

Guest Editorial for Special Section on the 13th International Conference on Intelligent Computing (ICIC)

De-Shuang Huang[✉], Vitoantonio Bevilacqua, and M. Michael Gromiha[✉]

THIS special section includes a selection of seven papers presented at the Thirteenth International Conference on Intelligent Computing (ICIC) held in Liverpool, UK, on August 7-10, 2017. ICIC was formed to provide an annual forum dedicated to the emerging and challenging topics in artificial intelligence, machine learning, bioinformatics, and computational biology, etc. It aims to bring together researchers and practitioners from both academia and industry to share ideas, problems, and solutions related to the multifaceted aspects of intelligent computing.

This year, the conference received 612 submissions from 21 countries and regions. All papers went through a rigorous peer review procedure and each paper received at least three review reports. Based on the review reports, the Program Committee finally selected 212 research papers for presentation at ICIC 2017. The authors of seven high quality papers were invited to submit an extended version to this special section. Following a rigorous review process, these seven papers were selected for publication.

The first paper in this section, “A Comprehensive Study on Predicting Functional Role of Metagenomes Using Machine Learning Methods” by Jyotsna Talreja Wassen, Haiying Wang, Fiona Browne, and Huiru Zheng, which proposed the application of embedded feature selection methods, namely, Extreme Gradient Boosting and Penalized Logistic Regression to determine important microbial species. The resultant feature set enhanced the performance of one of the most popular state-of-the-art methods, Random Forest (RF) over metagenomic studies. Experimental results indicate that the proposed method achieved best results in terms of accuracy, area under the Receiver Operating Characteristic curve (ROC-AUC), and major improvement in processing time. It outperforms other feature selection methods of filters or wrappers over RF and classifiers such as Support Vector Machine (SVM), Extreme Learning Machine (ELM), and k-Nearest Neighbors (k-NN).

- D.-S. Huang is with the Machine Learning and Systems Biology Laboratory, Tongji University, Zhabei, Shanghai Shi 200072, China. E-mail: shuang@tongji.edu.cn.
- V. Bevilacqua is with the Electrical and Electronics Department, Polytechnic of Bari Via Orabona, Bari 70126, Italy. E-mail: vitoantonio.bevilacqua@gmail.com.
- M. M. Gromiha is with the Department of Biotechnology, Indian Institute of Technology (IIT), Chennai, Tamil Nadu 600036, India. E-mail: gromiha@iitm.ac.in.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2019.2902324

The next paper, “A Novel Scaffolding Algorithm Based on Contig Error Correction and Path Extension” by Min Li, Li Tang, Zhongxiang Liao, Junwei Luo, Fang-Xiang Wu, Yi Pan, and Jianxin Wang, which presented a novel scaffolding algorithm which is on the basis of path extension Loose-Strict-Loose strategy and contig error correction, called iLSLS. iLSLS can help reduce the false relationships between contigs, and improve the accuracy of subsequent steps. iLSLS utilizes a scoring function, which estimates the correctness of candidate paths by the distribution of paired reads, and tries to conduct the extension with the path which is scored the highest. What’s more, iLSLS can precisely estimate the gap size. The experiments were conducted on two real datasets. And, the results show that the LSLS strategy is efficiently increasing the correctness of scaffolds, and iLSLS performs better than other scaffolding methods.

The paper “Efficiently Predicting Hot Spots in PPIs by Combining Random Forest and Synthetic Minority Oversampling Technique” by Xiaolong Zhang, Xiaoli Lin, Jiafu Zhao, Qianqian Huang, and Xin Xu presents a classification method combined with random forest classification and oversampling strategy to improve the training performance. A strategy with an oversampling ability is used to generate hot spots data so as to balance the given training set. Random forest classification is then invoked to generate a set of forest trees for this oversampled training set. The final prediction performance can be computed recursively after the oversampling and training process. This proposed method is capable to randomly select features and construct a robust random forest to avoid overfitting the training set. Experimental results from three data sets indicate that the performance of hot spots prediction has been significantly improved compared with existing classification methods.

The paper “Integration of Multi-Omics Data for Gene Regulatory Network Inference and Application to Breast Cancer” by Lin Yuan, Le-Hang Guo, Chang-An Yuan, Youhua Zhang, Kyungsook Han, Asoke K. Nandi, Barry Honig, and De-Shuang Huang proposed an approach, which uses biweight midcorrelation to measure the correlation between factors and makes use of nonconvex penalty based sparse regression for gene regulatory network inference (BMNPGRN). BMNPGRN incorporates multi-omics data (including DNA methylation and copy number variation) and their interactions in gene regulatory network model. The experimental results on synthetic datasets showed that BMNPGRN outperforms popular and state-of-the-art methods (including DCGRN, ARACNE, and CLR) under false positive control. Furthermore, we applied BMNPGRN on

breast cancer (BRCA) data from The Cancer Genome Atlas database and provided gene regulatory network.

