Heart Disease Diagnosis Using Hybrid Machine Learning Model in Layered Classification Approach

Tarun Kumar Agrawal[1], Dharmendra Gupta[2]

[1]Research Scholar (Ph. D), Vivekananda Global University Jaipur, Raj. India
[2]Asst. professor, Compucom Institute of Technology and Management Jaipur, Raj. India

Abstract: ***In recent decades, the incidence of heart disease has increased due to increasingly stressful lives and careless human behaviour, making it one of the major causes of mortality globally. Accurate and prompt prediction of heart disease is critical to successful prevention. Many techniques have been developed to help healthcare practitioners forecast this illness, but each algorithm has its own set of limitations. This research introduces a novel ensemble approach that improves forecast accuracy while lowering false alarms. The suggested technique was evaluated against both classic and new algorithms, and the findings show a considerable improvement in prediction accuracy.***

Keywords: AHA, HRV, IHDPS, SVM, DT, NB, K-NN, MLP, CFS, BFS

1.Introduction

Heart disease has been the top cause of mortality globally over the past many years. Researchers used a many-sidedness of data mining approaches to help healthcare practitioners diagnose this illness. With the healthcare business producing massive volumes of patient data, biomedical datasets are critical for identifying hidden patterns and linkages. These datasets are frequently analysed using intelligent healthcare information systems, which aid in the discovery of useful insights. The objective is to increase service quality by correctly identifying ailments and offering appropriate treatments at a reasonable price. When dealing with heart disease patient databases, clinicians may assign weight to factors that have a major influence on illness prognosis, allowing for more informed decisions. Heart disease risk factors include poor nutrition, smoking, and stress, lack of exercise, drug misuse, high blood pressure, alcohol, high blood sugar and cholesterol. Fatty meals damage veins, and increasing arterial pressure hardens heart walls, potentially leading to restricted blood flow and heart disease.

The human heart is a smooth muscle that uses regular contractions to pump blood via its arteries. Adults weigh between 250 and 350 grammes and their hearts consist of four chambers. The heart, which is mostly made up of cardiac muscle that operates automatically, is essential for blood circulation. Over a 66-year lifespan, the heart will beat at an average rate of 72 beats per minute and pump around 4.7-5.7 litres of blood every minute.

The American Heart Association (AHA) identifies the main risk factors for coronary artery disease as:

- Hypertension increases the burden on the heart and can cause thickening of the heart muscle.
- High cholesterol can result in plaque buildup in the arteries, which increases the risk of heart disease.
- Smoking damages the lining of the arteries and contributes to plaque accumulation.
- Diabetes significantly elevates the risk of heart disease due to high blood sugar levels, which can harm blood vessels.
- Excess weight places additional strain on the heart and is often linked with other risk factors such as hypertension and diabetes.
- Physical inactivity is associated with several risk factors, including obesity, high blood pressure, and elevated cholesterol levels.
- A diet rich in saturated fats, trans fats, cholesterol, and salt increases the likelihood of developing heart disease.
- Men are more likely to experience heart disease at a younger age, while the risk for women rises after menopause.
- A family history of heart disease increases your risk.
- Chronic stress can lead to behaviors and factors that heighten the risk of developing heart disease.

2.Literature Review

Sellappan Palaniappan and his team developed an Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques such as Naive Bayes, Neural Networks, and Decision Trees. The system highlighted the unique advantages of each method in meeting the mining objectives. Unlike traditional decision support systems, IHDPS could handle complex queries and uncover significant patterns and relationships between medical factors related to heart disease. It is web-based, user-friendly, scalable, reliable, and expandable.

In another study [3], the authors introduced a novel prediction mechanism. They developed a multi-parametric HRV feature set using statistical and classification techniques. Experiments were conducted on HRV indices, testing several classifiers, including Bayesian classifiers, CMAR, C4.5 and SVM [4, 5].

