Heart Disease Prediction Model

Muthulakshmi K[1*], Vignesh M[2a], Paramasivam P[3a], Vijayan S [4a]

[*]Associate Professor, Department of Information Technology, Panimalar Engineering College
[a]Student, B.Tech Information Technology, Panimalar Engineering College

---

Abstract

Heart disease, one of the major causes of impermanence worldwide, can be diminished by early heart disease diagnosis. Various studies gives only a glance into predicting heart disease with Machine Learning techniques. Two available datasets- Stat log and Cleveland both are used to build the model. In this paper, we present a unfamiliar method that targets at finding useful features by applying machine learning techniques results in improving the perfection in the prediction of cardiovascular disease. This prediction model is introduced with many known classification techniques and different sets of features. We build up high performance level with an accuracy level of 90% through this prediction model for heart disease. Therefore, early treatment is preferable to prevent the deaths caused by later heart disease diagnosis.

Keywords: Heart disease, disease prediction model, Machine Learning.

1. INTRODUCTION

Heart disease is a cardiovascular disease (CVD) and it is difficult to identify heart disease because of several contributory risk factors such as high blood pressure, diabetes, abnormal pulse rate, high cholesterol and several other factors. Heart disease is a condition when plate like on arterial walls can block the flow of blood and cause a heart attack or stroke and also several risk factors that can lead to heart disease include unhealthy diet, excessive use of tobacco and alcohol, and physically inactive. The severity of the heart disease is classified based on various methods like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), and Naive Bayes (NB) etc. We introduce neural networks using heart rate time series and this includes Right bundle branch block (RBBB), Left bundle branch block (LBBB). Machine learning based clinical decision making have been recently applied in healthcare area and machine learning algorithms (MLAs) are chaos firefly algorithm, Multi Layer Perceptron (MLP), Back Propagation Neural Network (BPNN)etc. We also introduce Computer Aided Decision Support System (CADSS) in the field of medicine and research.

- **Performance analysis and comparison with other models:**

we presented the statistical estimation to confirm the spouse of our model as compared to other models.

- **The Original case system development:**

Our developed system can be used as a practical guideline for the healthcare practitioners.

We conduct experiments used to identify the features of a Machine Learning Algorithm with a hybrid method and this hybrid method has stronger capability to predict heart disease compared to existing methods.

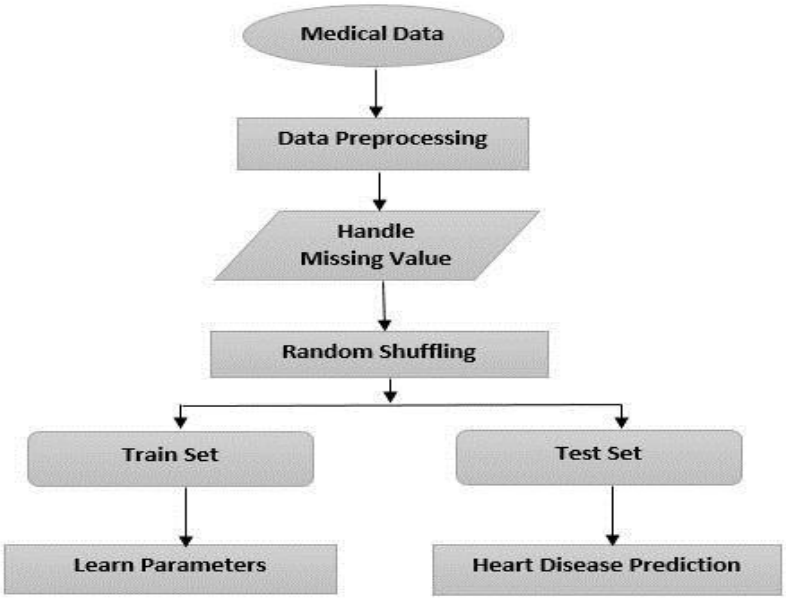

Figure 1: Machine Learning Algorithm with a hybrid method

2. STUDY REPORT

Several studies tells of the development of heart disease diagnosis based on machine learning models with the target of providing this model with advanced performance. Two available heart disease datasets such as Stat Log and

Cleveland, have used to compare the performance of prediction models among the researchers. The report says that females have less chance for heart disease compared to males. In heart diseases, correct diagnosis is main thing. But, the established approaches are insufficient for correct prediction and diagnosis. The developed model was then compared with other models such as NB, SVM and ANN.

