Pan Y, Huang DS, Wang JX *et al.* Special section on AI and big data analytics in biology and medicine. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 36(2): 231–233 Mar. 2021. DOI 10.1007/s11390-021-0002-3

Preface

It is our great honor to announce the publication of this special section on AI and big data analytics in biology and medicine in the Journal of Computing Science and Technology (JCST).

As more and more modern biological and medical data are produced, artificial intelligence (AI) and big data analytics are playing an increasingly important role in helping to draw meaningful and logical conclusions about biology and medicine. Understanding biological and medical data will help answer important life questions on Earth, find solutions to global health problems, and even help solve tough problems such as drug design and disease diagnosis. The information obtained from biology and medicine is not only very detailed, but also has unique properties such as low quality data, big data sizes, different complex formats, high dimensions, many duplications and much noise, and so on. They all require special skills or unique tools for analysis and interpretation. Thus, a lot of studies using AI and big data analytics on biological and medical data are becoming very popular and hot topics in the computer science research field.

To this end, we need to research and develop AI and big data analytics that can respond to these specific biological and medical data. Because of these unique properties of these data, traditional AI and big data analytics methods and tools are no longer appropriate for processing them. Therefore, specialized AI and big data analysis algorithms need to be developed to meet the conditions existing in these data and the needs of these two fields.

This special section includes 11 articles. In particular, the theme of big data analytics methods for biological and medical data includes three articles. The first article, “DeepHBSP: A Deep Learning Framework for Predicting Human Blood-Secretory Proteins Using Transfer Learning”, is authored by a team of researchers from Jilin University. This paper presents a novel deep learning model (DeepHBSP) combined with transfer learning by integrating a binary classification network and a ranking network to identify blood-secretory proteins from the amino acid sequence information alone, which is of great value to clinical application. The second article, “Effective Identification and Annotation of Fungal Genomes” is authored by a team of researchers from Harbin Institute of Technology. This paper introduces an automated bioinformatics pipeline called PFGI when facing the increase of fungal sequencing data, for the identification and annotation task based on the current approaches to the identification and annotation of fungal genomes and it has achieved high accuracy. The third article, “Synthetic Lethal Interactions Prediction Based on Multiple Similarity Measures Fusion”, is authored by a team of researchers from Tianjin University. This paper proposes a new approach based on similarity fusion to predict SL pairs due to urgent need to develop efficient computational methods of SL pairs identification for the cancer targeted therapy. This method shows good effectiveness, ease of use and expansibility in multiple experiments.

The theme of the Genome-wide association study (GWAS) consists of four articles. The first article, “Logistic Weighted Profile-Based Bi-Random Walk for Exploring MiRNA-Disease Associations”, is authored by a team of researchers from Qufu Normal University. In this paper, a logistic weighted profile-based bi-random walk (LWBRW) method is designed to infer potential MDAs based on known MDAs. And then case studies show the outstanding ability of the LWBRW method to explore potential MDAs. The second article, “Predicting CircRNA-Disease Associations Based on Improved Weighted Biased Meta-Structure”, is authored by a team of researchers from Shaanxi Normal University. The paper proposes a method named CDWBMS, which integrates a small number of verified circRNA-disease associations with plenty of circRNA information to discover the novel circRNA-disease associations. Case studies show that CDWBMS can predict unknown circRNA-disease associations. In conclusion, CDWBMS is an effective method for exploring disease-related circRNAs. The third article, “GAEBic: Novel Biclustering Analysis Method for miRNA-Targeted Gene Data Based on Graph Autoencoder”, is authored by a team of researchers from Jilin University. This paper proposes a novel biclustering method for miRNA-targeted gene data based on a graph autoencoder named as GAEBic. GAEBic applies graph autoencoder to capture the similarity of sample sets or variable sets and takes a new irregular clustering strategy to mine biclusters with excellent generalization. This biclustering method achieves comparable performance on the high throughput miRNA data of soybean and it can also be used for other species. The fourth article, "Collaborative Matrix Factorization with Soft Regularization for Drug-Target Interaction Prediction", is authored by a team of researchers from Central South University. This paper proposes a novel matrix factorization method, so-called collaborative matrix factorization with soft regularization (SRCMF) in order to rationally represent the correlation. SRCMF improves the prediction performance by combining the drug and the target similarity information with matrix factorization. SRCMF is robust leading to performance improvement in drug-target interaction prediction.

The theme of deep learning with Ultrasound, MRI, CT and X-Ray images consists of two articles. The first article, "Seg-CapNet: A Capsule-Based Neural Network for the Segmentation of Left Ventricle from Cardiac Magnetic Resonance Imaging", is authored by a team of researchers from Zhengzhou University. This paper proposes a capsule-based neural network, named Seg-CapNet, to model multiple regions simultaneously within a single training process. The Seg-CapNet model consists of the encoder and the decoder. The output vectors of Seg-CapNet contain low-level image features and semantic features, which is beneficial for improving segmentation accuracy. Experimental results show that the mean Dice coefficient of Seg-CapNet is increased by 4.7% and the average Hausdorff distance is reduced by 22%. The proposed model also reduces the model parameters and improves the training speed while obtaining the accurate segmentation of multiple regions. The second article, "Robust Needle Localization and Enhancement Algorithm for Ultrasound by Deep Learning and Beam Steering Methods", is authored by a team of researchers from Sichuan University. Based on current challenge of difficulty in clinical needle localization, this paper shows a novel robust needle localization and enhancement algorithm which is based on deep learning and beam steering methods. This algorithm can not only greatly reduce the processing time, but also significantly increase the needle localization accuracy and enhance the needle visualization for real-time clinical intervention applications.

