EasyChair Preprint

№ 8216

Novel Implementation of Cardio Vascular Disease Using Machine Learning Techniques

R Ramkumar and M.O Ramkumar

June 9, 2022

Novel Implementation of Cardio-vascular disease (CVD) using machine learning techniques

RAMKUMAR R
Computer Science Engineering
IFET College of Engineering
Gangrampalayam, Villupuram
ramkumar552001@gmail.com

RAMKUMAR M.O
Computer Science Engineering
IFET College of Engineering
Gangrampalayam, Villupuram
ramkumar.mo86@gmail.com

***Abstract*— The medical field is growing at a rapid pace with new diseases cropping up daily with the need for the invention of an appropriate course of treatment. The heart is a clenched human fist-sized muscular organ, which is responsible for the blood circulation. Though heart/cardiac disease is the name given to diseases affecting the heart in general, many diseases come under this name including coronary artery diseases (CAD), cardiomyopathy, Cardio Vascular Disease (CVD), and so on depending on the circulation of blood throughout the body. To support clinicians in the diagnosis of heart disease, heart disease data prediction has been so designed to analyze medical data with clinical expertise. Through improvement in these predicting systems, there can be an enhancement in the quality of medical diagnostic decisions for heart disease. Data mining plays a crucial part in the prediction of cardiac disease. In this work, the Naïve Bayes (NB) classifier, C4.5 classifier, and Artificial Neural Network (RNN)-Back Propagation (BP) methods are used. These traditional methods are utilized for predicting heart disease. When the dimensionality of the input is huge, the NB classifier method derived from the Bayesian theorem is used. Despite being simple, it performs better than other protocols. The C4.5 protocol builds decision trees from a training data set by utilizing data entropy perception. It is a well-known and used protocol and is also called the statistical classifier. For solving several decision modeling problems, RNN has been utilized as a tool in typical cases. The use of RNNs is evidenced in areas of modeling, pattern recognition, data processing, and sequence recognition systems.**

Keywords*— *EHR, Security, Sub Carrier, Data Leak Reduction

I. INTRODUCTION

The medical field is growing at a rapid pace since new diseases are cropping up frequently with the need for the invention of the appropriate course of treatment. An accurate and efficient modus operandi is required for proper diagnosis of the disease to provide appropriate treatment. If the system is automated then it can be very useful. But the issue is that in general, medical practitioners are not efficient enough in each sub-specialty making the availability of resources people sparse. So, if an efficient automatic medical diagnostic system is followed, it can be extremely beneficial to all strata involved in this process. The heart is a muscular organ that is responsible for blood circulation. The human heart works as a pump regulating the flow of blood in the human circulatory system. Through pumping of the heart, the deoxygenated blood is brought in from other parts of the body through veins and the oxygenated blood is pumped back from the lungs to various areas through arteries. Lungs help in the process of oxygenation. An electric impulse system available in the heart that controls the frequency of pumping is called the Sinus node which works as a natural pacemaker and is present at the top of the right atrium. The contraction and relaxation of the atrium and ventricles are actuated through the signals sent across 2 the heart muscle tissues described by Deepthi & Ravikumar (2014); Soni et al., (2011).

II. HEART DISEASE PREDICTION

The Heart Disease Data Prediction has been designed which analyses the medical data with a knowledge of clinical expertise. Through improvement in these predicting systems, there can be an enhancement in the excellence of medical diagnostic decisions for heart diseases. Proper medical records are maintained by most hospitals today which have an efficient hospital information system. The hospital information 5 system needs to be properly used or accessed so that the said system would serve to help physicians make the proper diagnosis and accurate clinical decisions reported by Patil (2014).

Delay in diagnosis can lead to a delay in commencement of treatment which might have drastic consequences in certain deadly diseases. In other instances, too many tests are conducted, which might have negative results leading to both losses of time and money. These are caused due to the lack of experience or expertise of the doctor. Various services are provided by healthcare organizations to enhance quality and efficiency, involve affected persons and their families, enhance care and coordination and public health, and maintain privacy and security regarding the health information of patients. The most common and deadly health issue to date is heart failure which affects specifically older patients due to lifestyle changes and the use of nonsteroidal anti-inflammatory drugs and ultimately leads to death.