In [14], S. Silvia Priscila and M. Hemalatha presented an effective categorisation strategy to diagnosis cardiac illness by categorising ECG signals into distinct rhythms, such as Normal, VE, VF, LBBB, and APB. The procedure starts with preprocessing the ECG data with a Morphological Filter to eliminate unnecessary baseline drift. Temporal, statistical, and morphological characteristics are then retrieved from several intervals, including the RR, TT, and

PR intervals. A Real Coded Genetic Algorithm reduces feature dimensionality, and the chosen features are categorised using the suggested PBNN network, which efficiently identifies arrhythmias. The scientists used the MIT-BIH arrhythmia database, which has 250 recordings, to categorise cardiac arrhythmia illnesses. The database contains 109, 000 beat labels for normal and atypical cardiac situations. The system's performance was tested using accuracy, sensitivity, and specificity.

SVM beat all other classifiers in the research. Niti Guru et al. [6] advocated utilising neural networks to predict heart disease, blood pressure, and diabetes. The system was evaluated and trained on a sample database including 13 input variables, including age, blood pressure, and angiography results. For diagnosis, a supervised neural network with a backpropagation algorithm was utilised, with potential illnesses identified through comparisons to trained data. Furthermore, in the article "Predicting Survival Causes After Out of Hospital Cardiac Arrest Using Data Mining Method" by Franck Le Duff et al. [7], a decision tree was constructed utilising a patient database to solve a medical issue. A decision tree may be swiftly created for any medical treatment or problem utilising data from a service or physician. Comparing conventional and data mining analyses revealed the advantages of data mining for selecting factors and determining their influence on the condition under consideration. A major drawback of the method is the difficulty in gathering enough data to construct an effective model. Kiyong Noh et al. [8] suggested an associative classification strategy for detecting cardiovascular illness that employs HRV (Heart Rate Variability) from ECG data, as well as data preparation and pattern analysis. Their investigation, which included a dataset of 670 patients separated into normal and heart disease groups, used an associative classifier.

Carlos Ordonez investigated limited association rules for predicting heart disease using a dataset that included medical records, risk variables, heart perfusion measures, and arterial constriction. To simplify the rule set, he imposed three constraints: attributes can only appear on one side of the rule, attributes are classified into uninteresting categories, and the number of attributes in a rule is restricted. These limitations drastically decreased the number of rules and execution time. The final criteria indicated the presence or absence of heart disease in four particular arteries. Data mining techniques can aid physicians in predicting patient survival and adjusting treatments. Boleslaw Szymanski presented a unique heuristic for efficient sparse kernel computing in SUPANOVA, which was evaluated on the Boston housing market dataset, as well as to improve heart disease identification utilising a new non-invasive magnetic field measurement of heart activity. This technique outperformed Support Vector Machines and related kernels, achieving 83.7% accuracy. The spline kernel also performed well on the Boston dataset.

Latha Parthiban et al. suggested a method for predicting cardiac illness that combines neural network adaptability, fuzzy logic, and genetic algorithms. Based on its training performance and accuracy, the CANFIS model showed promise for identifying cardiac disease. In his empirical investigation, John Peter looked at several classification algorithms for heart disease prediction, such as DT, NB, K-NN, and NN. Classification took longer due to the vast number of data; thus dimensionality was decreased using attribute selection methods before using several classification algorithms. With the CFS attribute selection technique, the Naive Bayes (NB) classifier had the highest accuracy in heart disease prediction. Maryam Tayef et al. created a prediction model for coronary heart disease using a decision tree algorithm based on a case-control study of patients at Ghaem Hospital in Mashhad, Iran. They employed 11 CHD risk variables to train the algorithm on 1, 640 records before evaluating it on 706 other records. The decision tree, which was pruned and used the Gini index for variable selection, attained a 94% accuracy rate by combining serum hs-CRP and other risk indicators.

3.Proposed Work

To enhance heart disease prediction accuracy, the proposed approach uses an ensemble classification technique combining Naïve Bayes and Multilayer Perceptron (MLP). Classification algorithms select the best hypothesis to fit observations, and data classification involves a two-step process: building a classifier (learning phase) by analyzing a training set, and then using it for prediction. Figure 1 illustrate the process flow of the proposed approach.