The results revealed that the proposed model achieved the highest performance among all the models with accuracy, sensitivity, and specificity of 86.8%, 82.9%, and 92.3%, respectively. The BPNN (RS-BPNN) was proposed by Nahato et al in 2015. Verma et al in 2016 developed a hybrid prediction model based on co relation feature subset (CFS), MLP, K-means clustering and particle swam optimization (PSO). The results showed that the proposed hybrid model achieved accuracy of up to 91.28%.

Finally, we also design and develop the HDCDSS to help doctors or clinicians diagnose the patient's heart disease status based on their current condition. Thus, earlier treatment could be conducted to prevent the risks further.

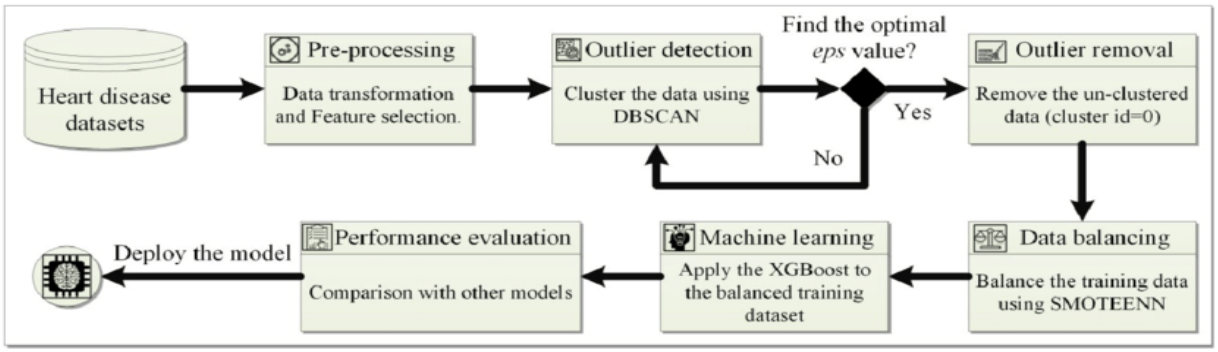

Figure 2: Architecture Diagram

3. DATASETS

The proposed Heart Disease Prediction Model was developed to provide high level performance prediction in the presence or absence of heart disease as given in the current condition of the datas. The UCI dataset is used for conducting the experiments of the proposed method, which resulted in 86.4% accuracies in the prediction of heart disease. The Probabilistic Principal Component Analysis (PPCA) method is proposed for evaluation, based on three data sets of Cleveland, Switzerland, and Hungarian in UCI respectively.

The feature selection with minimizing dimension is provided to a radial basis function. The results of the methods are 81.18%, 84.82% and 90.30% of UCI data sets of Cleveland, Switzerland and Hungarian respectively [29]. The hybrid method combining Linear regression (LR), ANN is introduced with rough set techniques and is the main contribution of this paper.

The detailed steps, including datasets and modules descriptions are presented in the following subsections. In addition, the performance of the proposed model with other models are evaluated and the results are presented in the results and discussion section. Finally, we ensure the applicability of the proposed model by converting the HDPM into the HDCDSS to diagnose heart disease status based on their condition.