The most commonly occurring heart disease is cardiovascular disease. This group of diseases can be predicted through suitable symptoms suggested by Venkata Lakshmi & Shivashankar (2014). For the prediction of heart diseases, various algorithms are used which include Decision Trees, Naïve Bayes, etc. Most hospitals use simple queries in decision support systems such as the average age in which a particular disease occurs, the gender in which it is more prevalent, and so on. Complex queries cannot be manipulated such as if a patient is afflicted by a particular disease or the type of treatment that will be more effective for patients with deadly diseases after crossing a particular stage and so on. 6 Currently, huge data are collected from patients may not be analyzed but just simply stored in the database. If knowledge hidden in these databases is utilized, many issues related to a particular disease condition can be solved and reported by Cherian & Bindu (2017).

There is a repetitive sequence of preprocessing data in the Knowledge Discovery process which includes cleaning, integration, correct data collection, data mining pattern identification, and knowledge presentation. Data Mining (DM) is extracting useful information from a huge dataset through classification, clustering, association, and so on, thus predicting or describing data. In the healthcare industry, data mining has extensive applicability in the classification of optimum treatment techniques, prediction

of risk factors of the disease, and finding out cost-conservative methods of inpatient care. Many studies have been conducted utilizing data mining models which apply to conditions such as diabetes, asthma, CVD, AIDS, etc., employed by Mythili et al. (2013).

There are three technical terminologies in data mining – data, information, and knowledge. Data can be facts, numbers, or texts which can be processed using a computer. A large quantity of data is accumulated by many organizations in various formats and databases. This information is then provided through patterns, associations, or relations. The information thus obtained is converted into knowledge about patterns.

III. Related Work

After going through many research papers, we selected several relevant papers for our literature review on healthcare information security issues. Basic concepts and problems, models, and architecture for the electronic health record have been reviewed which is discussed in [1][2][4]. We found the following directions for the review work.

Electronic health record (EHR) systems are in high demand to efficiently integrate all personally relevant medical information and present a lifetime document of medical histories. Internet-based EHR systems allow patients remote access to their entire medical history anytime. Hence an efficient protocol and architecture are required which is not standardized yet [3][11][12].

Anonymization is an important approach in healthcare information security, hence Hashing of medical data for privacy issues in healthcare is the central issue but an efficient and acceptable approach is not available yet. Several approaches have been proposed for assuring security issues related to healthcare security by this mechanism.

Access control mechanism and application related to e-prescription system and other consumer-related healthcare services requires a secure mechanism [5][6][14][15][17]. The above categories are the different dimensions of research in healthcare information security but are still interrelated. The concept of pseudonymization can be partially used in security and architecture solutions for ePrescription systems and EHR databases.

Bharti & Singh (2015) analyzed genetic algorithms, particle swarm optimization, artificial neural networks and found that they were useful in the prediction of heart diseases. The common approaches in these three algorithms had been defined and investigated as algorithms that had assisted in heart disease prediction.

Shaikh et al. (2015) have established an Intelligent System using a data mining method, Naive Bayes. It was executed as a Java application to answer the fixed questions. It would recover data from the mass record and would evaluate the user functions with trained data set. It could respond to difficult inquiries to diagnose heart diseases and support healthcare experts to make intellectual clinical decisions that could not sustain decision systems traditionally. Sonawane & Patil (2014) proposed a prediction scheme for heart diseases through a Learning Vector Quantization neural network algorithm. Here, the presented systems neural network had accepted 13 clinical features as input and predicted the existence or nonexistence of patient heart diseases together with various metrics.

Chen et al. (2011) improved a heart disease prediction scheme to aid medical experts to predict heart diseases position based on patient's 29 clinical data. The use of the Heart Disease Predict System (HDPS) was enhanced. The HDPS system would compose different attributes with input clinical data segment, ROC curve display segment, and prediction performance display segment. The projected concepts were efficient to predict patients' heart diseases. The HDPS system was a new concept to employ in heart disease classification.

