Heart Disease Prediction using an Ensemble Learning Method: A Study at King Abdullah Hospital in Bisha, Saudi Arabia

Ghalia A. Alshehri[1,*], Hajar M. Alharbi[1] and Husain H. Jabbad[2]

[1]*Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*

[2]*Department of Surgery, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract: The detection of diseases is essential to improving healthcare outcomes and saving lives. Thanks to technological advancements in medicine, machine learning has become a valuable tool for predicting future patient health outcomes. Despite the abundance of available patient data, accurately predicting cardiac disease has become increasingly challenging. In response, we developed an innovative ensemble learning approach (ELA) that combines three powerful machine learning (ML) techniques. Our ELA provides reliable predictions of cardiac disease that surpass those of the individual classification algorithms, resulting in higher accuracy. Our research yields a new combination of classification algorithms that significantly increases the prediction accuracy. We tested our model on a regional dataset collected from King Abdullah Hospital in Bisha, Saudi Arabia. We obtained the best results false negatives (FN) of 8, true positives (TP) of 70, true negatives (TN) of 72, false positives (FP) of 6, accuracy of 0.9113, sensitivity of 0.8839, specificity of 0.95, PPV of 0.9389, NPV of 0.8878, AUC of 0.9569, F1 of 0.9133 Kappa of 0.8220, MCC of 0.8277 with an ELA comprising logistic regression (LR), extra trees (ET) and support vector machine (SVM) with radial basis function (RBF) kernel. With our ELA, medical professionals can detect cardiac disease and provide timely interventions to prevent potentially life-threatening health issues.

Keywords: Machine learning, Ensemble learning, Classification, Disease prediction, Heart disease.

1. INTRODUCTION

Heart disease is the leading cause of death in the world, a very serious disease with devastating consequences. Heart disease according to statistics released by the World Health Organization and the annual statistical reports published by the Ministry of Health in 2010 claimed 42 percent of non-communicable disease-related deaths in Saudi Arabia [1]. Detecting heart disease early and providing proper treatment can significantly decrease mortality rates. Recent medical research has revealed crucial therapies that can be highly effective in treating heart-related issues. Additionally, advancements in artificial intelligence (AI) have made it possible to tackle various medical challenges. The most common methods for identifying heart disease include electrocardiograms (ECGs), angiography screening, and blood testing [2].

Heart disease danger elements include high blood pressure, elevated cholesterol, and hypertension, which may not be noticeable to the average person. Symptoms of heart disease can manifest as palpitations, dyspnea, and chest pain. The cardiac ailment caused by the heart not getting sufficiency of oxygen is sometimes referred to as angina pectoris. When the heart fails to pump blood efficiently, it may cause heart failure, which causes one to breathe with difficulty. Some heart conditions may not show any signs, especially in older individuals and diabetics. Therefore, the healthcare industry must track additional patient and pharmaceutical information to provide accurate diagnostic reports.

The medical field has experienced tremendous transformations as medical studies and technology have improved. Computer-aided systems are now crucial diagnostic and therapeutic tools, and AI provides new solutions to medical problems. Global medical experts continue to study and explore the potential applications of AI in medicine [3]. The predictive analysis of the long-term health and well-being of patients is determined through laboratory results, and electronic hospital information management systems have simplified this task by allowing the automated predictive analysis to be done. As contemporary medical procedures require knowledge-intensive systems, the demand for AI and knowledge-based systems is rising [4]. Computers are far more accurate than humans in recording and remembering information, with an accuracy rate of 99.9 %. We used several algorithms to develop an ELA, and used GridsearchCV to choose the best parameters to obtain the optimal results. We then applied the proposed ELA to a dataset collected from heart patients at King Abdullah Hospital in Bisha, Saudi Arabia. We used the wrapper method to select important features, thus improving the prediction

*Address correspondence to this author at the Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia; E-mail: gmesfaralshehri@stu.kau.edu.sa

results. This paper seeks to address some of the major research questions associated with predicting heart disease through ML methods. The first discusses the most effective algorithms to forecast heart disease, and ELA. In addition to this, the paper explains the way in which the accuracy of heart disease prediction can be enhanced by feature selection techniques to narrow down on the number of inputs to be used. And finally it will time how probable it would be on the output of the local hospital, to make an assumption of the success of the model, in the local regional environment. The paper consists of different sections. The section Related Work describes the previous studies on the prediction of heart diseases and the use of ML methods in this domain. Methodology section is devoted to describing the dataset employed and the methodology adopted in constructing the model. We provide details of our experiments and model results in the Experiments and Results section. Lastly, we conclude the paper with the findings of the research, and recommendations, and conclusions in the Conclusion section.

2. RELATED WORK

Pawlovsky (2018) [5] investigated the detection of heart diseases using the K-nearest neighbors algorithm and suggested a distance-based ensemble framework with two different configurations. The former setup involved three distance measures and the latter five. In addition, a weighted variant was also introduced, and the weights were determined based on the average performance of each distance measure when KNN is employed. The experimental evaluation was conducted on the Cleveland heart disease dataset given by UCI [6]. In all the setups tested, their collective decision had a steady average accuracy of about 85%.

Similarly, suggested a nested ensemble nu-support vector classification method to predict CAD by utilizing both conventional ML and ensemble learning methods. They tested the model on the Cleveland [6, 7] dataset using optimization and preprocessing software, including genetic search, ClassBalancer, Resample, and multi-level filtering. The proposed model showed high predictive performance, as it achieved maximum accuracies of 94.64% and 98.60% on the Z-Alizadeh Sani and Cleveland datasets, respectively-and can serve as an alternative non-invasive diagnostic.

Latha & Jeeva (2019) [8] investigated how the ensemble learning methods modify the predictive power of weak classification algorithms using a heart disease dataset. Their results showed that ensemble techniques, namely boosting and bagging, significantly improved the performance of otherwise poor-performing classifiers, and they can be useful in the initial detection of disease. The accuracy of the integrated ensemble methods were up to 7 percent higher than the standalone weak models. Atallah & Al-Mousa (2019) [9] applied majority voting to more than one machine learning model to improve the accuracy of the diagnosis. When they used their hard voting ensemble technique on the data of the UCI repository (Lapp, 2019) [10] the classification accuracy was 90% and it was better than the results of the individual model predictions.

Li *et al*. (2019) [11] designed a novel Adaboost + RF prediction model of cardiovascular disease with unbalanced data sets including missing data. The efficiency of the model was evaluated and measured through the recall, precision, F-measure, and ROC measures, and it was portrayed to be more efficient than other machine learning algorithms. In the study, a dataset of heart diseases presented by the UCI repository was utilized [11]. The Adaboost with Random Forest (RF) model assessment with five imbalanced classes presented a precision, recall, F1-score and AUC of 40.9%, 49.3%, 41.4%, and 71.6% respectively. Similarly, [12] uggested an artificial intelligence-driven framework that could identify heart diseases at an early stage to help them obtain timely medical assistance before the cardiac condition deteriorates. Their method employed gradient boosting-based sequential feature selection (GBSFS) that determined and selected pertinent features using the same UCI [10] Various machine learning algorithms like stacking, gradient boosting classifier (GBC), LR, KNN, DT, RF, multilayer perceptrons (MLPs), SVM and ET were used in the framework. The proposed model worked better where the test accuracy stood at 98.78, and this is superior to the benchmark techniques.