Attribute	Description	Type
Age	Patient's age in completed years	Numeric
Sex	Patient's Gender (male represented as1 and female as 0)	Nominal
Cp	The type of Chest pain categorized into 4 values: 1. typical angina, 2. atypical angina, 3. non-anginal pain and 4. asymptomatic	Nominal
Trestbps	Level of blood pressure at resting mode (in mm/Hg at the time of admitting in the hospital)	Numeric
Chol	Serum cholesterol in mg/dl	Numeric
FBS	Blood sugar levels on fasting > 120 mg/dl; represented as 1 in case of true, and 0 in case of false	Nominal
Resting	Results of electrocardiogram while at rest are represented in 3 distinct values: Normal state is represented as Value 0, Abnormality in ST-T wave as Value 1, (which may include inversions of T-wave and/or depression or elevation of ST of > 0.05 mV) and any probability or certainty of LV hypertrophy by Estes' criteria as Value 2	Nominal
Thali	The accomplishment of the maximum rate of heart	Numeric
Exang	Angina induced by exercise. (0 depicting 'no' and 1 depicting 'yes')	Nominal
Oldpeak	Exercise-induced ST depression in comparison with the state of rest	Numeric
Slope	ST segment measured in terms of the slope during peak exercise depicted in three values: 1. unsloping, 2. flat and 3. downsloping	Nominal
Ca	Fluoroscopy coloured major vessels numbered from 0 to 3	Numeric
Thal	Status of the heart illustrated through three distinctly numbered values. Normal numbered as 3, fixed defect as 6 and reversible defect as 7.	Nominal
Num	Heart disease diagnosis represented in 5 values, with 0 indicating total absence and 1 to 4 representing the presence in different degrees.	Nominal

Figure 3: Data set

3.1 HEART DISEASE DATASET:

We used two heart disease datasets Stat log and Cleveland are termed as datasets I and II respectively, to investigate how heart disease can be identified by applying the machine learning technique model. The proposed model is then applied to those two datasets and with the expectation of providing this prediction model.

The original value range from 0 to 4. The 0 value is used to represent the absence of heart disease while the values from 1 to 4 are used to represent the presence of heart disease with its stage condition. In this study, we followed previous studies converting the class value from a multi-class variable to a binary-class variable.

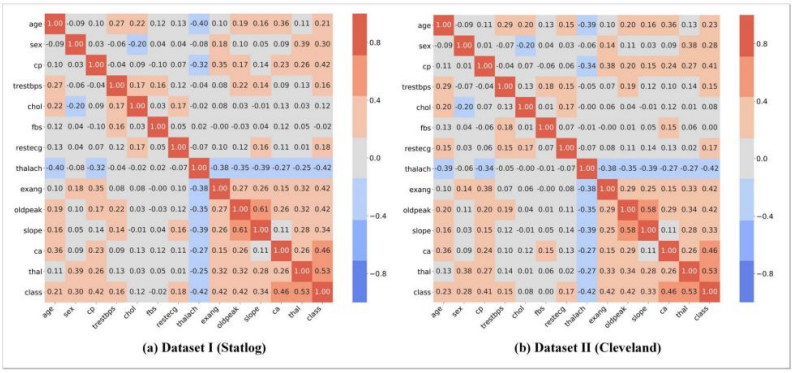

Figure 4: Comparision of Datasets

4. PROPOSED METHOD

4.1 Classification Modelling:

The bunch of datasets is done on the basis of the variables and Decision Tree (DT) features.

4.2 DECISION TREES:

These trees are simple and fast constructed in a top down recursive divide and conquer (DAC) approach. Tree pruning is performed to remove the irrelevant samples on D.

Entropy = −Xm j=1 pij log2 pij

Algorithm 1 Decision Tree-Based Partition

Require: Input: D dataset – features with a target class

for ∀features do

for Each sample do

Execute the Decision Tree algorithm

end for

Identify the feature space f1, f2, . . . , fx of dataset UCI.

end for

Obtain the total number of leaf nodes l1, l2, l3,....dn ln with its constraints (10) Split the dataset D into d1, d2, d3,.., dn based on the leaf nodes constraints.

Output: Partition datasets d1, d2, d3,..dn

5. RESULT

The proposed heart disease prediction model was applied to both datasets and showed positive results for correct predictions as compared to other models. The classification rules generated based on the rule after data pre-processing is done. After pre-processing, the data's best two techniques are chosen and the results are generated.

All the features selected in this model provides the best results.