Ordonez (2006) proposed an algorithm to employ search constraints to minimize rules, and association rules search on a training set and finally validated it on an autonomous experiment set. Search constraints and experiment set authentication considerably minimized the association rules and yielded rules that were set with greater precision. The author has shown significant rules with greater self-confidence, high lift, or both, to sustain effectively on the experiment set on multiple lopes. Such rules represented valued medical facts.

Raihan et al. (2016) enhanced a modest concept to predict enhancing ischemic heart disease (IHD) risk through the smartphone. Android model software had been improved by incorporating clinical data attained from patients stated with IHD. The projected research work was to create a simple concept to identify IHD risk and alert people to seek for themselves the assistance of a cardiologist to evade immediate death. Actively accessible devices had a few restrictions on making exploited by the population. The projected system product might minimize the restriction and support risk time assessment. An intellectual recommender scheme was established by Lafta et al (2015) that employed an inventive time series prediction algorithm to offer suggestions to heart disease patients using a telecommunication environment. The experimental consequences revealed this projected scheme had produced adequate recommendations with precision. The system also 30 provided a capable way to save the patient's capacity and help healthcare practitioners to carry out routine medical experiments. The projected research would minimize the capacity and health care cost and also assist in healthcare diligence transforming from the customary condition into a personalized scheme in a telehealth atmosphere.

Tavares et al. (2013) proposed a medical data investigation to determine whether children had cardiac problems. Here, unprocessed data was obtained at a Brazilian local hospital as preprocessed to construct the classification patterns. Non-invasive data was employed, like height, weight, gender, and birth date to make other derived variable sets like BMI (Body Mass Index) to upkeep the classification stage. The consequences revealed that this projected concept had performed well on other techniques in at least three out of four assessment measures.

Kantar et al. (2014) proposed a decision support algorithm to detect the normal sinus rhythm or other pathologies automatically. The enhanced algorithm would offer support to the doctor for educational applications. In the major code, thirteen functions were employed to diagnose eight multiple ECG pathology mechanically. Greater success was antedated for power prediction in the future in the established technique with the current research purpose.

Chauhan et al. (2015) showed a work structure practiced by the DM technique that assists the prediction and treatment of several diseases. The significant purpose of this projected work was to treat either their required Randomized Controlled Trials (RCT) or not with several attribute values. The projected system is composed of the patient's records union including clinical decision support that had low down medical faults, and undesirable process differences that assisted in creating effective patient security and results.

Wang et al. (2015) employed structured and unstructured data from Electronic Health Records (EHR) to predict Heart Failure (HF) onset by prediction accuracy diverse concerning time afore diagnosis. Models were evaluated by changing the prediction window beginning from 60 to 720 days before HF diagnosis. Since the prediction window reduced, the performance [AUC (95% Confidence Interval (CIs))] of the predictive HF models raised from 65% (63%-66%) to 74% (73%-75%) for the unstructured, from 73% (72%-75%) to 81% (80%-83%) for the structured, and from 76% (74%-77%) to 83% (77%-85%) for the combined data.

VI. Problem Definition

At present, the large amount of data collected from patients is simply stored in hospitals or health care centers and is not put forth for proper use. If the knowledge hidden in these databases is utilized, solutions can be found for many other problems that exist in the health care field regarding the services provided to the patients. The real-life data mining applications are interesting as they present data miners with various sets of problems, time and again. Dealing with heart disease patients' databases is one such real-life application. Classification is a significant data mining technique with broad applications and is also a challenging task with many applications. A neural network cannot be initialized with prior knowledge, and thus the network must learn from scratch. The learning process itself is time-consuming, and there is usually no assurance of success. Real-world optimization problems are challenging to solve, and many applications have to pact with Nondeterministic Polynomial (NP)-hard problems.

To solve such problems, optimization tools have to be used, though there is no guarantee that the optimal solution can be obtained. In this work, to optimize the weight of the NN structure, RNN techniques are proposed.

