Special Issue on Computational Intelligence for Healthcare

Gabriella Casalino *[ID] **and Giovanna Castellano ***[ID]

Computer Science Department, University of Bari Aldo Moro, 70125 Bari, Italy
* Correspondence: gabriella.casalino@uniba.it (G.C.); giovanna.castellano@uniba.it (G.C.)

The number of patient health data has been estimated to have reached 2314 exabytes by 2020. Traditional data analysis techniques are unsuitable to extract useful information from such vast quantity of data. Thus, intelligent data analysis methods combining human expertise and computational models for accurate and in-depth data analysis are necessary. The technological revolution and medical advances made by combining vast quantities of available data, cloud computing services, and AI-based solutions can provide expert insight and analysis on a mass scale and at a relatively low cost. Computational intelligence (CI) methods such as fuzzy models, artificial neural networks, evolutionary algorithms, and probabilistic methods have recently emerged as promising tools for the development and application of intelligent systems in healthcare practice. CI-based systems can learn from data and evolve according to changes in the environments by taking into account the uncertainty characterizing health data, including omics data, clinical data, sensor, and imaging data. The use of CI in healthcare can improve the processing of such data to develop intelligent solutions for prevention, diagnosis, treatment, and follow-up, as well as for analysis of administrative processes.

The present Special Issue on Computational Intelligence for healthcare is intended to show the potential and the practical impacts of CI techniques in challenging healthcare applications. The Special Issue received several submissions, all of which went through a rigorous peer-review process. After the review process, twelve papers were selected on the basis of the review ratings and comments. These selected papers range over main applications of CI in healthcare.

A special case of medical data is the data generated by omics technologies, which enable DNA decoding and genome sequencing. Such data represent the expression of genes or portions thereof in experimental subjects or in cell lines produced in the laboratory. The study of the complex interactions between genes makes it possible to understand their role in the course of a particular disease. In particular, the analysis of the different biological levels allows us to better address the knowledge of pathogenetic mechanisms at the molecular level, allowing the identification of biomarkers that allow us to improve the diagnosis or even to plan personalized therapies However, the complexity of the relationships and the uncertainty present in the data collected and in the phenomena studied make it necessary to use specific methods for the treatment of information with these characteristics. The first paper, entitled, "Explaining Ovarian Cancer Gene Expression Profiles with Fuzzy Rules and Genetic Algorithms" [1], addresses the analysis of gene expression data for approaching the study of expression profiles in ovarian cancer compared to other ovarian diseases. The work combines a feature selection among genes that is guided by the genetic algorithm into the creation of fuzzy if–then rules that explain how classes can be distinguished by observing changes in the expression of selected genes. After testing several parameters, a final model was obtained consisting of 10 genes involved in the molecular pathways of cancer and 10 rules that correctly classify all samples. Omics data of a different kind, namely sgRNA sequences, have been analyzed in the paper "CRISPRLearner: a deep learning-based system to predict CRISPR/Cas9 sgRNA on-target cleavage efficiency" [2]. In particular, ten datasets were considered. After a pre-processing step, including data standardization and augmentation, a convolutional neural network was used to predict

sgRNA cleavage efficiency. Experiments on benchmark datasets showed the effectiveness of the proposed method in correctly identifying the disease. Moreover, a comparison with state-of-the-art methods shows the superiority of the proposed deep-learning-based model.

In addition, detection and analysis of physiological data acquired from sensors are an essential process in smart healthcare applications. Indeed, with the advent of low-cost sensors and fast networks, a new discipline called Internet of Medical Things (IoMT) has emerged. It allows a continuous monitoring of patients through intelligent objects. Physiological data analysis can be performed in fog computing to abridge the excess latency introduced by cloud computing. However, the latency for the emergency health status and overloading in fog environment become key challenges for smart healthcare. The paper "Intelligent Fog-Enabled Smart Healthcare System for Wearable Physiological Parameter Detection" [3] presents a novel healthcare architecture for physiological parameter detection that resolves these problems. The overall system is built upon three layers. In the first layer, data from the wearable devices of patients are subjected to fault detection in a personal data assistant (PDA) via a rapid kernel principal component analysis (RK-PCA) algorithm. Then, in the second layer, the faultless data are processed to remove redundancy via a new fuzzy assisted objective optimization by ratio analysis (FaMOORA) algorithm. A two-level health-hidden Markov model (2L-2HMM) finds the user's health status from temporal variations in data collected from wearable devices. Finally, the user's health status is detected in the third layer through a hybrid machine learning algorithm called SpikQ-Net, and according to the user's health status, immediate action is taken. The proposed tri-fog health model is validated by a thorough simulation showing better achievements in latency, execution time, detection accuracy, and system stability.

