data transmission between clinical experts and AI for transferring DDig and drug RCS. CD is securely and immutably stored in BC, making it helpful for upcoming DM.

The clinical professionals that work in various healthcare institutions have involved the BC network. The DDig and drug RCS are verified by the clinical experts in the BC network. If the prediction is accurate, the block has been created and added to the BC network. This procedure increased the suggested technique's prediction accuracy. Likewise, this network effectively handles emergency situations. The accuracy of clinical care and DDig has been enhanced by the suggested technique.

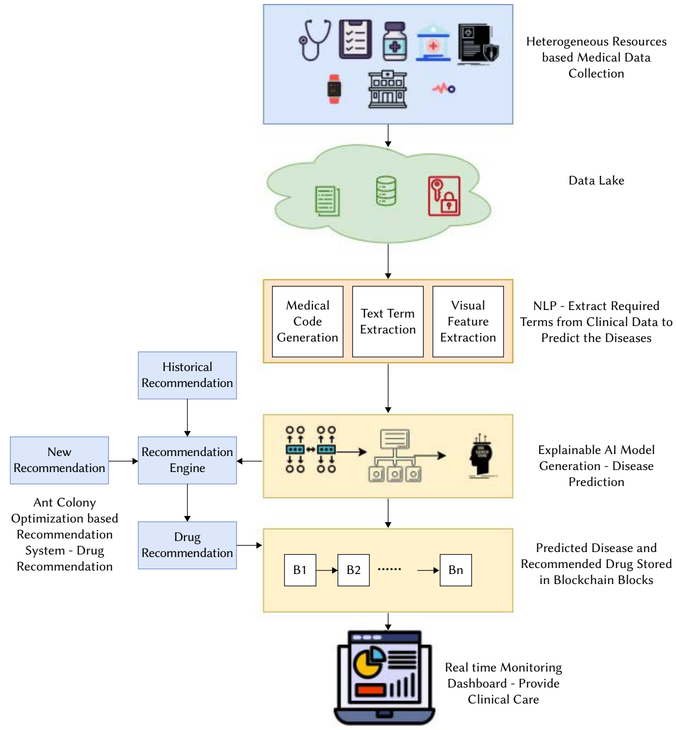


Fig. 1. Flow Diagram of the Proposed Work.

V. Methodology

A. Term Extraction From Clinical Record Using NLP

From free hand texts, images, SD, IoT sensor data, insurance providers, social media, and web knowledge, the CD are extracted. Diverse data scarcities are present in the data gathered from diverse sources. To extract relevant terms from this variety leads to high processing complexity. Therefore, the USD is converted into SD by using systematized nomenclature of medicine (SNOMED), international classification of diseases (ICD-CM) codes. Since, the SD analysis and validations are simpler than USD process. NLP helps to fix this issue. The simplest way to turn USD CD into SD is by using the NLP. The clinical caretakers benefit from the NLP conversion process in a variety of ways, including the reduction of manual analysis time, the production of precise and effective solutions, the provision of safety reviews, the ability to handle large-scale automated processing, and the management of large volumes. To transform USD into SD, various methods have been applied. Here is a list of them:

1. Medical Code Generation

The USD has been transformed into medical codes (MC) like unified medical language system (UMLs), ICD, and SNOMED for faster processing, risk and disease veracity prediction. MC is the process of converting diagnoses, treatments, services, and equipment used in healthcare into standard medical alphanumeric codes. Medical record material, like transcriptions, laboratory and radiologic results has been used to generate the Dig and procedure codes. The Health Care Procedural Coding System (HCPCS) level II categorization system and ICD-10-CM has been used to examine the USD CD and assign standard codes. The CD lists more than a thousand health conditions, diseases, injuries, and other causes. The classification, reporting, and tracking are made simpler by the usage of medical coding techniques (MCT). Every disease has various descriptors, acronyms and eponyms used in the healthcare industry. The MCT effectively standardizes the language and presentation of each component, making it simpler to comprehend, analyze, track and modify. Punctuation, special characters, mathematical symbols, and URLs are all have been deleted from USD during preprocessing. The stop-words have been eliminated for tokenization. The ICD-10 code book has been used to convert the key terms into MC. XAI uses this coded data to forecast diseases. Table I. Shows the sample ICD-10 code and its description.

TABLE I. Sample ICD-10 Code Format [29]

Code	Description
E08.220	Diabetes
E09.520	Drug or chemical
E10.110	Type 1 diabetics
E11.410	Type 2 diabetics
125.110	Arteriosclerotic heart disease
K50.013	Crohn's disease
L89.213	Pressure ulcer of right hip, stage III

2. Text Extraction From Semi-Structured Text Document

The PDF CD has been taken as the input into a semi-structured document (SSD). Equation (1) has been used for term extraction (TE). The semantic meaning of the text sequence has been captured by the word embedding technique.

$$\text{Term Extraction}(\mathrm{T_E}) = \begin{cases} \mathrm{P_M} \leftarrow \mathrm{P_{ID}}, \mathrm{Ph_{no}}, \text{Age} \\ \mathrm{S_M} \leftarrow \mathrm{P_{name}}, \mathrm{P_{Location}}, \mathrm{P_{add}}, \mathrm{P_{EID}} \\ \mathrm{K_M} \leftarrow \mathrm{P_{name}}, \mathrm{P_{SSN}}, \text{Gender} \end{cases} \tag{1}$$

Words with similar meanings have been given comparable numerical representations. The relevant terms in the CD have been found by the term frequency-inverse document frequency (TF-IDF). By using theTF-IDF technique the KE and simple text analysis have been performed. This TF-IDF method is unable to identify words that are semantic meaning-based. Thus, the semantic meaning-based data from the CD will be extracted by the Word2Vec algorithm. The SSD terms have been predicted using Algorithm 1. The values of TF-IDF have been calculated by using equations (2) and (3). The feature extraction (FE) extracts relevant data based on keywords, lemmas, and synonyms.

