International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

A Multi-Model Fusion Framework for Cardiovascular Risk Prediction

Dr. Meghna Utmal[1], Sakshi Singh[2], Kunti Uikey[3], Vaishali Gupta[4], Sajal Pandey[5]
Prof. & Head MCA[1] , Students Dept. of MCA [2,3,4,5]
Gyan Ganga Institute of Technology & Sciences, Jabalpur[1]
Gyan Ganga College of Technology, Jabalpur [2,3,4,5]

Abstract- — Heart disease remains a major health concern worldwide, affecting a large proportion of the global population. According to reports by the World Health Organization (WHO), approximately 17.9 million deaths occur annually due to cardiovascular diseases. In the context of the COVID-19 pandemic and its post-infection complications, cardiac failure has emerged as a commonly observed condition, highlighting the critical need for early diagnosis and prediction of heart disease to enable effective prevention. Timely detection can significantly reduce mortality rates. Recent advancements in machine learning techniques have greatly contributed to the healthcare sector, particularly in the prediction of heart diseases, thereby saving numerous lives. This paper presents an efficient ensemble-based machine learning approach for predicting heart-related disorders, achieving an accuracy of 88.52%.

Keywords: machine learning, prediction, heart disease, classification.

I. INTRODUCTION

Heart disease is among the most prevalent medical conditions today and poses a significant threat to human longevity. It accounts for approximately 17.5 million deaths worldwide each year. As the heart plays a vital role in sustaining life, its proper functioning is essential for survival, and any impairment can lead to serious health consequences. Heart disease refers to a range of conditions that negatively affect the normal functioning of the heart. Assessing an individual's risk of developing coronary heart disease is an important aspect of both preventive healthcare and clinical practice. Risk prediction models are commonly developed using multivariate regression analysis on data obtained from longitudinal studies. With the rapid advancement of digital technologies, healthcare institutions are accumulating vast volumes of complex medical data that are challenging to analyze using traditional methods. Consequently, data mining techniques and machine learning algorithms have become indispensable tools in medical research and healthcare analytics. These techniques can be applied directly to large datasets to build predictive models and extract meaningful patterns and insights.

Several demographic, clinical, and lifestyle factors contribute to the risk of heart disease, including age, sex, and gender. Clinical indicators such as type of chest pain, resting blood pressure, resting electrocardiogram (ECG) results, number of major vessels identified through fluoroscopy, ST-segment depression, chest pain location, maximum heart rate achieved (thalach), exercise-induced angina, tobacco consumption, and fasting blood sugar levels are significant predictors. Additionally, conditions and lifestyle factors such as hypertension, dietary habits, body weight, height, and obesity play an important role in the development of heart disease.

.

Figure 1: Leading Cause of Death (Source: American Heart Assoc.)

As per a survey by American heart association it has been found that out of 10 most dangerous disease the major leading cause of death is due to heart related disorders and around 23.4 % of people lose their lives in America due to heart disease.

Machine Learning

Machine learning is used to model human learning in computers and to familiarise ourselves with real-world knowledge. "A computer programme is said to learn from experience E with respect to particular classes of tasks T and

measurement of performance P, if its performance in tasks T, as measured by P, improves with experience E," according to [1]. ML was researched as a new field of study in the 1990s, despite the fact that its first ideas appeared in the 1950s [1]. ML algorithms are employed in a variety of industries, including business[2], advertising[3], and medicine[4]. The act of obtaining knowledge is known as learning. Humans inherently learn from experience because of their ability to reason. They may or may not be graded, depending on the learning technique utilised. There are four different types of learning: supervised, unsupervised, semisupervised, and reinforcement learning.

When algorithms are given training data and accurate responses, this is referred to as supervised learning. The machine learning method is applied to the training data, and the resulting model is then used to create predictions on the testing data. The actual value is now compared to the projected value.

II. RELATED WORK

Numerous researches on the diagnosis of cardiac disease have been conducted. They used various machine learning approaches for diagnosis and obtained varying probabilities for each strategy. This part of paper includes a quick review of the literature.

K. Polaraju et al [5] proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

R. Bhuvaneswari et al. [6] use the Naive Bayes classifier for medical use. The authors used two well-known algorithms, the Back Propagation Neural Network (BNN) and the Nave Bayesian (NB) data mining classification, to study the previous experience and to calculate the probability of an object among all objects. Bayesian techniques have been developed for probability concepts.