When different sensors are used for data acquisition, synchronization is a critical factor. In the article "A Synchronized Multi-Unit Wireless Platform for Long-Term Activity Monitoring", a time-synchronized multi-unit, multi-sensor, and multi-rate acquisition system for kinematic and static analysis is proposed [4]. It is a wearable multi-board acquisition system for offline activity monitoring. A master–slave architecture was used to syncronize measures acquired from different sensors. Moreover, a mobile Android application was developed in order to manage the data acquisition. The high modularity of the proposed platform makes it general-purpose. Indeed, experiments on different scenarios have been carried out to validate its performance. In particular, a case study of surface electromyography (sEMG) was used for monitoring muscle activities during walking. sEMG signals are also used in the article "A Deep Learning Approach to EMG-Based Classification of Gait Phases During Level Ground Walking" for gait phase classification [5]. Specifically, an Artificial Neural Network (ANN) was proposed to classify gait events and to predict foot–floor contact. The use of ANN has allowed the automatic selection of relevant features in data, thus avoiding the data engineering phase, which is necessary when using other machine learning algorithms. In vivo experiments have been conducted at the Movement Analysis Laboratory of Università Politecnica delle Marche to acquire gait signals. Raw signals were pre-processed to obtain the final labeled segments of the walking phases. Four different architectures of the multi layer perceptron algorithm (MLP) were proposed by modifying the model structure, and the results have been compared. The aim of the analysis was to detect the transitions between gait phases. Furthermore, a comparison with a feature-based (FB) method has shown that the best MLP model is more accurate in detecting phase changes.

A further essential and crucial task in healthcare is image analysis to support medical diagnosis. In particular, recent advances in neuroimaging techniques, such as diffusion tensor imaging (DTI), represent a crucial resource for brain image analysis in order to detect alterations related to severe neurodegenerative disorders, such as Alzheimer's disease (AD). The paper "An Ensemble Learning Approach Based on Diffusion Tensor Imaging Measures for Alzheimer's Disease Classification" [6] presents an ensemble learning approach for the automatic discrimination between healthy controls and AD patients, using DTI measures as predicting features and a soft-voting ensemble approach for the classification. The

proposed approach, efficiently combining single classifiers trained on specific groups of features, is able to improve classification performances with respect to the comprehensive approach of the concatenation of global features and at the same time reducing the dimensionality of the feature space and in turn the computational effort. A different task on images has been implemented in the article "Leukemia Image Segmentation by Using a Hybrid Histogram-Based Soft Covering Rough k-Means Clustering Algorithm" [7]. Image segmentation is the task of partitioning a given image in not-overlapped areas to detect regions of interest. Particularly, authors propose a leukemia diagnosis support system through nucleus segmentation. A soft set together with a rough set were used to represent the uncertainty in nucleus images. A four-step pipeline is proposed. Images are firstly pre-processed, and then a clustering algorithm is applied. A histogram-based method (HSCRKM) is proposed to identify the optimal number of cluster. Then, different features are extracted from the images, and the resulting data are used to predict the areas in the image as belonging to the leukemia tumor class or the healthy class. Several clustering and classification methods have been compared to identify the optimal pair. Results show that the proposed HSCRKM overcomes the compared clustering methods. Moreover, all the classification models increased their performance when trained on groups coming from HSCRKM. However, among all the considered prediction methods, logistic regression and neural network provided the best performance (average accuracy higher than 90%). Diabetic Retinopathy (DR) images are analyzed in the paper "Blended Multi-Modal Deep Convnet Features for Diabetic Retinopathy severity Prediction" for an early recognition of the disease [8]. Both uni-modal and multi-modal approaches, which combine data coming from different sources, were used to predict the severity level of the disease (healthy, mild, moderate, severe, and proliferative). To this aim, Deep Neural Networks (DNN) have been proposed. In particular, in the uni-modal approach, a single pre-trained ConvNet is used to extract the final feature representation. In the multi-modal approach, the final feature representation is obtained by blending deep features extracted from multiple ConvNets.

One main factor that hampers the effectiveness of CI methods in the medical domain is the imbalanced nature of medical data due to non-uniform distribution of the number of instances per class. The paper entitled "Integrating Enhanced Sparse Autoencoder-Based Artificial Neural Network Technique and Softmax Regression for Medical Diagnosis" [9] addresses the problem of unbalancement in medical datasets to create robust models for the prediction of different diseases. The authors propose an approach that integrates an enhanced sparse autoencoder (SAE) for effective feature learning and an optimized Softmax regression for robust classification. When employed for the prediction of three different diseases, namely chronic kidney disease, cervical cancer, and heart disease, the proposed method provides higher test accuracies compared to other machine learning algorithms. In addition, the paper "Two-Stage Monitoring of Patients in Intensive Care Unit for Sepsis Prediction Using Non-Overfitted Machine Learning Models" [10] addresses the problem of unbalanced clinical data. In this case, data concern patients in the Intensive Care Unit (ICU) to face the problem of early detection of sepsis, collected within the PhysioNet/Computing in Cardiology Challenge 2019. The labeled clinical dataset includes only 2% records with the sepsis label, leading to highly unbalanced dataset. To address these issues, the authors propose a method using two separate ensemble models to take into the account the amount of time the patients spent in the ICU. A total of 44 different methods, based on decision trees, naive Gaussian Bayes, SVM, and ensemble learners,are compared. Results show the effectiveness of the proposed method. Moreover, the considered machine learning models return comparable utility score values when the number of features is reduced, suggesting that feature engineering is necessary.