Algorithm 1. Information Extraction – Keyword and Sequence Extraction and Pattern Matching

Input: TD, Term Keyword List (TMKWL), Term Pattern (TMPAT)

Output: LT

Procedure: Term Identification (TD, TMKWL, TMPAT)

1. Initialize Term ⟵ 0
2. for all TextSeq, Seq. ∈ TD
    Matches each Pat∈ TMPATdo
    Term ⟵ <Seq.>
    end for
3. for all text Seq, Seq. ∈ TD
    Matches each Pat∈TMKWLdo
    Term ⟵ <Phrase><Seq.>
    end for
4. return Terms

$$\mathrm{TF(Term)} = \frac{\text{Number of Times Term Appears in a Document}}{\text{Total Number of Terms in the Document}} \tag{2}$$

$$\mathrm{IDF(Term)} = \log\left(\frac{\text{Total Number of Documents}}{\text{Number of Documents with Term}}\right) \tag{3}$$

Breaking down a data stream into a list of words, marking tokens related to parts of speech (POS), and associating useful lemma to POS, filtering the token list to obtain the most pertinent tokens, creating a list of features, and choosing the most pertinent features for the domain are performed by an above algorithm. Text tokenization, normalization, POS tagging, and lemmatization are the steps that make up FE. To DDig, these extracted traits are used.

The TF-IDF process extracts the general terms and not applicable for complex term extraction. In a word2vec algorithm, the CD is scanned entirely and the vector process is created for determining the often word identification. In this process, the semantic closeness between the words is identified by using the variables window, size and alpha. In word2vec, the unlabeled terms are identified by an artificial neural network. Table II shows the relevant term identification process of word2vec algorithm.Based on highest relevance, the terms are identified on the CD.

TABLE II. Word2Vec Similarity Identification – E.g Term – Great

Term	Relevance rate in percentage
Excellent	80
Fantastic	77
Perfect	74
Wonderful	70
Good	70
Amazing	63
Loved	62
Nice	62

3. Visual Feature Extraction

Medical images are used by doctors, radiologists, and pharmacists for drug discovery, DDig and treatment. The unique objects in the scanned images have been identified by VFE and extracted by scale-invariant FE technique. The convolution neural network (CNN) has been used for identifying abnormal object in medical images. The CNN extracts the features from an image's pixels and links them to the labels. The UMLS includes terms from the biomedical and health vocabularies. Hence, UMLS is used for identifying the extracted information. As a result, the visual features have been extracted from the images and then concepts are predicted from visual feature vector. The proposed NLP process is depicted in Fig. 2.

The CNN training complexity has been reduced by the 1 x 1024 length feature vector. With additional dimension reduction, CNN extracts highly relevant features. Two convolution layers (CL) and pooling layers, three dropout layers and one fully connected layer make up the proposed CNN network. The actual shape of the output is the same as the input using 64 and 128 filters, a four-size kernel, and a padding value of one for each of the two CL. Two max-pooling layers and one stride are used to acquire the most prominent features of the preceding feature map. The soft-max function and dropout layers address the over-fitting problems. Equation (4) describes how CNN output is expressed.

$$O_k^1 = f(s_k^1 + \sum_{i=1}^{N_{l-1}} VF_E(K_{ik}^{l-1}, t_i^{l-1})) \tag{4}$$

Where O_k^1 represents the output of kth neuron at the first layer, $f(\,)$ signifies the activation function, $VF_E(\,)$ represents the $VF_E(K_{ik}^{l-1}, t_i^{l-1})$

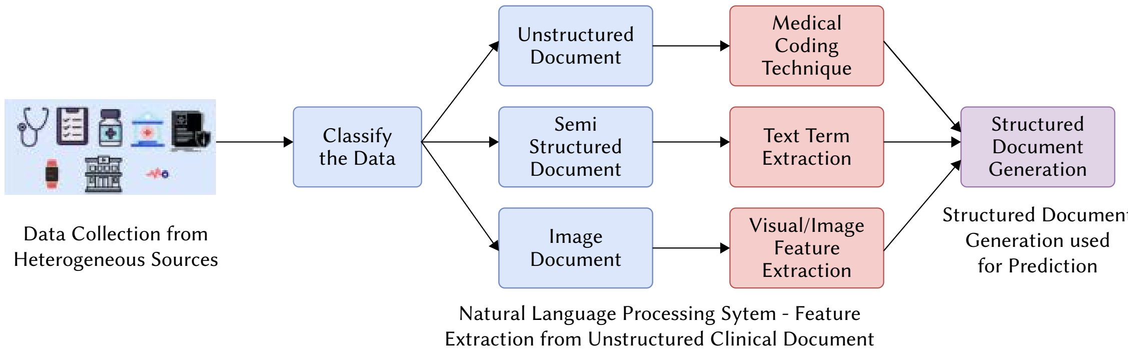


Fig. 2. NLP TEfrom USD CD.

denotes the kernel weight from the ith neuron at layer $l-1$ and its output. CNN encodes the data, and LSTM decodes it. The useful extracted terms are provided as an input to XAI to accurately predict diseases after being retrieved from USD, SSD and visual documents.

B. Explainable AI Based Disease Prediction

The prediction of the current AI is accurately predicting the results, but only little assistance given to the clinical expert (CE) in terms of DDig. Traditional methods for DDig include knowledge-based, data-driven, and hybrid approaches. These techniques fail to explain the disease forecasts. For the DM, high precision is required, and image specialists require much more information from the model than the binary prediction. ML may profit from advances in understanding that might lead to a clearer specification of its parameters in order to maintain impartiality in DM. ML accurately detects and explains the bias for training sets and tasks. The explanation can be obtained by describing which features are important for the output inference. Particularly in the medical industry, it is crucial that the interpretation of ML choices be paired with human interpretation. The previously trained ML model is used to predict and interpret new data. The evaluations of the layer, neuron, or prediction determine how to visualize the influence of pixels. The proposed XAI workflow is shown in Fig. 3.

