

Editor's Note

THE International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI - provides a space in which scientists and professionals can report about new advances in Artificial Intelligence (AI). On this occasion, for the last edition of the year, I am pleased to present a regular issue including different investigations covering aspects and problems in AI and its use in various fields such as medicine, education, image analysis, protection of data, among others.

On behalf of the editorial board, I must once again thank the reviewers and authors who trust and support the journal with their work.

I am going to briefly introduce each of the works that we will find in this edition.

The volume begins with a paper that focuses on mental health, Daus and Backenstrass carry out a pilot study with bipolar patients that assesses the viability of the new care approach through the recognition of emotions using mobile phones. The results indicate that it could be easy to analyze verbal emotions and facial recognition without compromising the privacy of the patient.

The next two articles are also focused on the health sector.

Amador-Domínguez et al. propose a hybrid system between Case Based Reasoning (CBR) and Deep Learning (DL) for the generation of medical reports. The authors take advantage of the CBR's explicability and the DL predictive power. The proposed system is fully modular and adaptable to various clinical scenarios, specifically the authors present a case focused on the development of radiology reports to illustrate the proposal.

On the other hand, Bareño-Castellanos et al. develop a prototype that supports medical decisions with patients with cardiovascular risk. Using easily obtained variables such as pressure force, weight and body mass index, different algorithms such as k-means, c-means and support vector machine (SVM) were tested with very good results, especially when combining these two last algorithms.

We continue with two works that refer to predicting human activities and gestures.

Li et al. implement a model based on deep neural networks to classify 4 affective stimuli: nervous, calm, happy and sad, using eye-tracking signals. Eye tracking has been applied in recent years in neuromarketing, neurocognition and user experience, among others. The results show that the accuracy of the model reaches 87.2%. In the future, affective computing will be applied in a wide range of fields such as medical rehabilitation, assistance for people with disabilities and education.

Verma and Singh investigate different architectures that allow to predict activities such as walk, sit down, standup, pickup, carry, throw, push, pull, wave hands and dap hands from images with depth and the sequence of skeletal joints. The proposed work is done in three different levels. In the first level, spatial-temporal features are extracted from different modalities, in the second level three independent SVMs are trained using the features extracted from the first level, and in the last level the probability scores are fused and optimized using two evolutionary algorithms such as a genetic algorithm (GA) and particle swarm optimization (PSO). They achieve 96.50% accuracy results.

Kim, Oh and Heo compare the two VGG architectures ResNet and SqueezeNet, with different convolutional filters, in order to identify different mosquito types using audio data and the circadian rhythm of insects. Adding activity circadian rhythm information to the networks showed an average performance improvement of 5.5%. The VGG13 network with 1D-ConvFilter achieved the highest accuracy of 85.7%.

In the next article, Suruliandi, Kasthuri, and Raja present a novel method named the Similarity Matrix-based Noise Label Refinement (SMNLR), which effectively predicts the accurate label from the noisy labeled facial images. In addition, they use Convolutional Neural Networks (CNN) for feature representation. The convolution and pooling layers of the CNN are able to get enough information, such as the edges, orientations, and corner features of facial images. Extensive experiments are carried out with three databases, surpassing the results of previous works.

Hernandez-Olivan et al. propose a general method to preprocess musical piece inputs for later identification. Preprocessing allows you to establish the most efficient combination of inputs to a CNN. They use a max-pooling of factor 6 at the beginning of the process as a pooling strategy, thus generating the self-similarity matrices that enter the CNN. This method reduces preprocessing and training time for the neural network.

In order to evaluate the quality of the software automatically, Gupta and Chug take into account the ISO / IEC 25010: 2011 standard where 8 characteristics of product quality are described, and maintainability is one of the most important. Maintainability refers to how easy it is for the software to deal with new requirements. To meet the objective, five Boosting Algorithms (BA) are compared using 7 empirically collected open source data sets. Based on the residual errors obtained, Gradient Boosting Machine (GBM) is the best performer, followed by Light Gradient Boosting Machine (LightGBM) for Root Mean Square Error (RMSE), whereas, in the case of Mean Magnitude of Relative Error (MMRE), eXtreme Gradient Boosting (XGB) performed the best for six out of the seven datasets.

There are also two works focused on speech. Let us see what they are about.

Nosek et al. work on synthesizing speech in multiple languages, in multi-speaker voices and multiple styles to train a neural network that identifies similarities and differences between speakers and establish relationships between the phonemes of different languages and produce high-quality synthetic speech. Using vocoders, it was shown to be capable of producing good quality synthetic speech even in languages in which it was not trained.

