INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 2057-5688

HEART DIESEASE PREDICTION

[1]BV. SREEDHAR REEDY, [2]M. THARUN REDDY, [3] HARISH SAYA, [4]DR.T. NALINI

UG Scholars, Department of CSE, DR.MGR Educational and Research Institution,India.

Professor, Department of CSE, DR.MGR Educational and Research Institution, India.

ABSTRACT

Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases. This paper presents a survey of various models based on such algorithms and techniques andanalyze their performance. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models are found very popular among the researchers .

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

INTRODUCTION

System using smart clothing for sustainable health monitoring. Qiu et al. [8] had thoroughly studied the het- erogeneous systems and achieved the best results for cost minimization on tree and simple path cases for heteroge- neous systems. Patients' statistical information, test results and disease history are recorded in the EHR, enabling us to identify potential data-centric solutions to reduce the costs of medical case studies.Wang et al. [9] proposed an efcient ow estimating According to a report by McKinsey [1], 50% of Americans have one or more chronic diseases, and 80% of American medical care fee is spent on chronic disease treatment. With the improvement of living standards, the incidence of chronic disease is increasing. The United States has spent an average of 2.7 trillion USD annually on chronic disease treatment. This amount comprises 18% of the entire annual GDP of the United States. The healthcare problem of chronic diseases is also very important in many other countries. In China, chronic diseases are the main cause of death, according to a Chinese report on nutrition and chronic diseases in 2015, 86.6% of deaths are caused by chronic diseases. Therefore, it is essential to perform risk assessments for chronic dis- eases. With the growth in medical data [2], collecting elec- tronic health records (EHR) is increasingly convenient [3]. Besides, [4] rst presented a bio-inspired high-performance heterogeneous vehicular telematics paradigm, such that the collection of mobile users' health-related real-time big data can be achieved with the deployment of advanced hetero- geneous vehicular networks. Chen et al.proposed a healthcare algorithm for the telehealth cloud system and designed a data coherence protocol for the PHR(Personal Health Record)-based distributed system. Bates et al. [10] proposed six applications of big data in the eld of health- care. Qiu et al. [11] proposed an optimal big data sharing algorithm to handle the complicate data set in telehealth with cloud techniques. One of the applications is to identify high-risk patients which can be utilized to reduce medical cost since high-risk patients often require expensive healthcare.

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

Moreover, in therst paper proposing health- care cyber-physical system [12], it innovatively brought for- ward the concept of prediction-based healthcare applications, including health risk assessment. Prediction using traditional disease risk models usually involves a machine learning algorithm (e.g., logistic regression and regression analysis, etc.), and especially a supervised learning algorithm by the use of training data with labels to train the model [13], [14]. In the test set, patients can be classified into groups of either high-risk or low-risk. These models are valuable in clinical situations and are widely studied [15], [16]. However, these schemes have the following characteristics and defects. The data set is typically small, for patients and diseases with specific conditions [17], the characteristics are selected through experience. However, these pre-selected characteristics maybe not satisfy the changes in the disease and its influencing factors.

With the development of big data analytics technology, more attention has been paid to disease prediction from the perspective of big data analysis, various researches have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk clas- sication [18], [19], rather than the previously selected characteristics. However, those existing work mostly considered structured data. For unstructured data, for example, using convolutional neural network (CNN) to extract text characteristics automatically has already attracted wide attention and also achieved very good results. However, to the best of our knowledge, none of previous work handle Chinese medical text data by CNN. Furthermore, there is a large difference between diseases in different regions, primarily because of the diverse climate and living habits in the region. Thus, risk classification based on big data analysis, the fol- lowing challenges remain: How should the missing data be addressed? How should the main chronic diseases in a certain

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

region and the main characteristics of the disease in the region be determined? How can big data analysis technology be used to analyze the disease and create a better model?

To solve these problems, we combine the structured and unstructured data in healthcare eld to assess the risk of disease. First, we used latent factor model to reconstruct the missing data from the medical records collected from a hospital in central China. Second, by using statistical knowledge, we could determine the major chronic diseases in the region. Third, to handle structured data, we consult with hospital experts to extract useful features. For unstruc- tured text data, we select the features automatically using CNN algorithm. Finally, we propose a novel CNN-based multimodal disease risk prediction (CNN-MDRP) algorithm for structured and unstructured data. The disease risk model is obtained by the combination of structured and unstructured features. Through the experiment, we draw a conclusion that the performance of CNN-MDPR is better than other existing methods.