S. Seema et al. [7] focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

P.Sai Chandrasekhar Reddy et al. [8] suggested the data mining ANN algorithm to predict heart disease. As the cost of diagnosing cardiovascular diseases has risen, a new approach for predicting cardiac diseases is needed. A prediction model can be used after an evaluation based on different parameters such as pulse rate, blood pressures, cholesterol, and so on.

Gudadhe et al.[9] proposed a method for detecting heart disease that uses an architecture comprising of SVM & a multilayer perceptron neural network. SVM model is applied after splitting dataset into training and testing for predicting heart disease. They were able to reach an accuracy of 80.41 percent.

III. PROPOSED METHODOLOGY

Figure 2 illustrates the step-by-step workflow adopted in the proposed methodology. The heart disease prediction system is implemented using Python on the Google Colab platform. For experimental analysis, a heart disease dataset obtained from the Kaggle data repository is utilized, with detailed information provided in the subsequent section. The proposed methodology begins with data preprocessing and exploratory data analysis to understand and prepare the dataset. This is followed by feature selection and partitioning of the data into training and testing sets for model development. Various supervised machine learning algorithms are then trained and their performances are evaluated individually. Finally, an ensemble-based approach employing the Random Forest algorithm is applied to develop a high-performance model for heart disease prediction.

Ensemble Methods in Machine Learning:

To enhance classification performance, ensemble learning techniques are employed, which combine multiple decision tree classifiers rather than relying on a single model. By integrating multiple weak learners, a strong learner is formed, resulting in improved accuracy and precision [10]. In this work, two ensemble strategies—Bagging and Boosting—are utilized to achieve more reliable and accurate prediction outcomes.

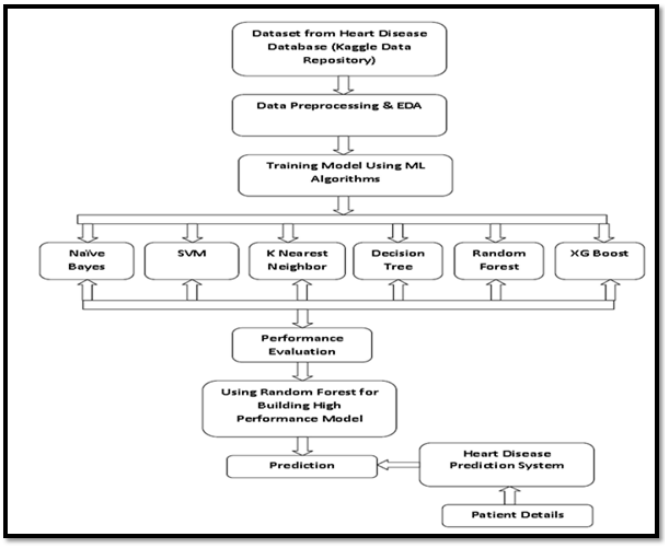

Figure 2: Workflow of our Proposed Ensemble Based Approach

Dataset

For our proposed work we have chosen dataset from well known kaggle dataset repository from where we took heart-disease dataset comprising of following 13 attributes. Link: https://www.kaggle.com/ronitf/heart-disease-uci

Table 1: Attributes of the Dataset

S.No	Attributes	Data Types	Description	Value Range
1	age	int64	Age in years	29 to 77
2	sex	int64	Gender instance	0 and 1
3	cp	int64	Chest pain type	0,1,2 and 3
4	trestbps	int64	Resting blood pressure in mm Hg	94 to 200
5	chol	int64	Serum cholesterol in mg/dl	126 to 564
6	fbs	int64	Fasting blood sugar > 120 mg/dl	0,1
7	restecg	int64	Resting ECG results	0,1 and 2
8	thalach	int64	Maximum heart rate achieved	71 to 202
9	exang	int64	Exercise induced angina	0,1
10	oldpeak	float64	ST depression induced by exercise relative to rest	0 to 6.2
11	slope	int64	the slope of the peak exercise ST segment	0,1 and 2
12	ca	int64	number of major vessels (0-3) colored by flourosopy	0 to 4
13	thal	int64	Defect types	0 to 3
14	target	int64	Diagnosis of heart disease	0,1

The initial phase involves data preprocessing and exploratory data analysis, during which missing values are identified using heat maps. Heat maps visually represent tabular data, where fully colored columns indicate the absence of missing values. If all cells corresponding to a particular attribute are completely filled, it confirms that no missing data exist for that feature. Subsequently, the age distribution of patients is analyzed. As illustrated in Figure 3, the results indicate that a significant proportion of patients affected by heart disease fall within the age range of 55 to 65 years.