In their article, Asif *et al*. (2023) [13] test hyperparameter optimization strategies and four ensemble learning models, ET, RF, XGBoost and CatBoost to achieve better heart disease predictions. By merging three Kaggle datasets [10, 14, 15] it was found that the ET method, the 80:20 split ratio, and the hyperparameters optimized by GridsearchCV could greatly enhance the accuracy of the model (to 98.15%). presented a system that uses ensemble algorithms without a feature selection approach to predict cardiac diseases efficiently. The Cleveland UCI dataset [15], was used. Dissimilar data were also addressed by using data balancing techniques and an isolation forest

and developed a prediction model using ensemble techniques, such as boosting, bagging, stacking, and voting. The suggested methodology was characterized by the following performance measures: accuracy 98.73%, sensitivity 98.7%, specificity 100%, PPV 100%, NPV 97%, F1 100% and AUC 100%. The authors of the article by Ganie *et al*. (2023) [16] performed research on heart disease prediction utilizing a dataset from the UCI ML repository [10]. Three boosting algorithms were applied-that is, AdaBoost, XGBoost, and GBC-and the feature significance approach was used to identify the presence of independent features of the outcome. The team also analyzed the data to find any missing data and outliers. The gradient boosting technique was much better than all others and reached an astonishing rate of 92.20%.

Cardiovascular diseases and cardiac arrest are major health problems in society, and Aziz *et al*. (2024) [17] focused on its early diagnosis with an innovative early warning system, which is very sensitive, and the false-positive rate is low. The paper examined data to reveal correlations between attributes and enhance the detection of cardiac arrest in its early phase, with an emphasis on a high F1-score. Several ML algorithms such as AdaBoost, LightGBM, CatBoost, and XGBoost were used. The most appropriate was the CatBoost because its highest accuracy value of 98.08%, its positive prediction of 98% and sensitivity of 97.8% and this model can be used in detecting early cardiac arrest. Narayanana (2024) [18] applied ML models to detect the initial signs of a heart disease. The analysis with the supervised ML classifiers were conducted to determine the key characteristics that will aid predicting heart disease to initiate early lifestyle change and medical intervention. The researchers used the SMOTE oversampling method to enhance model training so as to solve the imbalanced datasets problem. RF classifier has proved to be the most efficient in the early detection of heart disease with the greatest accuracy, sensitivity and specificity and precision of 96.6 percent accuracy, 90 percent sensitivity, and 100 percent specificity and precision [10] based on the available UCI heart dataset.

Despite the significant progress achieved in predicting cardiac illness through the use of ensemble learning techniques, a key limitation in previous studies is their dependence on publicly available datasets such as those from Kaggle and the UCI ML repository, which primarily reflect populations outside the Middle East. To date, there is a scarcity of research focused on predicting heart disease using patient data from Saudi Arabia or similar regions, where unique environmental, genetic, and lifestyle factors may influence disease incidence and outcomes. In this study, we address this gap by employing a dataset collected from King Abdullah Hospital in Bisha, Saudi Arabia. This local dataset offers valuable insights into the specific risk factors and predictive patterns relevant to the Saudi population, which could differ from those in widely used international datasets. Bridging this gap is crucial for developing more accurate and region-specific models for predicting heart disease, which could improve detection and treatment methods in Saudi Arabia.

3. METHODOLOGY

3.1. Data Collection

The features by which heart patients can be predicted are in our dataset that we gathered in King Abdullah Hospital located in Bisha, Saudi Arabia. Our data was gathered between 28 May and 13 July 2023. The information available to the authors could lead to identification of individual participants when collecting data. The final dataset comprises 21 features and 115 cases. The features contained in the dataset include chest pain (Chp), age, sex, electrocardiography (Ecg), blood pressure at rest, systolic (Bp-S; expected average of Bp-S = 90--120 mm Hg), resting blood pressure (diastolic (Bp-D; usual standard of Bp-D = 60--80 mm Hg), fasting blood sugar (Fbs; expected average of blood sugar levels during fasting = 70-100 mg/dL), glycated hemoglobin (HbA1c; usual standard of HbA1c = 4-5.6%), heart rate (HR; expected average of HR = 60-100 Bpm), height, weight, nationality (Na), family history of heart disease (FH), diabetes (DM), smoking (Sm), low-density lipoprotein (LDL; normal average of LDL = 2.5-3.34 mmol/L), high-density lipoprotein (HDL; normal average of HDL = 0-1.68 mmol/L), triglyceride (TG; normal average of TG = 0-2.26 mmol/L), dyslipidemia (DPL; a patient has DPL when LDL and TG are highly abnormal), rhythm (Rh), and patient diagnosed as a heart disease patient (class).

3.2. Handling Missing Data

We obtained 82 complete cases and, with the help of an expert, we completed some data that needed to be included. Some cases needed an Fbs test, for which we calculated the Fbs from HbA1c, The number of cases is 21. In some cases, HbA1c was lacking, for

which we calculated HbA1c from Fbs, The number of cases is 8. The following equation is the equation that was used to determine HbA1c by using Fbs [19].

$$HbA1c = 2.6 + 0.03 \times Fbs(mg/dL)$$

To compute Fbs using HbA1c in diabetics we inverted the earlier equation. Three cases were missing some data: namely, HDL, LDL, TG, DPL, and HbA1c. We completed these data with expected values, and calculated HbA1c from Fbs. In six cases some data were missing: namely, HDL, LDL, TG, DPL, HbA1c, and Fbs. We completed these with the expected values. In one case, HDL, LDL, TG, DPL, and Fbs were missing. We completed these with the expected values and calculated the Fbs from HbA1c. In four cases HbA1c and Fbs information was missing, and the patients were not diabetics; we completed them with standard values. In another case, HR data was lacking, which we completed by calculating the mean HR for all cases (HR = 81). While in a different case weight and height data were missing, which we obtained by calculating the mean for all cases (162 for height and 78 for weight). In addition, we completed the missing Chp data for one patient using the typical chest pain value. The standard values with which we completed the missing data were as follows: TG = 1.7 mmol/L, HDL = 1 mmol/L for men and 1.3 mmol/L for women, LDL = 3.4 mmol/L, Fbs = 99 mg/dL, and

Table 1: Attribute Information for the Dataset

Attribute	Description	Type of data	Domain
Age	Patient age (years)	Numerical	17-115
Sex	Gender	Binary	0 = female 1 = male
Chp	Chest pain (angina)	Nominal	1 = typical 2 = atypical 3 = no pain
Ecg	Electrocardiography	Nominal	1 = normal ECG 2 = abnormal ST segment specific 3 = abnormal ST segment non-specific 4 =abnormal with no ST segment changes
Bp-S	Resting blood pressure - Systolic (mmHg)	Numerical	80-241
Bp-D	Resting blood pressure - Diastolic (mmHg)	Numerical	41-133
Fbs	Fasting blood sugar (mg/dL)	Numerical	71-356
HbA1c	Glycated hemoglobin percentage (DCCT unit)	Numerical	1.61-14
HR	Heart rate (Bpm)	Numerical	45-122
Height	Height (Cm)	Numerical	79-190
Weight	Weight (Kg)	Numerical	37-172
Na	Nationality	Binary	0 = non-Saudi 1 = Saudi
FH	Family history of heart disease	Binary	0 = yes 1 = no
DM	Diabetes	Nominal	0 = non-diabetes 1 = NIDDM 2 = IDDM
Sm	Smoking	Binary	0 = yes 1 = no
DPL	Dyslipidemia	Binary	0 = yes 1 = no
LDL	Low-density lipoprotein (mmol/L)	Numerical	0.23-5.84
HDL	High-density lipoprotein (mmol/L)	Numerical	0.27-1.94
TG	Triglycerides (mmol/L)	Numerical	0.36--6.18
Rh	Rhythm	Binary	0 = normal 1 = abnormal
Class	Patient diagnosed as heart disease patient	Binary	0 = absence 1 = presence

