IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-8, Aug 2019 E-ISSN: 2347-2693

Analysis and Prediction of Heart Health using Deep Learning Approach

Yogita Solanki[1*], Sanjiv Sharma[2]

[1]Dept. of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India
[2]Dept. of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India

**Corresponding Author: solankiyogita53@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i8.309315 | Available online at: www.ijcseonline.org

Accepted: 22/Aug/2019, Published: 31/Aug/2019

Abstract— Medical data mining is a tremendously significant domain for exploration because of its importance in the expansion of innumerable applications in the medical domain. On the fact of briefing the deaths taking place globally, the heart disease seems as the foremost cause of death. The recognition of the chance of heart disease in an individual is a complex task for health specialists because it requires years of experience and intense medical tests to be conducted. In this research work, enhanced deep neural network (DNN) learning is introduced to treat patients accurately and for maintaining consistency in heart disease prediction system. So that anticipation of the loss of lives at the prior stage is possible. The results formulated ideally verify that the designed diagnostic scheme is able of calculating the risk level of heart disease efficiently when compared to other methodologies. The proposed model provides better results in heart diseases prediction compared to that of previous work. Early prediction of the disease reduces the costs and time of the treatment. The cost and time of treatment will be reduced due to the early prediction of heart disease.

Keywords— Machine learning, Medical Data Mining, Heart Disease, Tensor Flow, Deep Neural Network.

I. INTRODUCTION

Today, numerous medical facilities introduced various kinds of information frameworks to manage their social assurance or else patients' data. These data frameworks usually generate a lot of information which can be in distinctively organized in perfect manner; this database that consists of rich data is once in a while used for clinical choice making. Data mining methods have been extensively used to extract enlightened knowledge's from healthcare data repositories Medical data are large size in nature. If the data-set consists of duplicate & unrelated features, classification might yield less accurate result[1].

According to the study[2] of recent deaths, it has been found that heart diseases become a cause of the deaths. About 27 percent of deaths have occurred in the age group of 50-69 years due to heart disease. If all the age groups are included, then 31.8% of the deaths caused due to heart diseases. It is the foremost reason for death among males as well as females. Different risk factors for heart disease are blood pressure, smoking, cholesterol, age, etc. The only solution to this problem is early diagnosis. The early prediction of the disease decreases the mortality rate. In medical data mining, numerous algorithms used for heart disease prediction like- ID3 (Iterative Dichotomized 3), K-means, Logistic regression, neural network, CART (Classification and Regression Tree), C4.5, etc. Information technology has intruded into numerous fields, with the rapid growth of database technology. Today, most of the hospitals have established their own hospital information system (HIS), with the applications.

The amount of data in the database is enlarged rapidly, the database scale expands regularly and it increases complexity[3]. Therefore, data mining technologies are used for dealing with a large amount of Medical data. Diagnosis of cardiovascular disease is a crucial matter in the health care industry. After the survey, it was analyzed that machine learning approach provides better results [4].But there was also found out some drawbacks in the existing methodologies. In this research, advanced and improved methodology has applied that overcomes the drawbacks of existing methods. By analyzing, it has been found that those drawbacks can be overcome only with the help of deep neural network algorithm.

Deep learning[5] is an enhanced machine learning method for perception, learning of machines & feature extraction. Deep learning algos perform their operations by numerous consecutive layers. The layers are inter-linked & every layer get the output of the earlier layer as input. On the converse, Deep Learning techniques are able to capture composite relations. Deep learning and machine learning approaches are implemented by Tensorflow package in advanced manner.

In this paper, deep neural network has applied on Cleveland dataset for identifying the patient those who have any disease related to heart. The motive behind this system is to minimize the diagnosis time and improve the accuracy of diagnosis for supporting complicated Medical diagnosis decision process. The objective of researcher is to build up a medical decision support systems (MDSS) that assist physicians. The medical data mining plays a very important role in medical field. There are numerous data mining algorithm & methodologies that are utilized by the researchers for getting the accurate patterns that helps to diagnosis the diseases.

This paper proposes the model with deep neural network and Tensor flow package for predicting heart disease. In the remaining part of the paper, Section II portrays data source. Section III describes the Literature Survey. Proposed methodology covered in Section IV. In Section V, the experimental results and analysis have described. At last, the conclusion and future scopes of the research are explained in Section VI.

