mathematics

Editorial

Advances in Artificial Intelligence and Statistical Techniques with Applications to Health and Education

Carmen Lacave * and Ana Isabel Molina

Department of Information Technologies and Systems, Universidad de Castilla-La Mancha, Paseo de la Universidad, s/n, 13071 Ciudad Real, Spain; anaisabel.molina@uclm.es
* Correspondence: carmen.lacave@uclm.es

Citation: Lacave, C.; Molina, A.I. Advances in Artificial Intelligence and Statistical Techniques with Applications to Health and Education. *Mathematics* **2023**, *11*, 1344. https://doi.org/10.3390/math11061344

Received: 6 March 2023
Accepted: 8 March 2023
Published: 9 March 2023

The COVID-19 pandemic highlighted the importance of health and education and also revealed the need for innovative solutions relative to the challenges confronting these disciplines. The pandemic led to a greater demand for remote healthcare services and online education, which has highlighted the potential for technology to transform these fields. With the increasing availability of big data and advancements in artificial intelligence, there is an opportunity to develop more efficient and effective approaches to healthcare and education. The integration of technology in healthcare and education has already shown promising results. In healthcare, machine learning algorithms have been used to predict the likelihood of disease progression, identify at-risk patients, and personalize treatment plans. In education, learning analytics has been used to track student progress and identify areas where they may need extra support. These technologies can provide valuable insights that can inform decision making and lead to better outcomes.

Therefore, the proposed Special Issue on the application of artificial intelligence to healthcare and education is timely and important. The challenges facing these disciplines require a multidisciplinary approach that combines technological advances with expertise in statistics, data analysis, and other related fields. By sharing new methods, applications, and case studies, this Special Issue can contribute to the development of innovative solutions that improve healthcare and education for all. Topics addressed in this Special Issue include data mining, machine learning, learning analytics, prediction methods, pattern recognition, decision analysis, probabilistic reasoning, fuzzy systems, student or patient modelling, adaptive systems, collaborative systems, recommendation systems, experimental design, and empirical study cases. Specifically, there are twenty rigorously reviewed papers included in this Special Issue, with eleven specializing in the field of medicine and health [1–11] and nine specializing in the field of education [12–20]. In total, fifteen papers (43% of the received) were rejected for publication.

In the area of **health**, most papers aim at improving the *diagnosis of a disease*, such as influenza [3], Alzheimer's disease [7], and cardiac arrhythmias [11], or of a specific symptom, such as changes in the retinal vasculature [1], mental fatigue [4], gait variability [5] or vitamin D deficiency [8]. There is one for the *personalization* of rehabilitation exercises for stroke patients [6], and there are three theoretical papers [2,9,10] aimed at *improving the representation of models*, which can be applied to different and actual medical problems.

Regarding the techniques used, those related to artificial intelligence prevail. Specifically, the majority make use of different machine learning algorithms [1–4,6,8,11], although with distinct objectives. For example, some works propose new specific classification models. In [1], a novel semantic segmentation network is developed to detect retinal vessels from original images, thus providing accurate support during ophthalmological analysis. In [4], a combination of electrocardiogram signal feature extraction, principal component analysis, and random forest classification was used to propose a novel approach to detect mental fatigue with a 94.5% of accuracy. In [11], a discriminative convolutional sparse coding framework and a classification strategy based on a linear support vector machine were combined to diagnose abnormal arrhythmias, exhibiting an accuracy of 99.32%. The study in [8] compares different supervised learning algorithms to predict the anthropometric parameter that is most strongly associated with vitamin D deficiency. They conclude that the Naïve Bayes model showed the best area under the curve in four of six parameters and the logistic regression model for the other two. In [3], the diagnosis of influenza-like illness was attained by using a new sentiment classification model based on one-dimensional convolutional neural networks from Facebook posts and based on the user's sentiments. The experimental results obtained an accuracy of 96.6%.

