The Diagnosis of Heart Attacks: Ensemble Models of Data and Accurate Risk Factor Analysis Based on Machine Learning

[1,2]Shaymaa Hussein Nowfal, [3]Sudhakar Sengan, [4]Joel Sunny Deol, [5]Serwes Bhatta G, [6]Saravanan V and [7]Veeramallu B
[1]Department of Medical Physics, College of Science, University of Warith Al-Anbiyaa, Karbala, Iraq.
[2]Medical Physics Department, College of Applied Science, University of Kerbala, Karbala, Iraq.
[3]Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, Tamil Nadu, India.
[4]Department of Computer Science and Engineering-Artificial Intelligence and Machine Learning, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India.
[5]McCoy College of Science, Mathematics and Engineering, Midwestern State University, Wichita Falls, TX, USA.
[6]Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.
[7]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.
[1,2]shaymaa@uowa.edu.iq, [3]sudhasengan@gmail.com, [4]sunnydeolgosu@gmail.com, [5]serwesbhatta@gmail.com, [6]saravananv.sse@saveetha.com, [7]bvmallu@kluniversity.in

Correspondence should be addressed to Sudhakar Sengan : sudhasengan@gmail.com

Article Info
Journal of Machine and Computing (https://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202505046
Received 05 May 2024; Revised from 15 November 2024; Accepted 18 December 2024
Available online 05 January 2025.

Abstract – Recent studies in clinical studies have observed a rampant increase in the rate of heart attacks, even among the newer population. Medical experts compute a multitude of factors as origins of a heart attack. But, the medical community is not able to explain the exact reasons for the prediction of heart attacks. ML algorithms are now evading the healthcare sector to assist healthcare providers in diverse ventures. This work analyses the potential causes of heart attacks among different age groups besides predicting attacks from biological conditions. The proposed ensemble model constellates the prowess of Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Random Forest (RF), and Extreme Gradient Boost (XGB) to predict heart attacks. The performance of this ML is tested on a heart attack prediction dataset, and the results promise the model's power over its peers. The proposed system achieved a classification accuracy of 92.8% for the test set in the ensemble model.

Keywords – Ensemble Learning, Extreme Gradient Boost, Heart Attack, K-Nearest Neighbors, Meta Classifier, Random Forest, SVM.

I. INTRODUCTION

According to the World Health Organization (WHO), heart failure is one of the crucial diseases causing a high mortality rate of about 17.9 million people losing their lives every year. Cardio means heart, and the blood vessels that flow throughout the body are referred to as vascular. Heart failure is considered to be persistent and continues to occur for a long time. Related Heart Disease (HD) was declared a deadly clinical illness with similar indications like shortness of breath, joint swelling in ankles, and severe pulmonary cracks in vessels expressed by the European Society of Cardiology [1]. Different factors that influence heart failures, such as blood pressure, age, coronary disease, and mainly other lifestyle diseases like diabetes and viral infections, are also responsible for sudden failure in the heart. Generally, heart failure is a chronic disease that hardly affects the body's overall health and starts complications in the kidneys, heart valve problems, and liver damage. The prediction of HD is significant nowadays since there is an increase in mental stress in cities compared to rural areas. Hence, the evaluation of heart failure can control the mortality rate. The research aims to enhance the evaluation process in predicting HD accurately compared to human prediction. The work uses the dataset extracted from Kaggle with essential features to clean the outliers.

Further early diagnosis of heart failure will help reduce the difficulty and prevent heat from entering crucial loss phases. The alarming fact is that there is no medication or reverse process for making a heart retain its original stage. Fortunately, different measures can be taken to reduce the risk factors affecting the heart.

The advent studies have proved that predicting heart failure at its early stage can reduce mortality and prime the patient to revolutionize their lifestyle at the initial step. The tremendous advancement in healthcare technologies has allowed medical practitioners to make better decisions while identifying patient diseases. In this context, predicting the failure of the heart at its initial stage plays a significant role in reducing the adverse effects of the disease. Many Machine Learning (ML) tools help predict different lifestyle diseases**.** Support Vector Machine (SVM) works for the linear model and helps with regression and classification problems. K-Nearest Neighbour (k-NN) is a supervised algorithm that works well for simple problems, whereas Random Forest (RF) is widely used for complex problems by classification.

The considerable challenge faced by the classification methods in the healthcare field is that they contain a large amount of data from various patient medication histories. Medical practitioners make the inference, forming the attributes for prediction and classification. The dataset containing the risk factors responsible for heart failure is taken as the attributes, and real-world patient data for the corresponding features are considered in the samples. Since the examples from the real world contain redundant and noisy fields, data pre-processing is performed to make data consistent and relevant to the diagnosis. The risk factors that are associated with heart failure are based on lifestyle associations that are subject to change, like diabetes, high blood pressure, obesity, good habits like the regular practice of drinking, smoking which is very hard to give up, and the risk factors that are not supposed to change are responsible for heart failure are age, gender, heredity. With all the risk factors and test results of the blood sugar level considered, medical consultants will make predictions and diagnose the patient's condition. This decision needs highly technical experts and experienced doctors to compare the results with the previous cases for a highly accurate assertion [2].

Highlights of current studies showed that ML is used to diagnose heart problems, and the classification accuracy is high [3]. Also, many computational techniques like big data and the Internet of Things (IoT) have proven to be highly effective and accurate in decision-making, aiding medical experts in diagnosing patients' diseases with more comparative analysis and predictive algorithms [4]. These studies motivated the primary focus of the work to throw the limelight on ML in predicting a person's heart failure disease [5-6]. The results from the current model have obtained 92% accuracy and showed high performance in classifying the data. A heart failure prediction system was designed to assist less experienced healthcare doctors by framing different rules to handle different input types [7]. Some of the regulations farmed are Original Rules, classical Rules without duplicates, accurate Classified Rules, and ordered Sorted Rules, which give higher accuracy and perfect decision-making. The system is validated with the manual decision, and this can be used to predict HD at the 1st level. This helps medical experts validate a patient's decision and produce a robust result for the person affected by HD.