II. HEART DISEASE PREDICTION

Heart disease [6] is indicates all the diseases related to heart's disorder. On contrast cardiovascular disease defines difficulties with the, heart, circulatory system, blood vessels &, heart disease refers to the difficulties & distortion in the heart itself.

There are numerous kinds [7] of heart disease:-

- **Arrhythmia:** This occurs when there is an electrical impulse in the heart, which helps in balancing the heartbeat that stops working properly. It is an uneven heartbeat.
- **Coronary artery disease:** In this disease, plaque develops narrows the coronary arteries & because of this the heart obtains fewer oxygen & nutrients supply.
- **Dilated cardiomyopathy:** Due to weakening of the heart muscles, the heart cells become weak and can't properly pump the blood.
- **Myocardial infarction:** This disease destroys part of the heart muscle and generated bleeds from the intestine. It is generally caused because of the blood clot that establish in one of the coronary arteries.
- **Hypertrophic cardiomyopathy:** It is a hereditary disorder in which the left ventricle wall becomes thick, which make it difficult to get blood out of the heart. The sudden death in athletes is mainly occurred due to Hypertrophic cardiomyopathy.
- **Pulmonary stenosis:** The right ventricle of the heart makes it difficult to pump the blood into the pulmonary artery because the pulmonary valve is very tight.
- **Mitral regurgitation:** This disease makes blood to run back into the heart when it ought to go away. It is the disease that occurs when the mitral valve in the heart doesn't close tightly.

In this research, some common symptoms that causes heart diseases have taken for making intelligent system that can predict that a patient have heart disease or not. On the basis of common risk factors that cause disease related to heart, prediction of heart disease has been done.

III. RELATED WORK

There are numerous works has been done related to heart disease prediction system, the subsequent paragraphs review some of the papers on heart disease to present a comprehensive review of literature which may help to gain insight about various machine learning [8] techniques and their usages.

Rairikar et al. [9] have examine prediction systems for heart disease by a large numeral of input characteristics. The scheme utilizes medical conditions for example Gender, blood pressure, and cholesterol similar to 13 attributes for predicting the possibility of patient having Heart disease.

Dr. B. Umadevi et al. [10] mostly focuses on the study of various approaches of heart attack disease prediction research papers are analyzed and studied. The prediction accuracy of accessible systems is able to be enhanced. In future, novel methodologies are to be developed which can overcome the deficiency of the existing technique.

Sayali Ambekar et al. [11] implemented the CNN-UDRP algo for disease risk prediction by prepared data. The heart disease is predicting by naïve KNN algo & bayes algo. They compared the results b/w Naïve bayes & KNN algo. The accuracy of NB is 82% that is more than KNN algo. This system provided accurate disease risk prediction by given the recorded patient data as input. The risk is estimated in the form of high, medium & low risk of heart disease. Due to this system, there might be less time consuming & minimal cost might be probable for predicting disease risk. In the future scope, they will plan to include additional diseases and develop a system that can predict the risk from specific illness.

Aditi Gavhane et al. [12] proposed an application that predict heart disease on the basis old records of heart patients that contains the attributes such as age, sex, pulse rate etc. A neural network has proven to be the enhanced and more reliable algo than all other machine learning technique.

M.A. Jabbar et al. [13] proposed a model that trained by the Hidden Naïve Bayes (HNB) for predicting heart disease. Their outcome on heart disease dataset demonstrates that the

HNB classification technique has proven the more accurate algorithm than the other technique.

Purushottam et al.[14] have deliberate a system that can proficiently determine the rules for anticipating the risk level of patients that are based on the detailed parameter as per their health. The priority of the rules is decided on the basis of user's requirement. The achievement of the scheme is estimated in the basis of accuracy & the outcomes represent that the system has enormous potential to predict the heart disease precisely.

Sana Shaikh et al.[15] designed an intellectual System through data mining methods, that is, Naive Bayes. This model is implemented as a Java application. This application retrieves unseen data from the already collected database & analyzes the consumer values with trained dataset. It is capable of answering composite queries for analyzing heart disease & therefore supports healthcare practitioners to build intellectual medical decisions that can't be support by traditional decision systems.

R. Kannan et al.[16] implemented 4 different machine learning techniques that are SVM, random Forest, logistic regression & stochastic gradient boosting in R language by using Rstudio. All these machine learning methodologies are compared on the basis of accuracy. After comparing of all four algorithm, logistic regression achieved greatest accuracy of 87%.