Another approach was proposed by [2] with respect to improving the efficiency of classification algorithms when there are memory limitations. Then, an improvement in value-based potentials to better represent approximate information and positive results was obtained after verifying it on several Bayesian networks representing medical problems.

A different proposal is that of [6] in which fuzzy logic was used to propose a system capable of recommending rehabilitation exercises for stroke patients depending on the variation in the patient's level of progress as the rehabilitation process evolves.

Statistical techniques have also been used, such as from the most basic ones, making use of descriptive statistics, discriminant analysis, and correlations [5], to more sophisticated ones such as functional data analysis [7] or an optimal experimental design [9,10]. Thus, basic statistical techniques in [5] were used to analyse the data obtained from an experiment performed with adults with the aim to characterize the gait parameters affected by mobile-based dual-tasking and the impact of normal cognitive decline due to aging. In [7], functional data analysis was used to provide insight into the nature of the difference in movement patterns of Alzheimer's patients according to their stage of the disease. The results indicate that the distribution of acceleration differs between each stage of the disease, mainly due to scale differences. In [9], an analysis of the effect of misspecification in the probability distribution of the response variable was performed by comparing a normal distribution with the Poisson or gamma distribution. The results show that assuming a homoscedastic normal distribution when obtaining optimal designs may led to a great loss in efficiency depending on the true distribution of the response variable and on the model function chosen. In [10], new algorithms were developed to calculate D-optimal designs by describing the electrical behaviour of bioelectrodes, cell membranes, or biological tissues; as a result, a substantial amount of observations obtained the same results as the classic design is saved.

It should be noted that the data used for the above studies have been obtained from public repositories in papers [1,2,11]. The other papers obtain their data from different types of experimental studies, most of them conducted with individuals [3–8].

In the field of **education**, work is mainly oriented towards improving teaching [12–14] or learning [18] and predicting educational performance [15–17,20] or drop-out rates [19].

Within the first group, there are different proposals to *improve teaching*. Thus, in [12] an open-source software tool, called *OpenMarkov*, was used as a pedagogical tool to teach the main concepts of Bayesian networks and influence diagrams using two models to represent and deal with uncertainty in artificial intelligence systems, which are difficult to convey. In [13], a framework based on natural language processing techniques to analyse college-level subjects' didactic planning was presented. Such a framework was supported by an ontology that considers the information from didactic plans approved by an evaluation committee. The statistical analysis of the framework's evaluation results indicated an agreement of 92% with the expert and a reduction in the workload of instructors. Natural language processing techniques are also used in [14], in which a knowledge discovery model based on machine learning and text mining was proposed for a structured development consisting of analogies to teach programming concepts. The input of the proposal is a didactic teaching strategy with analogies, and it obtained a series of patterns for possible scenarios that can be used with a greater degree of assertiveness when presenting an analogy on a CS1 course. The obtained structure was validated by the professors and the students.

A different approach is that of [18], in which genetic algorithms were used to automatically form groups in collaborative learning scenarios, considering the students' personality traits of the big five model as a criterion for grouping. As a result, a strategy was provided to assess, in each context, which type of approach (homogeneous, heterogeneous, or mixed) is the most appropriate for organising groups.

The papers that predict *educational performance* make use of different techniques, such as techniques from artificial intelligence and statistics. For example, in [15], the prediction of high-risk students was performed by a new hybrid deep learning model with respect to a student's learning behavior. The model is based on a combination of a gated unit neural network, a deep neural network, and a neural attention mechanism. The experimental results show better performances than other prediction models based on deep learning. In [17], a statistical depth-based supervised classification technique was used to develop a methodology to allow teachers to define the most appropriate variables in order to measure the academic performance of students. The proposal takes as input the data recorded in the first tasks of the academic year. The experimental results show an 80% success rate from the two first tasks (of six), increasing to 90% after the first four, in predicting the final grade. Moreover, the proposal of [16] compared different statistical and artificial intelligence techniques to determine if the performance of a student in the continuous assessment determines the score obtained in the final exam of the course. Those techniques are based on linear, quantile, and logistic regression; artificial neural networks; and Bayesian networks. The results obtained show that although continuous assessments influence the final exam score, it was not decisive in predicting it. However, in [20], only statistical techniques based on quantile regression were used to predict the relations between students' academic performance and their after-school exercise. Using empirical data collected from middle school students, they proved that there exists a positive relation among both variables.

