Introduction to the Special Issue on Computational Intelligence for Biomedical Data and Imaging

Computational intelligence in the field of biomedical data includes working in the field of image processing, computer vision, machine learning, deep learning, big data analytics, and cloud computing. These tools play a significant role in the development of computer-aided diagnosis of medical problems and help the researchers working in the field of biomedical data analysis to understand complex medical data, which ultimately improves patient care. Major challenges associated with biomedical data analysis include high dimensionality, class imbalance, and small database size. Data acquisition is a major issue in the field of biomedical analysis due to problems in collaboration between medical experts and researchers. The acquired total number of samples of biomedical data is small, and there is even a smaller number of abnormal or disease samples, which leads to a high class imbalance. An efficient algorithm that deals with class imbalance is required to get high-performance models with a small database. There is a need to explore novel feature selection methods to improve predictive performance along with interpretation and to explore large-scale data in the biomedical sciences. Development of novel pattern recognition and machine learning methods specific to medical and genomic data is a challenge but extremely necessary. Healthcare and biomedical sciences have become data-intensive fields, with a strong need for sophisticated data-mining methods to extract the knowledge from the available information. Big data analytics has played an increasingly significant part in computational biomedical applications, specifically in examining large and varied datasets and uncovering relevant hidden patterns. Traditional biomedical data analysis methods have produced promising results but are surpassed by modern deep learning and machine learning approaches. Biologically inspired evolutionary computing algorithms have shown potential for better-performing systems in biomedical and bioinformatics applications. Good-quality high-dimensional biomedical data and images are required for efficient detection of various relevant abnormalities. The biomedical data are analyzed to create globally accessible systems leveraging data storage and computing power made available through cloud computing. Multi-modal analysis is currently being used for biomedical data for better-performing models. Fusion techniques are used to combine the information from different modalities, for example combining different image modalities or combining images with clinical information. These fusion methods are similar to the practices followed by biomedical experts in the real environment. Replication of these practices by researchers may lead to improvement of performance of developed biomedical models. Novel machine learning– and deep learning–based methods are needed to address the challenges to biomedical data. Special attention should be devoted to handle feature selection, class imbalance, and data fusion in biomedical and

machine learning applications. The aim of the special issue is to bring together medical experts who have interesting biomedical data and researchers from academia as well as industry to solve complex biomedical problems.

The first article of this special issue, authored by the guest editors, is a review on the use of machine learning for the diagnosis of Alzheimer's disease. The article "Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review," by M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, M. Prasad, and C. T. Lin, presents a detailed review of the challenges and developments in the diagnosis of Alzheimer's disease using various machine learning approaches. The article gives a wide view of the recent and ongoing research on automated diagnosis of Alzheimer's disease. Moreover, a review on the feature extraction techniques for Alzheimer's is also presented, with future directions for the researchers.

The subsequent articles in this special issue focus on the application of novel machine learning–based algorithms for biomedical applications. The machine learning algorithms include deep learning models, such as convolutional neural networks, and evolutionary algorithms, such as the Wolf–Bat algorithm.

The article "Exploring Disorder-aware Attention for Clinical Event Extraction," by Shweta Yadav, Pralay Ramteke, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya, aims to identify the attributes (severity, course, temporal expression, and document creation time) associated with the medical concepts from Electronic Medical Records. The article proposes an attention-based deep learning framework for medical disorder-level attribute classification. They have utilized the benchmark CLEF-2014 shared task dataset to train the model over four attribute mentions (Severity, Course, Temporal Expression, and DocTime), aiming to capture important aspects of the patient's health status over time. Experimental results illustrate that a disorder-aware attention framework has the capability to consider different parts of the sentence when different disorder mentions are taken into consideration, which results in state-of-the-art performance.

The article "Cell Nuclei Classification in Histopathological Images Using Hybrid O_LConvNet," by Suvidha Tripathi and Satish Kumar Singh, proposes a flexible and hybrid architecture, O_LConvNet, that integrates the interpretability of domain-specific handcrafted features and generalization of deep learning features to classify colon cancer cell nuclei images. The architecture allows integration of any recent deep architecture with application-specific handcrafted features.

The article "A Decision Support System with Intelligent Recommendation for Multi-Disciplinary Medical Treatment," by Nengjun Zhu, Jian Cao, Kunwei Shen, Xiaosong Chen, and Siji Zhu, proposes a framework of decision support systems for multi-disciplinary medical treatment. The framework includes an intelligent decision process as well as multiple customized techniques for treatment program recommendation. The former is a workflow diagram delicately combining human knowledge and machine knowledge. The latter bases traditional k-NN algorithms, however, with multiple practical considerations, such as disagreement in human decisions and customized feature engineering in the medical area via neural networks-based multi-task learning.