The organization of the paper is as follows: A brief description of the literature survey is shown, highlighting the challenges and limitations of previous work carried out in recent years, a description of the dataset which contains risk factors that can be changed according to the lifestyle and which cannot be changed, various classification techniques used in this paper, in addition to this the motivation factor for ensembling the results are highlighted in this Section II. Classification techniques, namely Naive Bayes, SVM, K-NN, and RF, are implemented for the dataset, and the risk factors responsible for the failure occurrence are highlighted in Section III. The ensemble model is implemented in Section IV. The experimental results are presented to measure the model's accuracy, which is given in Section V.

II. RELATED WORKS

Different algorithms evaluate the dataset with the ML at the abstract level; since then, a comparison study has been made with micro-level analysis [8]. All the possible combinations of the classifiers were analyzed, and a best-fitting predictive model was selected for reducing the mortality rate of the dreadful disease by early detection and suggested that the analysis of Chi-square with Principal Component Analysis (CHI-PCA) integrating with RF gives results with great vertical accuracy compared with the classifiers implemented independently. In this context, the classifiers are evaluated individually, and the performance is compared and validated with the proposed model. PCA was performed to reduce the dimensionality of the data. Gradient descent is used for finding the reducing the cost function in the model. In this view, the accuracy of the classifier was affected to a more considerable extent since the amount of sample deployed in the dataset was insufficient. The amount of data used is significantly less, leading to the poor diagnosis of patient disease even though the attributes assigned are adequate.

Literature witnesses the art of using ML for heart failure prediction and assessment [9]. In this review, many factors are insisted on by predicting and assessing the presence of the disease. The results also highlighted the adverse effect of failure in the heart with the analysis of the system's stability with the associated mortality level of a particular person. The different classifiers were developed to determine the accuracy of the sample given as the input [10]. Even though a highly complicated model is designed for handling the dataset, the accuracy of the result gradually decreases with the sample used in the dataset used for the prediction and analysis of heart failure disease [11].

The Feature Selections (FS) are incorporated to reduce redundant features and improve the quality of classification results. This is implemented through a Fast Correlation-Based FS to achieve higher-quality results [12]. The resulting output is indulged in extensive classification methods like Naive Bayes, RF, and SVM and further optimized by combining bio-inspired particle swarm optimization integrated with Ant colony optimization techniques that provide higher efficiency and reliable output in this hybrid approach. Another notable feature for predicting heart disease failure is building a prediction model based on the statistics of the patients who are readmitted due to heart failure and calculating the difference between the patients admitted and readmitted based on the heart failure and other reasons in a hospital [13]. Compared with other ML metrics and other classical methods, this work shows considerable improvements from the dataset collected. The existing method works on separate algorithms like SVM, K-NN, RF, and XGB, which give separate output; however, each algorithm will have its demerits. Combining these algorithms in the proposed method will reinforce the model and give maximum accuracy compared to the individual models' output.

Real-world datasets contain noisy and replicated data. Preprocessing data is mandatory for complicated datasets to fill in the missing data, avoid redundant values, and sample the dataset for higher performance and accuracy. Many software exist to solve challenges for the lost and replicated data. Rapid Miner is an efficient software that handles missing data and eliminates irrelevant data from the dataset [14]. The preprocessed data is then tested with various algorithms for classification, and the results are compared for accuracy analysis. The result showed that naive Bayes classifiers showed better performance when compared to other classifiers. Recent ML were taken and compared based on the heart disease evaluation. The assembling model combines various classifiers that give. Integrating all the classifier models will help increase accuracy and make decisions convenient at the end of the activation model. The problem is that the model suffers from overfitting of data, and the computational time is longer when compared to other models.

III. DESCRIPTIVE STATISTICS AND EXPLORATORY DATA ANALYSIS OF HD

The genesis of evidence-based clinical practice has its roots in the descriptive statistical analysis of clinical data. The multifaceted biological elements' inherent complexity and dynamic nature are highly temporal. The underlying, latent knowledge from these physical observations will be beneficial in arriving at quantitative conclusions [15]. These conclusions lay the foundations for clinical decisions. Medical practitioners must comprehend the statistical relations and their associations among the variables. Many risk factors contribute to the development of HD [16]. As heart ailments are steadily evolving as lifestyle diseases, it is pivotal for clinicians to understand the statistics and interrelation between the factors [17].

The extensive study was conducted on a dataset of nearly 304 subjects with rich features such as exercise-induced angina, blood vessel data, cholesterol, blood pressure, diabetic condition, and maximum heart rate. The dataset consists of 304 samples taken with 14 features described in the table. The dataset contains missing values and inconsistent data, so pre-processing methods clean it.

The inferences made from this study will be helpful for clinicians and ordinary people since it delves into the correlation among various biological parameters.

Table 1. Description of Features in the Dataset

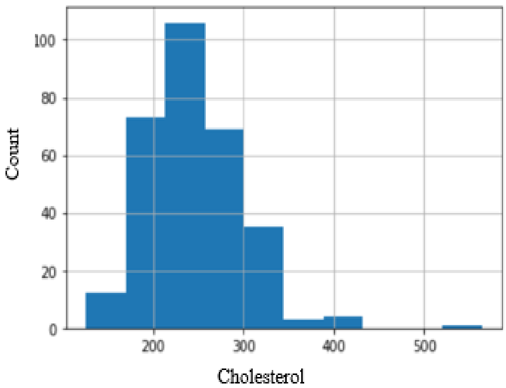
Features	Data Type	Data Range
Age	Numeric	[29,77]
Sex	Categorical	{Male, Female}
Chest Pain Type	Categorical	{0- Typical Angina, 1-Atypical Angina, 2- Non-Anginal Pain, 3-Asymptomatic}
Resting Blood Pressure	Numeric	[94, 200]
Cholesterol	Numeric	[126, 574]
Fasting Blood Sugar	Categorical	{Yes, No }
Rest ECG	Numeric	{0- Normal, 1- ST-T Wave Abnormality, 2-Probable or Definite Left Ventricular Hypertrophy}
Maximum Heart Rate	Numeric	[71, 202]
Exercise-Induced Angina	Categorical	{Yes, No}
Depression Status	Categorical	{ Yes, No}
Slope	Categorical	{Yes, No}
Number of Major Blood Vessels	Categorical	{0,1,2,3}
Thalassemia	Categorical	{1,2,3}
Target	Categorical	{Yes: More Chances of Heart Attack, No of Feeble Chances of Heart Attack}