Long-term electrocardiogram (ECG) is used to detect Premature Ventricular Contraction (PVC) in the paper "Searching for Premature Ventricular Contraction from Electrocardiogram by Using One-Dimensional Convolutional Neural Network" [11]. A one-dimensional convolutional neural network (CNN) has been used for the prediction tasks. It allows avoiding pre-processing phases on data, as required by common machine learning

algorithms, and it is able to directly analyze raw data while extracting the most relevant features from it. High performance values are returned by the diagnostic system (accuracy of 99.64%).

Finally, a multistage support vector machine model has been proposed for early recognition of Unipolar Depression (UD) disease in the article "Realizing an integrated multistage support vector machine model for augmented recognition of unipolar depression" [12]. A pre-processing phase for feature ranking is implemented in order to reduce the data dimensionality. Comparison with other machine learning algorithms has shown the effectiveness of the proposed approach in correctly identifying the disease, other than overcoming their performance. Moreover, the recursive feature selection method has proved to be able to improve the accuracy of the classifiers.

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

References

1. Consiglio, A.; Casalino, G.; Castellano, G.; Grillo, G.; Perlino, E.; Vessio, G.; Licciulli, F. Explaining Ovarian Cancer Gene Expression Profiles with Fuzzy Rules and Genetic Algorithms. *Electronics* **2021**, *10*, 375. [CrossRef]
2. Dimauro, G.; Colagrande, P.; Carlucci, R.; Ventura, M.; Bevilacqua, V.; Caivano, D. CRISPRLearner: A deep learning-based system to predict CRISPR/Cas9 sgRNA on-target cleavage efficiency. *Electronics* **2019**, *8*, 1478. [CrossRef]
3. Ijaz, M.; Li, G.; Wang, H.; El-Sherbeeny, A.M.; Moro Awelisah, Y.; Lin, L.; Koubaa, A.; Noor, A. Intelligent Fog-Enabled Smart Healthcare System for Wearable Physiological Parameter Detection. *Electronics* **2020**, *9*, 2015. [CrossRef]
4. Coviello, G.; Avitabile, G.; Florio, A. A synchronized multi-unit wireless platform for long-term activity monitoring. *Electronics* **2020**, *9*, 1118. [CrossRef]
5. Morbidoni, C.; Cucchiarelli, A.; Fioretti, S.; Di Nardo, F. A deep learning approach to EMG-based classification of gait phases during level ground walking. *Electronics* **2019**, *8*, 894. [CrossRef]
6. Lella, E.; Pazienza, A.; Lofù, D.; Anglani, R.; Vitulano, F. An Ensemble Learning Approach Based on Diffusion Tensor Imaging Measures for Alzheimer's Disease Classification. *Electronics* **2021**, *10*, 249. [CrossRef]
7. Inbarani H, H.; Azar, A.T. Leukemia image segmentation using a hybrid histogram-based soft covering rough k-means clustering algorithm. *Electronics* **2020**, *9*, 188. [CrossRef]
8. Bodapati, J.D.; Naralasetti, V.; Shareef, S.N.; Hakak, S.; Bilal, M.; Maddikunta, P.K.R.; Jo, O. Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. *Electronics* **2020**, *9*, 914. [CrossRef]
9. Ebiaredoh-Mienye, S.A.; Esenogho, E.; Swart, T.G. Integrating Enhanced Sparse Autoencoder-Based Artificial Neural Network Technique and Softmax Regression for Medical Diagnosis. *Electronics* **2020**, *9*, 1963. [CrossRef]
10. Abromavičius, V.; Plonis, D.; Tarasevičius, D.; Serackis, A. Two-Stage Monitoring of Patients in Intensive Care Unit for Sepsis Prediction Using Non-Overfitted Machine Learning Models. *Electronics* **2020**, *9*, 1133. [CrossRef]
11. Yu, J.; Wang, X.; Chen, X.; Guo, J. Searching for Premature Ventricular Contraction from Electrocardiogram by Using One-Dimensional Convolutional Neural Network. *Electronics* **2020**, *9*, 1790. [CrossRef]
12. Srinivasan, K.; Mahendran, N.; Vincent, D.R.; Chang, C.Y.; Syed-Abdul, S. Realizing an integrated multistage support vector machine model for augmented recognition of unipolar depression. *Electronics* **2020**, *9*, 647. [CrossRef]