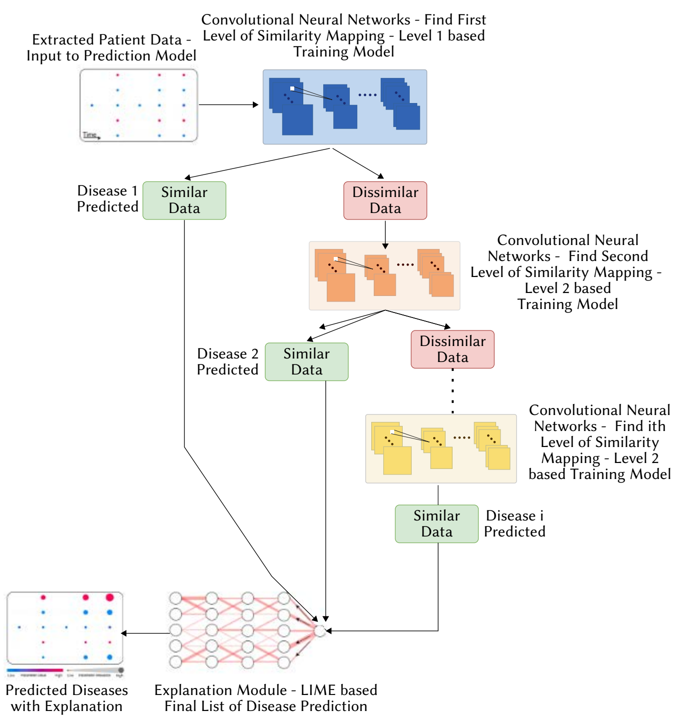


Fig. 3. Explainable AI Work Flow.

This phase took features from the NLP as input. On CNN, the first level similarity values have been predicted for identifying the "type 1 disease". The dissimilar data is transmitted to the next level of the CNN, and the level two mapped terms are classified from the dissimilar terms and labeled as "disease 2". With the use further disease data, the CNN next level has been trained and dissimilar values from the previous level is mapped to it. Certain terms have been mapped at each level, while others might not. This procedure continued, until all the retrieved terms have been mapped. Hence, this design is named as multi-level CNN model (CNNn). The CNNn prediction accuracy is high when compared to a single layer CNN model, and provides DDig information with sub-diseases. Only extracted features are examined for DDig. Hence, the minimal sized dissimilar terms are analyzed at each level. As a result, processing time of the proposed system is minimal. The XAI offers the prediction fairness, accessibility, interaction, causation, confidence, trustworthiness and privacy awareness. The explanations come in the form of text, images, local explanations, simplifications, and explanations that are relevant to the features. The LIME algorithm provides a justification based on expected outcomes. LIME analyzes the CNN output to determine how predictions alter as a result of various observations made during the subsequent procedure. The LIME technique's explanatory computation is performed using equation (5).

$$\delta(x) = argminL(f, g, \pi_x) + \omega(g), \qquad g \in G \tag{5}$$

Each model contained in the variable 'G' stands for decision trees, $\omega(g)$ denotes the number of trees and 'f' denotes the black box process. The variable 'L' is explained within the boundaries of the given locality π_x and the LIME approach aims to minimize 'L' and $\omega(g)$ to minimum value. Without any further requirement, the decision tree clearly self-explains the prediction.

C. Recommender System Based Disease Prediction

The RCS offers recommendations for products, services, and information to help users make better decisions. Based on user evaluations, the RCS determines recommendations. The RCS creates recommendations for new goods by applying a filtering algorithm to the input rating. Currently, RCS employs the Collaborative Filtering (CF) technique. Data sparsity and cold-start are two of the CF main drawbacks. These restrictions lead to the identification of bad recommendations. Hence, CF technique is not preferable to drug recommendation. Thus, the ACO RCS has been used to identify the drug recommendation in the proposed work. In an ACO, directed trust graph is utilized for finding the most reliable recommendations. The ACO selects the paths with the highest propagated trust values in order to determine the most effective suggestions. In order to properly manage and improve the accuracy of recommendations, the ant's updating approach is used in the proposed system. To raise the RCS accuracy rate, it is required to compute the semantic information

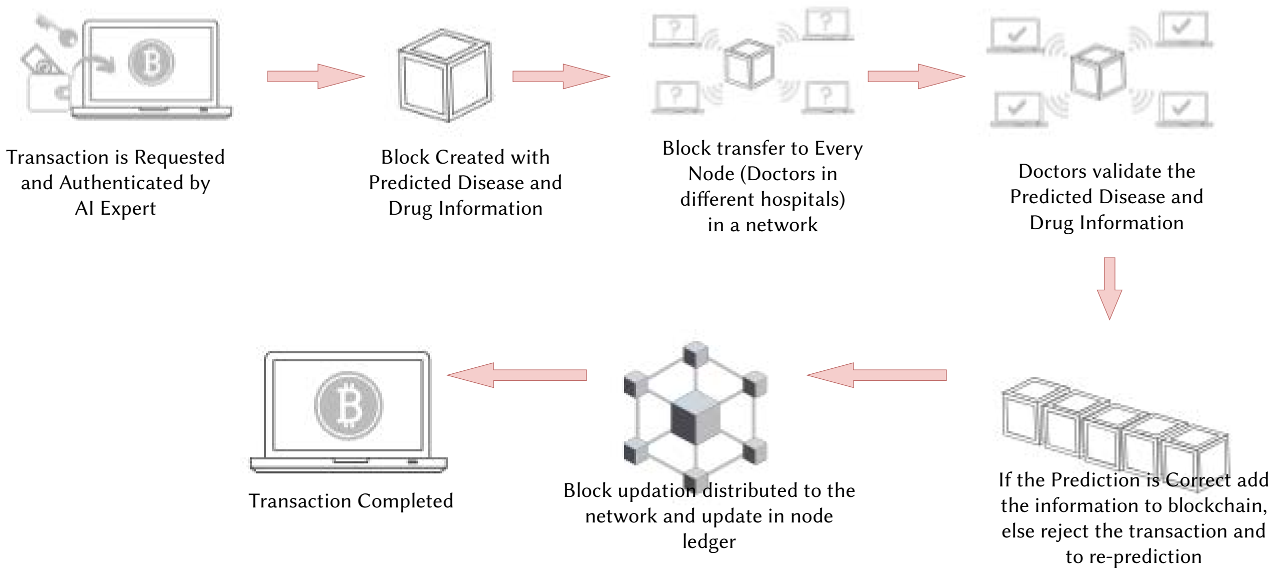


Fig. 4. Block and Blockchain Generation.

and cluster the items according to their semantic similarity. The recommended drug for the disease is determined by semantic data and user ratings. AI experts Dig diseases and prescribe drugs, while clinical specialists validate the predictions. Equation (6) is used to measure the semantic similarity between two values.

$$Sim(X,Y) = \sum_{i=1}^{|V|} \left(\frac{Common(X,Y,V[i])}{Max(\deg(X,V[i]), \deg(Y,V[i]))}\right) * Weight(V[i]) \quad (6)$$

A set of data type properties and disease properties are contained in the vector V. Degree (A, V) specifies the number times a particular item X is related via property V, Common (X, Y, V) denotes the number of times item X and Y are associated with each other via property V, and weight (V) shows the significance of the property (V). The rating of all active drugs is determined by using equation (7) to determine the recommendation.

$$A_{u,i} = A_u + \frac{\sum_{v \in Nib_{u,i(t)}} (A_{v,i} - A_v) * IW_{uv}^{c}(t)}{\sum_{v \in Nib_{u\,i(t)}} IW_{uv}^{c}(t)} \quad (7)$$

Reflects the relative importance of the user v's rating in the context of 'c' at a time 't', where Au and Av are the mean rating of an active user 'u' and a neighbor 'v' for items in a certain group 'c' respectively. Based on similarity values and active users, the ACO Dig the diseases and drugs.