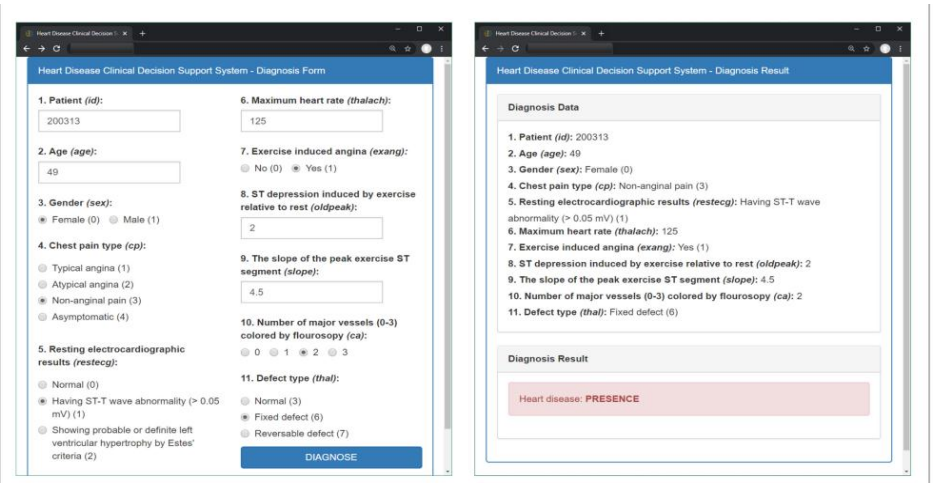

Figure 5: Result of the Proposed system

6. CONCLUSION

We proposed an effective heart disease prediction model (HDPM) for heart disease diagnosis. Identifying the process of raw healthcare data of heart information will help in the saving of human lives and early detection of abnormalities in heart conditions and also Machine learning techniques were used in this work to work on raw data and provide a new and new perception towards the heart disease. Further, with the concerns about privacy, security, many concepts could be further studied with the goal of improving the medical clinical decision support system. For this model, we have not got any feedback from heart specialist yet. In the future, dataset and comments are collected from heart specialist for verifying dataset and prediction model could be presented.

REFERENCES

[1] World Health Organization. (2017). Cardiovascular Diseases (CVDs). [Online]. Available: https://www.who.int/health-topics/cardiovascular diseases

[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ‘‘Using PSO algorithm for producing best rules in diagnosis of heart disease,’’ in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.

[3] C. Marrocco, R. P. W. Duin, and F. Tortorella, ‘‘Maximizing the area under the ROC curve by pairwise feature combination,’’ Pattern Recognit., vol. 41, no. 6, pp. 1961–1974, Jun. 2008, doi: 10.1016/ j.patcog.2007.11.017.

[4] S. H. Jee et al., ‘‘A coronary heart disease prediction model: The korean heart study,’’ BMJ Open, vol. 4, no. 5, May 2014, Art. no. e005025, doi: 10.1136/bmjopen-2014-005025.

[5] M. Sumithra and Dr. S. Malathi, “A Novel Distributed Matching Global and Local Fuzzy Clustering (DMGLFC) FOR 3D Brain Image Segmentation for Tumor Detection”, IETE Journal of Research, doi.org/10.1080/03772063.2022.2027284, 2021

[6] B.Buvanswari and T.Kalpalatha Reddy, “A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences” International Conference on Enenrgy, Communication,Data analytics and SoftComputing(ICECDS),CFP17M55-PRJ:978-1-5386-1886-8”,August 2017.

[7] Chethana, C., Subbiah Swaminathan, S. Sharanyaa, E. Sathish, R. Prathipa, and Anuradha Thakare. "Application Of Reverse Engineering in the Process of Utilization of Human Brain in Artificial Intelligence." Journal of Optoelectronics Laser 41, no. 3 (2022): 89-93.

[8] M. Sumithra and Dr. S. Malathi, " Modified Global Flower Pollination Algorithm-based image fusion for medical diagnosis using computed tomography and magnetic resonance imaging", International Journal of Imaging Systems and Technology, Vol. 31, Issue No.1, pp. 223-235, 2021

[9] K. Sridharan , and Dr. M. Chitra "SBPE: A paradigm Approach for proficient Information Retrieval , Jokull Journal" , Vol 63, No. 7;Jul 2013

[10] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques." 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020.

