An Analysis of The Outcome Of Heart Disease Using DNN And Data Mining Framework

B Shali Nayaki [1], C Subha Darathy [2]
[1, 2] Dept of CSE
[1, 2] Arunachala College of Engineering for Women

Abstract- *Mortality rate increases all over the world on daily basis. The reasons for this could be increase in the numbers of patient with cardiovascular disease. When considering death rates and large number of people who suffers from heart disease, it is revealed how important early diagnosis of heart disease. Traditional way of predicting heart disease is doctor's examination or number of medical tests such as ECG, Stress Test, and Heart MRI etc. Nowadays, Health care industry contains huge amount of heath care data, which contains hidden information. This hidden information is useful for making effective decisions. Computer based information along with advanced Data mining techniques are used for appropriate results. Neural network is widely used tool for predicting heart disease diagnosis. In this paper, a heart disease prediction system which uses artificial neural network backpropagation algorithm is proposed. 14 clinical features were used as input for the neural network and then the neural network was trained with backpropagation algorithm to predict absence or presence of heart disease.*

Keywords- Backpropagation, Data mining, Neural network

I. INTRODUCTION

Heart is made up of various Nerves and muscles. Any failure or defect of heart may lead to sudden death. Nowadays, in the world, heart disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the heart diseases. In 2008, 17.3 million people died due to heart disease. Over 80% of deaths in world are because of heart disease. WHO estimated by 2030, almost 23.6 million people will die due to Heart disease. This is one of the reasons why researcher has focus more in designing intelligent system that can be used to diagnose heart diseases with high accuracy, to avoid misdiagnosis. Besides, many people are living with heart disease without awareness. If heart disease could be predicted before, lots of patient deaths would be prevented and also a more accurate and efficient treatment way could be provided. Predication should be done to reduce risk of heart disease. Diagnosis is usually based on signs, symptoms and physical examination of a patient. Almost all the doctors are predicting heart disease by learning and experience. Developing a medical diagnosis system based on machine learning for prediction of heart disease provides more accurate diagnosis than traditional way and reduces cost of treatment. In this paper, prediction of heart disease by an automated medical diagnosis system based on machine learning is proposed to satisfy this need. Back propagation Algorithm which is commonly used Artificial Neural Network learning methodology was used for the prediction system. This system can help in diagnosing disease with less medical tests & effective treatments.

Heart Disease is one of the leading diseases around the world. The effective functioning of the heart plays a vital role in the body. There are many types of heart diseases such as Myocardial infarction, Myocardial ischemia, Congenital heart disease, Coronary heart disease, Cardiac arrest, Peripheral heart disease etc., There are various methodologies available in predicting the heart disease.

Machine learning is a type of artificial intelligence that makes the machines to learn from training data and makes predictions on the test data based on the learned data. The basic idea behind the machine learning is to find the patterns among the data and make the predictions. There are numerous applications in machine learning such as in recommender systems, medical diagnosis, bioinformatics etc., Basically, there are three types of learning in machine learning such as supervised learning, unsupervised learning and reinforcement learning.

Predictive analytics include various statistical techniques from predictive modeling, machine learning (ML) and data mining to make predictions based on the current or historical data. The use of predictive analytics are in the customer relationship management, healthcare industry and in many other fields. Deep learning has a significant impact on the predictive analytics. There are many models in the predictive modeling such as Naive bayes, Logistic regression, Neural networks, Support Vector Machine, Classification and Regression trees etc., Artificial neural network (ANN) is one of the mathematical or algorithmic approach. It is similar to the human brain neurons. The artificial neural network has connections, propagation direction and discrete layers. Each

layer is made up of nodes with the arrows that represents the interconnections between them. In the neural network, there is many nodes in the input layer. These input layer nodes are connected to the hidden layer nodes. Each input is assigned with the weights. The input nodes in the network passes the data to the nodes in the hidden layer which performs some tasks or computations and send the processed data to the output node. The output layer has the node which yields the final result. This is an overview of the process of neural network.