The article "Random Forest with Self-paced Bootstrap Learning in Lung Cancer Prognosis," authored by Qingyong Wang, Yun Zhou, Weiping Ding, Zhiguo Zhang, Khan Muhammad, and Zehong Cao, presents the improvement of lung cancer prognosis and classification based on gene expression data. They proposed ensemble learning with a random forest approach to improve the model classification performance by selecting multi-classifiers and investigated the sampling strategy by gradually embedding high- to low-quality samples via self-paced learning.

The article "Textual Entailment-based Figure Summarization for Biomedical Articles," by Naveen Saini, Sriparna Saha, Pushpak Bhattacharyya, and Himanshu Tuteja, presents an

unsupervised system for summarizing figures in biomedical articles. To improve the quality of the summary obtained, several qualitative measures are simultaneously optimized using a multi-objective evolutionary algorithm. A new way of measuring diversity among sentences in the summary is also proposed in this article. Extensive experiments on 94 figures related to biomedical articles are used to demonstrate the effectiveness of the proposed system.

The article "Pulmonary Nodule Detection Based on ISODATA-improved Faster RCNN and 3D-CNN with Focal Loss," by Chao Tong, Baoyu Liang, Mengze Zhang, Rongshan Chen, Arun Kumar Sangaiah, Zhigao Zheng, Tao Wan, Chenyang Yue, and Xinyi Yang, presents a a novel pulmonary nodule detection system. The system consists of two stages, pulmonary nodule candidate detection and false positive reduction. The authors introduce Iterative Self-Organizing Data Analysis Techniques Algorithm to Faster Region-based Convolutional Neural Network for candidate detection and propose a three-dimensional convolutional neural network trained with focal loss for false-positive reduction.

The article "Hybrid Wolf–Bat Algorithm for Optimisation of Connection Weights in Multi-layer Perceptron," by Jatin Arora, Utkarsh Agrawal, Rahul Singh, Deepak Gupta, Ashish Khanna, and Aditya Khamparia, proposes a hybrid Wolf–Bat algorithm, a novel optimization algorithm, for the purpose of finding the ideal set of weights for training a multilayer perceptron neural network. This novel approach is tested on 10 different datasets of the medical field as obtained from the UCI machine learning repository, and the results are compared with those of four recently developed nature-inspired algorithms: the Grey Wolf Optimization algorithm, Cuckoo Search, the Bat algorithm, and the Whale Optimization algorithm. The proposed method outperforms the other mentioned algorithms in terms of both speed of convergence and accuracy.

The article "Intelligent Classification and Analysis of Essential Genes Species using Quantitative Methods," by R. K. Rout, Sk. Sarif Hassan, S. Sindhwani, H. M. Pandey, and S. Umer, introduces an encoding method to decode the patterns of the DNA sequences of *Homo sapiens, Arabidopsis thaliana, Drosophila melanogaster*, and *Danio rerio* based on purine and pyrimidine base distributions. The essential genes are considered the necessary genes to measure the sustainability of different living organisms. These essential genes encode proteins that maintain central metabolism, DNA replications, translation of genes, and basic cellular structure and mediate transport process within and out of the cell. To discriminate essential genes from the other genes, the distribution of purine and pyrimidine in the essential genes has been thoroughly researched for *Homo sapiens, Arabidopsis Tthaliana, Drosophila melanogaster*, and *Danio rerio*. The Indigent classification method has been employed to classify the essential genes of these four species using proximity measures such as Shannon entropy, Fractal dimension, Hurst exponent, purine distribution, and pyrimidine distribution.

The article "Active Balancing Mechanism for Imbalanced Medical Data in Deep Learning-based Classification Models," by Hongyi Zhang, Haoke Zhang, Sandeep Pirbhulal, Wanqing Wu, and Victor Hugo C. de Albuquerque, develops a new under-sampling method, namely the active balancing mechanism (ABM), which is based on valuable information contained in biomedical data. The proposed ABM adopts the Gaussian Naïve Bayes method to estimate the object and entropy as a strategic query function to examine data under-sampling while retaining valuable data. The performance of the proposed method is evaluated by considering data from the Physikalisch Technische Bundesanstalt diagnostic electrocardiogram database, which includes 148 myocardial infarction subjects and 52 healthy subjects. Numerical results reveal that ABM takes less time compared to five conventional under-sampling methods of consumption, except RaUS.

M. Tanveer
Indian Institute of Technology Indore, India

P. Khanna
Indian Institute of Information Technology, Design & Manufacturing, Jabalpur, India

M. Prasad
University of Technology Sydney, Australia

C. T. Lin
University of Technology Sydney, Australia

Guest Editors