Qualitative Assessment of Data and Preprocessing

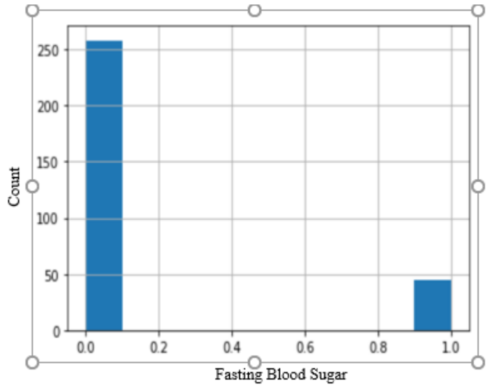
The raw medical data may contain duplications, missing values, and mismatched data types. Hence, the data is cleaned to remove missing values, and one hot encoding is used to categorize the data. Later, standard normalization is applied to the data to improve the cohesion among the data. Outlier detection is essential to pre-processing, typically performed through box plots showing out-of-boundary data [18][19].

Exploratory Data Analysis

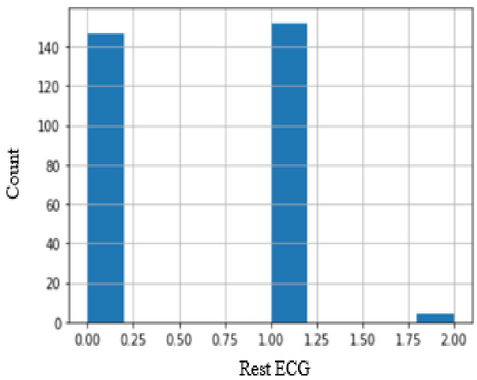
Exploratory Data Analysis (EDA) is a graphical summary of the data. It is the initial investigation of the data to detect outliers and discover underlying data patterns. The EDA over the heart attack prediction data is shown in **Fig 1**. The EDA analysis shows the versatile and robust distribution of data. Apart from this, this analysis sources significant clinical inferences. Most people in this study exhibited a blood pressure range from 130 to 135; the research was conducted on a sample population with low cholesterol and blood glucose levels. These inferences could generalize the hypothesis, serving as future research directions in clinical studies. The biological factors such as blood pressure, blood sugar, blood vessels, and thalassemia are described in **Table 1**, and the importance of using these features is compared and depicted in the graph in the result and discussion.