II. LITERATURE SURVEY

Their project explains the concept of Machine learning comprises of man- made consciousness which is utilized in tackling several problems in information science. The regular use of AI is the forecast of outcome dependent on previous information. The machine takes in examples from the current dataset, and afterward applies them to an obscure dataset to foresee the result. Order is an incredible AI method that is ordinarily utilized for expectation. Some characterization calculations foresee with acceptable exactness, while others show a restricted precision. Their project explores a strategy named outfit grouping, which is utilized for improving the exactness of powerless calculations by joining numerous classifiers. Examinations with their device were performed utilizing a coronary illness dataset. A similar logical methodology was done to decide how the outfit procedure can be applied for improving forecast precision in coronary illness. The focal point of their project isn't just on expanding the exactness of powerless order calculations, yet additionally on the execution of the calculation with a clinical dataset, to show its utility to foresee illness at a beginning phase. The consequences of the investigation show that troupe methods, for example, sacking and boosting are powerful in improving the forecast precision of feeble classifiers, and display acceptable execution in recognizing danger of coronary illness. A most extreme increment of 7% precision for powerless classifiers were accomplished with the assistance of outfit characterization.

Their project explains the concept of Different robotized choice emotionally supportive networks dependent on Artificial Neural Network (ANN) have been generally proposed for the recognition of coronary illness in past investigations. In any case, the greater part of these methods centre around the pre-preparing of highlights as it were. In their project, refinement of highlights and disposal of the issues presented by the prescient model, i.e., the issues of underfitting and overfitting. By evading the model from overfitting and underfitting, it can show great execute organization arrangement and superfluous highlights regularly bring about overfitting the preparation information. To dispose of unimportant highlights, they propose to utilize 2 statistical model while the ideally designed deep neural network (DNN) is looked by utilizing thorough exhausted strategy. The strength of the proposed half and half model named 2- DNN is assessed by contrasting its exhibition and ordinary ANN and DNN models, another best-in class AI models and recently announced techniques for heart disease expectation. The acquired outcomes are promising contrasted with the recently revealed strategies. The discoveries of the investigation recommend that the proposed symptomatic framework can be utilized by doctors to precisely anticipate heart disease.

Their project explains the concept of Automatic heart sound is one of the normally utilized procedures for cardiovascular infections discovery. A novel heart sound arrangement technique dependent on profound learning advances for cardiovascular infection expectation is presented, which is fundamentally included three sections: pre-preparing, 1-D waveform heart sound patches characterization utilizing a profound convolutional neural organization (CNN) with consideration component, and larger part deciding in favor of conclusive forecast of heart sound accounts. To upgrade the data stream of the CNNs, a square stacked style design with coterie blocks is utilized, and in every faction block a bidirectional association structure is presented in the proposed CNN. By utilizing the stacked faction and change obstructs, the proposed CNN accomplishes both spatial and channel consideration driving a promising characterization execution. Also, a novel detachable convolution with upset bottleneck is used to decouple the spatial and channel-wise significance of highlights proficiently. Analyses on Physio Net/CinC 2016 show that the proposed strategy acquires a predominant grouping result and dominates in utilization of boundary contrasting with best-in-class strategies.

Thrombospondin-1 (TSP1) is associated with atherosclerosis in animals with diabetes mellitus (DM), but the precise role of TSP-1 in human atherosclerosis remains unknown. Objectives: To investigate serum thrombospondin-1 level in patients with coronary artery disease with and without type 2 DM and its relationship to coronary artery scoring systems. Methods: The study of 180 patients recruited from those underwent coronary angiography for suspected coronary artery disease (CAD) was approved by Institutional Review Board and Institutional Ethical Committee for Human Research of Menoufia university hospital. They were divided according to presence of CAD and type 2 DM into 4 groups: Group I (n = 44 patients): Non diabetic subjects without CAD, Group II (n = 40 patients): Diabetic patients without CAD, Group III (n = 49 patients): Non diabetic patients with CAD and Group IV (n = 47 patients): Diabetic patients with CAD.

Serum level of TSP-1 was measured in all groups and coronary artery scoring analysis was done. Results: Serum TSP-1 levels were higher in patients with CAD and DM than in other groups ($P < 0.01$).

In their project, the various technologies of data mining (DM) models for the forecast of heart disease are discussed. Data mining plays an important role in building an intelligent model for medical systems to detect heart disease (HD) using data sets of the patients, which involves risk factors associated with heart disease. Medical practitioners can help the patients by predicting the heart disease before occurring. The large data available from medical diagnosis is analyzed by using data mining tools and useful information known as knowledge is extracted. Mining is a method of exploring massive sets of data to take out patterns which are hidden and previously unknown relationships and knowledge detection to help the better understanding of medical data to prevent heart disease. There are many DM techniques available namely Classification techniques involving Naïve bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI) and Clustering algorithms like KNN, and Support vector machine (SVM). Several studies have been carried out for developing prediction models using individual techniques and also by combining two or more techniques. Their project provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016. The comparison shows the accuracy level of each model given by different researchers.

