

Editor's Note

THE International Journal of Interactive Multimedia and Artificial Intelligence – IJIMAI – provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances in Artificial Intelligence (AI) tools or tools that use AI with interactive multimedia techniques. The present regular issue comprises different topics as generative AI, brain and main inspired computing, bird species identification, spam detection, recommendation systems, synthetic aperture radar automatic target recognition, hand gestures recognition, anomalies detection for video surveillance systems, disease detection, social networks analysis, or user experience. The collection of articles shows the wide use of deep learning techniques, although classical machine learning techniques, among others, are also present.

The issue starts with a topic that is getting a lot of attention from various media due to its popularity, because we can all easily use tools like chatGPT, Midjourney or Dall-E. These tools are considered generative AI that can generate content very similar to that generated by humans. The first article by Garcia-Peñalvo and Vázquez-Ingelmo, presents a literature mapping of AI-driven content generation, analyzing 631 solutions published over the last five years to better understand and characterize the generative AI landscape. Due to the concerns and acceptance issues that have arisen in society as a result of the emergence of this technology, the authors suggest more comprehensive understanding of what generative AI entails, so that the potential challenges are addressed more pragmatically and effectively.

The second article of this issue focuses on brain and mind inspired computing (BMC), an emerging research field. Liu and Wei design a model and framework for BMI theory. Based on the brain mechanism and mind architecture, they propose a semantic-oriented multimedia neural, cognitive computing model for multimedia semantic computing. Besides, they propose a hierarchical cross-modal cognitive neural computing framework for cross-modal information processing. Moreover, they propose a cross-modal neural, cognitive computing architecture for a remote sensing intelligent information extraction platform and unmanned autonomous system. Their research on remote sensing intelligent information extraction and cross-media information retrieval shows that the scene classification, target detection, target classification and target recognition based on the BMC algorithm work. The BMI theory proposed can be widely used in high-resolution earth observation systems and many other applications.

Next article describes a solution based on convolutional neural networks for bird species identification, which is crucial for avian diversity conservation. Das et al. explore the ability of deep transfer models such as VGG16, VGG19 and InceptionV3 to effectively extract and discriminate speech signals from different species of birds. The networks were trained using data from 37 bird species, obtaining accuracies equal to 78, 61.9 and 85%, respectively. In practical terms, the suggested method may be of great use to ornithologists by making the identification of bird species a straightforward process.

Also using convolutional neural networks, Vélez de Mendizabal et al. propose a new technique to decode Leetspeak. This is a type of slang writing that replaces some characters with visually similar symbols, preventing the spam classifiers recognising words, so that the spam emails are not detected. When messages are deobfuscated, the performance of the classifiers increases and reaches, in many cases, the values obtained when messages have not been obfuscated.

The authors propose a reliable convolutional neural network (CNN) design for Leetspeak deobfuscation processes and its evaluation, an image database used for training, and four datasets for evaluating Leetspeak deobfuscation processes.

Another type of neural networks, gated graph neural networks, are used in the next article. Seo and Kim target the problem of sequential recommendation to predict user's next action based on personal action sequences. Data sparsity is a challenge in these problems and translation-based recommendations, which learn distance metrics to capture interactions between users and items in sequential recommendations, contributes to overcome this issue. The authors propose an attentive flexible translation for recommendations (AFTRec) to predict the user's next item in sparse sequential recommendation datasets. Experiments using four sparse datasets and one dense dataset with different domains show that AFTRec outperforms the state-of-the-art baselines in terms of normalized discounted cumulative gain and hit rate on sparse datasets.

By also using attention mechanisms, Ukwuoma et al. target the problem of synthetic aperture radar (SAR) automatic target recognition (ATR). Their paper introduces a new attention based ResNet architecture appropriate for the SAR recognition task. They propose a simple channel attention mechanism into a ResNet architecture for SAR ATR involving only a handful of parameters while attaining clear performance gains. They also explore the One Policy Learning Rate on the ResNet-50 architecture and compare it with the proposed attention based ResNet-50 architecture. With the attention based model and the One Policy Learning Rate-based architecture, they were able to obtain recognition rates of 100% and 99.8%, respectively.

Next paper presents a solution that allows the recognition of hand gestures by analyzing three-dimensional landmarks using also deep learning technology. Osimani et al. propose the identification of 9 hand gestures by interpreting a cloud of 3D reference points obtained through a standard RGB camera. They introduce a neural network architecture that has a small number of hidden layers but high prediction hit rate of hand gestures. One of the main contributions, that considerably improves the performance, is a first layer of normalization and transformation of the landmarks. In their experimental analysis, they reach an accuracy of 99.87% recognizing 9 hand gestures.

In the field of video surveillance, Qasim Gandapur and Verdú present an automated deep learning model that detects and prevents anomaly activities. A real-world video surveillance system is designed based on the ResNet-50 architecture to extract the high-level features from input streams, while temporal features are extracted by a convolutional gated recurrent unit from the ResNet-50 extracted features in the time-series dataset. The UCF-Crime dataset is used to evaluate the proposed deep learning model, achieving 82.22% accuracy. In addition, the proposed model outperforms related deep learning models.

Next article, by Mariammal et al., examines the performance of classical machine learning methods as bagging, random forest, support vector machine, decision tree, Naïve Bayes and k-nearest neighbor classifiers using a crop dataset, a prisoners' dataset and the iris dataset. The results show that the bagging ensemble technique outperforms the rest.