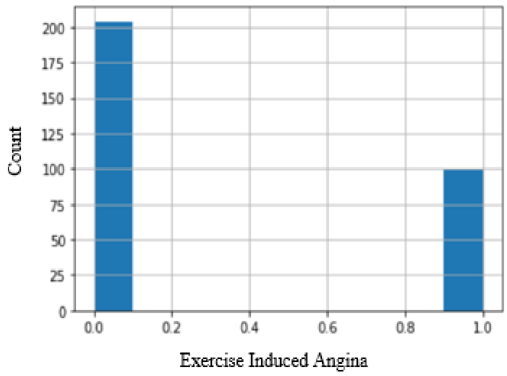
a) Cholesterol



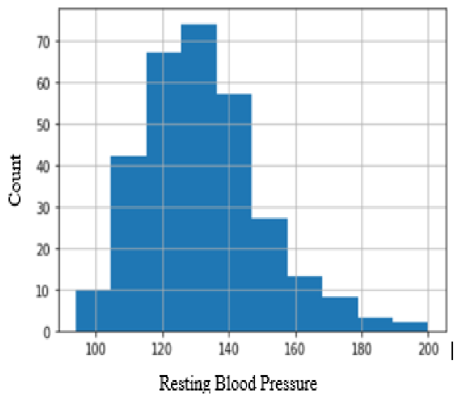
b) Fasting Blood Pressure



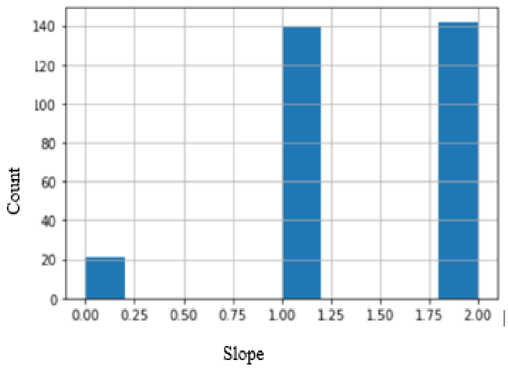
c) Rest ECG



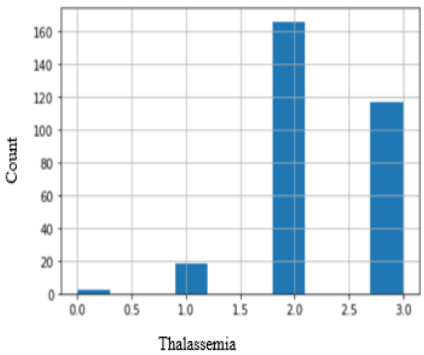
d) Angina



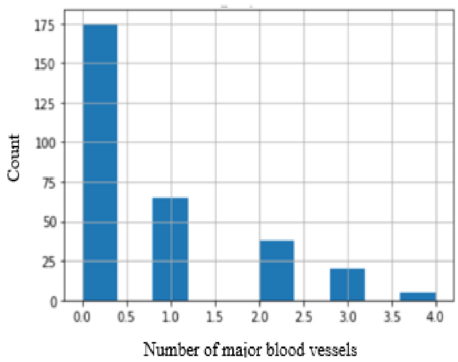
e) Blood Pressure



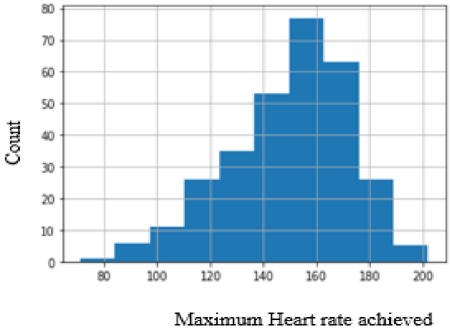
f) Slope



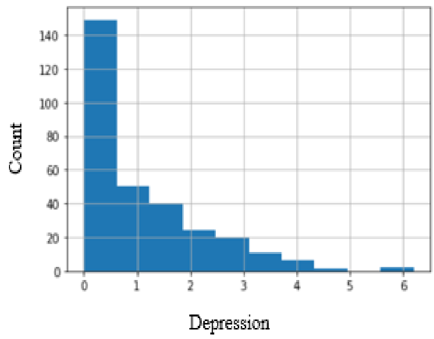
g) Thalassemia



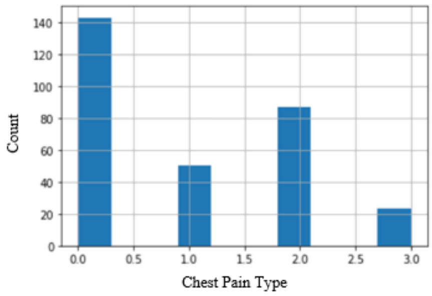
h) Flawed Blood Vessels



i) Max Heart Rate



j) Depression

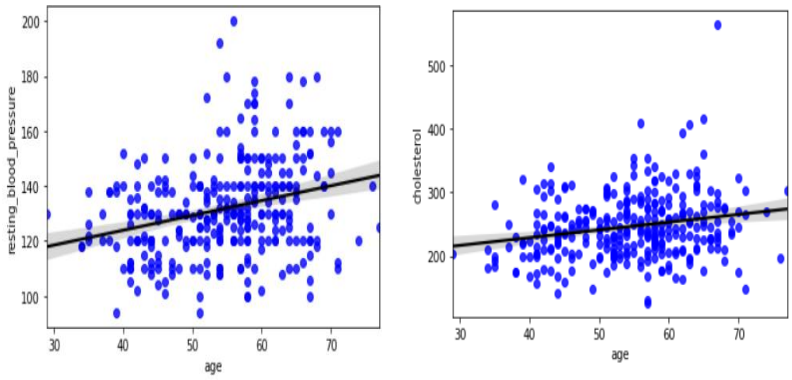


k) Chest Pain Type

Fig 1. EDA Analysis of Various Factors of Heart Attack.

Age-wise Regression Analysis

The chances of heart attack ascend along with age factors, and age and chances of heart attack are positively correlated. A more granular study of the biological factors could reveal interesting patterns. The dependencies between age and the various factors must be analyzed to determine the linear relationship. **Fig 2** shows the regression analysis of multiple characteristics of heart attacks concerning age.



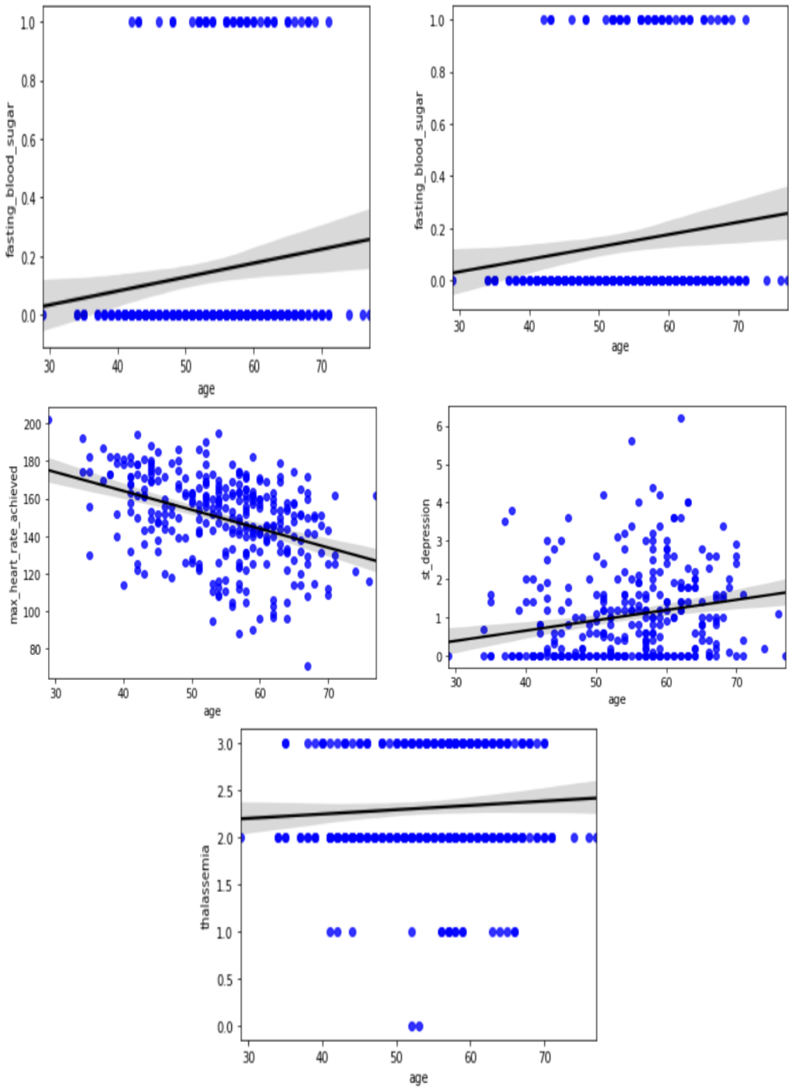


Fig 2. Regression Analysis of Age with Factors of Heart Attack.

When the blood pressure increases with age, then the chances of heart attack also increase. An important finding is that a moderate increase in blood cholesterol levels sharply increases the chances of a heart attack. People suffering from depression encounter many health ailments, among which increased chances of heart attack are more alarming. Some factors require further investigation. No profound correlation between high blood glucose levels, resting ECG, and heart attack could be found. Also, a decline in the heartbeat rate can be considered a predominant heart attack symptom.