Data mining techniques have been widely used to mine knowledgeable information from medical databases. In data mining, classification is supervised learning that can be used to design models describing important data classes, where class attributes are involved in the construction of the classifier. Nearest neighbor (KNN) is a very simple, most popular, highly efficient and effective algorithm for pattern recognition. KNN is a straight forward classifier, where samples are classified based on the class of their nearest neighbor. Medical databases are high volume in nature. If the data set contains redundant and irrelevant attributes, classification may produce less accurate results. Heart disease is the leading cause of death in INDIA. In Andhra Pradesh heart disease was the leading cause of mortality accounting for 32%of all deaths, a rate as high as Canada (35%) and USA. Hence there is a need to define a decision support system that helps clinicians decide to take precautionary steps. In their project they propose a new algorithm which combines KNN with genetic algorithms for effective classification. Genetic algorithms perform global search in complex large and multimodal landscapes and provide optimal solutions. Experimental results show that their algorithm enhances the accuracy in diagnosis of heart disease.

The diagnosis of heart disease using different features or symptoms is a complex activity. In their work two data mining classification techniques like Artificial Neural Network (ANN) and Naive Bayes are used to assist in the diagnosis of heart disease and medication is provided accordingly. It is very important to monitor various medical parameters and post operational days. Hence the latest trend in Health care communication method using IoT is adapted. In their work the AVR-328 microcontroller (Arduino board) is used as a gateway to communicate to the various sensors such as temperature sensor, heartbeat sensor, ECG sensor, sensor for keeping a track of drip levels (blood or saline) and a sensor to keep track of motion. The microcontroller picks up the sensor data and sends it to the network through Wi-Fi and hence provides real time monitoring of the health care parameters for doctors. The data can be accessed anytime by the doctor. The controller is also connected with a buzzer to alert the caretaker about variation in sensor output. At the time of extremity situation alert message is sent to the doctor through the android app connected to the cloud server. Hence quick provisional medication can be easily done by the doctor by using NFC tags without manually searching for history of the patient.

Their project explains the idea of the of heart disease prediction is the zones where machine learning is able to actualized. The Optimization calculations provide the benefit of managing difficult non-linear issues with better adaptability and versatility. In their proposal, the Fast Correlation-Based Feature Selection (FCBF) technique to channel repetitive highlights to increase the nature of heart disease characterization. At that point, grouping is performed which is dependent on various order calculations namely K-Nearest Neighbor, Naïve Bayes, Random Forest, Support Vector Machine and a Multilayer Perception Artificial Neural Network advanced by Particle Swarm Optimization (PSO) joined with Ant Colony Optimization (ACO) methods. The suggested methodology is utilized in heart disease dataset and the outcomes exhibit the adequacy and vigor of the suggested strategy in handling different sorts of information for heart disease prediction. Consequently, the proposed method analyzes the distinctive machine learning calculations and thinks about the outcomes utilizing diverse execution measures such as exactness, accuracy, review, f1-score, and so on. The most extreme characterization efficiency is obtained by utilizing the model suggested by FCBF, PSO and ACO and the result provide the presentation of the offered framework is better than that of the classification methods introduced previously.

Their project explains the idea of coronary heart disease (CHD) builds each year by a critical amount of morality. In addition, death from coronary heart disease acquires the most elevated predominance in Indonesia at 1.5 percent. The misdiagnosis of coronary heart disease is an urgent central which is the main consideration which create demise. To avoid wrong diagnosis of CHD the insightful framework has planned. Their method is suggested a simulation that can utilized for the analyze of coronary heart disease in preferable execution over conventional indicative strategies. A few analysts have built up a framework utilizing ordinary neural network and machine learning calculation. Yet, outcomes do not provide a decent execution. In light of a regular neural organization, deep neural organization (DNN) is suggested in work. The supervised learning neural network provides the calculation that great in the classification. In DNN model, the usage of double grouping was executed to analyze CHD present or CHD missing. To help execution examination utilizing the UCI machine archive coronary heart disease dataset, ROC Curve and its disarray grid was actualized in the proposed method.