D. Blockchain Based Disease and Drug Information Transfer and Storage

Massive amounts of CD are currently stored in the cloud at lower maintenance and storage expenses. Numerous security vulnerabilities, including as the access, alternation, or deletion of CD, have happened in cloud storage as a result of CSP management. Serious problems arise as a result of these procedures in healthcare organizations. Therefore, a different storage technique is required to store patients' personal and medical information in secured and immutable way. BC is now essential to all industries, due to immutable storage of hash value (HV). The immutable storage provides lots of benefits to healthcare sector, such as to take an accurate decision, refer the drug history in future, etc. Hence, DDig and drug RCS are stored in BC is preferable. The BC network is established amongst clinical experts (CE) from various organizations. To decide the disease type and drug RCS, the BC network shares, data within CE.

If more than 50% of the CE accepts the prediction, the new block has been created and added to the BC network. This confirmation procedure, aids in increasing prediction accuracy and enhancing clinical patient care. The confirmed drug and disease information is saved in a block and added to BC network. The blocks are kept in a decentralized, distributed ledger makes it easier to analyze the past data in the future. The HV has been generated by using the secured hash algorithm 256 (SHA256). The SHA256 offers collision-free HV and its' simple to build blocks in a BC is compared to the other HV produced technique. Each block includes a timestamp (TS) value and details about the patient, disease, and drug. In order to analyze past data and track patient treatment information in the quickest time intervals, the TS is helpful. The BC workflow is shown in Fig. 4. Fig. 4 clearly shows the transaction initialization, CE validation, block and BC construction process. The HV of the block generation is generated by equation (8).

$$H_V(\text{Patient}_i) = \text{SHA256}(\text{PB}_{HV} + \text{P}_{ID} + \text{Disease}_{info} + \text{Drug}_{info}) \quad (8)$$

CE from various organizations validates the patient's information and responds to AI experts in the form of acceptance or rejection. If every prediction is approved, the relevant disease and drug information for the particular patient will be finalized. If not (rejected prediction), re-prediction is performed on the extracted data. Through this approach, DDig accuracy is highly improved and patients receive better clinical care.

VI. Experimental Results

A. Experimental Setup and Dataset Description

To evaluate the performance of the proposed system, our technique was implemented in 128GB RAM capacity system with python 3. The XAI was implemented with the help of Tensor-Flow library files. The private BC network was built by using Ganache based Ethereum environment. The IPFS version 0.4.19 was used for storing the blocks in off-chain storage location. In a ML performance analysis, the training and test data ratio was taken as 20:80. The BC network is created between patients, CE and AI experts. The patient node, able to view and access the DDig and recommended drug information. The CE node, able to verify the prediction and add blocks to BC network. The AI node, analyze the prediction and send the results to CE node.

The NLP terms based datasets had been collected from biomedical named entity recognition from https://github.com/cambridgeltl/MTL-Bioinformatics-2016 which consists of NCBI-disease, Bio-creative Vs. Chemical Disease Relation Extraction, BC2GM and JNLPAB. This dataset contains a large number of entities, sentences, words, training, development and test documents. The disease coding data has been taken from https://github.com/suamin/multilabel-classification-bert-icd10. Large-scale automated ICD coding from CD notes is included in this collection. The database at https://dailymed.nlm.nih.gov/dailymed/spl-resources-all-drug-labels.cfm which contains drug labeling summaries of information for safe and effective use of the drug, proposed by the manufacturer and approved by Food and Drug Administration, is used to generate the drug identification and labeling-based dataset. The chest x-ray image dataset is taken from IU collection for chest x-ray studies in Open-I (http://openi.nim.nih.gov/). This dataset includes a collection of x-ray images and related radiology reports are in a textual form.

B. Term Extraction Using NLP

The PM, KM, and SM processes have been used to carry out the TE. The TE process is depicted in Table III based on different types of extractions. By utilizing PM, KM, and SM, Date, age, email, phone, zip code, street and CD keywords have been extracted with a 91.47% accuracy of F1-measure.

TABLE III. Term Extraction Using PM, KM and SM

Type of Extraction	Precision (%)	Recall (%)	F1-measure (%)
PM	97.12	85.98	91.47
KM	93.06	81.98	87.18
SM	94.49	82.82	88.27

Table IV discusses the precision, recall, and F1-measure of the terms for patient ID, doctor name, zip code, CD number, city name, and hospital name.

TABLE IV. Term Extraction Using PM, KM and SM

Entity Name	Precision (%)	Recall (%)	F1-measure (%)
Patient ID	86.75	84.79	85.79
Doctor Name	96.60	83.80	89.75
Zip Code	100	94.39	97.10
CD number	96.10	91.75	93.84
City Name	83.96	78.56	81.22
Hospital Name	77.10	76.89	78.85

The combined approach of the PM, KM and SM provides higher prediction accuracy than the ML algorithm prediction technique. The ML algorithm prediction accuracy depends on the training dataset. Some expected level of false positive and negative values have been occurred due to inconsistent gold-standard.

C. Disease Prediction Using EAI

Precision, recall and F1-measure are the metrics used to assess the accuracy rate of CNNn-LSTM. The number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) is used to calculate these metrics. For the proposed system's training and testing, 60% of CT scans and X-ray images have been randomly chosen. The CNNn-LSTM parameters configuration is shown in Table V.