This variety of techniques also appears in papers related to the prediction of *student dropout*. Thus, in [19], both machine learning models such as support vector machines, decision trees, and neural networks, together with logistic regression, were used to predict likely university dropout rates either at the beginning of the course of study or at the end of the first semester. The results show that dropout prediction cannot be made only with enrollment data, but that academic performance and the level of course preference must be considered. In general, machine learning techniques obtain similar predictions, although there are some areas in which those techniques offer worse results. Moreover, although logistic regression is never the best compared with other more sophisticated techniques, it can be used as a good approach in non-specialized audiences because of its ease of use.

In conclusion, the papers presented in this Special Issue provide a significant snapshot of the current research interests of the scientific community in the fields of medicine and education. In the medical field, there are clear efforts to develop and provide specialists with systems that can help make accurate diagnoses when problems are complex. This has led to the development of advanced technologies, such as machine learning algorithms, that can analyze vast amounts of medical data to help diagnose diseases more accurately and efficiently. Additionally, there is a crucial line of research focused on adapting treatments according to the needs and evolution of patients. This involves taking a personalized approach to medical care, where the individual needs of the patient are taken into account to create tailored treatment plans. This research is particularly important for conditions that have diverse symptoms and complex etiologies.

In the field of education, the focus is mainly on the early detection of school failure or dropout, with the aim of designing strategies that can address both problems in a timely manner and that can improve current failure rates. One significant area of research is the creation of tools or models that support students' learning based on their unique characteristics. This includes leveraging data-driven approaches to identify students' strengths and weaknesses, as well as creating personalized learning plans that are tailored to each student's learning style and abilities. Another important area of research in education is the creation of tools that support teachers in facilitating effective teaching. This includes the development of interactive learning tools and intelligent tutoring systems that can provide feedback to both teachers and students, as well as the use of data analytics to assess student progress and adjust teaching strategies accordingly.