C.Sowmiya et al.[17] implemented the potential of nine (9) classification techniques that was evaluated for predicting heart disease, that is naive Bayesian neural network, decision tree, ANN, KNN, SVM. By using medical profiles for blood pressure, chest pain type, age, sex, fasting blood sugar. This system can predict the patients that suffering from heart disease on the basis of their medical profiles. This analysis has proved that classification based technology contributes to high effectiveness and achieves higher accuracy than previous methods.

M. Sultana et al.[18] implemented SMO, KStar, J48, Bayes Net, Multilayer perceptron in WEKA tool for predicting heart disease. Bayes Net (87% of accuracy) & SMO (89% of accuracy) achieved most favorable performance than Multilayer perceptron, KStar & J48 method by applying k-fold cross-validation. The performance of these algos is still not satisfied. So, the performance has enhanced for diagnosing disease.

V. Chaurasia et al. [19] done the experiments of dissimilar classifier for finding the superlative classifier that predict those patients who have suffering from heart disease. The three classifiers ID3, DT & CART were mainly utilized for examination of patients those are having heart related diseases. CART methodology has proven better than other techniques with 83.49% of accuracy & the entire time for making the model is 0.23 seconds. The minimum average error is 0.3 compared to others in the CART classifier.

Yogita solanki et al.[20] presents various issues related to healthcare & different machines learning algos. A comprehensive evaluation of the literature has been summarized for analyzing the previous work done on this research topic. Numerous data mining methods like Naïve Bayes, Artificial Neural Network (ANN), Decision tree, K-Nearest Neighbor (KNN) & Support Vector Machine (SVM) are implemented for analyzing the best algorithm so far that can predict health disease.

Niyati I. Patel et al.[21] took an overview of the data mining techniques for the diagnosis of distinct diseases. As per their survey, hybrid data mining techniques provide accurate result for prediction of any disease. They considered all diseases for diagnosis.

H. Benjamin Fredrick David et al.[22] implemented three classification algos that are Naïve Bayes, Random Forest and Decision trees for predict the possibility of heart related diseases. The foremost objective of the researcher was to identify the best classification method that is mainly used for heart disease prediction. After obtaining the result of their experiment, they come to know that Random Forest achieved better accuracy than other two techniques with 81% accuracy.

IV. METHODOLOGY

A. Neural Network

Neural networks are usually arranged in layers. Layers are made up of several interconnected 'nodes' which have an 'activation function'. Patterns are accessible in the network throughout the 'input layer' that communicates with one or more 'hidden layers' wherever the definite processing is done via a system of weighted 'connections'. The hidden layers are finally associated to an 'output layer'. Artificial Neural Network & Deep Neural Network are the subsets of the Neural Network.

1) Artificial Neural Network

Artificial Neural Networks(ANN) are the computerized models encouraged by the human brain. ANNs are biologically motivated simulations on the computer to perform certain functions, such as classification, clustering, etc. The graphical representation of ANN is shown below in Figure 1.

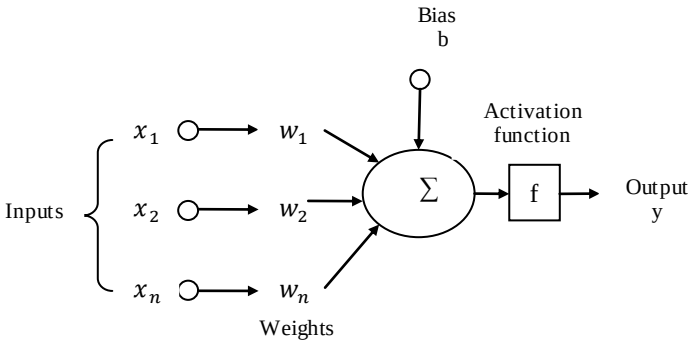

Figure 1. Artificial Neural Network

Each weight is multiplied by its associated inputs. For solving a problem, weights are the information utilized by the NN. If the value of weight is zero, then the addition of bias is done for making the output non-zero. The range of weighted sum is from 0 to infinity. Threshold value is determined to limit the response to the desired value. The computational sum is approved via an activation function. The activation function is set to the transfer function that is utilized to provide the favoured output.

$$Y = \text{Activationfunction}(\sum(\text{Weight} * \text{input}) + \text{bias}) \quad (1)$$

Some activation function are tanh, ReLU(rectifier linear unit), Sigmoid, Linear etc.