The sigmoid activation function, two fully connected layers, a batch size of 325, and 50 epochs, were used in the experiments. The training and testing losses for CNNn-LSTM accuracy have increased from 0.42 to 0.987 on the graph scale. From 1 to 5 epochs, the accuracy rate starts at 0.42 and rises to 0.987. The loss starts at 0.68 and gradually decreased until the 50th epoch. Table 6 compares the proposed CNNn-LSTM to DNN, GRU, RNN, LSTM, CNN1-LSTM, and CNN2-LSTM already in use.

TABLE V. CNNn-LSTM Parameter Configuration

Layer (Type)	Output Shape	Number of Parameters
Con_1	(None (N), 1023, 64)	256
M_pooling_1	(N, 512, 64)	64
Drop_out_1	(N, 512, 64)	64
LSTM_1	(N, 64)	131584
Droput_2	(N, 64)	64
Dense_Value_1	(N, 2)	258

TABLE VI. Performance Comparison of Proposed and Existing Techniques

Model Name	Precision (%)	Recall (%)	F1-Mesaure (%)	Accuracy (%)	Loss
DNN	91.78	87.96	86.78	87.64	1.78
GRU	94.89	94.87	94.85	94.86	0.13
RNN	94.69	94.23	93.96	93.95	0.10
LSTM	91.98	91.32	91.76	91.23	0.14
CNN1-LSTM	96.25	96.41	96.36	96.18	0.13
CNN2-LSTM	99.12	99.08	99.02	98.97	0.02
CNNn-LSTM	99.74	99.87	99.79	99.84	0.01

Till 24th epoch, the CNN1-LSTM model works well, after that, its performance suffers from the network's over-fitting. The CNN1-LSTM had an overall loss of 0.13. The CNN2-LSTM loss is below 0.02 and the proposed CNNn-LSTM loss is below 0.01. Based on various training and testing sets comparison, the suggested CNNn-LSTM outperforms than the CNN2-LSTM in terms of detection performance. Due to multiple levels in the proposed model, its accuracy rate is larger than other models.

D. Disease and Drug Recommendation Using ACO

In the medical industry, DDig's accuracy rate is essential. Drug recommendations are erroneous as a result of inaccurate DDig. The K-means, RF, and C4.5 algorithms as well as other popular clustering techniques are contrasted with the proposed ACO methodology. Fig. 5 illustrates a comparison of the proposed and current techniques. When compared to current techniques, the proposed ACO provides higher DDig accuracy and drug for each disease has now been identified.

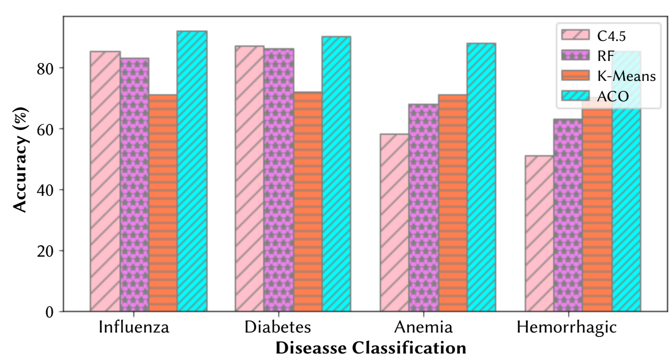


Fig. 5. Disease Classification Comparison.

TABLE VII. Drug RCS Analysis

Number of Patients	Ontology RCS				Proposed ACO RCS				Disease and Drug RCS
	Precision (%)	Recall (%)	Accuracy (%)	FMeasure(%)	Precision (%)	Recall (%)	Accuracy (%)	FMeasure(%)	
1	47.2	53.5	56.6	50.5	99.8	77.2	77.5	87.3	No
2	42.3	69.3	60.4	52.6	99.7	82.5	83.5	90.2	Yes, M
3	47.1	66.8	57.5	55.2	100	83.2	84.6	91.4	Yes, M
4	55.8	69.4	62.8	61.4	97.8	75.6	77.2	84.6	Yes, G
5	63.3	66.3	62.2	65.3	97.6	82.4	84.6	89.5	Yes, G
6	66.2	70.6	67.3	68.1	99.5	82.7	83.2	89.7	Yes, G
7	58.4	64.2	58.1	61.5	99.8	84.2	86.4	91.4	Yes, M
8	49.5	76.1	63.6	59.6	96.7	85.4	85.3	90.2	Yes, M
9	56.6	73.7	63.5	64.9	95.6	86.5	87.2	90.4	Yes, M
10	54.2	69.8	64.2	60.7	98.7	77.7	80.6	86.6	No
Average	54.4	67.2	61.4	59.5	97.8	81.5	83.7	89.4	-

TABLE VIII. DDigComparison - ML and BC Network

Parameters	LR	KNN	SVM	NB	DT	XGBoost	ANN	RF	Proposed BC
Precision (%)	74.58	81.36	78.65	65.24	80.75	82.54	82.53	84.98	99.56
Recall (%)	74.59	82.24	78.65	69.57	80.83	83.01	82.54	85.98	99.75
F1 Measure (%)	74.89	82.01	78.89	65.30	80.84	81.54	82.58	84.96	99.65
Accuracy (%)	74.25	82.05	78.65	62.78	80.84	82.45	82.58	84.79	99.58

For comparison, the diabetic drug RCS is analyzed in the proposed ACO and ontology-based RCS [30]. The results are demonstrated in Table VII. The results are diabetic [Yes, No] and drug [Metformin(M), Glindies (G)].

The accuracy rate increased from 61.4% to 83.7%, recall climbed from 67.2% to 81.5%, and F1-measure increased from 59.5% to 89.4% in the proposed method. The precision rate has been increased from 54.4% to 97.8%. By comparing the specific patient's prediction result with the average result, the information about the diabetes patient is precisely predicted for drug information. The drug is recommended to patients if their results are greater than average; otherwise, drug has not been recommended.

E. Disease and Drug Information Storage in B_c

The BC network receives the DDig and drug RCS from the previous stage. This information has been confirmed by more than 50% of the specialists on the BC network. After the DDig and drug RCS approval, the specialist adds the block to the BC network and notifies the requester. If not, the requester receives a message of refusal, and repeats the prediction based on other alternates. In table 8, the performance of the proposed BC network and ML algorithms like logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), naïve Bayes (NB), Decision tree (DT), artificial neural network (ANN), and random forest (RF) DDig performance have been compared.

