A Comprehensive Survey on Heart Disease Forecast using Machine Learning Algorithms

Monia[1], Sonal[2]

sonalkharb@gmail.com[1], mchahal@gmail.com[2]
M. Tech Scholar[1] – BPSMV, Department of CSE & IT, Sonepat, Haryana, India
Assistant Professor[2]– BPSMV, Department of CSE & IT, Sonepat, Haryana, India

Abstract

As per the latest report published by WHO**,** heart disease is increasing exponentially around the world in modern life style due to numerous reasons. The fatality rate due to cardiac arrest is more than 18 million per year in the globe. Due to the enormous population and insufficient infrastructure in healthcare it is not possible to find out heart disease in the initial stage and treatment initialization. But in recent times due to emerging fields like AI, DL, Soft Computing are competent enough to investigate these types of health issues in the initial stage. Therefore, the main motive of this research is to find out the solution to predict the heart disease well in advance with higher accuracy. In recent times machine learning is playing a vital role in healthcare, therefore numerous research papers have been published by researchers to investigate the appropriate algorithms to predict heart disease. After going through various research papers, it was found that different algorithms predict different accuracy on the same or different dataset. In our research work various machine learning approaches will be applied and analysed for better accuracy on validated dataset.

Keywords: Machine Learning, Heart Disease, Decision Tree, SVM, Logistic Regression, NN

1.   Introduction

Cardiac Cardiovascular diseases are at the prime reason of death around the world as per WHO report. Due to cardiovascular diseases so many people die per year than from any other illness. In 2016, there are approximately 18 million persons deceased from cardiovascular diseases in 2016, which is almost 30 % of all world-wide demises. 84 % death occurred due to stroke and heart attack out of these 30 %. The high death rate and costly surgical procedures already make heart illness become a grave hazard for numerous families in countless parts of the world, particularly for developing countries. Consequently, it is very important for persons to investigate connection among numerous kinds of features in human and the likelihood of misery from heart illness [6-7]. Therefore, there must be a robust system which can help us to predict that which type of persons are more probable to heart illness so that in well advance people can take precautions to avoid lethal disease. Machine learning is basically based on statistics, which prime objective is to use various optimization techniques so that it can able to deliver desired result as per estimation in various domain to solve the problem like in industry, medical, social, business and many more real-world applications. The advent of modern

technologies in electronic health enables humans and machines to generate an enormous amount of data in our day-to-day life [10]. With every online transaction, the data generated is at the rate of 2.5 quintillion bytes per day and can be stored online for future analysis. In order to infer knowledge from the large and complex datasets, they build a model for data extraction and classification using machine learning and statistical approaches. Clinical decision making is dominant to all patient care happenings which includes choosing a deed, between replacements. Besides executing actions, giving diagnosis, comfort and dealing of disease creates the grave capabilities for medical profession [4]. Figure 1 represents of cardiac situation among progressive countries.
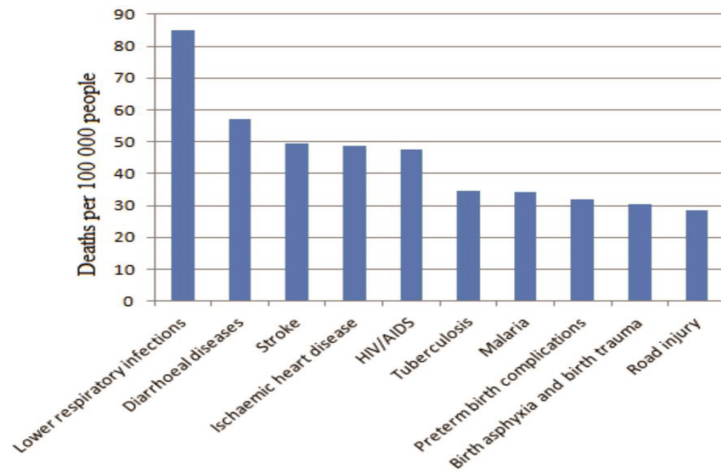

Fig.1: Different disease status in developing countries [12].