Overall, the scientific community in both medicine and education is leveraging techniques from statistics and artificial intelligence to drive significant advances in their respective fields. These approaches hold great promise for improving patient outcomes and enhancing the quality of education for students around the world.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arsalan, M.; Haider, A.; Koo, J.H.; Park, K.R. Segmenting Retinal Vessels Using a Shallow Segmentation Network to Aid Ophthalmic Analysis. *Mathematics* **2022**, *10*, 1536. [CrossRef]
2. Bonilla-Nadal, P.; Cano, A.; Gómez-Olmedo, M.; Moral, S.; Retamero, O.P. Using Value-Based Potentials for Making Approximate Inference on Probabilistic Graphical Models. *Mathematics* **2022**, *10*, 2542. [CrossRef]
3. Boulesnane, A.; Meshoul, S.; Aouissi, K. Influenza-like Illness Detection from Arabic Facebook Posts Based on Sentiment Analysis and 1D Convolutional Neural Network. *Mathematics* **2022**, *10*, 4089. [CrossRef]
4. Butkevičiūtė, E.; Michalkovič, A.; Bikulčienė, L. ECG Signal Features Classification for the Mental Fatigue Recognition. *Mathematics* **2022**, *10*, 3395. [CrossRef]
5. Carneros-Prado, D.; Dobrescu, C.C.; González, I.; Fontecha, J.; Johnson, E.; Hervás, R. Analysis of Dual-Tasking Effect on Gait Variability While Interacting with Mobile Devices. *Mathematics* **2023**, *11*, 202. [CrossRef]
6. Gmez-Portes, C.; Castro-Schez, J.J.; Albusac, J.; Monekosso, D.N.; Vallejo, D. A Fuzzy Recommendation System for the Automatic Personalization of Physical Rehabilitation Exercises in Stroke Patients. *Mathematics* **2021**, *9*, 1427. [CrossRef]
7. Nieto-Reyes, A.; Battey, H.; Francisci, G. Functional Symmetry and Statistical Depth for the Analysis of Movement Patterns in Alzheimer's Patients. *Mathematics* **2021**, *9*, 820. [CrossRef]
8. Patino-Alonso, C.; Gómez-Sánchez, M.; Gómez-Sánchez, L.; Sánchez Salgado, B.; Rodríguez-Sánchez, E.; García-Ortiz, L.; Gómez-Marcos, M.A. Predictive Ability of Machine-Learning Methods for Vitamin D Deficiency Prediction by Anthropometric Parameters. *Mathematics* **2022**, *10*, 616. [CrossRef]
9. Pozuelo-Campos, S.; Casero-Alonso, V.; Amo-Salas, M. Effect of Probability Distribution of the Response Variable in Optimal Experimental Design with Applications in Medicine. *Mathematics* **2021**, *9*, 1010. [CrossRef]
10. Sebastià Bargues, À.; Polo Sanz, J.-L.; Martín Martín, R. Optimal Experimental Design for Parametric Identification of the Electrical Behaviour of Bioelectrodes and Biological Tissues. *Mathematics* **2022**, *10*, 837. [CrossRef]
11. Zhang, B.; Liu, J. Discriminative Convolutional Sparse Coding of ECG Signals for Automated Recognition of Cardiac Arrhythmias. *Mathematics* **2022**, *10*, 2874. [CrossRef]
12. Díez, F.J.; Arias, M.; Pérez-Martín, J.; Luque, M. Teaching Probabilistic Graphical Models with OpenMarkov. *Mathematics* **2022**, *10*, 3577. [CrossRef]
13. Hernández-Carrasco, D.A.; Rose-Gómez, C.E.; González-López, S.; López-López, A.; García-Gorrostieta, J.M.; Borrego, G. A Framework to Assist in Didactic Planning at Undergraduate Level. *Mathematics* **2022**, *10*, 1355. [CrossRef]
14. Jiménez Toledo, J.A.; Collazos, C.A.; Ortega, M. Discovery Model Based on Analogies for Teaching Computer Programming. *Mathematics* **2021**, *9*, 1354. [CrossRef]
15. Liu, T.; Wang, C.; Chang, L.; Gu, T. Predicting High-Risk Students Using Learning Behavior. *Mathematics* **2022**, *10*, 2483. [CrossRef]
16. Morales, M.; Salmerón, A.; Maldonado, A.D.; Masegosa, A.R.; Rumí, R. An Empirical Analysis of the Impact of Continuous Assessment on the Final Exam Mark. *Mathematics* **2022**, *10*, 3994. [CrossRef]
17. Nieto-Reyes, A.; Duque, R.; Francisci, G. A Method to Automate the Prediction of Student Academic Performance from Early Stages of the Course. *Mathematics* **2021**, *9*, 2677. [CrossRef]
18. Revelo Sánchez, O.; Collazos, C.A.; Redondo, M.A. Automatic Group Organization for Collaborative Learning Applying Genetic Algorithm Techniques and the Big Five Model. *Mathematics* **2021**, *9*, 1578. [CrossRef]
19. Segura, M.; Mello, J.; Hernández, A. Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics* **2022**, *10*, 3359. [CrossRef]
20. Shin, K.; You, S. Quantile Regression Analysis between the After-School Exercise and the Academic Performance of Korean Middle School Students. *Mathematics* **2022**, *10*, 58. [CrossRef]