HbA1c = 5.1. The final number of data that we used after completing the data with the help of the expert was 108 cases. Table **1** describes the dataset. The method used to complete the missing data in our study was carefully crafted to ensure the accuracy and integrity of the datasetWe employed clinically informed imputation techniques to fill missing values in collaboration with domain experts, with a set of medical standards and predicted values serving as input parameters like TG, HDL, LDL, and others. In some instances, we used reversible equations, e.g. computing HbA1c based on Fbs, which are based on physiologically well-documented relationships. This approach would ensure that the imputed data meets with the accepted clinical patterns and hence the biological acceptability of the data set. In cases where we were not able to measure variables such as HR or height, we estimated with the use of relevant population means or values based on related measures and the difference between the forecasted health trends was minimal. Such imputation techniques are common in the literature of studies of heart disease prediction [20, 21], and are mentioned as a best practice medical data management approach. Generally, we have maintained the integrity of our dataset without bias, thereby facilitating the validity of the analysis that follows.

3.3. Data Augmentation

We used the Synthetic Minority Over-sampling Technique to eliminate the problem of class imbalance, a common problem in medical datasets. Imbalanced data can compromise the effectiveness of machine learning models by skewing predictions toward the majority class. SMOTE mitigates this issue by generating synthetic examples for the underrepresented class, rather than simply replicating existing samples. This method enhances the model's capacity to recognize instances from the minority class, thereby contributing to more accurate and balanced predictive performance. The effectiveness of SMOTE has been validated in various studies, notably in the work of Zhang *et al*. (2020) [22], who demonstrated its value when combined with data mining techniques to improve survival prediction in heart failure patients. In our study, the integration of SMOTE aimed to ensure balanced model performance across both the majority and minority classes, thereby enhancing the overall reliability and robustness of the predictive system. Importantly, SMOTE was applied strictly within the training folds during cross-validation, ensuring that no synthetic samples were introduced into the validation folds. This approach prevented data leakage and preserved the validity of our performance evaluation. The histograms of the collected dataset in Figure **1**.

3.4. Data Pre-Processing

Having the dataset of heart patients that was gathered at the King Abdullah Hospital, Bisha, Saudi Arabia, we pre-processed the data with the help of the normalization: a pre-processing step of the machine learning algorithms, which involves the use of such techniques as the decimal scaling, z-score, and min--max that will improve the performance of the obtained model [23]. Python's sklearn framework includes practical normalizing algorithms such as Normalizer,

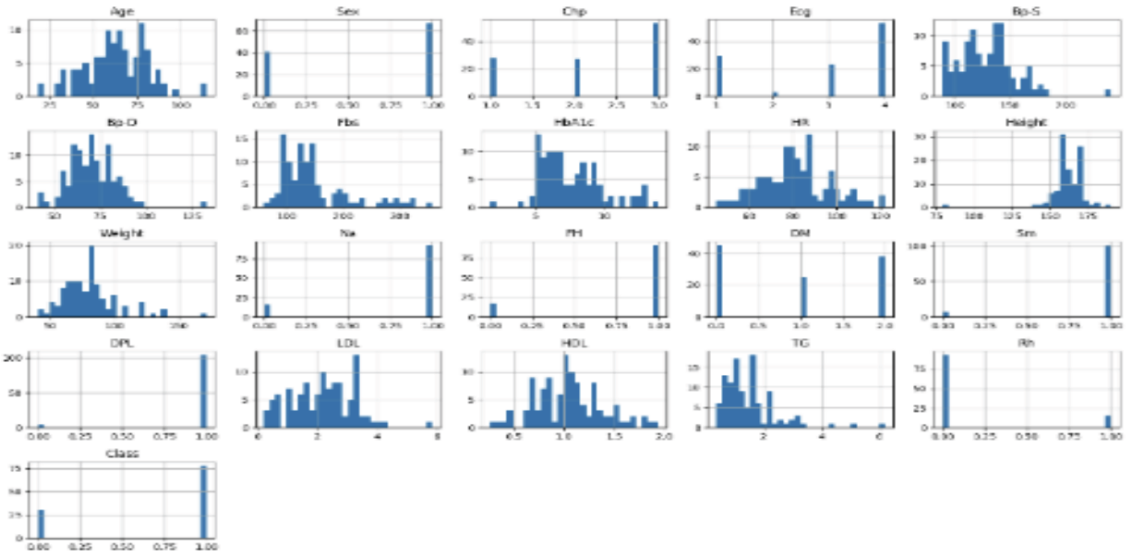


Figure 1: Histograms of the collected dataset.

RobustScaler, StandardScaler, MinMaxScaler, and MaxAbsScaler [24]. In this study, we employed MinMaxScaler as a normalization tool.

3.5. Feature Selection

In the feature selection, we used forward selection and backward elimination- popular dimensionality reduction techniques in machine learning. The method of feature selection is a baseline technique of dimensionality reduction to recognize a set of relevant features according to defined evaluation criteria. In addition to the increased efficiency of the learning process, it also decreases the computational cost and makes models interpretation simpler. Filter, wrapper and embedded methods are the most common [25]. Among them, wrapper methods follow a simple approach and evaluate the utility of features based on a specific classifier [26]. In particular, in backward elimination, the entire set of features is initially used, and the least useful feature is repeatedly removed until the model being developed has reached an acceptable performance [27]. Contrastingly, forward selection starts with an empty set and iteratively adds features one at a time, the ones that improve the performance.

3.6. Classification

The classifier algorithms used in our ELAs were SVM with four kernels radial basis function, linear, sigmoid and polynomial, AdaBoost, RF, decision tree (DT), LR [28], KNN [29], Naive Bayes (NB) [30], GBC [31], ET [32] and CatBoost. AdaBoost is a classification method that improves the effectiveness of machine learning algorithms by first building a committee of weak classifiers, and then integrating them into a strong classifier [33]. SVM is a binary data classification method that uses statistical theory in its study [34]. It has proved to be a useful tool in addressing cases with local minima and high dimensionality [35]. DT learning can be used to approximate the values of discrete target functions, the learned role is formulated as a decision tree. To simplify the learned trees so that humans can understand them, they may be rewritten as sets of if--then rules [36]. These are the most popular inductive inference algorithms. RF technique is a classification technique that employs multiple decision trees. It generates every tree using the bagging approach and randomization of the tree features, hence, generating a pool of loosely correlated trees capable of making committee predictions that are better than the individual tree predictions [37]. Ensemble learning aims to build a unified structure of data mining, data fusion, and data modeling based on diversified transformations [38]. We have applied GridsearchCV to pick the optimal hyperparameters of the models that we have used in the ELA. This approach is used to identify the best hyperparameters to employ in a specific algorithm and has been deployed as a part of a library within the sklearn model selection package. Simply put, the model is trained on the training set and the hyperparameters are optimized over. In addition to the best parameters, the number of repeats of cross-validation of each set of hyperparameters can also be ascertained [39]. In order to evaluate them, we used the AUC, positive predictive value, NPV (negative predictive value), precision, Matthew correlation coefficient and Kappa (the Cohen Kappa) measures. Note that positive predictive value is a measure of precision whereas recall is a measure of sensitivity. We used the model introduced in the study of Alshehri & Alharbi (2023) [40]; however, we further used the GridsearchCV method to select the optimal parameters for the ML models and ELA models to give optimal results. We tried the top two models, the top three models, and all. The best ELA was ELA (LR+ET+SVM(RBF)). The rationale for this combination lies in the diversity of the models it integrates. The logistic regression is linear and interpretable, the extra trees can model non-linear non-linear relationship with randomized decision tree groups, and the SVM using RBF kernel can have good capability to model non-linear decision boundary in high dimensional space. Since these algorithms are based on various mathematical concepts, their classification errors will be less correlated. By pooling them together in an ensemble, therefore, it decreases variance, increases robustness and generalization in comparison with any single model. The improved result of LR+ET+SVM (RBF) ensemble can be attributed to such theoretical complementing. The methodology framework is depicted in Figure **2**.