Cost is a very critical parameter which increases exponentially and affects the selecting medical strategies to apply on patients. In a short span of time doctors have to make their decision which would be the optimized treatment to cure the disease. The time available to make decisions is often brief, particularly with the shortening of the average patient visit. Figure 2 illustrates the status of developed countries regarding cardiac arrest.
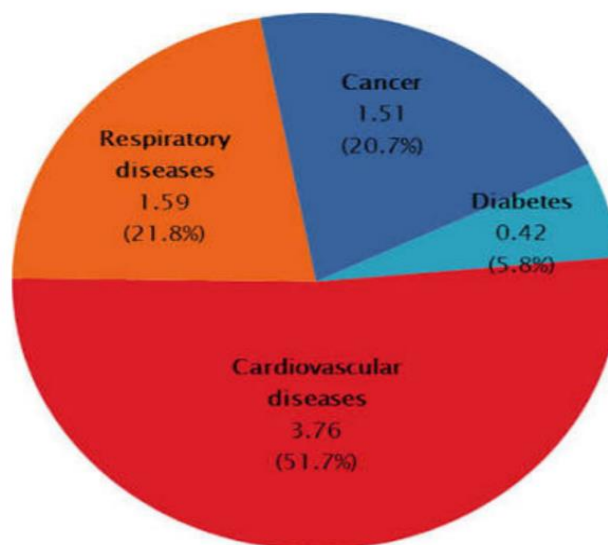

Fig.2: Death status caused by various illness in developed countries [2].

Types of Cardiac Disease: Heart disease is a prime concern in modern lifestyle. There are different stages of heart diseases. As per medical conditions figure 3 depicts different types of heart disease. Heart disease has various attributes which unregulated the proper function of the heart. These classifications are mainly segmented into following ways as per clinical evidence.

· Heart Failure

· Myocardial infarction

· Heart arrhythmia

· Atrial Fibrillation

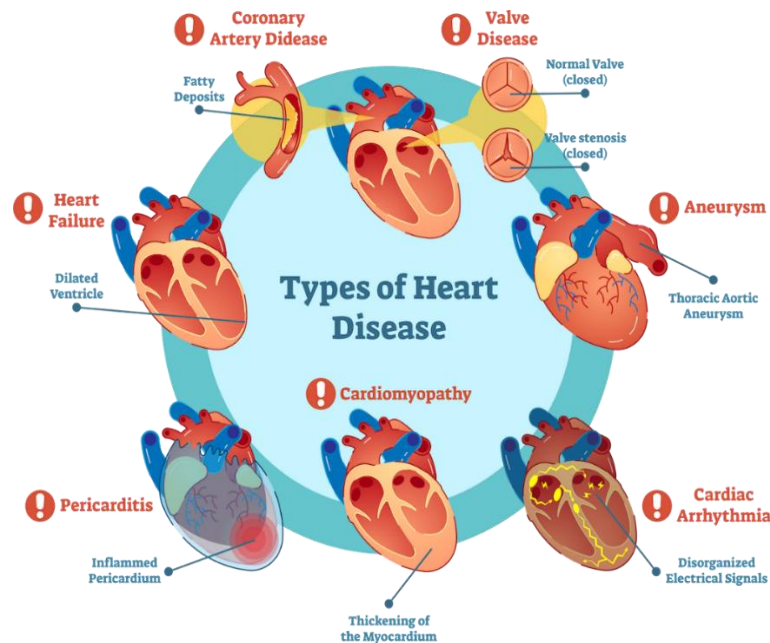· Angina pectoris

· Cardiomyopathy

Fig.3: Types of Cardiac Disease [University Diagnostic Medical Imaging].

Coronary Artery Disease: In our body, proper circulation of blood is a very important action. When blood circulation reduces then a discomfort is induced in coronary artery. Due to this phenomenon, veins are damaged by arteries and affect the normal heart function by generating restless changing regular systolic and diastolic parameters of the heart.

Acute myocardial infarction: In medical terms cardiac arrest occured due to acute myocardial infarction. The primary cause of heart/cardiac attack is the presence of unctuous material which affects the proper circulation of blood in the body and heart. This results in blockage of arteries due to tissue damage [5-6]. This leads to a severe condition in which blockage arteries unable to circulate oxygenated rich blood supply to the body parts which result in abnormal functions of various organs.