2) Deep Neural Network

A deep neural network(DNN) is a neural network that has a definite level of complexity. Neural networks that have more than two layers are deep neural network. It is a class of machine learning algos that utilized multiple layers, processing units for extraction & feature transformation. The required GPU (Graphical Processing Unit) for deep learning should be based on the socket type, the number of cores and the cost of the processor. Nowadays, Tensor flow package is mainly used for implementing deep neural network.

It is a software library that is mainly developing for Machine Intelligence. There are numerous packages that are available in the library such as Keras, Tensorflow etc. The R interface to Tensor Flow consists of a suite of R packages that provide a variety of interfaces to Tensor Flow. Tensor flow is a low-level interface to the Tensor Flow computational graph.

In the proposed methodology the dataset acquired from Cleveland put as input into the deep neural network & the outcome achieved in the appearance of a binary output (i.e. 0 or 1) where 1 shows the existence while 0 shows the nonexistence of the heart disease. To calculate the required intensive neural network, R is used as a development language.

Initially Cleveland dataset has pre-processed by the removing the rows that contains missing values. The Cleveland datasets is divided into two sets: testing and training dataset in the ratio of 7:3.After that, the training and testing dataset has normalized or scaled by using equation 1.

$$\text{Normalized value} = (\text{Value} - \text{Mean})/\text{sd} \quad (2)$$

Where sd= standard deviation

The training dataset utilized to prepare the DNN and the testing dataset used to recognize the accuracy. A deep neural network is used with TensorFlow and keras package. The model has made by three hidden layers with 2 activation function (tanh, ReLU) but for overcoming the difficulty of overfitting, some percentage of neurons are dropout at each hidden layer. Dataset has splited 20 percent for validation. After training the model, it validates the data by using 20 percent dataset. In this paper, 32 batch size and 100 epochs have been taken. At every epoch, loss is decreases while increases the accuracy.

PROPOSED ALGORITHM:

Step 1. Select Dataset.

Step 2. Pre-process the dataset by removing the rows that contains missing values

Step 3. Divide dataset into two part(70% training data & 30% testing data)

Step 4. Train the model by applying a deep neural network with the tensorflow Package

Step 5. Made 3 hidden layers in deep neural network

Step 6. Apply Tanh activation function in first hidden layer

Step 7. Apply ReLu activation function in second and third hidden layer

Step 8. Cross validate the training dataset

Step 9. Test the model by 30% testing dataset

Step 10. Obtained predicted outcome

Step 11. End

The flowchart representation of the proposed methodology is shown in Figure 2.