4. EXPERIMENT AND RESULTS

In this experiment, we used algorithms: SVM with four different kernels (RBF, linear, sigmoid, and poly), AdaBoost, RF, LR, KNN, NB, GBC, ET, CatBoost, and DT. We divided the dataset using cross-validation CV=10. We will choose the best ELA for the confusion matrix, specifically FN because it is important in predicting diseases [41]. After all, we do not want a model that diagnoses heart patients as healthy people because this poses a risk to their health. Therefore, we will choose the model with a lower FN number to

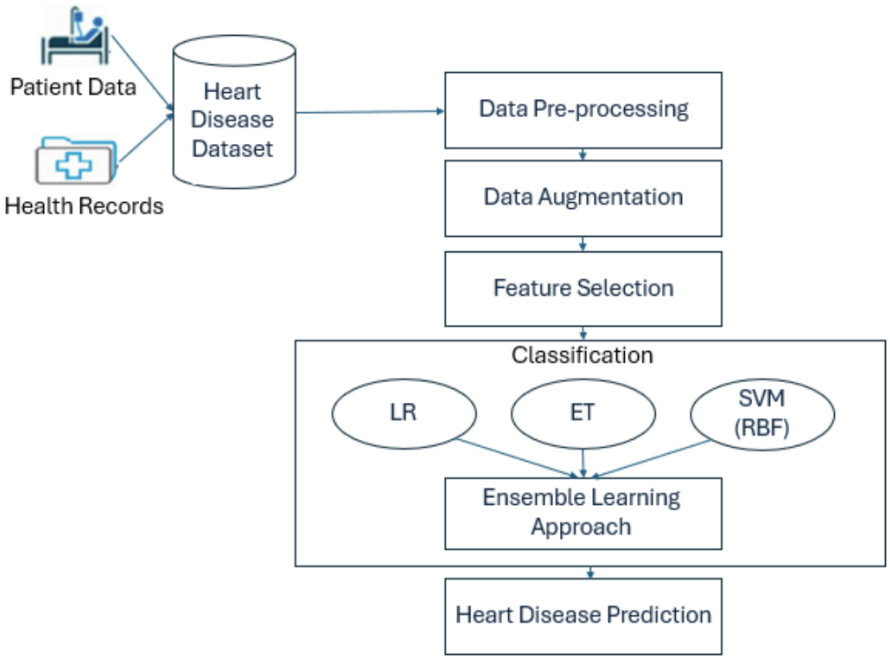


Figure 2: Methodology framework.

preserve the safety of patients. We train the models with gridsearchCV to find the best hyperparameters The optimal hyperparameters of each model were: DT(criterion= entropy, max_depth= 4, min_samples_leaf = 2, min_samples_split= 3), RF (bootstrap = False, max_depth= 4, max_features= 'sqrt', min_samples_leaf= 1, min_samples_split= 2, n_estimators= 10), ET (max_depth= 10, min_samples_split = 10, n_estimators= 50), SVM(linear) (kernel='linear', C= 1, gamma= 1), SVM (RBF) (C= 10, gamma= 0.5, kernel= rbf), SVM (poly) (C= 1, gamma= 0.5, kernel= poly, degree= 3), SVM (sigmoid) (C= 10, gamma= 0.1, kernel= sigmoid), AdaBoost (algorithm= SAMME.R, learning_rate= 0.1, n_estimators= 50), CatBoost (verbose=0, depth= 4, iterations= 200, learning_rate= 0.1), GBC (learning_rate = 0.01, max_depth = 5, n_estimators = 50), KNN (metric = manhattan, n_neighbors = 9, weights = distance)LR (C= 0.1, penalty= l1, solver= liblinear)., We applying SMOTE to handle the problem of class imbalance. The cases of heart disease were 78 and healthy cases were 30. Following the application of SMOTE, the population with heart disease was 78 and healthy population was 78. In Table **2**, we used the collected dataset after applying SMOTE to it, and we applied wrapper methods to it to obtain the best feature set to use with the prediction model. We used all models without using CatBoost because it is impossible to show the results using R^2. The best method was Backward with SVM (RBF) obtaining the highest R^2 of 0.82. wherefore we will use this set of features with different models to get the best prediction result.

Then we used SMOTE with several ML models used in disease prediction, and the results of the models are in Table **3**. To choose the models that we will use for the ELAs, in the first we decided on four models that obtained the lowest FN, which are LR and ET for the first model ELA 1 because they were the models with the lowest FN, where LR obtained 11, ET obtained 12, Then in the second model, which is ELA 2, we added SVM (RBF) because it was the third lowest for FN and also AdaBoost, where they got 13. However, we chose SVM (RBF) because it was lower than AdaBoost in FP, where it got FP = 3, while AdaBoost got 1. FP = 24. As for the third model ELA 3, we added AdaBoost.

In Table **4**, the results of ELAs after SMOTE, At this stage, we decided to try all models to create ELAs to search for the best combination. Therefore, we decided to try all models in ELAs to obtain the best result for FN. We will mention all ELAs and the models they contain in Table **4**, ELA 1(LR + ET), ELA 2(LR+ET+SVM (RBF)), ELA 3(LR + ET + SVM (RBF) + AdaBoost), ELA 4(LR + ET + SVM (RBF) + AdaBoost + KNN), ELA 5(LR + ET + SVM (RBF) + AdaBoost + KNN + RF), ELA 6(LR + ET + SVM (RBF) + AdaBoost + KNN + RF + GBC), ELA 7(LR + ET+ SVM (RBF) + AdaBoost + KNN + RF + GBC + CatBoost), ELA 8(LR + ET + SVM (RBF) + AdaBoost + KNN + RF + GBC + CatBoost + DT), ELA 9(LR + ET + SVM (RBF) +

Table 2: Results for Wrapper Methods with Different Reduced Feature Sets after SMOTE