2. Literature Review

In the field of medical there are many tasks which are challenging for example forecast of pattern to avoid and control illness. To execute the proposed task, supervised machine learning algorithms implemented like LMT, SVM, Navie Bayes etc. First to develop a model, dataset is required and that dataset further classified into training and testing dataset. After that dataset is preprocessed so that prime features can be extracted and rest of the features which are not prime can be neglected so that accuracy can be improved. Out of implemented algorithm, random forest forecast best accuracy [2]
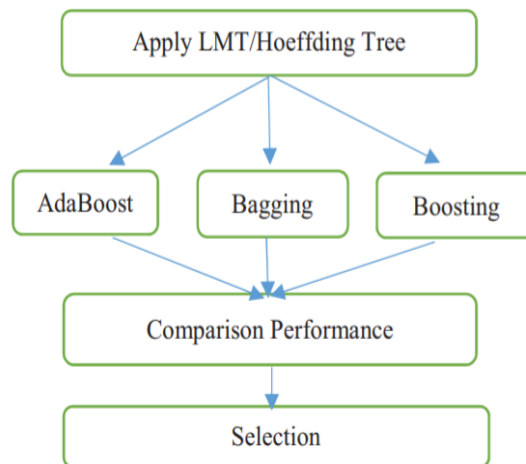

Fig.4: Steps for heart diseases prediction [2]

This research paper is primarily focused on predicting likelihood of patients for having a heart disease depending upon numerous medical traits. They have prepared heart disease prediction system for identifying which patient is more likely to be diagnosed with a heart ailment utilizing medical antiquity of patients. They have utilized diverse ML algorithms like logistic regression and KNN for prediction and classification of patients with heart diseases. A fairly Cooperative method was utilized for regulating the working of model for improving accurateness of forecast of Heart Attack in any patient. The asset of projected model is quite substantial and it was able to foresee evidences of heart diseases in particular individuals by utilizing KNN and Logistic Regression. The results showed upright exactness comparing to formerly utilized classifiers like naive bayes [1].

Therefore, the prime motive of this research is formation of a ML model for heart diseases predictions dependent upon the interrelated parameters. The author has utilized a standard dataset of UCI Heart disease prediction for research work having 14 dissimilar constraints related with heart diseases. ML algorithms like Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree are utilized for developing model. In research the author has also tried to discover relationships between diverse parameters available in datasets utilizing standard ML techniques and based on the parameters chance of cardiac diseases can be found in an efficient manner. From the results Random Forest technique is found to be more accurate and it also give prediction results rapidly. This is also supportive for the medical consultants at their clinics as decision supporting systems [3].

In the given research, ML technique has been proposed that utilizes many different characteristics for the prediction of cardiac diseases. BMI is the primary method used for estimation of heart diseases. It is an imperative procedure for prediction of cardiovascular diseases. The article mainly focuses on the effects of BMI on predictions of cardiovascular diseases. This model has planned many dissimilar features as well as relapse and organizing methods. The author has concluded that BMI is a substantial aspect when prediction of cardiovascular diseases is done [4].

In this research article, authors suggested a technique whose prime objective is to find important features from dataset and then by applying ML algorithms resulting into better accuracy in forecasting of cardiovascular illness. In the world, as per various reports mortality rate due to heart illness is one of the prime reasons. In medical healthcare system, to forecast cardiovascular illness is very tedious task in the domain of medical data examination.
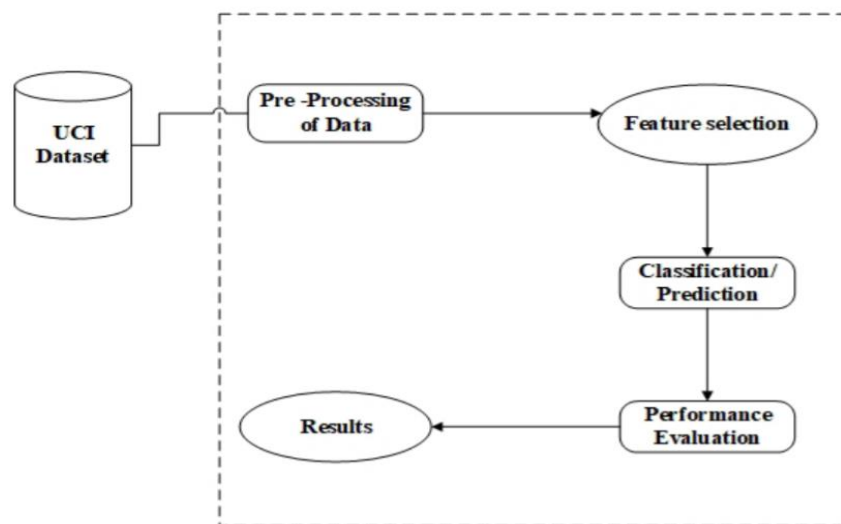

Fig.5: Experiment workflow with UCI dataset [5].