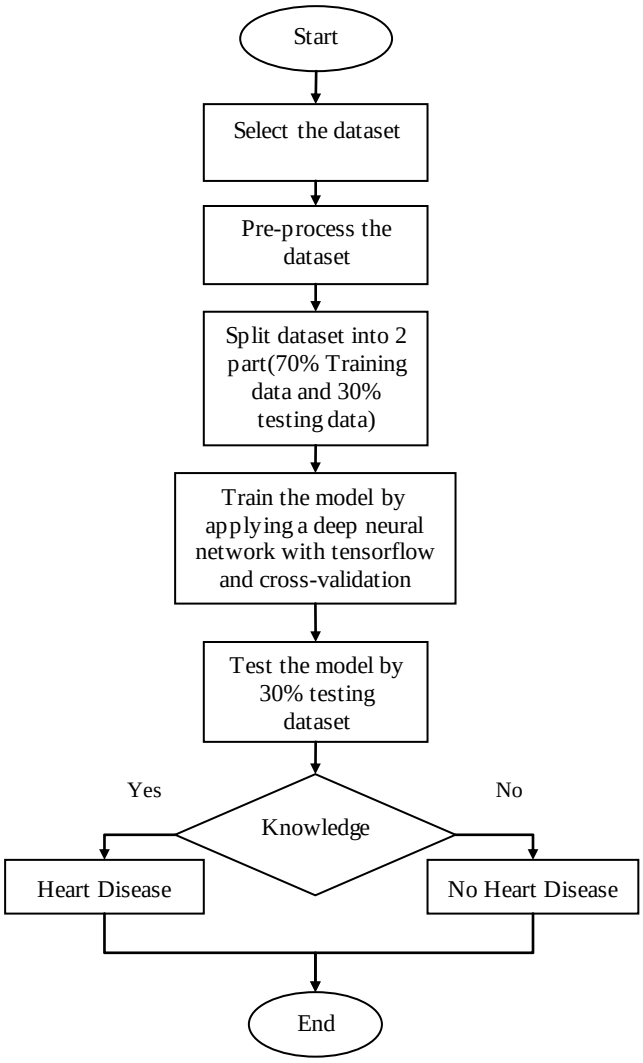

Figure 2: Flowchart of Proposed Methodology

V. RESULTS AND DISCUSSION

A novel model has developed for better prediction, after examining the existing method. In the proposed methodology, the deep neural network has utilized with the tensorflow package in RStudio. It generates a deep neural network by using the neuralnet package as demonstrated in Figure 3.

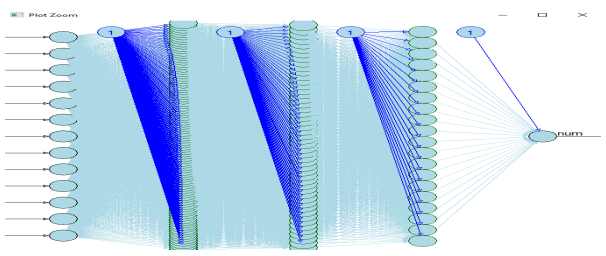

Figure 3: Deep neural network

This neural network have 13 attributes as a input and it represents 3 hidden layers in which first layer have 100 neuron from which 40% neuron are dropout for preventing overfitting, second layer have 50 neurons with dropout of 30% and third hidden layer have 20 neurons with dropout of 20%.

The accuracy of the proposed approach over the existing approach is compared below in Figure 4.

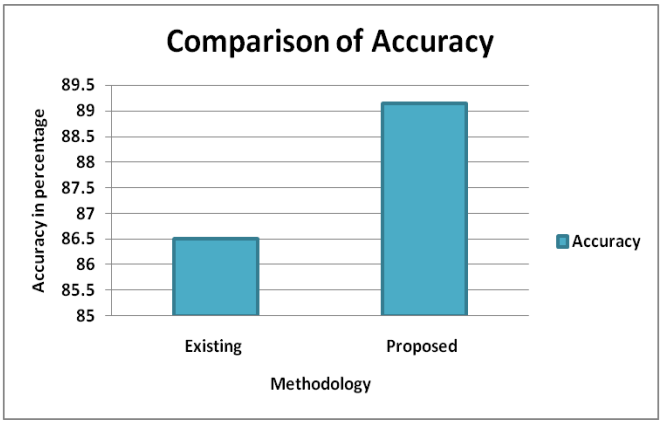

Figure 4: Comparison of accuracy

Table 1: Comparison Result of Algorithm

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Logistic Regression	86.51	87.18	82.92	84.99
Deep Neural Network	89.15	88.23	85.71	86.95

The above Table 1 demonstrated the comparative result of data mining algorithm namely Logistic regression and Deep learning approach in terms of performance evaluation(Precision, Recall, F1 Score and Accuracy). The more percentage of precision value represented that more accurate result are obtained with less error.The recall represented that this model contain more true positive results(having heart disease) from all the patients. F1 Score shows as a harmonic mean of both recall and precision.

The comparison of performance measure such as recall, precision & F1-Score of the proposed methodology over existing methodology by using some performance evaluation measure be shown in Figure 5.

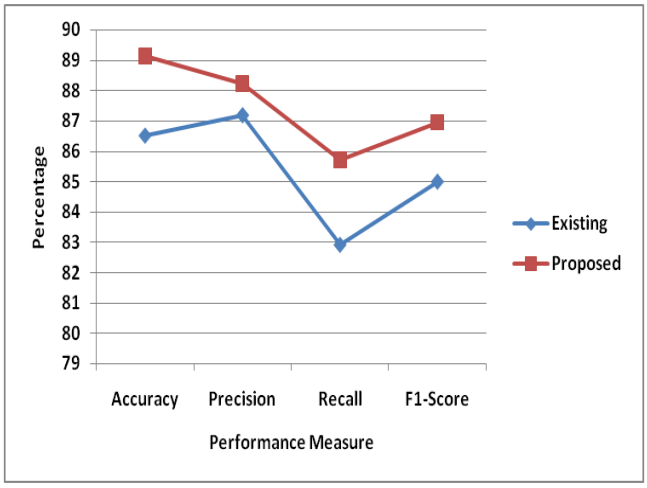

Figure 5: Performance Evaluation

The epoch is a hyper parameter that describes the number of times the learning algo will effort throughout the whole training dataset. One epoch means that every sample in the training dataset modernizes the internal model parameters. One or more batches are come under an epoch. The graphical representation of accuracy and loss at each and every epoch is represented in Figure 6.