V. METHODOLOGY

Classifications tend to make life easier in the case of a supermarket when things are placed on a shelf randomly it can make it an unpleasant experience to shop. There are various techniques of classification 76 in data mining and

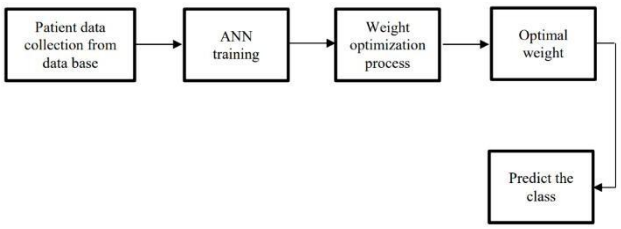

Figure 5.1 Architecture diagram of Novel implementation on heart disease prediction

RNN is one of them. The RNN architecture, the actual number of codes to be chosen, and how the weights have to be set between the nodes at the time of training and evaluation of results are all completely covered. The function of activation is mentioned together with the rate of learning, the momentum, and the pruning. The BP algorithm is a very popular RNN algorithm that was demonstrated by Cilimkovic (2015). The RNN can work on errors better than that of the traditional computer programs (like in a scenario of a faulty statement in the program which can halt everything when the RNN will handle errors better). Here in this work, the optimized RNN-BP along with the GSO and the PSO is proposed.

Cleveland Database Cleveland data set obtained from the internet was used as the database. This dataset classifies people as normal and abnormal in connection with cardiac disease. Only fourteen raw attributes were used though there were seventy-six of them in the databases. Four distinct classes were represented by the two-bit value of the output.

These classes are- 0 represents a normal individual, 1 represents the first stroke, 2 represents a second stroke and 3 represents death. Data Representations: Quantity of instances:

Quantity of features: 14 as well as a class feature.

Class 0: Normal Person,

Class 1: first stroke,

Class 2: second stroke,

Class 3: end of life.

Dataset for classification process: As per the statistics given by the World Health Organization, death is commonly caused by the cardiac disease throughout the world. The usage of RNN and data selection is the chief component for categorizing heart disease. Four distinct sets of UCI are used for obtaining data, as well as the Centre for Machine Learning and Intelligent Systems.

VI. PERFROMANCE EVALUATION

Alneamy, JSM Et al., [3] On the basis of a comparison of the present system, which uses the TLBO and FWNN to train and evaluate datasets gathered from diverse sources. It has the lowest accuracy compared to the system we offer. By rapidly training and testing on their models, our suggested approach is completely implemented in the GSO and RNN to increase the efficiency and accuracy of the patient record. It is the most straightforward way for efficiently training and testing data from multiple datasets. It is for this reason that the simplest technique is more accurate than other model ways. It is a simple webpage to engage with the patient and

the doctor can grasp and comprehend the requirements. It may analyze and process each data point in the dataset from the data source to guarantee that the patient's specific symptom is predicted using GSO and RNN.