To handle such a large data manually, which is produced by healthcare industry each day is big task and an effective examination is not possible. But AI playing very prime role in different domain especially in healthcare system. With help of ML, a data scientist can do this tough task without any hassle. In proposed system using integrated technique, random forest algorithm gives better accuracy in forecasting heart attack [5].

3. Machine Learning

The advent of modern technologies in electronic health enables humans and machines to generate an enormous data in our day-to-day life. With every online transaction, the data generated is at the rate of 2.5 quintillion bytes per day and can be stored online for future analysis. In order to infer knowledge from the large and complex datasets, they build a model for data extraction and classification using machine learning and statistical approaches.
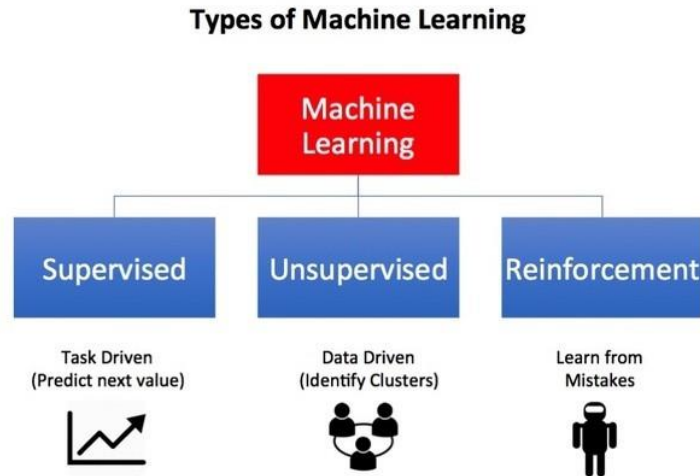
Types of Machine Learning

Fig.6: Machine Learning Classification.

Maintaining the medical data applications are incredibly varied, and at the higher level, one can examine a patient's sensitive information such as sugar level, BP, glucose and temperature for recognition and forestalling illness in the primary stage. In the low level, a single feature such as BP level can be examined to provide proper medications In all the aspects various methods and techniques are utilised to get the meaningful knowledge prevailing actual ecosphere unrefined data. In this paper, a combinational tactic of traditional statistical methods with model-based machine learning techniques will be used to efficiently classify the heart illness from datasets with reduced attributes and highly improved accuracy measures.

Supervised Techniques

These types of techniques build the training model using known evidence with little uncertainty. Figure 7 shows the classification of supervised learning. The forecasting models are developed using supervised learning with classification and regression techniques. Some of the conventional classification algorithms include SVM, naive bayes classification techniques, NN, LR (Logistic Regression), KNN, decision trees and discriminant analysis.
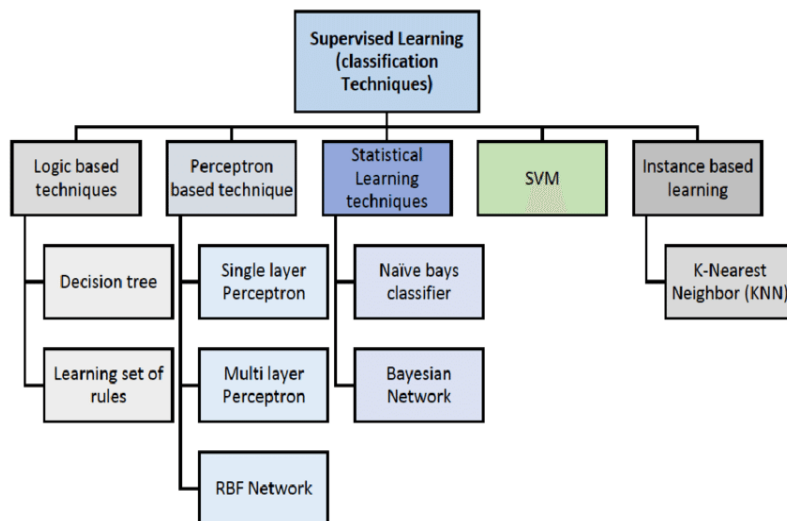

Fig.7: Supervised Learning Techniques.