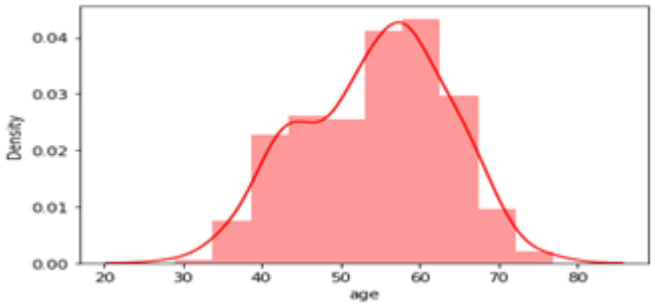

Figure 3: Bar Graph depicting age distribution

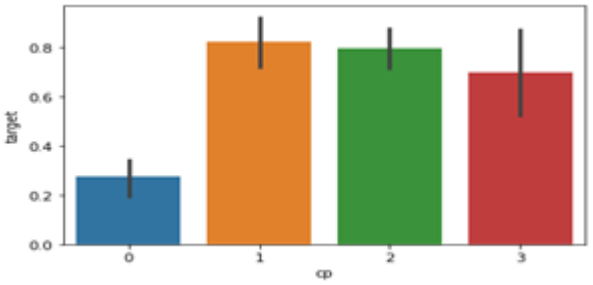

Figure 4: Bar Graph chest pain (cp) wrt target

From the figure 4 above we can interpret that chest pain of '0' that is the ones with typical angina are much less likely to have heart problems.

Now in order to apply machine learning techniques we split our dataset into Testing & Training with 80:20 where 80% of data comprise of Training data & 20% comprise of testing data.

After splitting the dataset we get the shape of our dataset as (242, 13) for Training & (61, 13) for Testing.

IV. RESULTS & PERFORMANCE EVALUATION

Perfomance Metrics Analysis

Accuracy: Accuracy is used to find the correct values; it is the sum of all true values divided by total values

(True Positive + True Positive) / (True Positive +True Negative +False Positive +False Negative)

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad \text{..... (i)}$$

Precision: How often a model predicts a positive value is correct? It is all the true positives divided by the total number of predicted positive values. True Positive/True Positive + False Positive.

$$Precision = \frac{(TP)}{(TP+FP)} \quad \text{..... (ii)}$$

Recall: It used to calculate the models ability to predict positive values. How often does the model actually predict the correct positive values? It is true positives divided by the total number of actual positive values.

True Positive/True Positive + False Negative

$$Recall = \frac{(TP)}{(TP+FN)} \quad \text{..... (iii)}$$

F-1 Score: F1 measure is used when we need to take both Precision and recall.

$$F1 = \frac{2*Precision*Recall}{(Precision+Recall)} \quad \text{..... (iv)}$$

Table 2: Comparative Analysis of ML Algorithms wrt Precision, Recall & F1 Score

Algorithm	Precision	Recall	F1 Score
Naïve Bayes	84	90	87
SVM	88	88	88
KNN	75	71	73
Decision Tree	78	93	85
XG Boost	84	87	86
Random Forest	91	88	89