The ratio of training and testing data for ML algorithms has been taken as 80:20. The proposed BC-based DDig predicts diseases with the highest accuracy rate when compared to ML algorithms. The accuracy rate of the proposed technique dependent on more than 50% of CE knowledge, not dependent on a single CE. Thus, the proposed technique's DDig prediction accuracy rate outperforms than ML algorithm's performance. Likewise, the drug RCS has been evaluated by the ML and BC network.

Now the block has been added to the BC network. The BC network's performance is affected by changes in the quantity of request or the patient-to-patient ratios. As a result, the performance of the BC networks has been analyzed by throughput, latency, and response time. The computational complexity of the CNNn-LSTM and ACO techniques are investigated in the following subsection.

VII. Performance Analysis

A. Computational Complexity

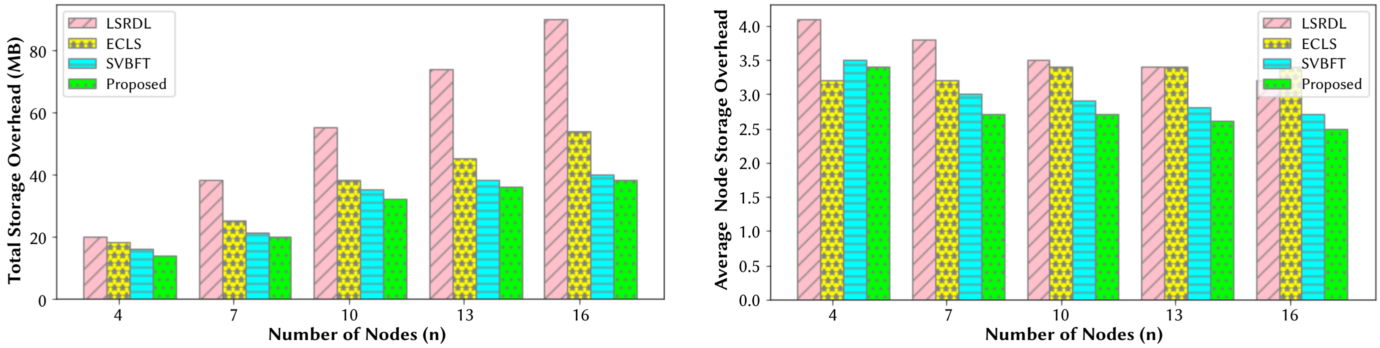
The CNNn-LSTM computational cost depends on the number of layers in CNNn. If $n_{c,1} = n/2$ and $n_{c,2} = n/4$ and the computational complexity of the CNN is $O(\lambda n^2)$ Here, λ denotes the size of the training set in sequential order. The complexity of the LSTM network is $O(4n_l\lambda(\frac{n^2 n_l}{16} + n_l)$ where nl is the number of neurons in each gate. Equation (9) computes the CNNn-LSTM network computational complexity of a single data sample.

$$O(\lambda n^2\left(4n_l\left(\frac{n_l}{16} + n_l\right)\right) \tag{9}$$

Prediction complexity of the RCS for a 'n' active users and 'm' objects is O(n*m). The time required to compare the active users to the specific context of the other users is O(n*m). The ACO RCS has an O(nm+nm) computational complexity overall.

B. Storage Space Requirement Analytics

The limitation of BC is the block storage size. Each block can hold up to 1MB of data. The CD contains a variety of data types requires a lot of storage space because they are kept on the block in their original format. In order to securely store, the predicted results, recommended drug information and expert decisions are kept in the caretaker's locations. This storage technique requires less than 1MB of storage capacity for each patient. The data for every patient had been kept in a distinct block. The suggested system's storage overhead has been compared with conventional storage methods such as score-voting based Byzantine fault tolerance (SVBFT), erasure code-based low storage (ECLS) and low storage room distributed ledger (LSRDL). The comparison of the proposed and existing methodologies is shown in Fig. 6a and b [31]. The number of nodes has been regulated at 4 to 16 and the block size is set at 400 to 2800 transactions.



(a) (b)

Fig. 6. (a) Total storage Costs. (b) Node storage overhead on average.

As can be seen in fig. 6a and b, the proposed strategy has lower average and total overheads than existing techniques now in use. Under different storage techniques, the overall storage space requirement grows as the number of nodes does as well.

C. Storage B_c Throughput and Latency Time Analysis

Throughput (TP) Analysis: The average number of transactions that can be handled successfully per second (TPS) in the BC network is used to calculate TP. Total transactions divided by total time (time in seconds) have been used to calculate the BC network transaction time. For TP analysis, the total number of TPS has been used. The proposed work's TP analysis based on the operations such as opening the transaction, query processing and transfer data has been shown in Fig. 7.

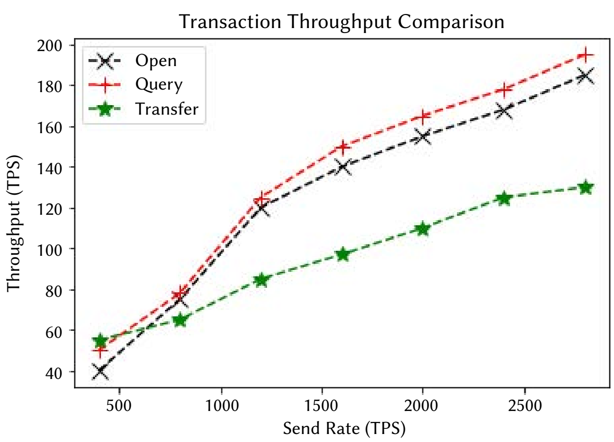


Fig. 7. Transaction Throughput Analysis.

Latency (LT) Analysis: The LT has been calculated from transaction submission confirmation. The propagation and processing times are included for determining the LT time. A processing rate of 400 to 2800 TPS has been used for measuring the BC network's performance, with 50 TPS has been taken into account for data opening, querying, and transferring. The LT begins when a transaction is submitted and ends when it is added in a BC network. With varying sending rates of 400, 800, 1200, 1600, 2000, and 2800 TPS, the average latency had been measured from 400 to 2800. The average LT rate increased along with the transmission rate. Fig. 8 shows the LT analysis of the proposed work.