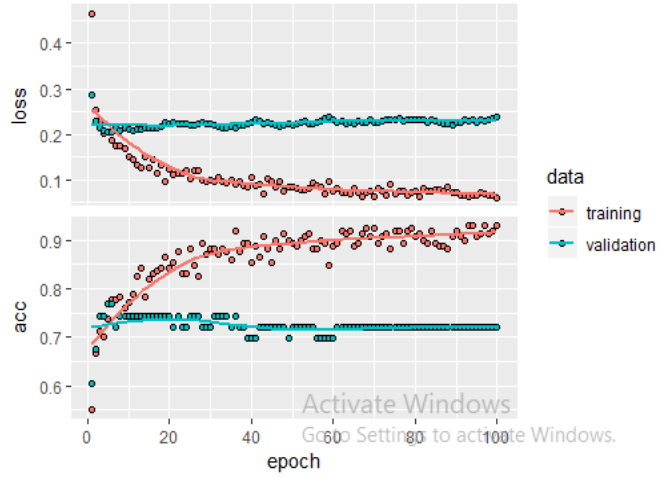

Figure 6. Visualization of model performance

At every epoch the loss of the prediction decreases with increases the accuracy of the heart disease prediction. Two graphs are shown in figure one is between loss and epoch and another is between accuracy and epoch.

The confusion matrix and overall performance of the deep neural network(proposed methodology) visualized in the Rstudio by using the R programming language. The visualization of the overall performance is represented as shown in Figure 7.

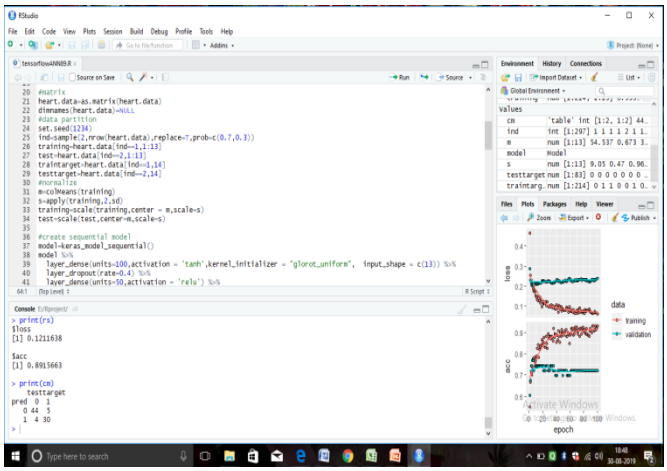

Figure 7. Overall result of Proposed Methodology

VI. CONCLUSION AND FUTURE SCOPE

The objective of this research was to create a model that deals with non-linear features, a huge amount of data and provide more accurate result than existing research. Thirteen predictor variables and one target variable from the UCI repository Cleveland dataset are utilized for predicting the diagnosis of heart disease. 30% of the data has been taken as a test dataset, which has not been seen during the data training phase. The dataset is trained by Deep Neural Network with keras package then validate the model. This deep learning with keras package has increased the speed of prediction. This model performs better than the existing approach with 85.72% precision, 88.24% recall, 12.1% loss and 89.15% accuracy.

REFERENCES

[1] Mai Shouman, Tim Turner and Rob Stocker, "Using data mining techniques in heart disease diagnosis and treatment", Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC), 2012.

[2] Pagidipati, N. J., & Gaziano, T. A., "Estimating Deaths From Cardiovascular Disease: A Review of Global Methodologies of Mortality Measurement. Circulation", 127(6), 749–756, 2013. doi:10.1161/circulationaha.112.128413.