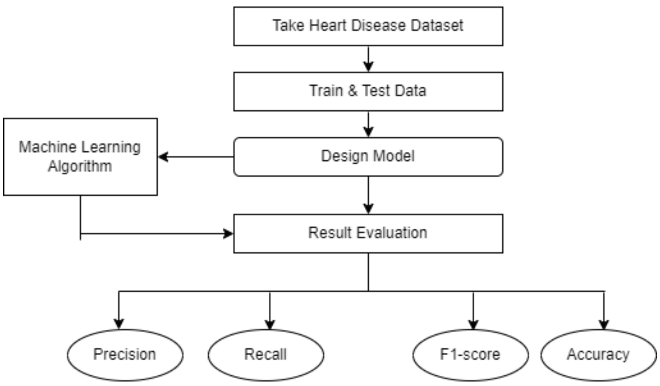


Figure 1: Proposed Flow Chart

The suggested system selects appropriate characteristics and applies Naïve Bayes for initial illness prediction. If the prediction is true, the data is updated by eliminating the signature; otherwise, a neural network is used for a second try. The procedure compares outcomes to current techniques using a tiered approach, with each layer employing a unique prediction algorithm. The first layer employs Naïve Bayes with a limited feature set. Correctly predicted data is deleted from the dataset. The leftover data is employed in the following layer for additional prediction, which improves total accuracy through this tiered approach.

The processing phases of the suggested technique can be described as follows:

- Take heart disease and normal person datasets from the UCI repository.
- Use 10-fold cross-validation to divide the dataset into training and testing sets.
- Select the best feature set by combining CFS and BFS.
- Analyze data with the Naïve Bayes prediction algorithm.

- If the prediction is right, delete the data from the dataset.
- If erroneously predicted, send it to the MLP for additional analysis.
- Repeat steps 3 through 5 until the full dataset has been analyzed.
- Determine the precision, recall, f1-score and accuracy of the procedure utilized.

4.Experimental Results

Extensive simulations were performed on three distinct datasets: Statlog (Heart), Z-Alizadeh Sani, and Hungarian to compare the performance of the proposed approach to traditional and contemporary approaches.

Table 5.1 Dataset Used for Experiments

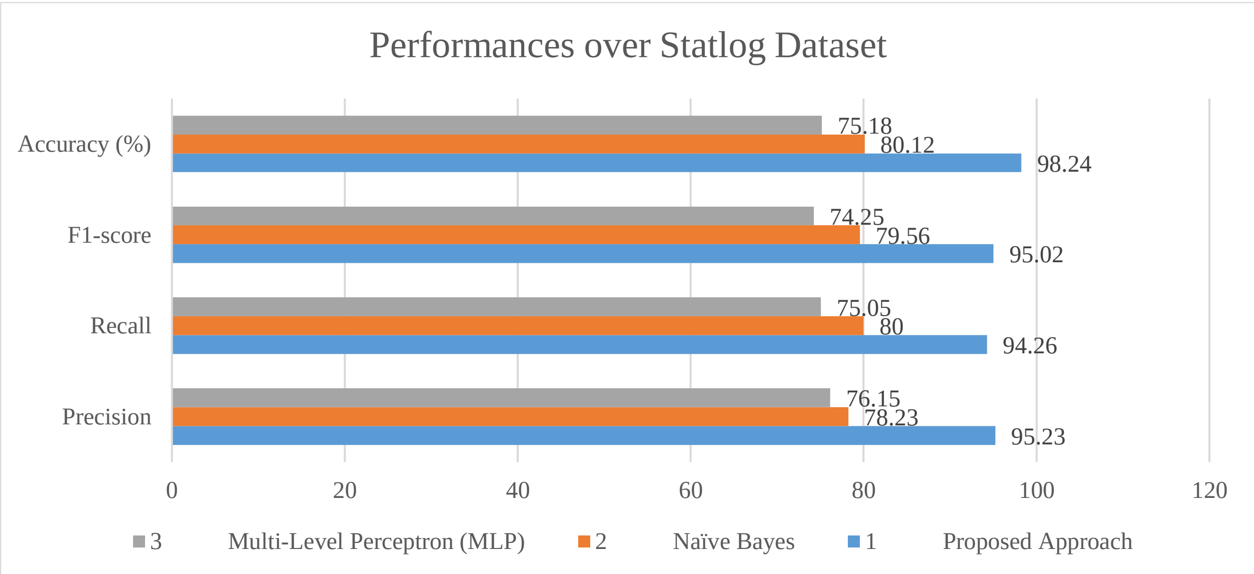
S. No.	Data Collection	Attributes	Instances
1.	Statlog (Heart) dataset	14	270
2.	Hungarian dataset	14	294
3.	Cleveland dataset	14	303