The next paper, "Identifying Key Genes of Liver Cancer by Networking of Multiple Data Sets" by Su-Ping Deng and Wei-Li Guo. To identify key genes of the liver cancer with high accuracy, we integrated multiple microarray gene expression data sets to compute common differentially expressed genes, which will result in more accuracy than those from individual data set. To find the main functions or pathways that these genes are involved in, some enrichment analyses were performed including functional enrichment analysis, pathway enrichment analysis, and disease association study. Based on these genes, a protein-protein interaction network was constructed and analyzed to identify key genes of the liver cancer by combining the local and global influence of nodes in the network. The identified key genes, such as TOP2A, ESR1, and KMO, have been demonstrated to be key biomarkers of the liver cancer in many publications. All of the results suggest that the proposed method can effectively identify key genes of the liver cancer. Moreover, the proposed method can be applied to other types of data sets to select key genes of other complex diseases.

The paper "Classification of a DNA Microarray for Diagnosing Cancer Using a Complex Network Based Method" by Peng Wu, and Dong Wang exploited a Complex Network (CN) classifier to implement the classification task. An algorithm was used to initialize the structure, which allowed input variables to be selected over layered connections and different activation functions for different nodes. Then, a hybrid method integrated the Genetic Programming, and the Particle Swarm Optimization algorithms was used to identify an optimal structure with the parameters encoded in the classifier. The single CN classifier and an ensemble of CN classifiers were tested on four bench data sets. To ensure diversity of the ensemble classifiers, the authors constructed a base classifier using different feature sets, i.e., Pearson's correlation, Spearman's correlation, Euclidean distance, Cosine coefficient and the Fisher-ratio. The experimental results suggest that a single classifier can be used to obtain state-of-the-art results and the ensemble yielded better results.

This section ends with the paper "An Efficient Ensemble Learning Approach for Predicting Protein-Protein Interactions by Integrating Protein Primary Sequence and Evolutionary Information" by Zhu-Hong You, Wen-Zhun Huang, Shanwen Zhang, Yu-An Huang, Chang-Qing Yu, and Li-Ping Li proposed a highly efficient method to detect PPIs by integrating a new protein sequence substitution matrix feature representation and ensemble weighted sparse representation model classifier. The proposed method is demonstrated on *Saccharomyces cerevisiae* dataset and achieved 99.26 percent prediction accuracy with 98.53 percent sensitivity at precision of 100 percent, which is shown to have much higher predictive accuracy than the state-of-the-art methods. Extensive contrast experiments are performed with the benchmark data set from Human and *Helicobacter pylori* that our proposed method can achieve outstanding better success rates than other existing approaches in this problem. Experimental results illustrate that our proposed method presents an economical approach for computational building of PPI networks, which can be a helpful supplementary method for future proteomics researches.

We would like to express our sincere thanks to the ICIC 2017 Program Committee members for their invaluable

effort in making ICIC 2017 such a success. We would like to thank the external reviewers for volunteering their time to review the submissions to the conference and the special section. We would like to thank Aidong Zhang, the current editor-in-chief of the *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, for offering this opportunity for wider dissemination of the research presented at ICIC 2017 in *TCBB*. Last but not least, we would like to thank the authors of these seven articles for their time and effort in submitting their high-quality work to ICIC and *TCBB*.

De-Shuang Huang
Vitoantonio Bevilacqua
M. Michael Gromiha
Guest Editors



De-Shuang Huang received the BSc degree from the Institute of Electronic Engineering, Hefei, China, in 1986, the MSc degree from the National Defense University of Science and Technology, Changsha, China, 1989, and the PhD degree from Xidian University, Xian, China, in 1993, all in electronic engineering. Currently, he is a visiting professor at Liverpool John-Moore University, UK. In addition, he is the director of the Institute of Machines Learning and Systems Biology, Tongji University. He is currently the fellow of the International Association of Pattern Recognition, and a senior member of the IEEE and the International Neural Network Society (INNS). He was the 2015 International Joint Conference on Neural Networks (IJCNN 2015) General Chair, July 12-17, 2015, Killarney, Ireland. He has published more than 190 journal papers. His current research interest includes bioinformatics, pattern recognition, and image processing.



Vitoantonio Bevilacqua received the bachelor's and the PhD degrees in electronic engineering, from Polytechnic of Bari, in 1996 and 2000, respectively. He is currently a tenured assistant professor in computing systems in the Department of Electrical and Electronic Engineering of Polytechnic of Bari. Since 1996, he has been working and investigating in the field of computer vision and image processing, neural networks, evolutionary algorithms, and hybrid expert systems. The main applications of his research are in real world, in biometry, in medicine, and recently in bioinformatics and systems biology. He has published more than 70 papers in refereed journals, books, and international conferences proceedings and chaired several sessions such as speech recognition, biomedical informatics, intelligent image processing, and bioinformatics in international conferences. He won the Best Paper Award at the International Conference on Intelligent Computing held in Shanghai (ICIC 2008), he was program chair of ICIC 2009, publication chair of ICIC 2010, tutorial chair of ICIC 2011, and publication chair of ICIC 2012.



M Michael Gromiha received the PhD degree in physics from Bharathidasan University, Tiruchirappalli, India, in 1994. He pursued his postdoctoral research at the International Center for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy, and the Institute of Physical and Chemical Research (RIKEN), Tsukuba, Japan. Currently, he is working as a senior research scientist at the Computational Biology Research Center (CBRC) of the National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan. His main research interests are structural analysis, prediction, folding and stability of globular and membrane proteins, protein interactions and development of databases, and online tools in bioinformatics. He has published more than 120 research articles, 20 reviews, and a book.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.