A comprehensive plot of the occurrence of heart attacks among numerous ages is portrayed in **Fig 3**. The startling observation is that the rate of heart attacks is highly correlated with people aged 70 and above. Also, the incident of a heart attack is expected even in people with normal blood pressure, as shown in **Fig 4**.

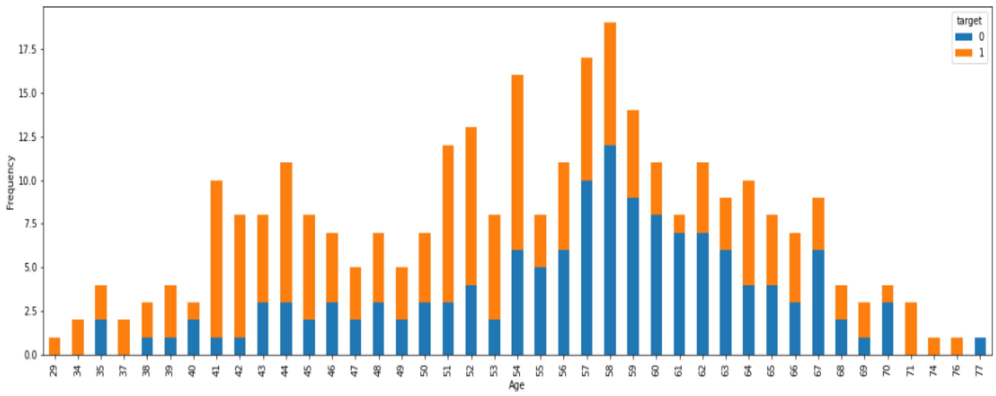


Fig 3. Incidence of Heart Attacks on Different Ages.

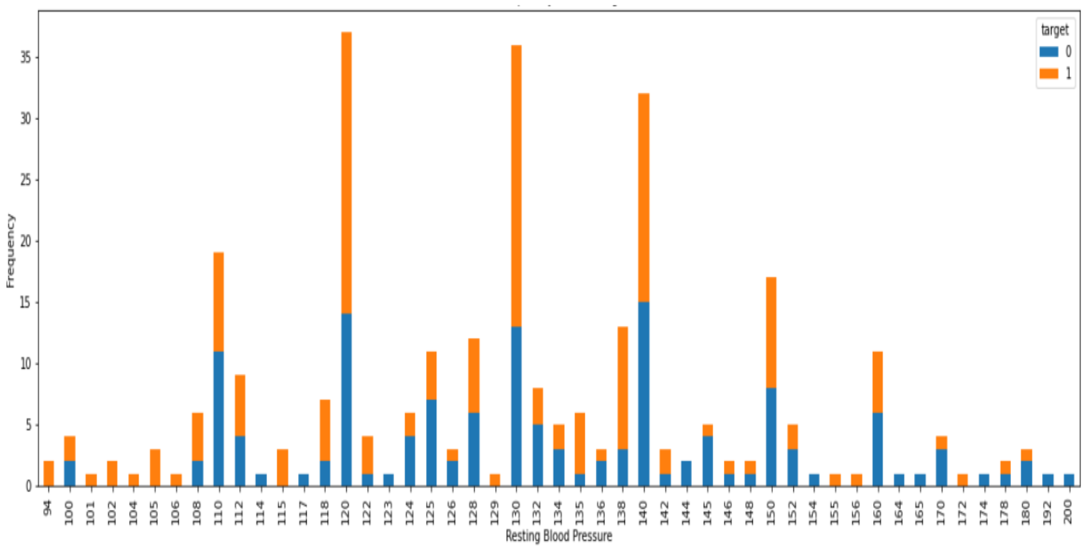


Fig 4. Distribution Of Blood Pressure Among the Patients.

Ranking The Factors

The dataset has a rich set of features with mixed associations with the incident of a heart attack **Fig 5**. Expressing the relationships on a quantitative scale of [0,1] is done using correlation analysis of the various biological factors with the occurrence of a heart attack. **Table 2** summarizes the degree of associativity exhibited by the four major factors measured in the dataset.

Table 2. Summarizes The Degree Of Associativity

Factors	Degree of Associativity
Presence of Chest Pain	0.43
Heart Rate	0.42
Slope Rate Measured in ECG	0.35
ECG	0.14

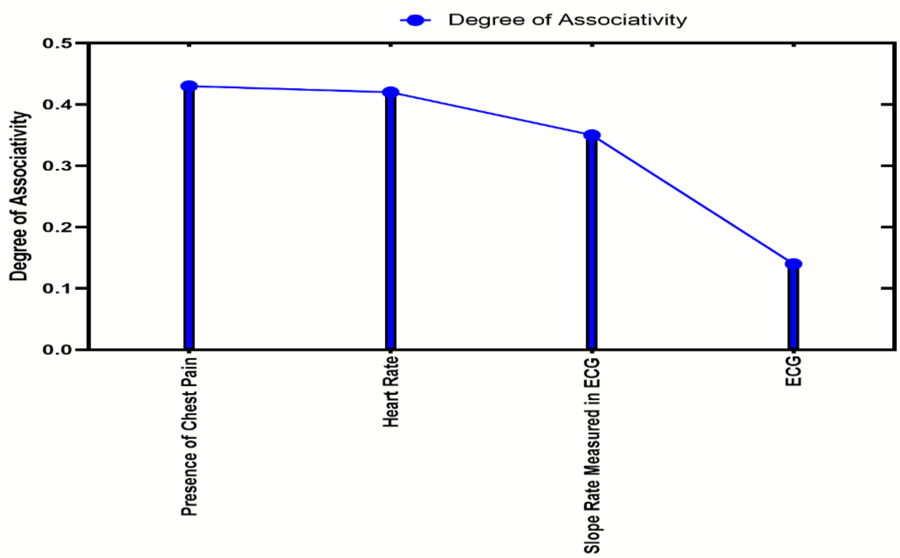


Fig 5. Factors Measured in The Dataset.