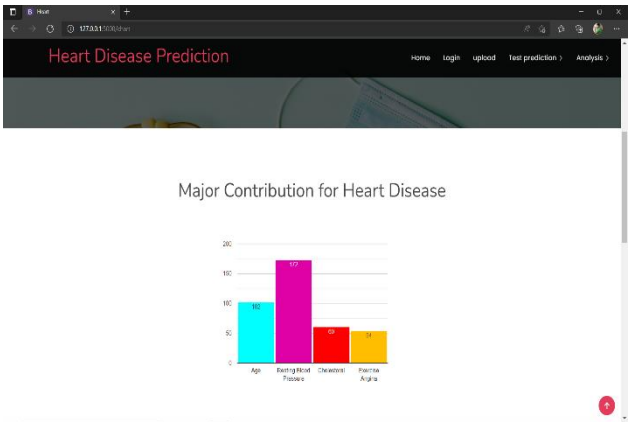

Figure 6.1 Major contribution for heart disease

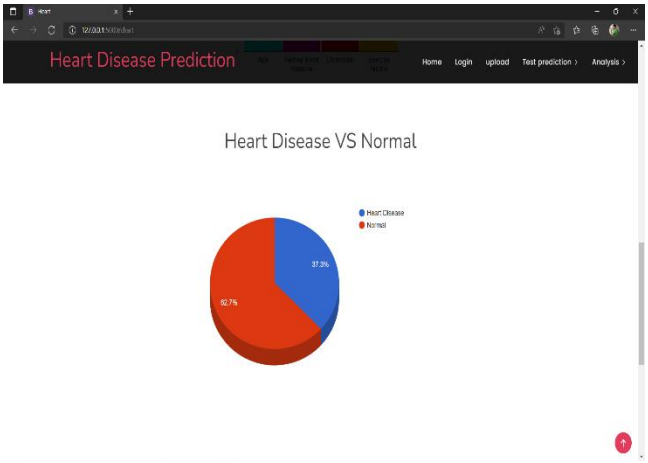

Figure 6.2 heart disease vs normal

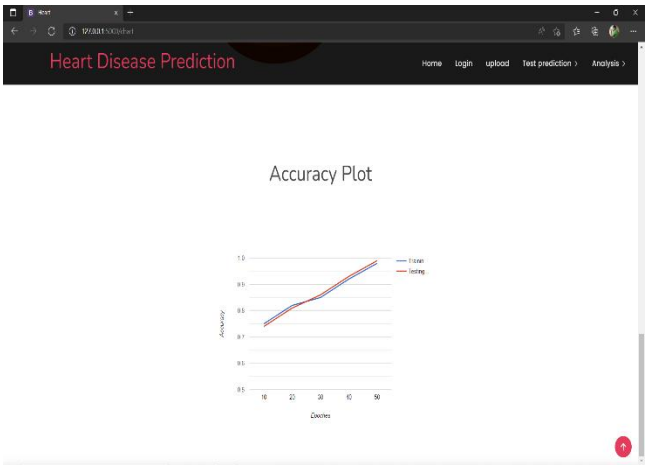

Figure 6.2 heart disease vs normal

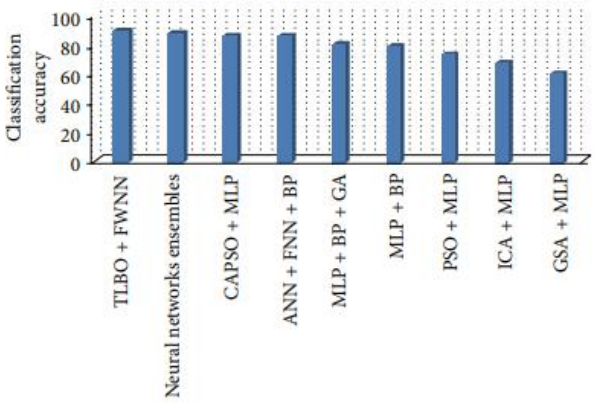

Figure 6.3 Performance evaluation of Existing system

VII. RESULT AND DISCUSSION

Based on the major contribution for heart disease of various fields such as age, blood pressure, cholesterol, over anxiety to be processed based on the dataset to be collected on the Kaggle using GSO and KNN to calculate and improve the efficiency and simplicity of the project to be enhanced in the real-world problems to be calculated and tested