LITERATURE SURVEY

Senthil kumar Mohan have suggested mine to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced ML techniques can help remedy this situation. This research concludes with various models that can be used for prediction. Anjan N. Repaka stated the performance of prediction for two classification models, which is analysed and compared to previous work. Experimental results show the improved accuracy percentage of risk prediction of our proposed method compared to other works. Aditi Gavhane addresses the issue of prediction of heart disease according to input attributes on the basis of various data mining techniques and represented them with their accuracy in tabular format. It proposes to develop an application which can predict the vulnerability of a heart disease given basic

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

symptoms like age, sex, pulse rate etc. The machine learning algorithm neural networks has proven to be the most accurate and reliable algorithm and hence used in the proposed system. Santhana Krishnan predicts the arising possibilities of Heart Disease. The outcomes of this system the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using four main Machine Learning Algorithm Namely Decision Tree, Logistic Regression ,Vector Machine and Naive Bayes Algorithm which shows the best algorithm among these two in terms of accuracy level of heart disease.

PROPOSED SYSTEM

Dimensionality Reduction involves selecting amathematical representation such that one can relate the majority of, but not all, the variance within the given data, thereby including only most significant information. The data considered for a task or a problem, may consists of a lot of attributesor dimensions, but not all of these attributes may equally influence the output. A large number of attributes, or features, may affect the computational complexity and may even lead to overfitting which leads to poor results. Thus, Dimensionality Reduction is a very important step considered while building any model. Dimensionality Reduction is generally achieved by two methods -Feature Extraction and Feature Selection.

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

METHODOLOGY

Decision Tree

There are dissimilar kinds of decision trees. The only difference is in scientific ideal that they use to first-rate the class of feature through rule mining. A gain ratio decision tree is very common and fruitful category. It is the association amongst information gain and classified information. In entropy system, the characteristic that reduces entropy and exploits information gain is nominated as tree root. For selecting tree root, it is first essential to estimate information gain of all attributes. Later, the attribute that exploits information gain will be nominated.

KNN

This is one of the simplest and fundamental methods of classification where the user does have a little knowledge or no understanding of the dissemination of the data. While carrying out Discriminant examination when some dependable parametric controls of probability densities are not known or found challenging to understand this classification method was developed to perform such calculations. The exact location of the K-nearest neighbor should be decided with the help of the training dataset. To find how much close each fellow of the training dataset is from the target how row that is to be examined, we make use of Euclidean distance. Discovery of the k-nearest neighbors and allocating the group to the row that is being inspected. Now repeat the technique for the rows outstanding in the target set. We can also select the maximum value of K in this software after that the software automatically builds a parallel model on the values of k upto the maximum specifies value. The first phase

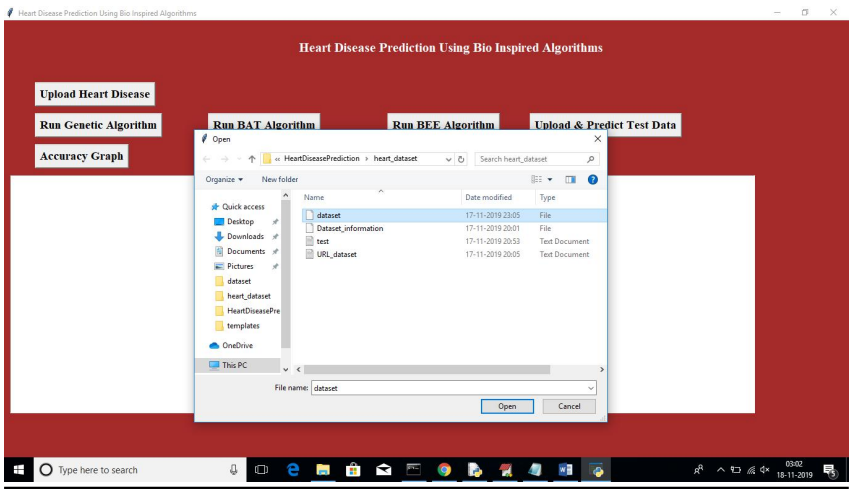
INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 2057-5688