Unsupervised Techniques

It is not necessary that there always be labelled data sets to predict prime information. For unlabelled dataset, unsupervised machine learning is implemented successfully to forecast the outcome. Clustering is the most common method of unsupervised learning technique. Figure 8 shows the classification of unsupervised learning techniques. hidden markov models, Fuzzy clustering, hierarchical clustering, K-means clustering algorithm and many more are the prime algorithms of clustering. Researchers developed enormous algorithms which fall under unsupervised and supervised categories but its application drastically varies from one scenario to another. Thus, selection of an appropriate machine learning algorithm provides better results in prediction and classification processes. However, choosing the right algorithm is often a difficult task.
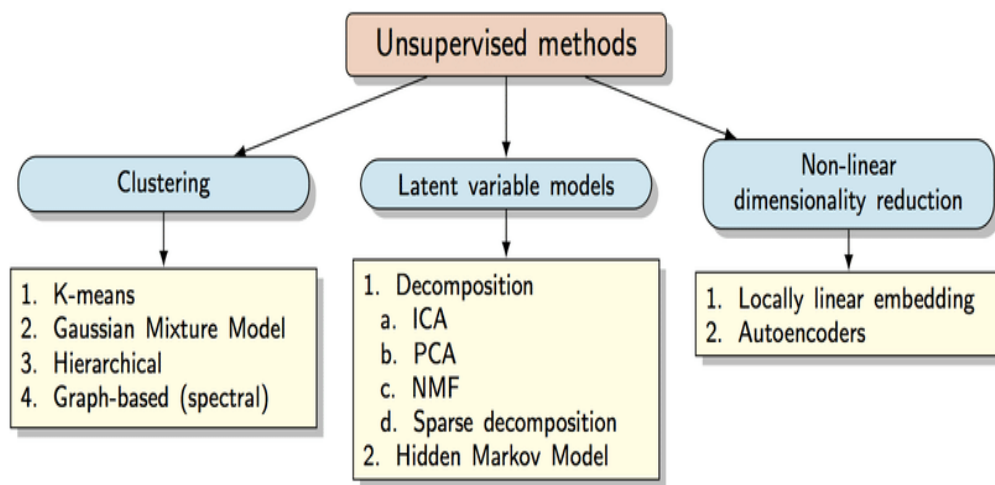

Fig.8: Un-supervised Learning Techniques.

4. Discussion

The primary task of this review paper is to analyse various machine learning algorithms which help to predict heart disease in effective manner. After going through various research paper, it was found that different algorithms predict different accuracy on same or different dataset. In modern lifestyle, healthcare issue is prime concern and to diagnose the disease in primary stage reduce the fatality rate. Emerging field like Soft Computing, ML and DL playing crucial role in healthcare to predict the disease in effective manner. For better accuracy in heart disease prediction more attributes in data set can be added along with integrated machine learning algorithms.

References

[1]. Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath, "Heart disease prediction using machine learning algorithms", ICCRDA 2020, IOP Conf. Series: Materials Science and Engineering, 1022 (2021) 012072 IOP Publishing DOI:10.1088/1757-899X/1022/1/012072

[2]. Pranav Motarwar, Ankita Duraphe, G Suganya, M Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning", 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), IEEE, 2020, DOI: 10.1109/ic-ETITE47903.2020.242

[3]. Vijeta Sharma; Shrinkhala Yadav; Manjari Gupta, "Heart Disease Prediction using Machine Learning Techniques", 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), IEEE, 18-19 Dec. 2020, DOI: 10.1109/ICACCCN51052.2020.9362842