The theme of applications using AI and big data analytics for biological and medical data includes two articles. The first article, "CytoBrain: Cervical Cancer Screening System Based on Deep Learning Technology", is authored by a team of researchers from Wuhan University. This paper introduces an artificial intelligence (AI) system, named CytoBrain, to automatically screen abnormal cervical cells to help facilitate the subsequent clinical diagnosis of the subjects. Compared with three sophisticated deep learning models, CompactVGG consistently achieves the best classification performance. The results illustrate that the system based on the key part CompactVGG is efficient and effective and can support large-scale cervical cancer screening. The second article, "An Efficient WRF Framework for Discovering Risk Genes and Abnormal Brain Regions in Parkinson's Disease Based on Imaging Genetics Data", is authored by a team of researchers from Hunan Normal University. This paper proposes a Weighted Random Forest (WRF) model as the feature screening classifier for the study of complex brain diseases such as Parkinson's disease (PD). The authors' team implements sample classification and optimal feature selection based on WRF, and constructs a multimodal analysis framework for exploring the pathogenic factors of PD.

Guest Editors

Yi Pan, Professor, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen & Georgia State University, Atlanta yipan@gsu.edu

De-Shuang Huang, Professor, Institute of Machine Learning and Systems Biology, Tongji University, Shanghai dshuang@tongji.edu.cn

Jian-Xin Wang, Professor, School of Computer Science and Engineering, Central South University, Changsha jianxinwang@csu.edu.cn

Fa Zhang, Professor, Institute of Computing Technology, Chinese Academy of Sciences, Beijing zhangfa@ict.ac.cn

Yi Pan is the dean and chair professor of computer science and control engineering at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, and Regents' Professor Emeritus at Georgia State University, Atlanta. He is a fellow of AIMBE, RSPH, IET and JSPS and a member of EUAS. Dr. Pan received his B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, Beijing, in 1982 and 1984, respectively, and his Ph.D. degree in computer science from the University of Pittsburgh, Pittsburgh, in 1991. His research interests include bioinformatics and health informatics using big data analytics, cloud computing, and machine learning technologies. Dr. Pan has published more than 450 papers including over 250 journal papers with more than 100 papers published in IEEE/ACM Transactions/Journals. In addition, he has edited/authored 43 books.

De-Shuang Huang is a professor in Department of Computer Science and the director of Institute of Machine Learning and Systems Biology at Tongji University, Shanghai. He is currently a fellow of the International Association of Pattern Recognition (IAPR Fellow), fellow of the IEEE (IEEE Fellow) and senior member of the INNS, Bioinformatics and Bioengineering Technical Committee Member of IEEE CIS, Neural Networks Technical Committee Member of IEEE CIS, the member of the INNS, co-chair of the Big Data Analytics section within INNS, and associated editors of IEEE/ACM Transactions on Computational Biology & Bioinformatics, and Neural Networks, etc. He founded the International Conference on Intelligent Computing (ICIC) in 2005. ICIC has since been successfully held annually with him serving as General or Steering Committee Chair. He also served as the 2015 International Joint Conference on Neural Networks (IJCNN 2015) General Chair and the 11th IEEE Computational Intelligence in Bioinformatics and Computational Biology Conference (IEEE-CIBCBC) Program Committee Chair. His main research interest includes neural networks, pattern recognition and bioinformatics.

Jian-Xin Wang received his B.S. and M.S. degrees in computer science and application from Central South University of Technology, Changsha, and his Ph.D. degree in computer science and technology from Central South University, Changsha. Currently, he is the dean and a professor in the School of Computer Science and Engineering, Central South University, Changsha. He is also a leader in Hunan Provincial Key Laboratory on Bioinformatics, Central South University, Changsha. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics and computer network. He has published more than 200 papers in various international journals and refereed conferences. He is a senior member of IEEE. He has been on numerous program committees and NSFC review panels, and served as editors for several journals such as IEEE/ACM Trans. Computational Biology and Bioinformatics (TCBB), Journal of Computer Science and Technology.

Fa Zhang is a professor in Institute of Computing Technology, Chinese Academy of Sciences, Beijing. He received his Ph.D. degree in Institute of Computing Technology, Chinese Academy of Sciences, Beijing. His research interests include bioinformatics, biomedical data integration analysis, bio-medical image processing, and high-performance computing. Currently his group concerns computational methods on biological and medical data analysis, in particular 3D reconstruction of Cryo-electron microscopy and pathological image analysis. His group developed the first tomography reconstruction software based on GPU platform in China, named AuTOM. Also, they developed several softwares and tools in CryoEM and medical image processing. All of softwares have been used and helped biologists to obtain some significant biological discoveries. He has published over 100 papers in international journals such as Cell Research, Science Advances, Bioinformatics and conference proceedings such as ISMB.