III. PROPOSED SYSTEM

Deep Learning is a technology of which mimics a human brain in the sense that it consists of multiple neurons with multiple layers like a human brain. The network so formed consists of an input layer, an output layer, and one or more hidden layers. The network tries to learn from the data that is fed into it and then performs predictions accordingly. The most basic type of neural network is the

ANN (Artificial Neural Network). The ANN does not have any special structure, it just comprises of multiple neural layers to be used for prediction.

An ANN also called as neural network is a mathematical model based on biological neural networks. Artificial neural network is based on observation of a human brain. Human brain is very complicated web of neurons. Neuron has axons, dendrites and synapses. The designed ANN has three layers: namely an input layer, a hidden layer and an output layer. The dataset was taken from the UCI machine learning repository. The heart disease dataset is made up of 75 raw features from which 13 features were published. These features are very vital in the diagnosis of heart diseases. The features include fasting blood sugar test which must indicate < 120mg / dl for a patient with absent test result and test result of >120mg / dl for a patient that has heart disease. Also, a patient that has serum cholesterol greater than 180mg/dl is also considered as heart disease present.

A Heart Disease Prediction System (HDPS) using data mining and Deep Neural Network (DNN) techniques has been presented. From the DNN, a multilayer perceptron neural network along with back propagation algorithm is used to develop the system. Because this model proves the better results and helps the domain experts and even person related with the field to plan for a better diagnose and provide the patient with early diagnosis results as it performs realistically well even without retraining. The experimental result shows that using neural networks the system predicts heart disease with higher accuracy.

Heart disease prediction using Deep Neural Networks (DNN) is an advanced application of machine learning that leverages the power of deep learning algorithms to accurately predict the risk of heart disease in patients. DNNs are a subset of Artificial Neural Networks (ANN) with multiple hidden layers, allowing them to learn complex patterns and representations from the input data.

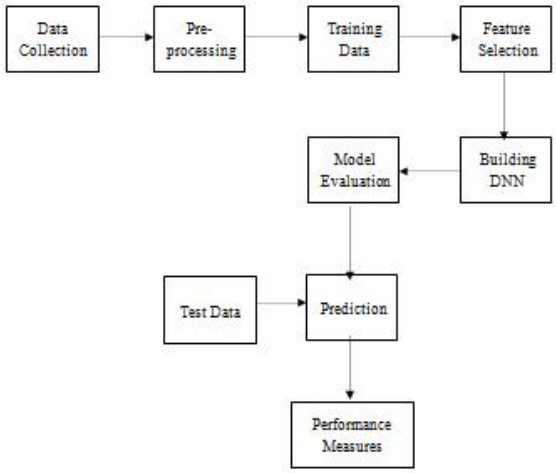


Fig 1 Block Diagram of Proposed System

IV. MODULES DESCRIPTION

Data Collection

The first step in heart disease prediction using DNN involves collecting a dataset with relevant information about patients, including features such as age, sex, blood pressure, cholesterol levels, blood sugar, family history, and other medical indicators. This dataset should also include labels indicating whether each patient has heart disease or not.

Data Pre-processing

Before feeding the data into the DNN, it undergoes preprocessing steps, such as handling missing values, normalization, and feature scaling. Data preprocessing ensures that the data is in a suitable format for training the deep neural network.

DNN Architecture Design

The architecture of the DNN is designed with multiple hidden layers, allowing for the extraction of complex features and patterns from the input data. The number of hidden layers and nodes in each layer can be adjusted based on experimentation and hyperparameter tuning.

Training the DNN

The DNN is trained using the prepared dataset, where the input features act as input nodes, and the output layer predicts the likelihood of heart disease. During training, the DNN learns from the data by adjusting the weights and biases in each layer through backpropagation and gradient descent algorithms.

Validation and Testing

After training, the DNN's performance is evaluated using a separate validation dataset to check for overfitting and fine-tune the model's hyperparameters. Finally, the model is tested on a completely unseen test dataset to assess its accuracy in predicting heart disease risk for new patients.

Model Evaluation

Model evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are used to assess the DNN's performance in heart disease prediction. These metrics help determine how well the model is generalizing to new data and making accurate predictions.

Deployment

Once the DNN is trained and evaluated, it can be deployed in real-world applications to assist healthcare professionals in predicting heart disease risk for patients. It can be integrated into medical systems or mobile applications to provide quick and reliable predictions.

Continuous Improvement

To ensure the DNN's accuracy and effectiveness, continuous monitoring and updates can be performed as more data becomes available. This allows the model to adapt to new patterns and improve its predictions over time.