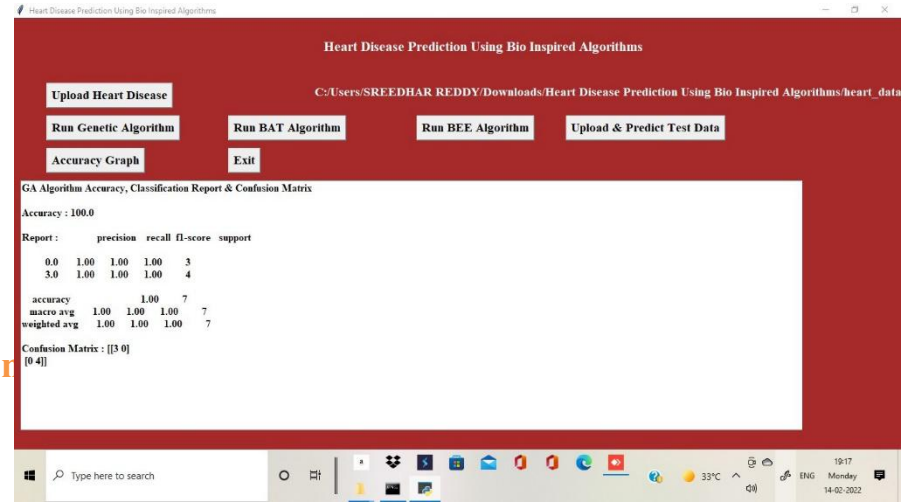
by means of K-nearest Neighbor classification technique with the support of WEKA tool is to decide the training dataset and then the input and output variables must derive in. Standardizing the data is the second step it guarantees that the distance degree allocates identical weight to each variable is the second phase in this course. The best score achieved of k between 1 and the given value is chosen that helps building parallel models on all values of k up to the extreme identified value for which k=9 was selected and scoring is done using the finest models from the available ones. Finally the data needed for classification is enter.

RESULT

To Run This Project 'Upload Heart Disease' button and upload heart disease dataset



After that we click on the run "Genetic Algorithm"