Torres et al. focus on the development of a system for the detection of downy mildew disease in roses through image analysis using convolutional neural networks and the correlation of environmental

variables through an experiment in a controlled environment. Besides, they develop an IoT platform that integrates the artificial intelligence module. The model is validated comparing with three different models of neural networks. According to the tests and the analysis of the results obtained with the microclimatic variables, it is observed that the relative humidity variable can influence the development and appearance of downy mildew disease when its value is above 85% during an extended period of the system.

Also targeting disease detection, the article by Intriago-Pazmiño et al. describes a study of four metrics to find their best parameters. The metrics are Contrast Improvement Index (CII), Enhancement Measurement Estimation (EME), Entropy EME (EMEE), and Entropy. The metrics are applied to two cases of studies: fundus and mammography images, on five datasets. These datasets contain healthy and pathological images, and some are poor quality images. Based on the experimental results provided, the conclusion is that EMEE, EME, and CII metrics are valuable for measuring the enhancement of the studied medical images.

A clustering validation index (CVI) is a metric used to evaluate the results of a clustering algorithm. In the next article, Kumar Sharma et al. propose an internal CVI to be used as a complementary measure to the available internal CVIs. These are used frequently in clustering to measure the goodness of the clustering algorithms without taking any external inputs. The proposed index uses a modified compactness measure and an updated separation measure, based on the notion of S-divergence. A total of 10 databases of two classes, synthetic and realworld ones, are considered in this work to prove the effectiveness of the proposed metric over some of the most popular existing internal CVIs. Empirical results with four popular clustering algorithms show that the proposed index is proficient in determining the number of clusters and the best partition for several of the databases, including the database with arbitrary cluster shapes.

In the following paper, Debbi proposes a causal explanation technique for conjunctive queries in probabilistic databases. While query answer explanation in relational databases focuses on why is a tuple in a query result, in probabilistic databases, it should also explain why the tuple has a certain probability. Based on the notions of causality, responsibility and blame, the author addresses explanation for tuple and attribute uncertainties in a complementary way. Through an experiment on a real-dataset, the framework shows to be helpful for explaining complex queries results. Comparing to existing explanation methods, the method could be also considered as an aided-diagnosis method through computing the blame, which helps to understand the impact of uncertain attributes.

A platform that allows the automatic detection of irregular swimming pools is proposed in the next article by Sánchez San Blas et al. The platform uses geographic information tools (GIS) based on orthophotography, combined with advanced machine learning techniques for object detection, as well as a multi-agent architecture, which allows distributed computing and the evaluation of different algorithms combined to improve the detection process. The system has been validated by testing it in different towns in Spain, showing that it is possible to determine the presence of a pool in an image with an accuracy better than 97%.

Focusing on social networks area, the article by Cavaliere et al. presents an emotion-aware solution to analyse users' reactions towards topics constantly discussed over time or in a specific brief period in Twitter. The rationale behind the approach is the combination of emotional analysis of tweet content with the time frequent analysis of relevant topic itemsets, and tweet spread to detect those topics that may have the highest impact. First, the method extracts topics as frequent itemsets from tweets, then the support over time and RoBERTa-based

sentiment analysis are applied to assess the current topic spread and the emotional impact, next a time-grid-based approach allows a granule-level analysis that serves to predict future users' reactions towards topics. Finally, a score function allows building comparative ranked lists of the most relevant topics according to topic sentiment, importance and spread. Experiments demonstrate the potential of the framework on the IEEE COVID-19 Tweets Dataset.

In the same area of social networks, the work by Muñoz et al. develops a methodology for analysing tourism through complex mathematical algorithms, based on unstructured data extracted from social networks. Specifically, graphic and textual data from the profiles of Instagram users feed the classification models. These mine user demographic information and gain insight on what the users were doing in each of their posts, trying to classify that information into any of the categories discovered in the article, acting as a discovery tool for the tourism industry. This has great potential for comparisons on larger amounts of data and even between tourism profiles between cities.

Strukova et al. examine diverse technology-mediated environments that can generate rich data sets through users' interaction and where data can be used to explicitly or implicitly perform a data-driven evaluation of different competencies and capabilities. Their survey revealed four key multimedia environments that include sites for content sharing and consumption, video games, online learning and social networks. The authors found that all these environments are highly correlated with the measurement and development of capabilities such as expertise, language proficiency and soft skills. According to the surveyed studies, this measurement was done with the application of different methods (machine learning, network analysis, natural language processing, statistics and experimental design), which the authors also discuss in detail.

The User Experience Questionnaire (UEQ) is one of very few standard user experience questionnaires available in many different languages. Next article by Hernández-Campos et al. analyses changes in some items of the UEQ for use in the context of Costa Rican culture. Although a Spanish version of the UEQ exists, the authors use a double-translation and reconciliation model for detecting the most appropriate words for Costa Rican culture. These resulted in 7 new items that were added to the original Spanish version. 161 participants participated in a study that examined the original items and the new ones. The results show that the Costa Rican version is neither better nor worse than the original Spanish version, therefore the UEQ is very robust to some changes in the items.

The last article of this issue also relates to user experience area. The objective of the study by Alonso-Virgós et al. is to know if there are "user response" guidelines that a developer with no training or usability experience applies intuitively. Besides the study aims to know the most important recommendations and guidelines, according to the web developers themselves. Knowing the most forgotten recommendations by web developers helps to design effective and specific training in this field.

In closing, I would like to thank both the authors and the reviewers for their commitment to quality assurance and improvement of the articles, aiming for the best reader experience.

Dr. Elena Verdú
Editor-in-Chief

Universidad Internacional de La Rioja