Heart disease prediction using DNNs holds significant potential in improving healthcare outcomes by enabling early detection and timely intervention. However, it's essential to ensure the model's ethical use, data privacy, and adherence to medical guidelines when deploying such predictive systems in clinical settings.

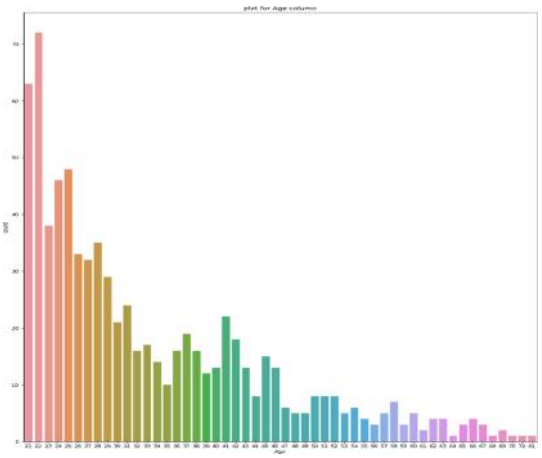
V. **SCREEN SHOTS**

	num_preg	glucose_con	Diastolic_bp	Thckness	Insulin	BMI	Diab_pred	Age	Target
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

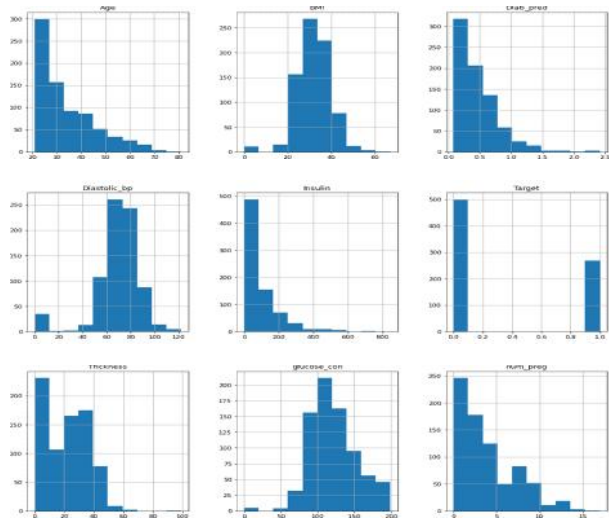
Head rows of dataset

	num_preg	glucose_con	Diastolic_bp	Thckness	Insulin	BMI	Diab_pred	Age	Target
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

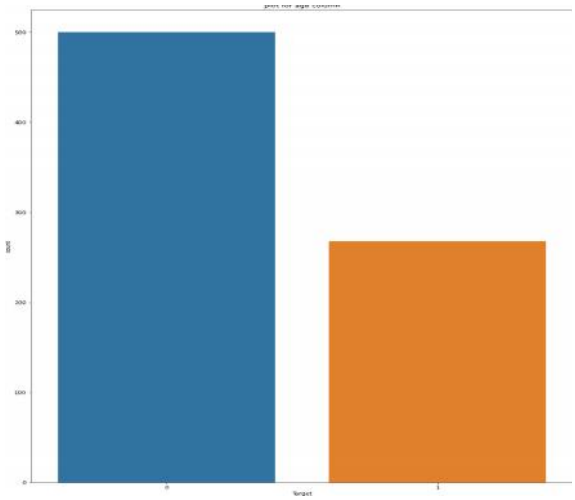
Tail rows of dataset



Plot of Age column



Data Visualization



Plot of Target column

Classification Report:

	precision	recall	f1-score	support
Diabetes	0.78	0.92	0.84	355
No Diabetes	0.75	0.49	0.60	182
accuracy			0.77	537
macro avg	0.76	0.70	0.72	537
weighted avg	0.77	0.77	0.76	537

Confusion Matrix:
[[325 30]
[92 90]]

Classification report of Training Phase

VI. **CONCLUSION**

In this paper we have proposed a new feature selection method for heart abnormalities classification using DNN and various feature selection methods for Andhra Pradesh Population. We applied different feature selection methods to rank the attributes which contribute more towards classification of heart abnormalities, which indirectly reduces the no. of diagnosis tests to be taken by a patient. Our experimental results indicate that on an average with DNN and feature subset selection provides on the average better classification accuracy and dimensionality reduction. Proposed method eliminates useless and distortive data. This research will contribute reliable and faster automatic heart abnormalities diagnosis system, where easy diagnosis of heart abnormalities will save lives. Coronary heart abnormalities can be handled successfully if more research is encouraged.