A. Performance Analysis. Using the Statlog (Heart) dataset

This experiment phase employed the Statlog dataset, which has 14 characteristics and 270 occurrences. Each instance reflects a person's record, whether they have heart disease or are normal. The collection comprises records for 120 persons with heart disease and 160 healthy people.

Comparative Performance Analysis of the Statlog Dataset

S. No.	Techniques	Precision	Recall	F1-score	Accuracy (%)
1	**Proposed Approach**	**95.23**	**94.26**	**95.02**	**98.24**
2	Naïve Bayes	78.23	80.00	79.56	80.12
3	Multi-Level Perceptron (MLP)	76.15	75.05	74.25	75.18

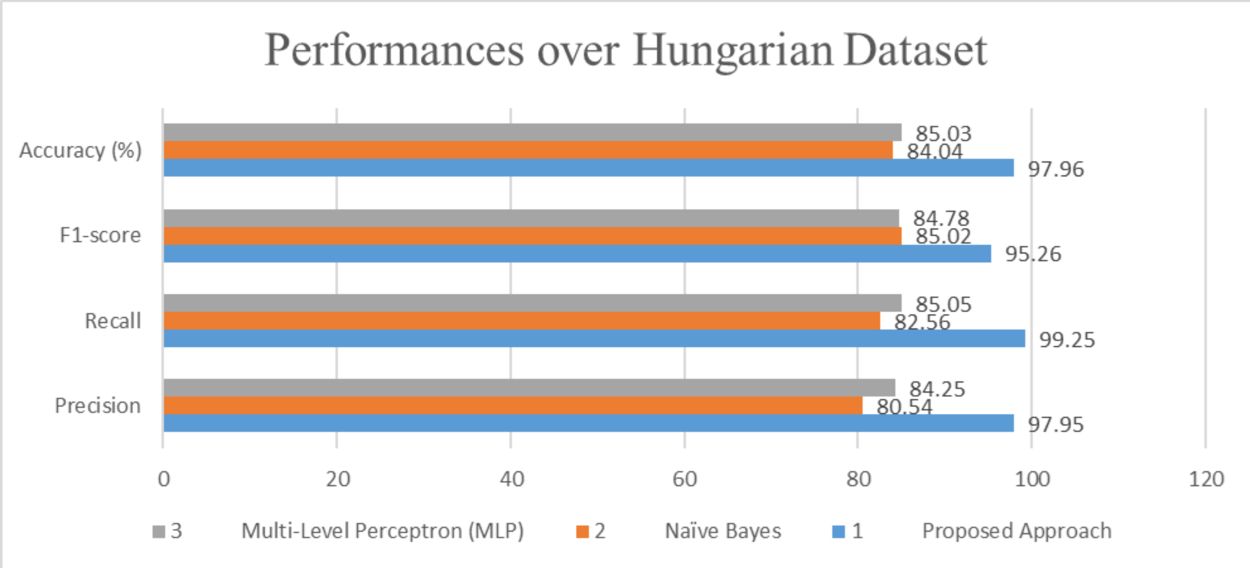


B. Performance Analysis using Hungarian and Cleveland Datasets.

A further experiment was carried out utilising the Hungarian and Cleveland datasets to test the suggested approach's efficiency with varied parameters. Table 3 compares the results, which are comparable to the prior examination. The data demonstrate that, as previously stated, the suggested prediction strategy outperforms independent prediction techniques, indicating its real-time efficiency in data prediction.