Method	Feature count	Names of features	R2
Forward + DT	8	Chp, Ecg, HR, Weight, DM, Sm, DPL, Rh	0.31
Backward + DT	9	Age, Ecg, Bp-D, Fbs, HR, FH, DM, HDL, TG	0.34
Forward + RF	8	Chp, Ecg, HR, Height, Na, FH, DM, LDL	0.48
Backward + RF	15	Sex, Chp, Ecg, Bp-S, Bp-D, Fbs, HR, Height, Na, FH, DM, Sm, DPL, LDL, Rh	0.51
Forward + SVM (linear)	14	Age, Sex, Chp, Ecg, Bp-S, Fbs, HR, Height, Na, DM, Sm, LDL, TG, Rh	0.39
Backward + SVM (linear)	8	Sex, Chp, Ecg, Bp-S, HbA1c, HR, LDL, Rh	0.34
Forward + SVM (RBF)	15	Chp, Ecg, Bp-S, Fbs, HR, Height, Na, FH, DM, Sm, DPL, LDL, HDL, TG, Rh	0.74
Backward + SVM (RBF)	14	Age, Sex, Chp, Ecg, Bp-S, Bp-D, HR, Na, DM, Sm, DPL, LDL, HDL, Rh	0.82
Forward + SVM (poly)	17	Age, Sex, Chp, Ecg, HbA1c, HR, Height, Weight, Na, FH, DM, Sm, DPL, LDL, HDL, TG, Rh	0.65
Backward + SVM (poly)	9	Age, Sex, Chp, Ecg, HR, Na, DM, LDL, Rh	0.79
Forward + SVM (sigmoid)	15	Age, Chp, Ecg, Bp-D, Fbs, HbA1c, HR, Height, Weight, FH, DM, DPL, LDL, TG, Rh	0.28
Backward + SVM (sigmoid)	7	Chp, Ecg, HbA1c, HR, LDL, TG, Rh	0.36
Forward + AdaBoost	5	Age, Sex, Chp, Ecg, Bp-S	-0.06
Backward + AdaBoost	5	Age, Sex, Chp, Ecg, Bp-S	-0.06
Forward + ET	9	Age, Sex, Chp, Ecg, Bp-D, HbA1c, HR, Na, DPL	0.57
Backward + ET	12	Age, Sex, Chp, Ecg, Bp-D, Fbs, HR, Height, Na, DPL, HDL, TG	0.6
Forward + GBC	7	Chp, Ecg, Height, Na, FH, DM, DPL	0.42
Backward + GBC	13	Age, Sex, Ecg, Bp-S, Bp-D, Fbs, Height, Na, FH, DM, Sm, DPL, LDL	0.54
Forward + KNN	8	Age, Sex, Ecg, Na, FH, DM, HDL, Rh	0.51
Backward + KNN	11	Age, Sex, Chp, Ecg, Bp-D, Fbs, HbA1c, Weight, Na, DM, Rh	0.54
Forward + NB	7	Age, Ecg, Bp-S, HbA1c, HR, Height, LDL	0.42
Backward + NB	7	Age, Ecg, Bp-S, HbA1c, HR, Height, LDL	0.42
Forward + LR	6	Age, Sex, Chp, Ecg, Bp-S, Bp-D	-0.32
Backward + LR	5	Age, Sex, Chp, Ecg, Bp-S	-0.32

Table 3: Compare Several Models after SMOTE

Model	FN	TP	TN	FP	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
DT	24	54	68	10	0.7825	0.6786	0.8446	0.8419	0.7412	0.8269	0.7424	0.5500	0.5635
RF	16	62	66	12	0.8217	0.7821	0.8357	0.8525	0.8166	0.9134	0.8195	0.6299	0.6778
ET	12	66	72	6	0.8675	0.8214	0.9375	0.8992	0.8546	0.9444	0.8636	0.7345	0.7416
SVM (linear)	18	60	65	13	0.8017	0.7679	0.8321	0.8214	0.7915	0.8699	0.7908	0.6016	0.6064
SVM (RBF)	13	65	75	3	0.8979	0.8321	0.9625	0.9625	0.8569	0.9571	0.8876	0.7946	0.8065
SVM (poly)	13	65	75	3	0.8975	0.8321	0.9625	0.9607	0.8596	0.9433	0.8858	0.7941	0.8070
SVM (sigmoid)	17	61	61	17	0.7821	0.7821	0.7821	0.7972	0.7889	0.8359	0.7795	0.5640	0.5746
AdaBoost	13	65	54	24	0.7642	0.8357	0.6893	0.7476	0.8215	0.7625	0.7788	0.5258	0.5459
CatBoost	18	60	72	6	0.8475	0.7679	0.925	0.9082	0.8092	0.9304	0.8270	0.6937	0.7046
GBC	16	62	65	13	0.8546	0.8107	0.8339	0.8571	0.8014	0.8610	0.8412	0.6446	0.6927
KNN	15	63	76	2	0.8925	0.8089	0.975	0.9714	0.8512	0.9470	0.8715	0.7847	0.8018
NB	56	22	72	6	0.6033	0.2804	0.925	0.9	0.5615	0.8760	0.3907	0.2050	0.3014
LR	11	67	44	34	0.7117	0.8607	0.5589	0.6714	0.8195	0.8076	0.7487	0.4193	0.4518

Table 4: ELAs after SMOTE

ELA	FN	TP	TN	FP	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
ELA1	11	67	66	12	0.8489	0.8589	0.8357	0.8324	0.8470	0.9304	0.8194	0.6942	0.6723
ELA2	8	70	72	6	0.9113	0.8839	0.95	0.9389	0.8878	0.9569	0.9133	0.8220	0.8277
ELA3	13	65	70	8	0.8604	0.8357	0.8571	0.9028	0.8549	0.9393	0.8580	0.7214	0.7293
ELA4	14	64	75	3	0.8988	0.8339	0.9232	0.9514	0.8633	0.9333	0.8782	0.7845	0.8114
ELA5	14	64	72	6	0.8671	0.8214	0.925	0.9232	0.8477	0.9460	0.8614	0.772 0	0.7720
ELA6	13	65	68	10	0.8742	0.8357	0.925	0.8927	0.8509	0.9444	0.8756	0.7472	0.7678
ELA7	15	63	74	4	0.8738	0.8089	0.925	0.9121	0.8492	0.9350	0.8500	0.7470	0.7755
ELA8	12	66	70	8	0.8671	0.8339	0.925	0.9204	0.8397	0.9382	0.8667	0.7472	0.7653
ELA9	15	63	74	4	0.8796	0.8071	0.9375	0.9482	0.8411	0.9304	0.8648	0.7722	0.7725

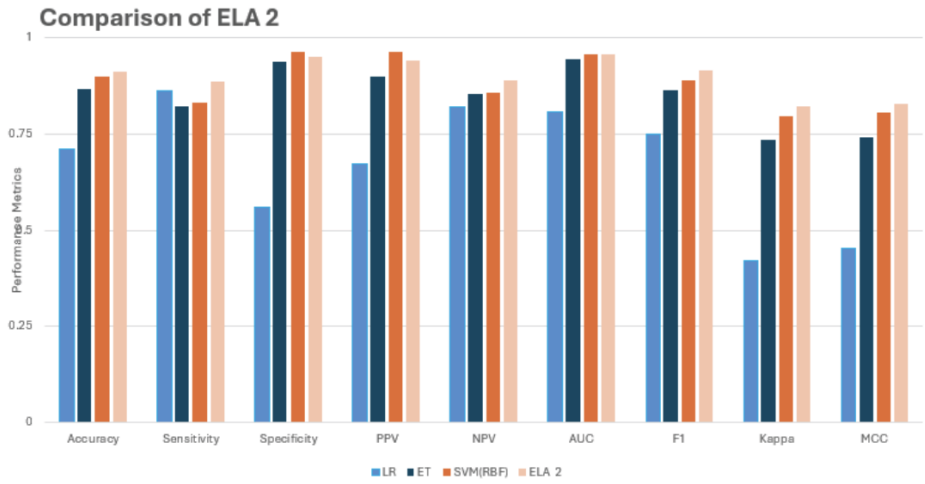


Figure 3: Comparison of ELA 2 with it's single models.

AdaBoost + KNN + RF + GBC + CatBoost + DT + NB). The best ELA that received the lowest FN is ELA 2, which obtained FN = 8, the highest TP = 70, the highest accuracy = 0.9113, the highest sensitivity = 0.8839, the highest specificity = 0.95, the highest NPV = 0.8878, the highest AUC = 0.9569, and the highest F1 = 0.9133 and higher Kappa = 0.8220 and higher MCC = 0.8277 compared to other ELAs.