In conclusion, Deep Neural Networks (DNNs) have shown remarkable promise in heart disease prediction, offering a powerful tool for improving healthcare outcomes. Through the utilization of large and diverse datasets, DNNs can effectively learn intricate patterns and representations from patient data, enabling accurate risk assessments for heart disease.

The successful implementation of DNNs in heart disease prediction is attributed to their ability to handle complex and non-linear relationships among various medical indicators. With multiple hidden layers, DNNs can extract high-level features and insights from input data, contributing to their superior predictive capabilities.

By leveraging DNNs for heart disease prediction, healthcare professionals can benefit from timely and reliable risk assessments, enabling early detection and intervention. Identifying patients at high risk for heart disease allows for proactive measures, personalized treatment plans, and lifestyle recommendations to prevent potential complications.

Despite the advantages, the deployment of DNNs in healthcare requires careful consideration of ethical considerations and data privacy. Ensuring the security and confidentiality of patient information is paramount in building trust between patients, healthcare providers, and AI-driven predictive models.

Future work, continued research and advancements in DNN architectures, interpretability techniques, and model optimization will further enhance the accuracy and effectiveness of heart disease prediction. Collaborations between medical experts and AI specialists will be crucial in developing robust and reliable DNN models for real-world clinical applications.

Overall, DNNs represent a transformative approach to heart disease prediction, offering the potential to revolutionize patient care and pave the way for more proactive

and personalized healthcare strategies. As this technology continues to evolve, it has the capacity to make a significant impact on global health by reducing the burden of heart disease and improving patient outcomes.

REFERENCES

[1] Ahmed, Rizgar Maghdid, and Omar Qusay Alshebly. "Prediction and factors affecting of chronic kidney disease diagnosis using artificial neural networks model and logistic regression model." Iraqi Journal of Statistical Sciences 16.28 (2019): 140-159.

[2] Ali, Farman, et al. "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." Information Fusion 63 (2020): 208-222.

[3] Ali, Liaqat, et al. "An Automated Diagnostic System for Heart Disease Prediction Based on Chi-Square Statistical Model and Optimally Configured Deep Neural Network." IEEE Access 7 (2019): 34938-34945.

[4] Banu, NK Salma, and Suma Swamy. "Prediction of heart disease at early stage using data mining and big data analytics: A survey." Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 016 International Conference on. IEEE, 2016.

[5] Burse, Kavita, et al. "Various pre-processing methods for neural network based heart disease prediction." Smart innovations in communication and computational sciences. Springer, Singapore, 2019. 55-65.

[6] Cp, Prathibhamol, Anjana Suresh, and Gopika Suresh. "Prediction of cardiac arrhythmia type using clustering and regression approach (P-CA-CRA)." Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on. IEEE, 2017.

[7] Darmawahyuni, Annisa, Siti Nurmaini, and Firdaus Firdaus. "Coronary Heart Disease Interpretation Based on Deep Neural Network." Computer Engineering and Applications Journal 8.1 (2019): 1-12.

[8] de Carvalho Junior, Helton Hugo, et al. "A heart disease recognition embedded system with fuzzy cluster algorithm." Computer methods and programs in biomedicine 110.3 (2013): 447-454.

[9] Galit, S., Nitin, R. P., & Peter, C. B. (2016). Data Mining for Business Intelligence: Concepts,Techniques and Applications in Microsoft Excel® withXLMiner.

[10] Kester, Q. A. (2016, July). Using Formal Concepts Analysis Techniques in Mining Data from Criminal Databases and Profiling Events Based on Factors to Understand Criminal Environments. In International Conference on Computational Science and Its Applications (pp. 480-496). Springer, Cham.

[11] "Data mining based fragmentation and prediction of medical data." Computer Research and Development (ICCRD), 2011 3rd International Conference on. Vol. 2. IEEE, 2011.

[12] Khalil, Ahmed J., et al. "Energy Efficiency Predicting using Artificial Neural Network." (2019).

[13] Khourdifi, Youness, and Mohamed Bahaj. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization." Int. J. Intell. Eng. Syst. 12.1 (2019): 242-252.

[14] Kumar, B. Senthil, and R. Gunavathi. "Comparative and Analysis of Classification Problems." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 7.8 (2017).

[15] Latha, C. Beulah Christalin, and S. Carolin Jeeva. "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques." Informatics in Medicine Unlocked 16 (2019): 100203.