[4]. Atharv Nikam; Sanket Bhandari; Aditya Mhaske; Shamla Mantri, "Cardiovascular Disease Prediction Using Machine Learning Models" 2020 IEEE Pune Section International Conference (PuneCon), IEEE, 16-18 Dec. 2020, DOI: 10.1109/PuneCon50868.2020.9362367

[5]. Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, Special Section on Smart Caching, Communications, Computing And Cybersecurity For Information-Centric Internet Of Things, DOI: 10.1109/ACCESS.2019.2923707

[6]. Dinesh Kumar G, Santhosh Kumar D, Arumugaraj K, Mareeswari V, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms", Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India

[7]. Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar, "Prediction of Heart Disease Using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), IEEE Conference, IEEE Xplore ISBN:978-1-5386-0965-1

[8]. Zhang, Shuai, Y.-L. S. A. (2017), 'Deep learning-based recommender system: a survey and new perspectives', Journal of ACM Computing Surveys 1(1), 1–35.

[9]. Amin Khatami, Abbas Khosravi, C. L. (2017), 'Medical image analysis using wavelet transform and deep belief networks', Journal of Expert Systems With Applications 3(4), 190–198.

[10]. Ashwini Shetty, Naik, C. (2016), 'Different data mining approaches for predicting heart disease', International journal of innovative research in science, engineering and technology 3(2), 277–281.

[11]. Aydin, S. (2016), 'Comparison and evaluation data mining techniques in the diagnosis of heart disease', Indian journal of science and technology 6(1), 420–423.

[12]. Bayasi, N. and Tekeste (2016), 'Low-power ECG-based processor for predicting ventricular arrhythmia', Journal of IEEE transactions on very large-scale integration systems 24(5), 1962–1974.

[13]. Berikol, B. and Yildiz (2016), 'Diagnosis of acute coronary syndrome with a support vector machine', Journal of Medical System 40(4), 11–18.

[14]. Zhiyong Wang, Xinfeng Liu, J. G. (2016), 'Identification of metabolic biomarkers in patients with type-2 diabetic coronary heart diseases based on metabolomic approach', 6(30), 435–439.

[15]. Singh, M. and Martins (2016), 'Building a cardiovascular disease predictive model using structural equation model and fuzzy cognitive map', Journal of fuzzy system 02(6), 1377–1382.

[16]. S.Prabhavathi (2016), 'Analysis and prediction of various heart diseases using DNFS techniques', International journal of innovations in scientific and engineering research 2(7), 678–684.

[17]. Sali, R. and Shavandi, M. (2016), 'A clinical decision support system based on support vector machine and binary particle swarm optimisation for Cardiovascular disease diagnosis', International Journal of Data mining and Bio-informatics 15(1), 312–327.

[18]. Ghadge, Prajakta, K. (2016), 'Intelligent heart attack prediction system using big data', International journal of recent research in mathematics computer science and information technology 2(2), 73–77.

[19]. G.Purusothaman and K.Krishnakumari (2015), 'A survey of data mining techniques on risk prediction: heart disease', Indian journal of science and technology 8(5), 643–651.

[20]. Richter, A., Listing, J., Schneider, M., Klopsch, T., Kapelle, A., Kaufmann, J., Zink, A. and Strangfeld, A. (2015), 'Impact of treatment with biologic dmards on the risk of sepsis or mortality after serious infection in patients with rheumatoid arthritis', Annals of the rheumatic diseases pp. 147–153.

[21]. Sairabi, Mujawar, D. (2015), 'Prediction of Heart Disease using Modified K-means and by using Naive Bayes', International Journal of Innovative Research in Computer and Communication Engineering 3.

[22]. Vafaie, Ataei, M. (2014), 'Heart diseases prediction based on ECG signals classification using a genetic-fuzzy system', Journal of biomedical signal processing and control 14(5), 291–296.

[23]. Wang (2014), 'Study on qi-deficiency syndrome identification modes of Coronary heart disease based on metabolomic biomarkers', Journal of evidence-based complementary and alternative medicine 24(16), 192–198.

[24]. Yang, X., Li, M., Zhang, Y. and Ning, J. (2014), 'Cost-sensitive naive bayes classification of uncertain data', Journal of Scientific World 9(8), 1897–1904.

[25]. Chandna, D. (2014), "Diagnosis of heart disease using data mining algorithm", International journal of computer science and information technologies 5(2), 1678–1680.