Automatic audio-visual speech recognition is an emerging research field where visual speech is recognized through face detection, Region of Interest (ROI) detection, and lip tracking. Debnath and Roy propose a new method for the extraction of visual characteristics using Pseudo Zernike Moment (PZM) to follow the movement of the lips. Subsequently, they weigh the importance of the characteristics with statistical analysis of Analysis of Variance (ANOVA), the Kruskal-Wallis test and the Friedman test, demonstrating that this step is important to overcome the results of other speech recognition methods. Finally, they recognize audio-visual speech using machine learning algorithms such as SVM, artificial neural networks (ANN) and Naive Bayes, with very good results.

Basavaraju et al. present a Neighborhood Structure-Based Model that locates the region where there is text in both images and videos. The tests are performed on 5 different image datasets and video frames where there is low contrast, composite background and lighting effects. In the tests we can see how the probable textual spaces are successfully localized in multiple languages.

Seal et al. propose a non-Euclidean similarity measure that is based on Jeffreys nonlinear divergence (JS). They analyze the method with

real and synthetic databases, demonstrating the superiority of the method over other c-means algorithms. It could be very useful for designing new clustering algorithms.

The next article is a contribution to the area of copyright protection, so important in our times. Kumari et al. optimize Discrete Wavelet Transform (DWT) for embedding an imperceptible and a robust non-blind image watermarking. The extraction is processed by a Recurrent Neural Network based Long Short-Term Memory (RNN-LSTM), obtaining the original image.

Kadry et al. recommend a methodology to solve the multi-thresholding problem of RGB scale images using entropy value. The authors propose a random search along with a novel Multiple-Objective-Function (MOF), to maximize MOF. The tests confirm that the performance of MFO is better than PSO, BA and FA, and approximately similar to MA and AOA.

Abdulkareem et al. conduct a review and analyze the most relevant studies in the image dehazing field. They conducted an objective image quality assessment experimental comparison of various image dehazing algorithms and reflect different observations that can serve as a useful guideline for practitioners who are looking for a comprehensive view on image dehazing.

Garrido et al. analyze different multi-agent systems, which implement creative methods such as establishing analogies with known problems, brainstorming, lateral and parallel thinking. The authors propose different guides for each of these methods that help determine if a solution meets the requirements of the creative task. In addition, they define a conceptual model to implement a creative computational system.

The following work is an advance towards the metaverse that we have heard so much today. Lopez et al. have thought of democratizing virtual reality experiences for the educational field. They use the concept of mixed reality that combines virtual reality and the real world in the same scenario. The proposed solution enables content creators to design, build and publish training experiences in the cloud, using Microsoft HoloLens2. The user can see an object in 3D in the same point of view as in the real environment and interact with the object thanks to the recognition of gestures and speech of HoloLens2.

Fei Xu et al. measure the similarity between works of art or paintings, extracting characteristics using the Sparse Metric Learning-based Kernel Regression (KR-SML) algorithm. Although its main objective is to improve the teaching-learning in the identification of works of art, the work also seeks to predict the genre, the artist and the style of painting. The model obtains good prediction results by combining SVM with ANN, and Human-Computer Interaction (HCI).

Cobos-Guzman et al. present a virtual assistant, capable of interacting with the public to improve communication. The system can recognize the level of attention from audiovisual resources and synchronizes the assistant to increase the level of attention of the audience. The system is composed of two large modules, one that captures audio and images to extract characteristics that represent behavior and attitudes; and another that represents knowledge through ontologies, and with the help of reinforcement learning, the assistant is able to decide the best strategy.

The following work arises as a concern resulting from the online tests used in colleges and universities, forced by the Covid-19 pandemic. Balderas et al. propose a model to detect fraudulent collaborations between students who present online assessments, based on the time of presentation of the test and the grade obtained. The software tool proved to be very useful for teachers to detect fraud.

Tlili et al. applied Learning Analytics (LA), which focuses on understanding students' in-game behavior trajectories and personal

learning needs during the game. This systematic literature review examined how LA in educational games has evolved. The results indicate that factors such as student modeling, iterative design, and customization must be taken into account. Furthermore, the use of LA creates several technical challenges such as data management and ethics that are still unsolved.

And finally, in the last work Amo et al. raise a framework with 7 principles: legality, transparency, data control, anonymous transactions, responsibility, interoperability and local first. This framework is proposed to be considered in the development and adoption of educational technology that collects, stores, manages and analyzes educational data; improving the privacy and security of the data.

Dr. Xiomara Patricia Blanco Valencia
Managing Editor
Universidad Internacional de La Rioja