Figure 7.1 heart disease prediction

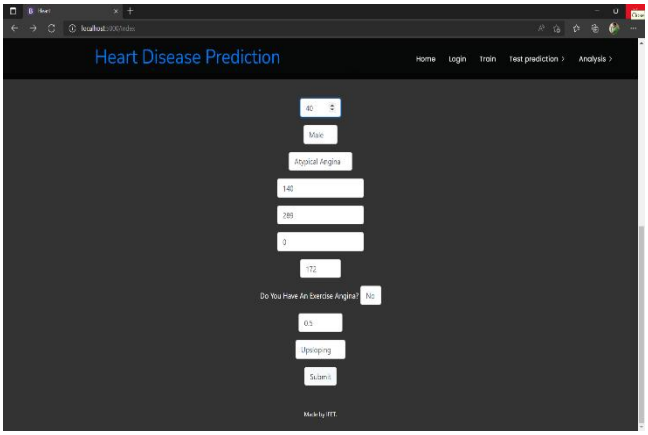

Figure 7.2 heart disease prediction with I/P

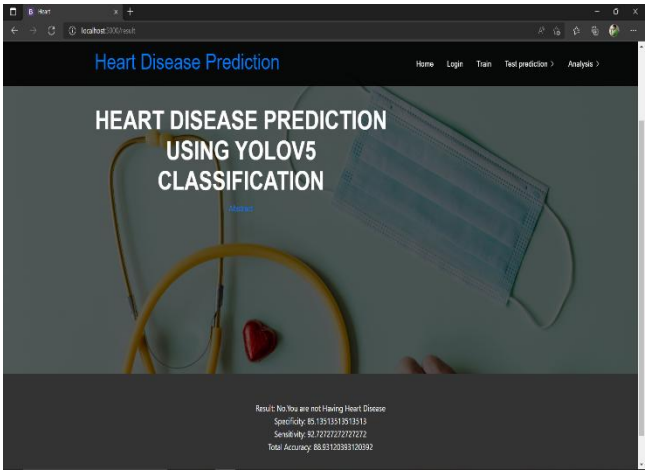

Figure 7.3 heart disease prediction with O/P

The outcome demonstrates that each unique data set is processed to provide specificity, sensitivity, and total accuracy based on data categorization utilising the GSO and RNN algorithms. It predicts heart illness based on analysis of training and testing data to create heart disease with improved accuracy and performance based on Figure 7.1,7.2,7.3,7.4,7.5

.

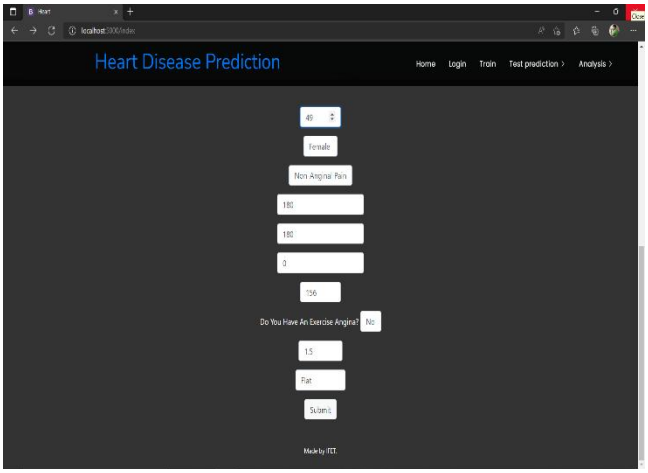

Figure 7.4 heart disease prediction with SERIOUS CASE I/P

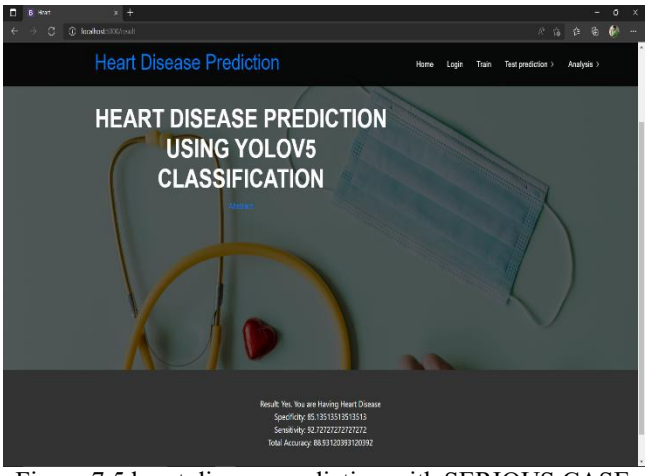

Figure 7.5 heart disease prediction with SERIOUS CASE O/P

VIII. CONCLUSION AND FUTURE WORK

Heart diseases are a severe threat and typically occur when arteries that provide oxygen and blood to the heart are entirely blocked or made narrow. There is a vast amount of data produced in medical organizations which are not appropriately utilized. The classification problem of designating various observations into various disjointed groups would play an important role in making business decisions among others. RNN is the mathematical simulation of biological neurons responsible for human brain functioning. RNN model is structured with inter- connected computational neurons utilized for executing mathematical mapping at the time of the learning procedure. Results show that the naive Bayes has higher classification accuracy by 8.76% for C4.5 and by 2.23% for RNN-BP. Predicting heart disease is intended to aid cardiologists in their diagnosis.

