AN EFFECTIVE EARLY WARNING ATTEMPT OF HEART FAILURE WITH SIGNIFICANT FEATURES AND PROMISING COMBINATION METHODS OF MACHINE LEARNING

BY

Ananda Sutradhar
ID: 191-15-2404
AND

Mustahsin Al Rafi
ID: 191-15-2415

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Mohammad Jahangir Alam
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Hasin Rehana
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

FEBRUARY 2023

APPROVAL

This Project/internship titled **"An effective early warning attempt of heart failure with significant features and promising combination methods of machine learning"**, submitted by Name: Ananda Sutradhar, ID No:191-15-2404 and Name: Mustahsin Al Rafi, ID No:191-15-2415 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *04/02/2023*.

BOARD OF EXAMINERS

7/2/23

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. S. M. Aminul Haque
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dewan Mamun Raza
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

4/2/2023

External Examiner

Dr. Shamim H Ripon
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mohammad Jahangir Alam, Senior lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Mohammad Jahangir Alam
Senior lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Hasin Rehana
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Ananda Sutradhar
ID: 191-15-2404
Department of CSE
Daffodil International University

Mustahsin Al Rafi
ID: 191-15-2415
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing making us possible to complete the final year research paper successfully.

We are really grateful and wish our profound indebtedness to **Mohammad Jahangir Alam**, **Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "IOT, Machine Learning, and Deep Learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Hasin Rehana**, **Lecturer**, Department of CSE Daffodil International University for her kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to extend our sincere gratitude to Dr. Touhid Bhuiyan, Professor and Head, Department of CSE, as well as to the other professors and employees of the CSE department of Daffodil International University, for your kind assistance in finishing our research.

We would like to thank our entire course mate Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Heart failure (HF) is currently the leading cause of morbidity and mortality worldwide. Diagnosis of a medical condition is difficult and time-consuming in medical science. Whereas Machine learning (ML) techniques can help reduce HF's mortality rate by providing early warnings. It would be more promising and accurate when we have significant data and features. In this paper, we incorporate different ML methods with significant features which can serve as warnings at the early stages. Initially, general preprocessing techniques are applied in the Kaggle heart failure dataset and introduce the SMOTETOMEK-BOOST method for handling imbalanced class problems. Then two well-known feature selection techniques Feature Importance by Random Forest and Information Gain are applied purpose of reducing the dimensions of the data and selecting the most significant features. All different feature sets are trained with Decision Tree (DT), Extra Tree (ET), Gradient Boost (GB), and Support Vector Machine (SVM), along with presenting a hybrid classifier named CBCEC by combining the best-performing classifier with two ensemble methods. Experimental results demonstrate that the proposed CBCEC model performs the highest results of 93.67% accuracy with Feature Importance (FI) based feature selection. Finally, explain the global behaviors of the best-performing features set by applying an explainable method named the Partial Dependence Plot (PDP).

Keywords: Heart failure, SMOTETOMEK-BOOST, Feature selection, Ensemble method, Explainable AI.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Recently heart disease is the most common and becoming the leading cause of disease worldwide [1]. The healthcare systems are facing tremendous risk and burden due to the rise in heart disease with high death rates. Various risk factors such as diabetes, thyroid, high blood pressure, high cholesterol, and abnormal pulse are contributing to heart disease is difficult to identify and growing rapidly [2]. Several non-lifestyle risk factors, such as age, gender, family history, and high levels of fibrinogen, must also be considered [3]. The symptoms include weight gain, nausea, extreme fatigue, shortness of breath, chest pain, anxiety, weakness in the legs or arms, etc [4].

Women are felling heart failure (HF) more than men [5] and older are at absolute risk than younger for developing HF [6]. Worldwide every year approximately 17.9 million people die cause of the cardiovascular disease (CVD) which has a higher prevalence in Asia [7]. One person dies every 36 seconds in the United States due to CVD [8]. Basically, the death rate for individuals with heart failure after discharge from the hospital is 10.4% at 30 days, 22% at 1 year, and 42.3% at 5 years [9].