IV. A NOVEL ENSEMBLE MODEL WITH EGB AS META CLASSIFIER

Prediction of heart attacks from biological parameters is a challenging task in the medical domain. Many ML, DL, and IoT have been proposed and successfully deployed in healthcare. A novel ensemble model that merges the predictive power of base learners, namely SVM, k-NN, XGB, and RF, with XGB as meta classifiers, is used to predict the chances of a heart attack. SVM will work for high-dimensional data. The k-NN will have significantly less computational time and is suitable for simple problems. XGB produces an optimized result for the given task. RF is used to fill in the missing values in the data. These base learners show proven efficacy in predicting diseases from the health parameters.

Data Preparation

The most crucial step before applying any ML to the data is data preprocessing or data preparation. As medical data tend to be noisy, inconsistent, and highly unstructured, it is essential to enhance the quality of the data for more accurate results. The data preprocessing activities performed on the heart attack prediction dataset are detecting missing values, normalization, and checking for any duplication in data.

Dealing with Missing Values

The highly heterogeneous dataset comprises data that is both unconditional and continuous-valued. The missing values present in each of the types must be handled uniquely. The missing values in continuous-valued attributes are credited with the mean value of the corresponding point. In contrast, the missed categorical value is packed with the label with the highest frequency.

Handling Duplicate Data

The duplicate record is detected and removed from the dataset. As medical data is extracted from Electronic Health Records (EHR), which are highly unstructured, there is a higher chance of duplicate records. The presence of too much duplicate data will make the prediction biased.

Data Normalization

The biological data will be measured on different scales. Normalization smoothens the variation in the scaling without distorting the variation in data ranges. Z-score normalization is applied to the dataset.

Proposed Ensemble Model

An ensemble algorithm integrates the results of multiple homogeneous or heterogeneous-based learners. The motivation for developing ensemble algorithms is that the learning algorithms generally output a single hypothesis likely to suffer from three bottlenecks: statistical, computational, and symbolic. Algorithms affected by statistical issues exhibit high variance, while extended problems are pigeonholed by high bias. The computational bottleneck manifests high computational conflict. These effects can be effectively mitigated by deploying ensemble algorithms apart from delivering high-performance results with excellent robustness.

To predict heart attacks, the proposed ensemble algorithm combines the predictive power of four base learners: SVM, k-NN, XGB, and RF. XGB is a meta classifier that picks the meta-features from the four classifiers. The complete model is given in **Fig 6.**

The model is trained on accurate timer data, which is pre-processed. Each base learner used in the model is a robust learning algorithm. To further enhance the accuracy of the ensemble model, the XGB is used as a meta-classifier. The literature reveals that boosting algorithms are best for handling noisy data. Apart from this, the XGB can effectively mitigate the effects of bias and variance in the final prediction.

Let BL={BL_1, BL_2, BL_3, BL_4} be the base learners in the proposed ensemble model. The final predicted output (Y') is estimated according to EQU (1).

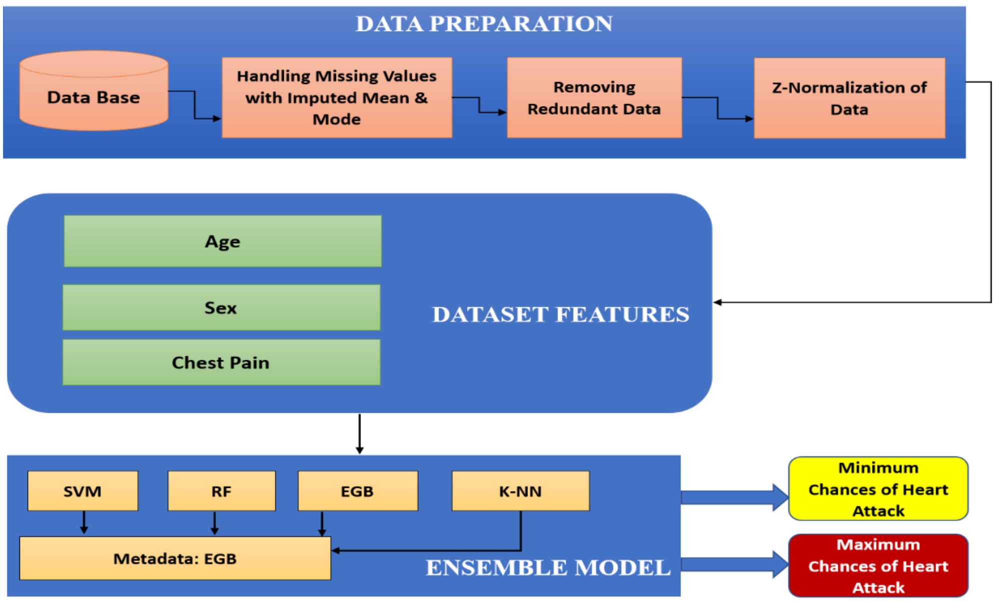


Fig 6. Ensemble Model for Heart Attack Prediction.

$$Y_i' = \sum_{i=1}^{n} bl(X_i) \tag{1}$$

X_i is the feature from the dataset, and the proposed ensemble model has four base learners who individually predict heart attack chances with consistent performance. This heterogeneous ensemble model reduces overfitting, and the meta classifier ensures that the model picks the suitable meta feature, which is the predicted output of the base learners. XGB can output more accurate approximations from the base learners by using the second-order partial derivatives to minimize the loss function, and is expressed as EQU (2). Also, the inbuilt L_1 and L_2 regularizations increase the model's generality.

$$Loss^{(t)} = \sum_{i=1}^{n} [g_i bl_t (X_i) + \frac{1}{2} w_i bl_t^{2} (X_i)] + \Omega(bl_t) \tag{2}$$

The values of g_i and w_i are the 1st and 2nd-order gradients of the loss function. These gradients are computed in EQU (3) and EQU (4).

$$g_i = \partial_{y'} (t-1) lo(y_i, y'^{(t-1)}) \tag{3}$$

$$w_i = \partial^2 y'^{(t-1)} lo(y_i, y'^{(t-1)}) \tag{4}$$

The objective function (l_o) for the computation of the 1st and 2nd-order gradient is given EQU (5).

$$lo(x) \approx lo(x_i) + lo'(x_i) bl_t (X_i) \tag{5}$$

V. EXPERIMENTS SETUP

The dataset comprises real-time data collected from 303 people in different age groups. The dataset description is given in **Table 1**. Prediction of the chances of heart attack was made by considering 13 features. The dataset is cleaned as discussed in Section IV. One hot encoding is performed on the categories like rest ECG and chest pain type before using the ensemble approach. One hot encoding ID converts categorical data into numerical data for an algorithm to better predict the model. Each numerical vector is considered to be a separate value that is added as the feature.