We notice in Figure **3**, that ELA 2 outperformed the rest of its component models in accuracy, sensitivity, NPV, F1, Kappa, and MCC. However, its results were slightly lower than those of SVM (RBF) in specificity, PPV, and AUC.

In the study of Alshehri & Alharbi (2023) [40], they used ELA consisting of four algorithms: AdaBoost, SVM (linear), RF and DT and applied them to three public datasets Z-Alizadeh Sani dataset [7], StatLog UCI [42, 43] CVD dataset. We will apply their model to the collected dataset, and we will apply our proposed model to the three public datasets. The best model that [40] obtained was ELA (AdaBoost+ SVM (linear)+ RF+ DT) with the Z-Alizadeh Sani dataset [7]. With the wrapper feature selection model Backward+ SVM (linear), so we will use this model with the collected dataset. First, when using Backward+ SVM (linear), we obtained an R^2 of 0.34 and the number of features is 8, which is Sex, Chp, Ecg, Bp-S, HbA1c, HR, LDL, Rh.

Table 5: ELA from the Study of Alshehri & Alharbi with the Collected Dataset

ELA	FN	TP	TN	FP	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost + SVM (linear) + RF+ DT	17	62	63	16	0.814	0.7875	0.825	0.7956	0.8160	0.9152	0.7772	0.6349	0.5883

We applied ELA (AdaBoost+ SVM (linear)+ RF+ DT) to the collected dataset after the feature selection step and used SMOTE to make the data balanced.

When comparing the results of their model in Table **5** with our proposed model ELA 2 in Table **4**, we notice that our model excelled, as it obtained better results in FN = 8, TP = 70, TN= 72, FP= 6 and the highest accuracy= 0.9113, the highest sensitivity= 0.8839, the highest specificity= 0.95, the highest PPV= 0.9389, the highest NPV=0.8878, the highest AUC= 0.9569, and the highest F1= 0.9133 and highest Kappa=0.8220 and higher MCC=0.8277. That is, our proposed model excels in all metrics.

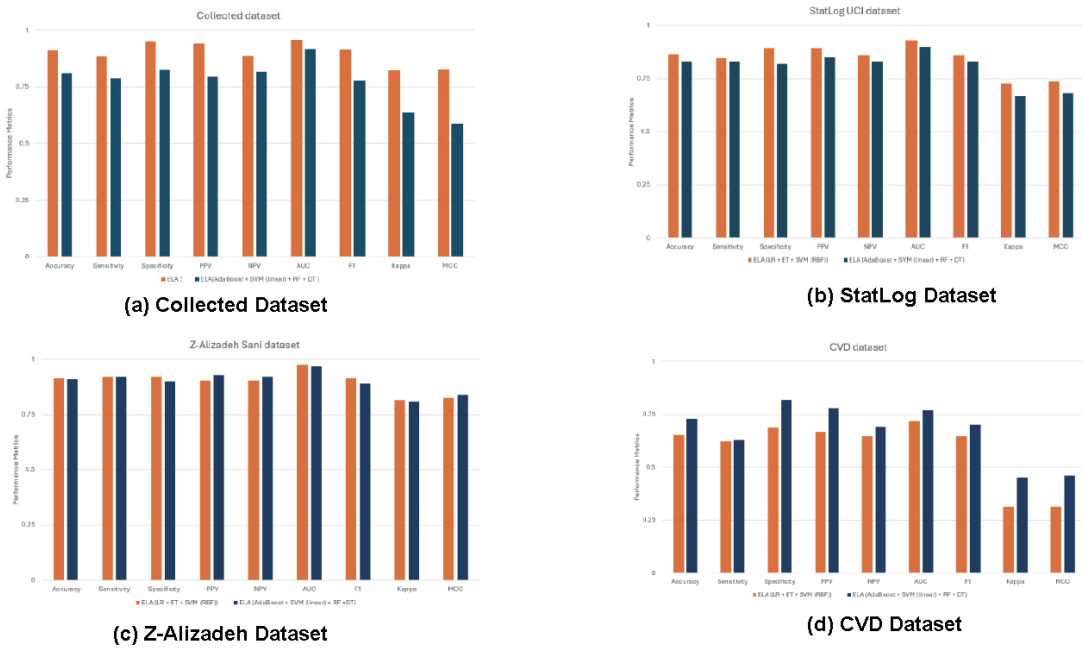
Then we applied our proposed model to public datasets. First, we used Backward + SVM (RBF). For the feature selection step, StatLog UCI dataset [42] was obtained R^2 of 0.49, with nine features: Age, Sex, Chp, Mhrt, Exian, Opk, Slope, Vessel, and Thal. As for the Z-Alizadeh Sani dataset [7], after applying the feature selection step, obtained an R^2 of 0.83 with 25 features: Age, Weight, BMI, DM, HTN, Current Smoker, FH, Obesity, CVA, DLP, BP, Edema, Systolic Murmur, Typical Chest pain, Dyspnea, Atypical, Nonanginal, St Depression, Tinversion, LVH, TG, Na, WBC, Region RWMA and VHD. As for the CVD dataset [43], The feature selection step was used, followed by an R^2 of -0.38 with nine features; age, height, weight , ap_hi, ap_lo, cholesterol, gluc, alco, and active. StatLog UCI data results [43], the Z-Alizadeh Sani dataset [7] and the CVD dataset [43] after applying the proposed model, which is ELA (LR+ET+SVM (RBF)) are shown in Table **6**. When comparing the results of StatLog UCI dataset [42] in the study by Alshehri & Alharbi (2023) [40], The result obtained when using the ELA (AdaBoost+ SVM (linear)+ RF+ DT) was the accuracy of 0.83, the sensitivity of 0.83, specificity of 0.82, PPV of 0.85, NPV of 0.83, AUC of 0.90, F1 of 0.83, Kappa of 0.67, and MCC of 0.68. However, when we applied our proposed model ELA (LR+ET+SVM (RBF)) to the same dataset, and its results are shown in Table **6**, we noticed that the proposed model excelled in all metrics. As for Z-Alizadeh Sani dataset [7]. In the study by Alshehri & Alharbi (2023) [40], The result obtained when using the ELA model (AdaBoost+ SVM (linear)+ RF+ DT) was the accuracy of 0.91, the sensitivity of 0.92, specificity of 0.90, PPV of 0.93, NPV of 0.92, AUC of 0.97, F1 of 0.89, Kappa of 0.81, and MCC of 0.84. However, when we applied our proposed model ELA (LR+ET+SVM (RBF)) to the same dataset, and its results are shown in Table **6**, We note that our proposed model excelled in specificity of 0.9208, AUC of 0.9767, and F1 of 0.9144, while obtaining equal results in accuracy, sensitivity, and Kappa. As for CVD dataset [43] In the study by Alshehri & Alharbi, [40] The result obtained when using the ELA model (AdaBoost+ SVM (linear)+ RF+ DT) was the accuracy of 0.73, the sensitivity of 0.63, specificity of 0.82, PPV of 0.78, NPV of 0.69, AUC of 0.77, F1 of 0.70, Kappa of 0.45, and MCC of 0.46. However, when we applied our proposed model ELA (LR+ET+SVM (RBF)) to the same dataset, and its results are shown in Table **6**, it gave lower results in all metrics.