[11] M. Sumithra and Dr. S. Malathi, "3D Densealex NET Model with Back Propagation for Brain Tumor Segmentation", International Journal OfCurent Research and Review, Vol. 13, Issue 12, 2021.

[12] B.Buvaneswari and Dr.T. Kalpalatha Reddy,"EEG signal classification using soft computing techniques for brain disease diagnosis",Journal of International Pharmaceutical Research ,ISSN : 1674-0440,Vol.46,No.1,Pp.525-528,2019.

[13] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease." In Inventive Communication and Computational Technologies, pp. 33-48. Springer, Singapore, 2022.

[14] K. Sridharan , and Dr. M. Chitra "Web Based Agent And Assertion Passive Grading For Information Retervial", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 16, September 2015 pp:7043-7048

[15] M. Sumithra and Dr. S. Malathi, "Segmentation Of Different Modalitites Using Fuzzy K-Means And Wavelet ROI", International Journal Of Scientific & Technology Research, Vol. 8, Issue 11, pp. 996-1002, November 2019.

[16] Sharanyaa, S., S. Lavanya, M. R. Chandhini, R. Bharathi, and K. Madhulekha. "Hybrid Machine Learning Techniques for Heart Disease Prediction." International Journal of Advanced Engineering Research and Science 7, no. 3 (2020).

[17] M. Sumithra and S. Malathi, " A Survey of Brain Tumor Segmentation Methods with Different Image Modalitites", International Journal of Computer Science Trends and Technology (IJCST) – Vol. 5 Issue 2, Mar – Apr 2017

[18] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "High Performance Hybrid Cognitive Framework for Bio-Facial Signal Fusion Processing for the Disease Diagnosis", Measurement,ISSN: 0263-2241, Vol. 140, Pp.89-99,2019.

[19] Sharanyaa, S., and M. Shubin Aldo. "Explore places you travel using Android." In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 4796-4799. IEEE, 2016.

[20] M. Sumithra and Dr. S. Malathi, "A Brief Survey on Multi Modalities Fusion", Lecture Notes on Data Engineering and Communications Technologies, Springer, 35, pp. 1031-1041,2020.

[21] Sharanyaa, S., and Madhumitha RP. "Eyeball Cursor Movement Detection Using Deep Learning." RP, Madhumitha and Rani. B, Yamuna, Eyeball Cursor Movement Detection Using Deep Learning (July 12, 2021) (2021).

[22] M. Sumithra and S. Malathi, "A survey on Medical Image Segmentation Methods with Different Modalitites", International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.

[23] B.Buvaneswari and Dr.T. KalpalathaReddy,"ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial",International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091,Vol.8,No.1,Pp. 12-17,2019

[24] K. Sridharan , and Dr. M. Chitra , Proficient Information Retrieval Using Trust Based Search On Expert And Knowledge Users Query Formulation System, Australian Journal of Basic and Applied Sciences, 9(23) July 2015, Pages: 755-765.

[25] Sharanyaa, S., and K. Sangeetha. "Blocking adult account in osn's using iterative social based classifier algorithm." International Journal of Scientific Engineering and Science 2, no. 1 (2018): 33-36.

[26] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "ACPT- An Intelligent Methodology for Disease Diagnosis",Journal of Advanced Research in Dynamical and Control Systems,ISSN : 0974-5572,Vol.11,No.4,Pp.2187-2194,2019.

[27] Sumithra, M., Shruthi, S., Ram, S., Swathi, S., Deepika, T., "MRI image classification of brain tumor using deep neural network and deployment using web framework", Advances in Parallel Computing, 2021, 38, pp. 614–617.

[28] K. Sridharan , and Dr. M. Chitra "RSSE: A Paradigm for Proficient Information Retrieval using Semantic Web" , Life Science Journal 2013;10(7s), pp: 418-425

[29] Sharanyaa, S., S. Vijayalakshmi, M. Therasa, U. Kumaran, and R. Deepika. "DCNET: A Novel Implementation of Gastric Cancer Detection System through Deep Learning Convolution Networks." In 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 1-5. IEEE, 2022.