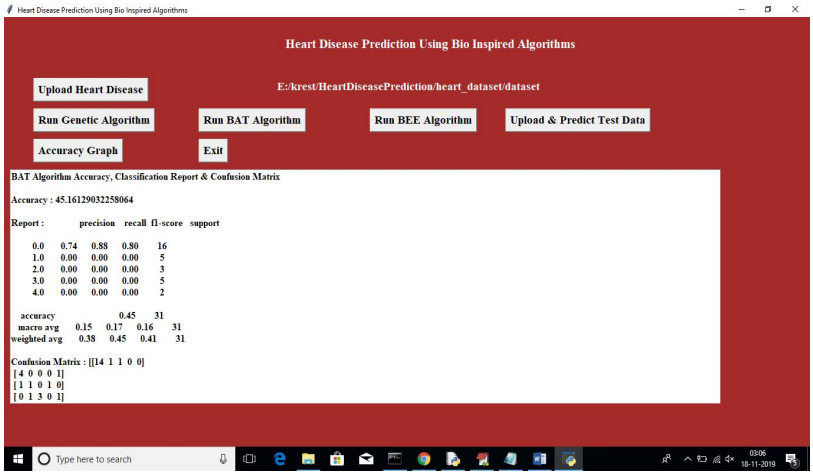


INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

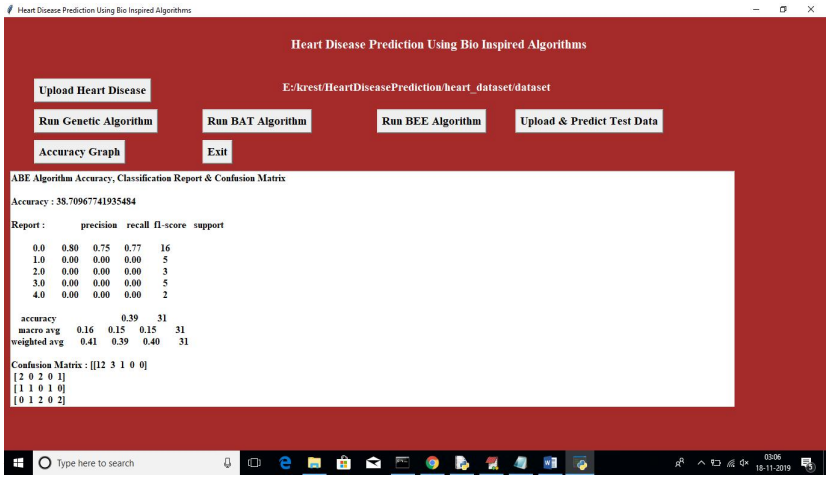
ISSN : 2057-5688

In above screen for GA accuracy, precision and recall we got 100% result

Now click on 'Run Bat' algorithm button to get its accuracy



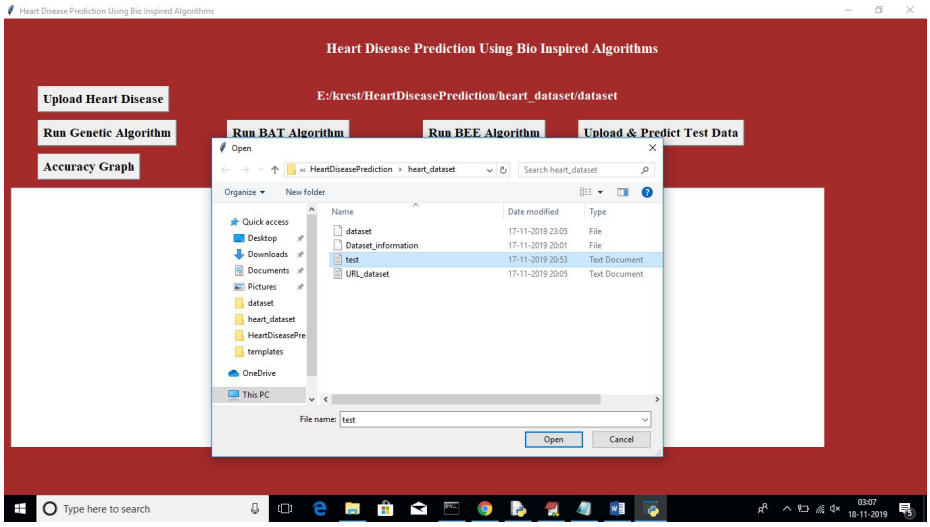
In above screen for BAT we got 45% accuracy, now click on 'Run BEE Algorithm' button to get BEE accuracy



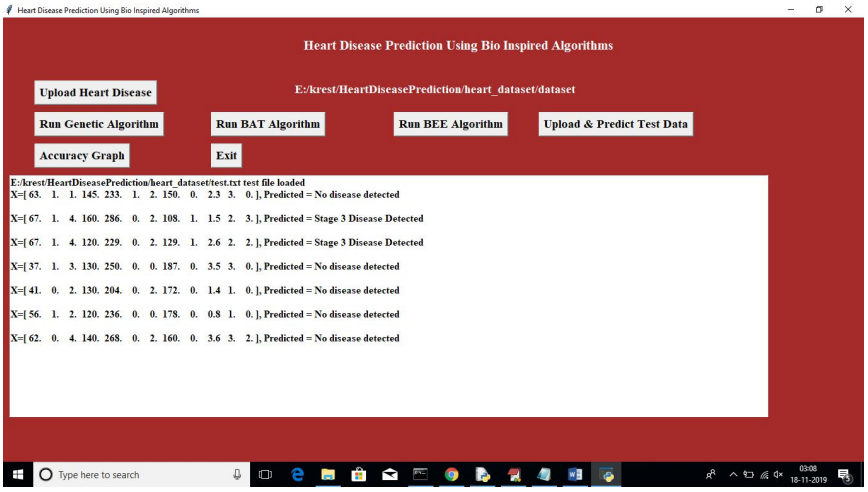
INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 2057-5688

In above screen for BEE we got 38% accuracy, now click on 'Upload & Predict Test Data' button to upload test data and to predict it class .