The poorer results on the CVD dataset can be explained by mismatches between the chosen feature set and data variables, and differences in demographic and clinical factors of the populations under study. The CVD dataset (as opposed to the regional dataset and other benchmark datasets) contains variables of different distributions and definitions, which could have constrained the generalizability of our optimized feature subset. This shows the relevance of selecting features and model-training in data-specific settings.

In Figure **4**, Comparing the results of the collected dataset and three public datasets when applied to two models, the first ELA (LR+ET+SVM (RBF)) and the second is ELA (AdaBoost+ SVM (linear)+ RF+ DT), The first is our proposed model and the second is a model from a study by Alshehri & Alharbi (2023) [40]. In Figure **4a** compared to the collected dataset, we notice that our proposed model significantly outperforms ELA2 (LR+ET+SVM (RBF)) over the other model in all metrics. In Figure **4b** compared to the StatLog UCI dataset [42], we also notice that our proposed model significantly outperforms all metrics. In Figure **4c**

Table 6: Result of the Proposed ELA with Public Datasets

Dataset	FN	TP	TN	FP	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
StatLog UCI	24	126	133	17	0.8667	0.8467	0.8933	0.8925	0.8605	0.9307	0.8594	0.7267	0.7376
Z-Alizadeh Sani	16	200	194	22	0.9121	0.9212	0.9208	0.9052	0.9048	0.9767	0.9144	0.8148	0.8229
CVD	13209	21770	24115	10864	0.6553	0.6221	0.6892	0.6667	0.6459	0.7191	0.6438	0.3117	0.3120



(a) Collected Dataset

(b) StatLog Dataset

(c) Z-Alizadeh Dataset

(d) CVD Dataset

Figure 4: Comparison of different datasets used in the study.

comparing the Z-Alizadeh Sani dataset [7], We note that our proposed model excelled in specificity, AUC, and F1, while obtaining equal results in accuracy, sensitivity, and Kappa. In Figure **4d** Compared to the CVD dataset [43], our proposed model gave lower results than the other model.

5. DISCUSSION

This study demonstrates that the proposed ensemble learning algorithm, a combination of logistic regression, extra trees, and SVM with the RBF kernel, possesses a superior predictive accuracy compared to the other individual classifiers and other ensemble classifiers. Interestingly, the model has put greater emphasis on clinically significant variables that included age, blood pressure, diabetes, dyslipidemia, smoking, chest pain, and electrocardiographic abnormalities. These risk factors are firmly known as key determinants of heart disease and are very common among the Saudi population, with diabetes and obesity playing a significant role in the occurrence of heart diseases [1, 2]. The fact that model-selected features are correlated with recognized regional risk factors establishes the clinical relevance of the findings and hints at the possibility that the proposed model could become a decision-support tool implemented in hospitals.

The proposed ensemble was compared to other published research conducted at competitive performance measures against similar work, and in some few cases, the proposed ensemble was seen to perform better. Abdar *et al*. (2019) [44] obtained the highest accuracy of 94.66% and 98.60% with a nested ensemble SVM model on the benchmark datasets, and 98.78% with sequential feature selection boosting methods. Similarly, Asif *et al*. (2023) [13] demonstrated that the hyperparameter tuning method as an ensemble approach raised the prediction accuracy to 98.15%. Nevertheless, these studies were performed only on publicly available data like Cleveland and Z-Alizadeh Sani, and our research fills an important gap as it is concerned with a local Saudi dataset. That our proposed model attained an accuracy of 91.13% on a relatively small yet region-specific dataset highlights its possible generalizability to real-world clinical practice. Moreover, it is justifiable to elaborate that the characteristics that depend on the nature of the data and the ability to generalize the models to the nonhomogeneous population are necessary and also present in other studies of this nature because the models of the public CVD data only produce inferior results [8, 45].

These positive results are not achieved without some limitations to the research. It is a rather limited dataset (n=108), and although the problem of the imbalanced distribution of classes was addressed using SMOTE, synthetic data augmentation might cause overfitting and deteriorate generalizability. Furthermore, the data were gathered in a relatively small-time range in one hospital, which will not likely reflect the entire range of the patients in the area. The

limitations may be addressed in the future with multicentric data and by using the model on more heterogeneous and larger populations. This study has two implications. First, the findings indicate the viability of ensemble machine learning techniques in supplementing clinical decision-making to diagnose heart diseases. Second, the study will assist in creating more region-specific tools to predict patients and positively impact their outcomes as it will incorporate region-specific data that could be used in Saudi Arabia. The expansion of the clinical variables, the progress of the sample and the subsequent verification of the model in the hospital will become the second step of the work.

6. CONCLUSION

Our study aimed to gather a local dataset comprising heart patients and predict cardiac illness using an ensemble learning approach (ELA). For this purpose, we employed several ML algorithms: SVM four kernels, AdaBoost, RF, DT, LR, KNN, NB, GBC, ET, CatBoost. The method used in feature selection was SMOTE to deal with class imbalance, then two wrapper algorithms were employed: back elimination and forward selection. GridsearchCV was further utilized to optimize hyperparameters for the ML algorithms, which were subsequently integrated into the ELA. The best-performing model was ELA (LR+ET+SVM (RBF)), which achieved strong predictive outcomes after feature selection with Backward + SVM (RBF). This best model received a FN of 8, a TP of 70, a TN of 72, a FP of 6, an accuracy of 0.9113, a sensitivity of 0.8839, a specificity of 0.95, a PPV of 0.9389, a NPV of 0.8878, an AUC of 0.9569, a F1 of 0.9133, a Kappa of 0.8220 and a The proposed model performed well when it was applied to public datasets. On the StatLog UCI dataset, it performed well with all metrics with an accuracy of 0.8667, a sensitivity of 0.8467, a specificity of 0.8933, PPV of 0.8925, NPV of 0.8605, AUC of 0.9307, F1 of 0.8594, Kappa of 0.7267, and MCC of 0.7376. In case of the Z-Alizadeh Sani data, the model with best specificity (0.9208), AUC (0.9767), and F1 (0.9144) is once again an indication of its strength. Although the dataset collected from King Abdullah Hospital comprised 108 patients, which is relatively small, the rigorous use of SMOTE, cross-validation, and feature selection ensured meaningful and reliable results. This focused dataset represents an important first step in developing region-specific predictive models for heart disease. Future work will involve expanding the dataset across multiple hospitals and larger populations to strengthen generalizability and confirm the clinical applicability of the proposed approach. Data access will be made available to reasonable request by the respective author and will be under the approval of the Aseer IRB, Ministry of Health, Saudi Arabia.

REFERENCES

[1] Ministry of Health, Saudi Arabia. Heart disease is the cause of 42% of deaths from non-communicable diseases in the Kingdom [Press release] 2013. https://www.moh.gov.sa

[2] Rath A, Mishra D, Panda G, Satapathy SC. Heart disease detection using deep learning methods from imbalanced ECG samples. Biomedical Signal Processing and Control 2021; 68: 102820. https://doi.org/10.1016/j.bspc.2021.102820

[3] Devi AD, Xavier S. Enhanced prediction of heart disease by genetic algorithm and RBF network. International Journal of Advanced Information Engineering and Technology 2015; 2(2): 29-37.

[4] Djam XY, Wajiga GM, Kimbi YH, Blamah NV. A Fuzzy Expert System for the Management of Malaria. International Journal of Pure & Applied Sciences & Technology 2011; 5(2): 84-102.