Comparative Performance Analysis of the Hungarian Dataset

S. No.	Techniques	Precision	Recall	F1-score	Accuracy (%)
1	**Proposed Approach**	**97.95**	**99.25**	**95.26**	**97.96**
2	Naïve Bayes	80.54	82.56	85.02	84.04
3	Multi-Level Perceptron (MLP)	84.25	85.05	84.78	85.03



5.Conclusion & Future Work

This study developed a hybrid data prediction approach that combines the strengths of Naïve Bayes and Multilayer Perceptron algorithms, with an attribute selection procedure. The approach's usefulness and efficiency were shown using multiple datasets from the UCI repository, and all findings were good. Future research might focus on shortening the evaluation time to improve the method's performance.

References

[1] Humar Kahramanli, Novruz Allahverdi "Design of a hybrid system for the diabetes and heart diseases", ELSEVIER, Expert Systems with Applications 35 (2008). pp. no.82-89.

[2] Anil Jain and Douglas Zongker "Feature Selection: Evaluation, Application, and Small Sample Performance" IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol.19, No.2, February 1997. pp. no.153-158.

[3] Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV, " LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp.56-66, May 2007.

[4] Li, W., Han, 1., Pei, 1. : "CMAR: Accurate and Efficient Classification Based on Multiple Association Rules". In: Proc. of 2001 International Conference on Data Mining, 2001.

[5] Mrudula Gudadhe, Kapil Wankhade and Snehlata Dongre "Decision Support System for Heart Disease based on Support Vector Machine and Artificial Neural Network" International Conference on Computer and Communication Technology (ICCCT), IEEE.2010. p. p no. -741-745.

[6] Niti Guru, Anil Dahiya and NavinRajpal "Decision support system for heart diseases prediction using neural networks" Delhi Business Review Vol.8, No.1, PP 1-6 (January-June 2007).

[7] Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol.107, No. Pt 2, pp.1256-9, 2004.

[8] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol: 345, pp: 721-727, 2006.

[9] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules, " Seminar Presentation at University of Tokyo, 2004.

[10] Boleslaw Szmanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel, Lijuan Zhu, "Using Efficeint Supanova Kernel For Heart Disease Diagnosis". Proc. ANNIE 06, intelligent engineering systems through artificial neural networks, vol.16, pp: 305-310, 2006.

[11] Latha Parthiban and R. Subramanian "Intelligent heart Diseases Prediction System Using CANFIS and Genetic Algorithm" International Journal of Biological, Biomedical and Medical Sciences 3.3.2008.

[12] John Peter "An Empirical Study On Prediction Of Heart Disease Using Classification Data Mining Techniques" IEEE-International Conference On Advances In Engineering, Science And Management PP.514-517 (ICAESM-2012) March 30, 31, 2012.

[13] Maryam Tayefia, Mohammad Tajfard, Sara Saffar, Parichehr Hanachi, Ali Reza Amirabadizadeh, Habibollah Esmaeily, Ali Taghipour, Gordon A. Ferns, Mohsen Moohebati, Majid Ghayour-Mobarhan "hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm" Computer Methods and Programs in Biomedicine, ELSEVIER, 141 (2017) pp. no-105-109

[14] S Silvia Priscila, M Hemalatha "Diagnosis of heart disease with particle bee-neural network" Special Section: Computational Life Sciences and Smarter Technological Advancement. Biomed Res-India 2017 Volume Special Issue. Pp. no-1-7

[15] Xiao Liu, Xiaoli Wang, Qiang Su, Mo Zhang, Yanhong Zhu, Qiugen Wang and QianWang —A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method, Hindawi Computational and Mathematical Methods in Medicine Volume 2017, pp. no.1-11.