1.2 Motivation

It is crucial to examine the sign of heart disease as soon as possible to start management with counseling and medicines. Electrocardiograms and angiograms are considered the most standard and noninvasive tests to examine heart disease. Although it is quick and simple to do, it may overlook asymptomatic people and instead diagnose those with a normal electrocardiogram rhythm. Additionally, there are certain drawbacks to using the electrocardiogram as a prognostic tool to forecast future CHD [3]. On the other hand, angiograms are quite expensive, limiting accessibility to low-income families. Some additional tests might be required including blood tests, chest X-rays, echocardiogram, blood pressure monitoring, and stress tests to examine heart disease, these are very complex, time-consuming, and expensive.

As a result, researchers are nowadays concentering on machine learning (ML) techniques to diagnose heart disease, which can save time, money, resources, many lives, and burden on clinicians. Early identification of HF would make it possible to explore pharmacological and lifestyle changes that might reduce the course of the disease and enhance patient outcomes. ML algorithms are one of the most significant developments in recent years to avoid many diseases by providing early warnings. It is used in the process of examining data to uncover hidden information that can be utilized to make critical decisions in the future.

Numerous methods have been investigated by researchers to predict heart failure from the same dataset that we selected. Such as Zahid et al. [11] proposed two different gender-based models to predict mortality. Chicco et al. [12] analyzed the performance of machine learning with only two features. However, should also concenter on other significant features which have a high impact on heart failure. Abid et al. [8] explore the most risk factor using a feature selection technique. They utilized SMOTE to overcome data imbalance techniques. But SMOTE might generate noisy and uninformative samples [10], where ML algorithms are more effective when trained on proper data. Afterward, Minh [13], and Saurav Mishra [14] worked with this dataset for proper survival prediction. However, still, numerous improvements are needed in this area, including selecting the most significant features, improving results with an effective classifier, and exploiting the hidden factors using explainable AI, though none of the existing studies utilized any explainable technique.

1.3 Rationale of the Study

The main objective of this research is to warn at the early stages which reduce the mortality rate of HF by an effective classifier with proper data and significant features. Hence balance the target class, we propose SMOTETOMEK-Boost, a combined method of SMOTETomek (SMOTE for over-sampling and Tomek links for under-sampling) and boosting. Where SMOTETomek is a fantastic technique to get away from SMOTE's drawbacks [30]. Substantially, combining over and under-sampling strategies with an ensemble classifier increases the effectiveness of data [21]. Feature Importance [8] [16] and Information Gain [14] [18] are utilized to extract significant features of the disease. For training data, we employ four well-known traditional classifiers named Decision Tree, Gradient Boost, Support Vector Machine, and Extra Tree classifiers. Specially, we proposed one hybrid classifier, which combined two ensemble classifiers name Bagging (BG) and Voting (VT) with the best-performing general classifier. While the BG method is useful for reducing variance with maintaining bias [29]. On the other hand, the VT method would be more significant when using two or more classifiers as base estimators [31], here we have done it as. Alongside, the main motive for applying these two combination methods (SMOTETOMEK-Boost and CBCEC) is existing studies [3] [8] [17] are recommended using multiple combinational machine learning and ensemble models in the future to predict HF. However, the major contributions of this research are as follows:

- Overcome data imbalance issues by SMOTETomek, which is a hybrid method of oversampling and under-sampling. At the same time injects this method at each boosting iteration by Ada Boost classifiers.
- Propose a hybrid classifier by combining the best-performing classifiers with some ensemble methods of Bagging and Voting.
- Generate the global explanations of each feature based on the output features by PDP, so that stakeholders are notified to know the riskiest value range.