[5] Pawlovsky AP. An ensemble based on distances for a kNN method for heart disease diagnosis. In 2018 International Conference on Electronics, Information, and Communication (ICEIC) IEEE 2018; pp. 1-4. https://doi.org/10.1109/ICEIC.2018.8333207

[6] Janosi A, Steinbrunn W, Pfisterer M, Detrano R. UCI machine learning repository-heart disease data set [Data set]. University of California, Irvine 1988.

[7] Alizadehsani Z, Alizadehsani R, Roshanzamir M. Z-Alizadeh Sani data set [Data set]. UCI Machine Learning Repository 2017. https://archive.ics.uci.edu

[8] Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked 2019; 16: 100203. https://doi.org/10.1016/j.imu.2019.100203

[9] Atallah R, Al-Mousa A. Heart disease detection using machine learning majority voting ensemble method. In 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS) IEEE 2019; pp. 1-6. https://doi.org/10.1109/ICTCS.2019.8923053

[10] Lapp D. Heart Disease Dataset [Data set]. Kaggle 2019. https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

[11] Li R, Shen S, Chen G, Xie T, Ji S, Zhou B, Wang Z. Multilevel risk prediction of cardiovascular disease based on Adaboost+RF ensemble learning. IOP Conference Series: Materials Science and Engineering 2019; 533(1): 012050. https://doi.org/10.1088/1757-899X/533/1/012050

[12] Chaurasia V, Chaurasia A. Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique. Biomedical Materials & Devices 2023; 1(2): 932-941. https://doi.org/10.1007/s44174-023-00085-w

[13] Asif D, Bibi M, Arif MS, Mukheimer A. Enhancing heart dis-ease prediction through ensemble learning techniques with hyperparameter optimization. Algorithms 2023; 16(6): 308. https://doi.org/10.3390/a16060308

[14] Ahmed R. Heart Disease [Data set]. Kaggle 2020. https://www.kaggle.com/datasets/data855/heart-disease

[15] Cherngs. Heart disease Cleveland UCI [Data set]. Kaggle 2020. https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci

[16] Ganie SM, Pramanik PKD, Malik MB, Nayyar A, Kwak KS. An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms. Computer Systems Science and Engineering 2023; 46(3): 3993-4006. https://doi.org/10.32604/csse.2023.036209

[17] Aziz S, Afreen N, Akram F, Ahmed M. A Framework for Cardiac Arrest Prediction via Application of Ensemble Learning Using Boosting Algorithms. Procedia Computer Science 2024; 235: 3293-3304.

[18] Narayanana J. Implementation of Efficient Machine Learning Techniques for Prediction of Cardiac Disease using SMOTE. Procedia Computer Science 2024; 233: 558-569.

[19] Musa IR, Omar SM, Sharif ME, Ahmed ABA, Adam I. The calculated versus the measured glycosylated haemoglobin (HbA1c) levels in patients with type 2 diabetes mellitus. Journal of Clinical Laboratory Analysis 2021; 35(8): e23873. https://doi.org/10.1002/jcla.23873

[20] Beaulieu-Jones B, Greene CS, Consortium P. A new analy-tical framework for missing data imputation and classification with uncertainty. PLOS One 2022; 17(3): e0264238. https://doi.org/10.1371/journal.pone.0264238

[21] Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, Marrero J, Zhu J, Higgins PDR. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 2013; 3(8): e002847. https://doi.org/10.1136/bmjopen-2013-002847

[22] Zhang Y, Zhang J, Gong C, *et al*. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access 2020; 8: 182459-182472. https://doi.org/10.1109/ACCESS.2020.3028736

[23] García S, Luengo J, Herrera F. Data preprocessing in data mining. Springer 2015. https://doi.org/10.1007/978-3-319-10247-4

[24] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011; 12: 2825-2830.

[25] Tang J, Alelyani S, Liu H. Feature selection for classification: A review. In Data classification: Algorithms and applications. CRC Press 2014; pp. 37-64.

[26] Hasan N, Bao Y. Comparing different feature selection algorithms for cardiovascular disease prediction. Health and Technology 2021; 11(1): 49-62. https://doi.org/10.1007/s12553-020-00493-8

[27] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research 2003; 3: 1157-1182.

[28] Dey A, Ashour AS, Bhattacharya S. Machine learning techniques for heart disease prediction: A comparative study. In 2017 International Conference on Electronics, Communication and Aerospace Technology (ICECA) IEEE 2017; pp. 547-550. https://doi.org/10.1109/ICECA.2017.8203711

[29] Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications 2009; 36(4): 7675-7680. https://doi.org/10.1016/j.eswa.2008.09.013

[30] Pattekari SA, Parveen S. Prediction system for heart disease using Naive Bayes. International Journal of Advanced Computer and Mathematical Sciences 2012; 3(3): 290-294.

[31] Rajkumar A, Reena GS. Diagnosis of heart disease using machine learning algorithms. International Journal of Research in Engineering and Technology 2010; 2(6): 741-744.

[32] Ali L, Zhu C, Zhou M, Javeed A. Reliable Parkinson's disease detection by using an intelligent system based on L2-regularized logistic regression and extra trees classifier. Future Generation Computer Systems 2019; 97: 238-252. https://doi.org/10.1016/j.future.2019.02.012

[33] Schapire RE. Explaining adaboost. In Empirical Inference. Springer 2013; pp. 37-52. https://doi.org/10.1007/978-3-642-41136-6_5

[34] Soofi AA, Awan A. Classification techniques in machine learning: Applications and issues. Journal of Basic and Applied Sciences 2017; 13: 459-465.

[35] Jakkula V. Tutorial on support vector machine (SVM) (Technical Report). School of EECS, Washington State University 2006.

[36] Mitchell TM. Machine learning. McGraw-Hill 1997.

[37] Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. Expert Systems with Applications 2019; 134: 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

[38] Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. Frontiers of Computer Science 2020; 14(2): 241-258. https://doi.org/10.1007/s11704-019-8208-z

[39] Patil S, Bhosale S. Hyperparameter tuning based performance analysis of machine learning approaches for prediction of cardiac complications. In International Conference on Soft Computing and Pattern Recognition. Springer 2020; pp. 605-617. https://doi.org/10.1007/978-3-030-73603-3_55

[40] Alshehri GA, Alharbi HM. Prediction of heart disease using an ensemble learning approach. International Journal of Advanced Computer Science and Applications 2023; 14(8). https://doi.org/10.14569/IJACSA.2023.0140819

[41] Ullah T, Ullah SI, Ullah K, Ishaq M, Khan A, Ghadi YY, Algarni A. Machine learning-based cardiovascular disease detection using optimal feature selection. IEEE Access 2024. https://doi.org/10.1109/ACCESS.2024.3361234

[42] Sumbria S. Statlog (Heart) Data Set [Data set]. Kaggle 2019. https://www.kaggle.com/datasets/shubamsumbria/statlog-heart-data-set

[43] Ulianova S. Cardiovascular Disease dataset [Data set]. Kaggle 2019. https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

[44] Abdar M, Acharya UR, Sarrafzadegan N, Makarenkov V. NE-nu-SVC: A new nested ensemble clinical decision support system for effective diagnosis of coronary artery disease. IEEE Access 2019; 7: 167605-167620. https://doi.org/10.1109/ACCESS.2019.2953920

[45] Yewale D, Vijayaragavan SP, Bairagi VK. An Effective Heart Disease Prediction Framework based on Ensemble Techniques in Machine Learning. International Journal of Advanced Computer Science and Applications 2023; 14(2). https://doi.org/10.14569/IJACSA.2023.0140233

Received on 26-04-2025 Accepted on 29-07-2025 Published on 16-09-2025

https://doi.org/10.6000/1929-6029.2025.14.52