Performance Evaluation

The performance of the proposed model is evaluated based on the classification Accuracy, Precision, Recall, Support, and F1-score. The ensemble model shows superior performance over other models. However, to appraise the performance of the integrated ensemble model with XGB as a meta-classifier, assessing the operation of individual base learners is imperative. **Table 3** summarizes the results of the various assessments conducted during the experimentation of the proposed model.

Table 3. Performance Comparison of Classifiers with the Proposed Model

ML Classifier	Meta Classifier	Accuracy	Precision	Recall	Support	F1-Score
RF	NA	86.89	87	87	61	87
EGB	NA	83.6	84	84.5	61	84.2
K-NN	NA	88.5	89	89	61	89
SVM	NA	88.5	88	88	62	88
RF+K-NN+SVM	K-NN	89.91	89	89	61	89
EGB+K-NN+SVM	SVM	88.56	89.3	89.3	60	89.4
RF+EGB+SVM	EGB	89.67	89	89.1	61	89.0
EGB+K-NN+SVM	EGB	88.67	89.5	89.3	61	89.6
RF+K-NN+EGB	K-NN	86.76	86.1	86.5	60	86.3
SVM+RF+K-NN+SVM	SVM	91.86	91	91	61	91
SVM+RF+K-NN+SVM	RF	92.56	92.69	92.6	61	92.6
SVM+RF+K-NN+SVM	K-NN	92.98	91.97	91	61	91.4
RF+K-NN+SVM	**EGB**	**93.19**	**92.88**	**92.8**	**62**	**92.8**

It is evident from **Fig 7** that the ensemble models perform better than the base classifiers. The prediction accuracy and other supporting metrics steadily increase as more models are combined. The performance of the proposed model is investigated with different meta-classifiers. These results imply that the model outputs a better understanding when EGB is used as a meta-classifier.

The graphical comparison of the metrics of the algorithms is shown in **Fig 6**. The proposed ensemble model shows improved performance than the other state-of-the-art techniques. Also, the model's performance is assessed by training and testing with different meta-classifiers. The results demonstrate that the ensemble model with XGB as a meta-classifier gives enhanced results.

The model's accuracy is compared with the output of the manual decision taken from medical experts and is validated with many inputs and outputs. The model is designed, and the performance is validated on a heart attack prediction dataset downloaded from the Kaggle repository. The results have achieved accuracy when compared to other ML.

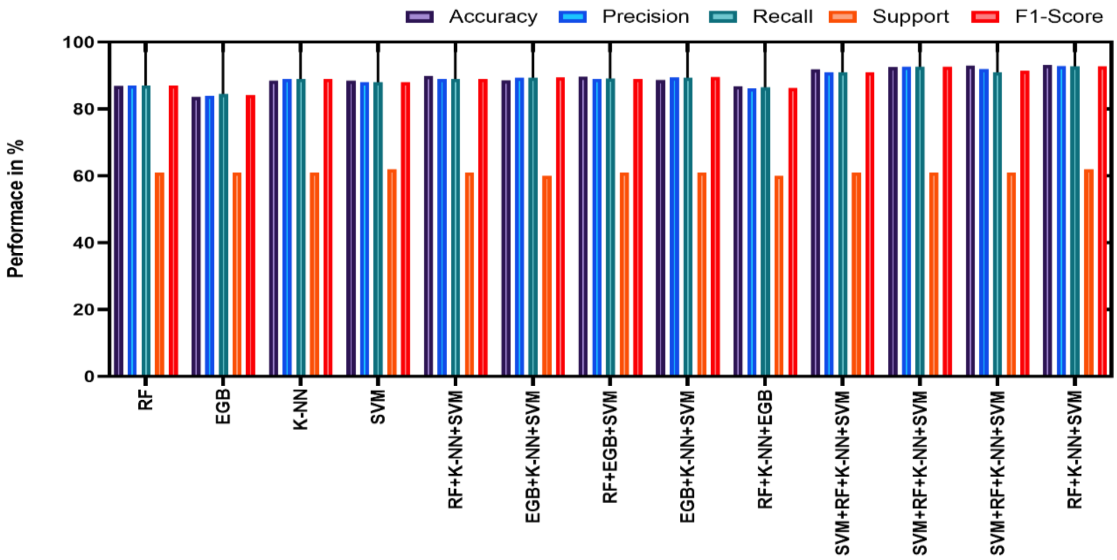


Fig 7. Comparison of Ensemble Model.

VI. CONCLUSION AND FUTURE WORK

The lifestyle change has increased the rate of heart attacks among the human population. The information about the multitude of factors that contribute to the occurrence of heart attacks is frequently misleading. Clinicians face many challenges in diagnosing the onset of heart attacks from the early signs and symptoms. The proposed system achieved a classification accuracy of 92.8% for the test set in the ensemble model. This system forecasts heart failure risk and provides proficient guidance to medical experts**.** This article assesses the correlation between various biological factors and heart attacks. Most importantly, the ensemble model with EGB as a meta-classifier is used to predict the chances of heart attack with considerable accuracy from the real-world data set. This model can help clinicians decline the mortality rate by diagnosing heart attacks from early warnings so that the patients can receive preventive treatment.

Future research directions of this work would be to test the model in the factual environment, such as medical centers, and hypothesize the relation among the biological factors contributing to heart attacks. Since the ensemble model contains a combination of many MLs, the computational time is high, leading to the model's overfitting.

Since the model has trained well in the training phase, the model suffers from overfitting rather than underfitting. It may decrease the accuracy rate during the validation of the model. Hence, future work can handle this issue for better enhancement.