[3] K. Subhadra, Vikas B, "Neural Network Based Intelligent System for Predicting Heart Disease", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5,pp. 484-487, March 2019

[4] Zhi-Gen Hu, Jian-Ping Li, "Research And Application Of Data Warehouse And Data Mining Technology In Medical Field", 12th International Computer Conference On Wavelet Active Media Technology And Information Processing (ICCWAMTIP), IEEE 2015.

[5] Kamran Farooq et al., "Clinical Decision Support Systems: A Visual Survey".

[6] J. C. Prather et al., "Medical data mining: knowledge discovery in a clinical data warehouse", Proc AMIA Annu Fall Symp. 1997: 101–105.

[7] M. Thiyagaraj, G. Suseendran, "Survey on heart disease prediction system based on data mining techniques", Indian

Journal of Innovations and Developments Vol 6(1), January, 2017.
[8] Jaymin Patel et al., "Heart Disease Prediction Using Machine learning and Data Mining Technique", volume 7, pp. 129-137, March 2016.
[9] Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. , " Heart disease prediction using data mining techniques", 2017 International Conference on Intelligent Computing and Control (I2C2).
[10] Dr. B. Umadevi, "A Survey on Prediction of Heart Disease Using Data Mining Techniques", International Journal of Science and Research (IJSR), 2015.
[11] Ambekar, S., & Phalnikar, R., "Disease Risk Prediction by Using Convolutional Neural Network", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018. doi:10.1109/iccubea.2018.8697423.
[12] Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, P. K., " Prediction of Heart Disease Using Machine Learning. 2018 Second International Conference on Electronics", Communication and Aerospace Technology (ICECA), 2018.doi:10.1109/iceca.2018.8474922.
[13] Jabbar, M. A., & Samreen, S., "Heart disease prediction system based on hidden naïve bayes classifier", 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), 2016.doi:10.1109/cimca.2016.8053261.
[14] Ritika, Chadha, Mayank, Shubhankar, "Prediction of heart disease using data mining techniques", Springer 2016. DOI: https://doi.org/10.1007/s40012-016-0121-0.
[15] Shaikh, S., Sawant, A., Paradkar, S., & Patil, K., "Electronic recording system-heart disease prediction system", 2015 International Conference on Technologies for Sustainable Development (ICTSD), 2015.doi:10.1109/ictsd.2015.7095854.
[16] R. Kannan, V. Vasanthi, "Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease", Springer Briefs in Forensic and Medical Bioinformatics, pp 63-72, 14 June 2018. DOI: https://doi.org/10.1007/978-981-13-0059-2_8.
[17] Sowmiya, C., & Sumitra, P., "Analytical study of heart disease diagnosis using classification techniques", 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017.doi:10.1109/itcosp.2017.8303115.
[18] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction", 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2016.
[19] Vikas Chaurasia, Saurabh Pal, "Early Prediction of Heart Diseases Using Data Mining Techniques", Caribbean Journal of Science and Technology, Vol. 1, pp. 208-21, December 2013.
[20] Yogita Solanki, Sanjiv Sharma, "A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.11, No.4, pp. 43-51, 2019. DOI: 10.5815/ijieeb.2019.04.05.
[21] Niyati I. Patel , Hiren R. Patel, "A Survey on Prediction of Disease with Data Mining", International Journal of Computer Sciences and Engineering(IJCSE), Survey Paper Vol.-7, Issue-2, E-ISSN: 2347-2693,pp. 289-293, Feb 2019.
[22] H. Benjamin Fredrick David and S. Antony Belcy, "Heart Disease Prediction Using Data Mining Techniques", Department of Computer Science and Engineering, Manonmaniam Sundaranar University, India, ICTACT JOURNAL ON SOFT COMPUTING, OCTOBER 2018.

Authors Profile

Sanjiv Sharma, he is currently employed as an assistant professor in the Department of Computer science Engineering and Information Technology in MITS, Gwalior. He has 12 years of teaching and research experience. He did his bachelor of engineering from Pt. Ravishankar University, Raipur. He completed his Master of technology from RGPV, Bhopal, and Ph. D from Bansthali Vidhya Peeth, Jaipur. His areas of research are Social Network Analysis, Data Mining and Network security.

.

Yogita Solanki**,** she is currently pursuing Master of technology in Computer Science and Engineering (CSE) from MITS, Gwalior, Madhya Pradesh, India. She completed her Bachelor of Engineering (B.E) from ITM GOI, Gwalior, Madhya Pradesh, India. Her area of interests is Data Mining and Social Network Analysis.

