Editorial

Data Mining Techniques in Medical Informatics

The advent of high-performance computing has benefited various disciplines in finding practical solutions to their problems, and our health care is no exception to this. Signal processing, image processing, and data mining tools have been developed for effective analysis of medical information, in order to help clinicians in making better diagnosis for treatment purposes.

Data mining has become a fundamental methodology for computing applications in medical informatics. Progress in data mining applications and its implications are manifested in the areas of information management in healthcare organizations, health informatics, epidemiology, patient care and monitoring systems, assistive technology, large-scale image analysis to information extraction and automatic identification of unknown classes. Various algorithms associated with data mining have significantly helped to understand medical data more clearly, by distinguishing pathological data from normal data, for supporting decision-making as well as visualization and identification of hidden complex relationships between diagnostic features of different patient groups. There are nine papers in this Special issue, covering different areas in medical informatics.

Paper 1 proposes a metabonomic study applied to medical diagnosis. Metabolomics and metabonomics belong to the "-omics" sciences. Particularly, metabonomic correlates the metabolic fingerprint to characteristics of specific patient categories. Usually, metabonomic studies are conducted by in-vitro spectroscopy. The aim of this study was to apply data-mining metabonomic techniques to the clinical diagnosis of genetic mutations in migraine sufferers. This is one of the first applications of advanced data-mining techniques to a mixed database consisting of hematochemical, instrumental, and genetic variables.

There has been an effort to use motion-related surface vibration, to detect independent finger motions is in practice. Accelerometers have been used in a finger tapping experiment to collect the finger motion related mechanical vibration patterns. The extracted time-domain and frequency-domain features were fed to back-propagation neural networks, to classify different finger motions. The insights provided in paper 2 will be helpful for prosthetic hand control.

Microscopic imaging is ubiquitous in several medical informatics disciplines, including but not limited to cancer informatics, neuro-informatics, and other emerging health informatics disciplines. The decision support applications frequently require the sensitive and specific detection of pathological changes in cells, which further require the accurate measurement of their geometric parameters. In paper 3, Du *et al*. have suggested that due to the complex nature of cell issues and problems inherent to microscopy, unsupervised mining approaches of clustering can be incorporated in the segmentation of cells. They have evaluated the performance of multiple unsupervised data- mining techniques in cell image segmentation. The authors adopted four distinctive, yet complementary, methods for unsupervised learning, including those based on k-means clustering, expectation maximisation (EM), Otsu's threshold, and Galois Message Authentication Code (GMAC). These methods are comparatively evaluated, both quantitatively and qualitatively using synthetic simulated and real images.

In image processing, medical decision support applications frequently require the ability to identify and locate sharp discontinuities in an image, for feature extraction and interpretation of image content. In the paper 4, the authors have proposed a new edge detection technique based on the regional recursive hierarchical decomposition, using quadtree and post-filtration of edges by means of a finite difference operator. The authors have shown that in medical images of common origin, focal and/or penumbral blurred edges can be characterized by an estimated intensity gradient. The authors have also rigorously evaluated the performance of their algorithm on retinal and CT-scan images, and demonstrated promising results. Their algorithm efficiently decreases false dismissals of predominantly significant edges, and significantly lowers the false alarms found in classical approaches.

Face Recognition technology has been gaining prominence with proliferation of images and video. The Medical field has seen a huge change in data collection, from text to images to video. Paper 5 reports the application of face recognition technology in the medical field, to classify the images of the esophagus into three grades of esophagitis (inflammation of the esophagus). Herein, Principal Component Analysis, Fisher Face method and Independent Component Analysis, are used for classification of the images.

The growing healthcare burden and suffering due to life threatening diseases such as cancer and the escalating cost of drug development can be significantly reduced by design and development of novel methods in translational bioinformatics and allied medical informatics disciplines. Functional genomics is rapidly becoming the cornerstone of transformative medical research, and the resulting ability to interpret gene expression data from large genotype databases and associating it with the phenotype data is becoming increasingly relevant. Paper 6 proposes an associative pattern-mining based approach for feature extraction, in terms of two weighted similarity measures for the clustering of similar gene expression profiles. The authors evaluate the usability and efficacy of their methods, by applying the proposed techniques on three publically available multiclass cancer gene expression datasets, and support their results with an online biomedical literature search. Such Data mining approaches for the analysis of microarray gene expression offer promise for precise, accurate, and functionally robust

analysis of genomics data in cancer classification. The efficiency and scalability of the presented technique also makes it well suited to the domains of medical image analysis for feature extraction and clustering of similar feature based rules.

Mining information from EMG signals to detect complex motion intention has attracted growing research attention, especially for upper-limb prosthetic hand applications. Paper 7 investigates the possibility to relate the around-shoulder muscle activity with the forearm motions. Experiments were conducted to record the EMG signals of different arm and hand motions. Data were analyzed to decide the contribution of each sensor, in order to distinguish the arm-hand motions as a function of the reaching time. Results showed that it is possible to differentiate hand grips and arm position while performing a reaching and grasping task.

Exercise heart rate has diagnostic implications, and the heart rate in onset and offset exercise conditions are compared in Paper 8. The raw heart rate is modelled using a first order system, so that this comparison can be made by using the gain and time constants of modelled heart rate recordings. This approach and results shown in this paper are of practical meaning to physiological monitoring and regulation during exercise and training.

Tissue microarray (TMA) technique is a high throughput technique, to provide a standardized set of images which are uniformly stained, facilitating effective automation of the evaluation of the specimen images. The TMA technique is widely used to evaluate hormone expression for diagnosis of breast cancer. If one considers the time taken for each of the steps in the tissue microarray process workflow, it can be observed that the maximum amount of time is taken by the analysis step. Paper 9 proposes a data mining approach, using colour analysis and neural networks for classification, in order to automate the analysis step and remove the bottleneck in the TMA workflow.

In this issue, we have made an earnest effort to provide various data mining methodologies applied to medical informatics, for the benefit of researchers, professionals and teachers.

U. Rajendra Acharya

(*Co-Guest Editor*)

Department of Electroincs and Computer Engineering

Ngee Ann Polytechnic

Singapore

E-mail: aru@np.edu.sg

Wenwei Yu

(*Co-Guest Editor*)

Department of Medical System Engineering

Graduate School of Engineering

Chiba University

Chiba

Japan

E-mail: yuwill@faculty.chiba-u.jp