A method is proposed to classify the data on heart ailment. Patients' medical history predicts any previous symptom through data mining since the data is a major characteristic to select the method applied for dataset plummeting. When most unrelated and redundant structures are removed from the data, the choice of structure helps enhance the presentation of learning models when decreased data goes for classification. In this work, the use of the RNN-BP method is optimized with PSO and GSO algorithms. A population 108-based stochastic optimization system is a healthy and active PSO and is based on the swarms' movement besides intelligence. Training RNN can be facilitated using the GSO algorithm to obtain the real outcome for solving the three real-world standard issues. Ideally, the animal searching attitude motivates the GSO which is built on the creator-debtor concept in which, the group members search for opportunities of the creator or the debtor. This device assists in addressing uninterrupted augmentation issues. Results show that the GSO-RNN-BP has a higher classification accuracy of 11.11% for RNN-BP and 4.12% for PSO-RNN-BP.

REFERENCES

1. AbuKhousa, E & Campbell, P 2012, 'Predictive data mining to support clinical decisions: An overview of heart disease prediction systems', In Innovations in information technology (IIT), 2012 international conference on IEEE, pp. 267-272.

2. Abushariah, MA, Alqudah, AA, Adwan, OY & Yousef, RM 2014, 'Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (RNN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches', Journal of Software Engineering and Applications, vol. 7, no. 12, p. 1055.

3. Alneamy, JSM & Alnaish, RAH 2014, 'heart disease diagnosis utilizing hybrid fuzzy wavelet neural network and teaching-learning based optimization algorithm', Advances in Artificial Neural Systems.

4. Alqaraawi, A, Alwosheel, A & Alasaad, 2016, 'Towards efficient heart rate variability estimation in artifact-induced Photoplethysmography signals', In Electrical and Computer Engineering (CCECE), 2016 IEEE Canadian Conference on IEEE, pp. 1-6.

5. Al-Shayea, QK 2011, 'Artificial neural networks in medical diagnosis, International Journal of Computer Science Issues, vol. 8, no. 2, pp. 150-154.

6. Alshurafa, N, Eastwood, JA, Pourhomayoun, M, Liu, JJ & Sarrafzadeh, M 2014, 'Remote health monitoring: Predicting outcome success based on contextual features for cardiovascular disease', In Engineering in Medicine and Biology Society (EMBC), 2014 36th annual International Conference of the IEEE, pp. 1777-1781.

7. Alzahani, MS, Althopity, A, Alghamdi, A, Alshehri, B & Aljuaid, S 2014, 'An Overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction', vol. 2, no. 4, pp. 310-315.

8. Amato, F, López, A, Peña-Méndez, EM, Vaňhara, P, Hampl, A & Havel, J. 2013, 'Artificial neural networks in medical diagnosis', pp. 47-58. 111

9. Amin, SU, Agarwal, K & Beg, R 2013, 'Genetic neural network-based data mining in the prediction of heart disease using risk factors', In Information & Communication Technologies (ICT), 2013 IEEE Conference on IEEE, pp. 1227-1231.

10. Amma, NB 2012, 'Cardiovascular disease prediction system using genetic algorithm and neural network', In Computing, Communication and Applications (ICCCA), 2012 International Conference on IEEE, pp. 1-5.

11. Anbarasi, M & Saleem Durai, MA 2016, 'A survey on swarm optimization techniques for heart disease prediction', vol. 8, no. 2, pp. 4130-4141.

12. Atherton, JJ, Bauersachs, J, Carerj, S, Ceconi, C, Coca, A, Erol, Ç & Guazzi, M 2016, '2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure, European Heart Journal, vol. 37, pp. 2129-2200.

13. Banu, MN & Gomathy, B 2013, 'Disease Predicting System Using Data Mining Techniques,

International Journal of Technical Research and Applications, vol. 1, no. 5, pp. 41-45.