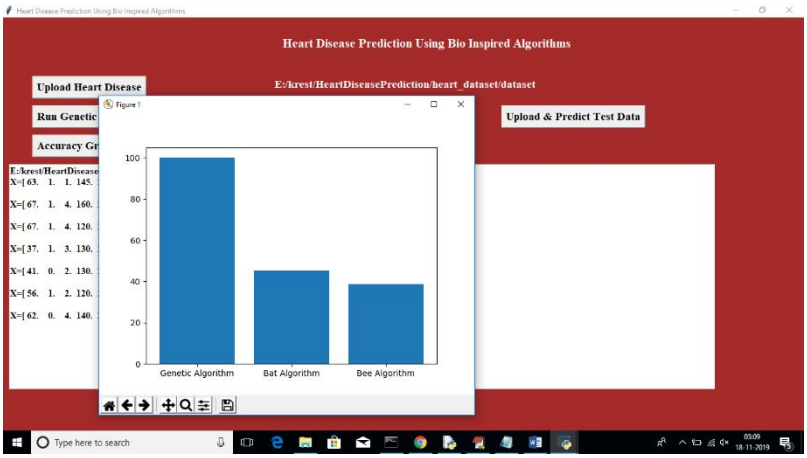
In above screen I am uploading test file which contains test data without class label, after uploading test data will get below screen .



INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 2057-5688

In above screen application has predicted disease stages. Now click on 'Accuracy Graph' button to view accuracy of all algorithms in graph format .



In above graph x-axis represents Algorithm Name and y-axis represents accuracy of those algorithms

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

CONCLUSION

In this paper, we propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

REFRENCES

[1]P.Groves,B. Kayyali,D. Knott,and S.van Kuiken, The 'Big Data'Revolution in Healthcare:Accelerating Value and Innovation. USA:Center for US Health System Reform Business Technology Ofce,2016.

[2] M. Chen, S. Mao, and Y. Liu, ``Big data: A survey," Mobile Netw. Appl., vol. 19, no. 2, pp. 171209, Apr. 2014.

[3] P. B. Jensen, L. J. Jensen, and S. Brunak, ``Mining electronic health records: Towards better research applications and clinical care," Nature Rev. Genet., vol. 13, no. 6, pp. 395405, 2012.

[4] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, ``A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics," IEEE Trans. Intell. Transp. Syst.,

ISSN : 2057-5688

vol. 16, no. 6, pp. 30333049, Dec. 2015.

[5] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, ``Wearable 2.0:

Enable human-cloud integration in next generation healthcare system,"

IEEE Commun., vol. 55, no. 1, pp. 5461, Jan. 2017.

[6] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, ``Smart clothing: Con-

necting human with clouds and big data for sustainable health monitor-

ing," ACM/Springer Mobile Netw. Appl., vol. 21, no. 5, pp. 825845,

2016.

[7] M. Chen, P. Zhou, and G. Fortino, ``Emotion communi-

cation system," IEEE Access, vol. 5, pp. 326337, 2017,

doi: 10.1109/ACCESS.2016.2641480.

[8] M. Qiu and E. H.-M. Sha, ``Cost minimization while satisfying hard/soft

timing constraints for heterogeneous embedded systems," ACM Trans.

Design Autom. Electron. Syst., vol. 14, no. 2, p. 25, 2009.

[9] Franck Le Duff, CristianMunteanu, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, page no. 1256-1259, 2004.

[10] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, KarstenSternickel,Lijuan Zhu, "Using Efficient Supanova Kernel For Heart Disease

Diagnosis", Proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol. 16,page no. 305-310, 2006.

[11] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer 2006,Vol:345, page no. 721- 727.

[12] Hongyu Lee, Ki Yong Noh, Keun Ho Ryu, "MiningBiosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, May 2007, page no. 56-66.

[13] Niti Guru, Anil Dahiya, NavinRajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1, January - June 2007.

[14] Hai Wang et. al.,"Medical Knowledge Acquisition through Data Mining", Proceedings of 2008 IEEEInternational Symposium on IT in Medicine and Education 978-1-4244- 2511-2/08©2008 Crown.

[15] SellappanPalaniappan, RafiahAwang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", (IJCSNS), Vol.8 No.8, August 2008.

[16] LathaParthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3,Page No. 3, 2008.

[17] Chaitrali S. Dangare, Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 888)Volume 47No.10, June 2012.

INTERNATIONAL JOURNAL OF TECHO-ENGINEERING
ISSN : 2057-5688

[18] S. Vijiyarani et. al., “An Efficient Classification Tree Technique for Heart Disease Prediction”,International Conference on Research Trends in Computer Technologies (ICRTCT - 2013) Proceedings published in International Journal of Computer Applications (IJCA) (0975 – 8887), 2013 (pp 6-9).

[19] Harsh Vazirani et. al.," Use of Modular Neural Network for Heart Disease", Special Issue of IJCCT Vol.1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5 August 2010 (pp 88-93).