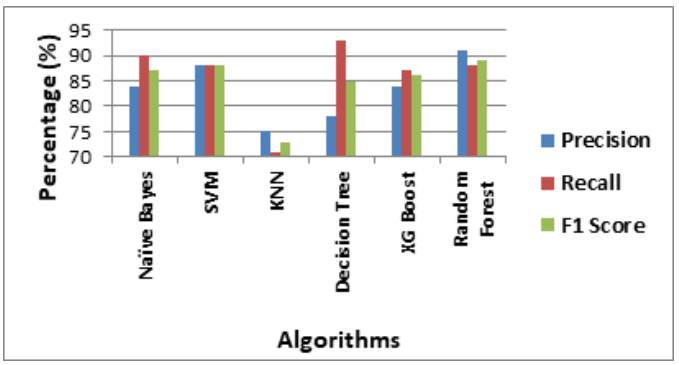

Figure 5: Bar graph depicting comparison of performance metrics in ML algorithms

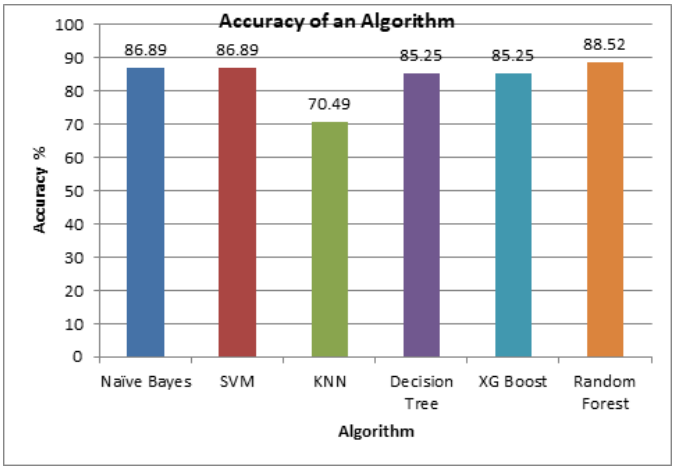

Figure 6: Comparison Bar Graph of accuracy of all Algorithms

The graph above shows that of all the proposed algorithms ensemble based approach comprising of Random Forest gives the best accuracy of 88.52 % for heart disease prediction.

V. CONCLUSIONS

In this study, various machine learning algorithms are compared to predict whether an individual is likely to develop heart disease based on personal characteristics and clinical symptoms. The primary objective of this work is to evaluate and compare the accuracy of different algorithms and to analyze the factors contributing to their performance variations. The heart disease dataset used for this research is obtained from the Kaggle data repository and consists of 303 instances. Machine learning techniques are applied to divide the dataset into training and testing subsets [5]. A total of 13 attributes are considered, and six different algorithms are implemented to

assess their predictive performance. The experimental results indicate that an ensemble-based approach utilizing the Random Forest algorithm achieves the highest accuracy of 88.52% on the given dataset. While alternative algorithms may yield better results for different datasets or problem instances, Random Forest proved to be the most effective in this case. Additionally, incorporating a larger number of attributes could potentially improve prediction accuracy; however, this would increase computational complexity, processing time, and system overhead. Taking these factors into account, the current approach was selected as the most suitable for this study.

REFERENCES:

1. Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). Machine learning: An artificial intelligence approach. Springer Science & Business Media.
2. Apte, C. (2010). The role of machine learning in business optimization. In Proceedings of the 27th International Conference on Machine Learning (ICML-10) (pp. 1-2).
3. Cui, Q., Bai, F. S., Gao, B., & Liu, T. Y. (2015). Global Optimization for Advertisement Selection in Sponsored Search. Journal of Computer Science and Technology, 30(2), 295-310.
4. Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in medicine, 23(1), 89-109.
5. K. Polaraju, D. Durga Prasad, “Prediction of Heart Disease using Multiple Linear Regression Model”, International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.
6. R. Bhuvaneswari and K. Kalaiselvi, “Naive Bayesian Classification Approach in Healthcare Applications”, International Journal of Computer Science and Telecommunications [Volume 3, Issue 1, January 2012].
7. Dr. S.Seema Shedole, Kumari Deepika, “Predictive analytics to prevent and control chronic disease”, https://www.researchgate.net/punlication/31653078 2, January 2016.
8. Mr. P.Sai Chandrasekhar Reddy, Mr.Puneet Palagi, S.Jaya, “Heart Disease Prediction using ANN Algorithm in Data Mining”, International Journal of Computer Science and Mobile Computing, April 2017, pp.168-172.
9. M. Gudadhe, K. Wankhade and S. Dongre, "Decision support system for heart disease based on support vector machine and Artificial Neural Network," 2010 International Conference on Computer and Communication Technology (ICCCT), Allahabad, Uttar Pradesh, 2010, pp. 741-745.
10. C. B. C. Latha and S. C. Jeeva, ``Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informat. Med. Unlocked, vol. 16, no. 2, 2019, Art. no. 100203.