VIII. Security Analysis

Data privacy for patient CD is offered by the proposed BC-based DDig and drug RCS storage technique. HV of CD, is stored in BC instead of actual data. For various CD sizes, an equal size of HV is

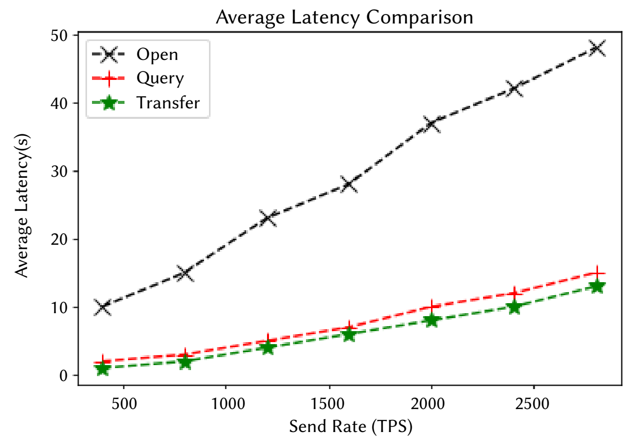


Fig. 8. Latency Time Analysis.

produced as an output. This characteristic prevents data tracking of individuals. Instead of private storage, the CD is kept at the decentralized storage location. The decentralized storage technique avoids a single point of failure and authorized user access denying. The proposed technique differs from attribute based encryption (ABE) and key-aggregate cryptosystem (KAC) in terms of security support for CD. The proposed BC storage is compared with existing ABE and KAC techniques, as shown in table IX.

TABLE IX. Comparison of ABE, KAC and Proposed BC Technique

Security Parameter	ABE	KAC	Proposed BC
Tamper Resistance	√	√	√
Information Traceability	√	√	X
Secure Storage	√	√	√
Privacy Protection	√	√	√
Access Log	X	X	√
Access Period	X	X	√
Reliance on Trusted Third Parties	√	√	X
User Anonymity	X	X	√
Control of CD	Incomplete	Incomplete	Complete

Attack analysis of proposed technique:

- Sybil Attack: In a proposed work, the BC network was created between the registered and authorized users in a permissioned network. The attackers are unable to participate in the conversion and access the information. Hence, unauthorized access is completely avoided with this technique.

- Spoofing attack: The authorized user devices are synchronized with the network services. Hence, the authorized users are able to access the data through the registered devices. If an unregistered device is trying to enter into the network, the alter message is sent to the caretakers. Hence, the spoofing attacks are impossible in this proposed technique.
- Man-in-the-middle attack: the information is stored in the form of HV. If the attacker tries to access information at the time of data transfer, they are unable to identify the information. Hence, middle-in-the-middle attack is not possible.
- Denial of service attack: In a proposed network, the registered users can only initiate the transaction and transfer the information. Other users are unable to enter into the network. Hence, the denial of service attack is eliminated in the proposed technique.

IX. Conclusion

To facilitate easy maintenance, analysis, and transfer, the health information is kept in digital form. The knowledge, accessibility, and experience of the professionals determine how well manual disease diagnosis and drug prediction perform in the medical area. The specialist's availability has a significant impact on the patient's quality of life. The artificial intelligence helps in making decisions that are accurate but does not reveal decision details. The specific hospital determines whether it can handle emergency patients. The clinical data is in the form of structured and unstructured forms like numerical results, text prescriptions, scanned images, and other types. It is important to manage unstructured data carefully while making decisions. With more than 91% accuracy, a natural language processing system analyses the unstructured data to distinguish between distinct data types, and the CNNn-LSTM technique accurately diagnosed diseases with higher than 99% accuracy. An ant colony optimization-based recommender system analyses the prediction and more precise drug for the disease. The drug decision and disease information have been saved in a permissioned blockchain after expert approval. More than 50% of clinical experts validate the decisions. As a result, judgments are made in emergency situations as quickly and precisely as feasible. The computational complexity and storage space requirements of the proposed system are lower than those of the existing techniques. Similarly, several aspects of blockchain throughput, latency, and security analysis have been examined. In a proposed work, the patient and clinical experts are having complete control over the clinical data. In future, the proposed technology will be extended for dynamic data collection in remote locations and decision-making.

References

[1] Gazali, S., S. Kaur, and I. Singh, "Artificial intelligence based clinical data management systems: A review," *Informatics in Medicine Unlocked*, vol. 9, pp. 219–229, 2017.

[2] A. Martinez-Millana, A. Saez-Saez, R. Tornero-Costa, and N. Azzopardi-Muscat, "Artificial intelligence and its impact on the domains of universal health coverage, health emergencies and health promotion: An overview of systematic reviews," *International Journal of Medical Informatics*, vol. 168, pp. 1–12, 2022.

[3] S. M. Sumathi, N. Vijayaraj, S. P. Raja, and M. Rajakamal, "Internet of things based confidential healthcare data storage, access control and monitoring using blockchain technique," *Computing and Informatics*, vol. 41, no. 5, pp. 1–31, 2022.

[4] S. M. Sumathi, S. Sangeetha, and A. Thomas, "Generic cost optimization and secured sensitive attribute storage model for template-based text document on cloud," *Computer Communications*, vol. 150, pp. 569–580, 2020.

[5] V. Bellini, P. Pelosi, M. Valente, A. V. Gaddi, and M. Baciarello, "Using artificial intelligence technique to support clinical decisions in perioperative medicine," *Perioperative Care and Operating Room Management*, vol. 28, pp. 1–8, 2022.

[6] J. R. Parikh, C. A. Genetti, A. Aykanat, C. A. Brownstein, and K. Schmitz-Abe, "A data-driven architecture using natural language processing to improve phenol-typing efficiency and accelerate genetic diagnoses of rare disorders," *HGG Advances*, vol. 2, no. 1, pp. 1–10, 2021.

[7] J. Lve, N. Viani, J. Kam, L. Yin, S. Verma, S. Puntis, R. N. Cardinal, A. Roberts, R. Stewart, and S. Velupillai, "Generation and evaluation of artificial mental health records for natural language processing," *Digital Medicine*, vol. 3, no. 1, pp. 1–10, 2020.

[8] E. H. Houssein, R. E. Mohamed, and A. A. Ali, "Machine learning techniques for biomedical natural language processing: A comprehensive review," *IEEE Access*, vol. 9, pp. 1–26, 2021.

[9] C. Comito, D. Falcone, and A. Forestiero, "AI-driven clinical decision support: Enhancing disease diagnosis exploiting patient similarity," *IEEE Access*, vol. 10, pp. 1–12, 2022.