CRediT Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Shaymaa Hussein Nowfal, Sudhakar Sengan, Joel Sunny Deol G, Serwes Bhatta, Saravanan V and Veeramallu B; **Methodology:** Joel Sunny Deol G, Serwes Bhatta and Saravanan V; **Software:** Shaymaa Hussein Nowfal and Sudhakar Sengan; **Data Curation:** Serwes Bhatta, Saravanan V and Veeramallu B; **Writing- Original Draft Preparation:** Joel Sunny Deol G, Serwes Bhatta, Saravanan V and Veeramallu B; **Visualization:** Shaymaa Hussein Nowfal and Sudhakar Sengan; **Investigation:** Serwes Bhatta, Saravanan V and Veeramallu B; **Supervision:** Shaymaa Hussein Nowfal and Sudhakar Sengan; **Validation:** Joel Sunny Deol G, Serwes Bhatta, Saravanan V and Veeramallu B; All authors reviewed the results and approved the final version of the manuscript.

Data Availability
No data was used to support this study.

Conflicts of Interests
The author(s) declare(s) that they have no conflicts of interest.

Funding
No funding agency is associated with this research.

Competing Interests
There are no competing interests

References
[1]. K. Zafar et al., "Deep Learning-Based Feature Engineering to Detect Anterior and Inferior Myocardial Infarction Using UWB Radar Data," IEEE Access, vol. 11, pp. 97745–97757, 2023, doi: 10.1109/access.2023.3312948.
[2]. U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak, "Deep Learning Applications in ECG Analysis and Disease Detection: An Investigation Study of Recent Advances," IEEE Access, vol. 12, pp. 126258–126284, 2024, doi: 10.1109/access.2024.3447096.
[3]. X. Wang, J. Hu, H. Lin, W. Liu, H. Moon, and Md. J. Piran, "Federated Learning-Empowered Disease Diagnosis Mechanism in the Internet of Medical Things: From the Privacy-Preservation Perspective," IEEE Transactions on Industrial Informatics, vol. 19, no. 7, pp. 7905–7913, Jul. 2023, doi: 10.1109/tii.2022.3210597.
[4]. Z. Liu, Y. Cheng, and S. Tamura, "Multi-Label Local to Global Learning: A Novel Learning Paradigm for Chest X-Ray Abnormality Classification," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 9, pp. 4409–4420, Sep. 2023, doi: 10.1109/jbhi.2023.3281466.
[5]. F. S. Butt, M. F. Wagner, J. Schafer, and D. G. Ullate, "Toward Automated Feature Extraction for Deep Learning Classification of Electrocardiogram Signals," IEEE Access, vol. 10, pp. 118601–118616, 2022, doi: 10.1109/access.2022.3220670.
[6]. S. Deepika and N. Jaisankar, "Detecting and Classifying Myocardial Infarction in Echocardiogram Frames With an Enhanced CNN Algorithm and ECV-3D Network," IEEE Access, vol. 12, pp. 51690–51703, 2024, doi: 10.1109/access.2024.3385787.
[7]. A. Degerli et al., "Early Detection of Myocardial Infarction in Low-Quality Echocardiography," IEEE Access, vol. 9, pp. 34442–34453, 2021, doi: 10.1109/access.2021.3059595.
[8]. A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and Robust Machine Learning for Healthcare: A Survey," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 156–180, 2021, doi: 10.1109/rbme.2020.3013489.
[9]. L. Yao et al., "Enhanced Automated Diagnosis of Coronary Artery Disease Using Features Extracted From QT Interval Time Series and ST–T Waveform," IEEE Access, vol. 8, pp. 129510–129524, 2020, doi: 10.1109/access.2020.3008965.
[10]. J. Chillapalli, S. Gite, B. Saini, K. Kotecha, and S. Alfarhood, "A Review of Diagnostic Strategies for Pulmonary Embolism Prediction in Computed Tomography Pulmonary Angiograms," IEEE Access, vol. 11, pp. 117698–117713, 2023, doi: 10.1109/access.2023.3319558.
[11]. C. P. Marini et al., "Incidence and effects of deep vein thrombosis on the outcome of patients with coronavirus disease 2019 infection," Journal of Vascular Surgery: Venous and Lymphatic Disorders, vol. 10, no. 4, pp. 803–810, Jul. 2022, doi: 10.1016/j.jvsv.2021.10.013.
[12]. Y. Fares et al., "Systematic screening for a proximal DVT in COVID-19 hospitalized patients: Results of a comparative study," JMV-Journal de Médecine Vasculaire, vol. 46, no. 4, pp. 163–170, Jul. 2021, doi: 10.1016/j.jdmv.2021.05.003.
[13]. S. Soffer et al., "Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis," Scientific Reports, vol. 11, no. 1, Aug. 2021, doi: 10.1038/s41598-021-95249-3.
[14]. F. Shi et al., "Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 4–15, 2021, doi: 10.1109/rbme.2020.2987975.
[15]. X. Yang et al., "A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images," IEEE Access, vol. 7, pp. 84849–84857, 2019, doi: 10.1109/access.2019.2925210.
[16]. H. Yuan, Y. Shao, Z. Liu, and H. Wang, "An Improved Faster R-CNN for Pulmonary Embolism Detection From CTPA Images," IEEE Access, vol. 9, pp. 105382–105392, 2021, doi: 10.1109/access.2021.3099479.
[17]. L. Schmuelling et al., "Deep learning-based automated detection of pulmonary embolism on CT pulmonary angiograms: No significant effects on report communication times and patient turnaround in the emergency department nine months after technical implementation," European Journal of Radiology, vol. 141, p. 109816, Aug. 2021, doi: 10.1016/j.ejrad.2021.109816.
[18]. J. C. González Sánchez, M. Magnusson, M. Sandborg, Å. Carlsson Tedgren, and A. Malusek, "Segmentation of bones in medical dual-energy computed tomography volumes using the 3D U-Net," Physica Medica, vol. 69, pp. 241–247, Jan. 2020, doi: 10.1016/j.ejmp.2019.12.014.
[19]. Ö. Özdemir and E. B. Sönmez, "Attention mechanism and mixup data augmentation for classification of COVID-19 Computed Tomography images," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 8, pp. 6199–6207, Sep. 2022, doi: 10.1016/j.jksuci.2021.07.005.