[10] S. Sim and M. Cho, "Convergence model of AI and IoT for virus disease control system," *Personal and Ubiquitous Computing*, vol. 25, pp. 1–11, 2021.

[11] H. Lu, S. Uddin, F. Hajati, M. A. Moni, and M. Khushi, "A patient network-based machine learning model for disease prediction: The case of type-2 diabetes mellitus," *Applied Intelligence*, vol. 51, pp. 4667–4680, 2021.

[12] F. Amato, L. Coppolino, G. Cozzolino, G. Mazzeo, F. Moscato, and R. Nardone, "Enhancing random forest classification with NLP in DAMEH: A system for data management in e-health domain," *Neurocomputing*, vol. 457, pp. 79–91, 2021.

[13] E. L. Romm and I. F. Tsigelny, "Artificial intelligence in drug treatment," *Annual Review of Pharmacology and Toxicology*, vol. 60, pp. 353–369, 2020.

[14] E. Khodabandehloo, D. Riboni, and A. Alimohammdi, "HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline," *Future Generation Computer Systems*, vol. 125, pp. 168–189, 2021.

[15] T. Ploug and S. Holm, "The four dimensions of contestable AI diagnostics: A patient-centric approach to explainable AI," *Artificial Intelligence in Medicine*, vol. 107, pp. 1–5, 2020.

[16] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, pp. 115710–115726, 2020.

[17] R. F. Mansour, A. E. Amraoui, I. Nouaouri, V. G. Diaz, D. Gupta, and S. Kumar, "Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems," *IEEE Access*, vol. 9, pp. 163512–163521, 2021.

[18] V. Singh and D. Jain, "A hybrid parallel classification model for the diagnosis of chronic kidney disease," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 4, pp. 8–15, 2021.

[19] S. Dinakaran and P. Anitha, "An efficient drug compound analysis using spectral deep feature classification based compound analysis model for drug recommendation," *Neuroscience Informatics*, vol. 2, no. 3, pp. 1–9, 2022.

[20] F. Ali, S. El-Sappagh, S. M. R. Islam, and A. Ali, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Future Generation Computer Systems*, vol. 114, pp. 23–43, 2021.

[21] F. Ali, S. M. R. Islam, D. Kwak, P. Khan, and N. Ullah, "Type-2 fuzzy ontology-aided recommendation systems for IoT-based healthcare," *Computer Communications*, vol. 119, pp. 138–155, 2018.

[22] L. F. Grandamorales, P. Valdiviezo-Diaz, R. Reategui, and L. Barba-Guaman, "Drug recommendation system for diabetes using a collaborative filtering and clustering approach: Development and performance evaluation," *Journal of Medical Internet Research*, vol. 24, 2022.

[23] Q. Ye, C.-Y. Hsieh, Z. Yang, Y. Kang, J. Chen, D. Cao, S. He, and T. Hou, "A unified drug-target interaction prediction framework based on knowledge graph and recommendation system," *Nature Communications*, vol. 12, no. 1, pp. 1–12, 2021.

[24] G. Zhang, X. Chen, L. Zhang, B. Feng, X. Guo, J. Liang, and Y. Zhang, "STAIBT: Blockchain and CP-ABE empowered secure and trusted agricultural IoT blockchain terminal," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 5, pp. 56–65, doi: 10.9781/ijimai.2021.12.002.

[25] T. Hovorushchenko, A. Moskalenko, and V. Osyadlyi, "Methods of

medical data management based on blockchain technologies," *Journal of Reliable Intelligent Environments*, vol. 9, no. 1, pp. 5–16, 2022.

[26] S. Singh, S. K. Sharma, P. Mehrotra, P. Bhatt, and M. Kaurav, "Blockchain technology for efficient data management in healthcare system: Opportunity, challenges and future perspectives," *Materials Today: Proceedings*, vol. 62, pp. 5042–5046, 2022.

[27] R. Cerchione, P. Centobelli, E. Riccio, S. Abbate, and E. Oropallo, "Blockchains coming to hospital to digitalize healthcare services: Designing a distributed electronic health record ecosystem," *Technovation*, vol. 117, pp. 1–16, 2022.

[28] G. Lin, H. Wang, J. Wan, L. Zhang, and J. Huang, "A blockchain-based fine-grained data sharing scheme for e-healthcare system," *Journal of Systems Architecture*, vol. 127, 2022.

[29] K. Johnson, "Implementation of ICD-10: Experiences and lessons learned from a Canadian hospital," in *IFHRO Congress, AHIMA Convention*, Oct. 2004. [Online]. Available: www.ahima.org.

[30] D. Riano, F. Real, J. A. López-Vallverdu, F. Campana, S. Ercolani, P. Mecocci, R. Annicchiarico, and C. Caltagirone, "An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients," *Journal of Biomedical Informatics*, vol. 45, no. 3, pp. 429–446, 2012.

[31] C. Li, J. Zhang, X. Yang, and Y. Luo, "Lightweight blockchain consensus mechanism and storage optimization for resource-constrained IoT devices," *Information Processing and Management*, vol. 58, no. 6, pp. 1–24, 2021.

Sumathi M.

Sumathi M. completed her B. Eng. in computer science and engineering in 2003 from the Shri Angalamman College of Engineering and Technology, Tiruchirappalli. She completed her M.Tech. in information technology in 2008 from the Bharathidasan University, Tiruchirappalli. She completed her Ph.D. in 2021 in the area of data security from the National Institute of Technology, Tiruchirappalli. Currently, she is working as Assistant Professor in the School of Computing at SASTRA Deemed University, Thanjavur, Tamil Nadu, India.

S. P. Raja

S. P. Raja is born in Sathankulam, Tuticorin District, Tamilnadu, India. He completed his schooling in Sacred Heart Higher Secondary School, Sathankulam, Tuticorin, Tamilnadu, India. He completed his B. Tech in Information Technology in the year 2007 from Dr. Sivanthi Aditanar College of Engineering, Tiruchendur. He completed his M.E. in Computer Science and Engineering in the year 2010 from Manonmaniam Sundaranar University, Tirunelveli. He completed his Ph.D. in the year 2016 in the area of Image processing from Manonmaniam Sundaranar University, Tirunelveli. Currently he is working as an Associate Professor in the School of Computer Science and Engineering in Vellore Institute of Technology, Vellore, Tamilnadu, India.