1.4 Research question

- What is the status and impact of HF worldwide?
- Which phases are in the existing research??
- How are the SMOTE flaws addressed by the Tomek-link and Boosting methods?
- Which features are most significant for interpreting HF results?
- For which set of parameters the general classifiers are performed best?
- How does the combined ensemble classifier get multiple advantages compared to the individual classifier?
- What is the riskiest range of features, which affected HF mostly?

1.5 Expected Outcomes

- Proposed a combined method to balance data and overcome the drawbacks of SMOTE.
- Identify the most significant features of HF**.**
- Determine the best-performing general classifier for combining our intended proposed classifier**.**
- Proposed an effectively combined classifier that can able to outperform the general classifiers.
- Generate the proper dependence between the target and input feature, so that stockholders would aware of his cases.
- The major aim is effectively to warn HF at the early stages to reduce mortality.

1.6 Research Objectives

- To make classification smooth and efficient
- To generate better outcome comparing to the other related works
- To raise public awareness to the heart failure patient

1.7 Report Layout

The contains of this research paper are as follows:

- In chapter 1 we have discussed about introduction, motivation, rationale, question, and objective of our research.
- Chapter 2 examines the related works as well as comparative analysis and summary of existing research.
- Research methodologies are discussed in chapter 3, which includes the research subject and instruments, data collection, data preprocessing, feature selection, ML classifiers, and hyperparameter tuning.
- The experimental results and discussion are held in chapter 4.
- The 6^{th} chapter described the conclusion and future work part.

CHAPTER 2
BACKGROUND STUDY

2.1 Related Works

Nowadays ML can serve to solve a vast number of problems in the healthcare area. There are some previous studies already on HF, where researchers commonly used some ML techniques to compare different methods, try to generate high-risk features, and detect or predicted HF.

For example, Lorenzoni et al. [15] compared the performance of eight machine learning classifiers including support vector machine (SVM) for the prediction of heart failure patients. They find the highest accuracy from the GLMN classifier as 81.2%. Minh et al. [13] compared the results of seven machine learning classifiers including SVM and Decision Tree (DT) after applying the grey wolf optimization feature selection method. They got the highest accuracy of 85% from the random forest classifiers. After comparing all results, it is observed that the decision tree generates the highest accuracy of 85.33%.

ABID et al. [8] tried to find significant features with some effective data mining techniques for boosting the accuracy of HF patients. SMOTE and feature importance methods were employed for data balancing and finding the most risk features respectively. Then applied a variety of classification models, including Decision Tree, Gradient Boost, Extra Tree, and SVM, where Extra Tree outperforms other models and achieved a 0.9262 accuracy. Lal Hussain et al. [16] used high-rank features to detect HF or normal class. They employ a variety of highly potent machine learning approaches, where SVM achieves the top detection performance in terms of 88.79% accuracy when using all multimodal features.

Dafni et al. [18] used some ML approaches to address the HF diagnosis. This work was ongoing on several preprocessing steps including missing values removal, outlier detection, and balance data. LMT and ROT classifiers with various feature combinations produced the best accuracy results even though when using simple clinical features.

A few researchers presented some hybrid and ensemble methods in their studies. Such as Mohan et al. [17] presented a hybrid model (HRFLM) for the prediction of heart disease. In order to enhance the training of machine learning models, the authors also proposed a novel feature selection strategy. The presented model was found to be 88.7% accurate. The two-tier ensemble model, devised by Tama et al. [3] uses some classifiers as base classifiers for another ensemble. Extreme Gradient Boosting, Random Forest, and Gradient Boosting Machine class labels are used to create the suggested stacking architecture. Four different types of datasets are used to evaluate their suggested detection model, where this model generated robust results. Pronab et al. [7] employed some ensemble methods like bagging and boosting for the effective prediction of heart disease. Random Forest, Decision Tree, Gradient Boost, KNN, and Ada Boost were employed, along with integrating these classifiers with the bagging and boosting method. Comparison of all classifiers RFBM produced the highest accuracy. Raza et al. [19] proposed an ensemble architecture with a majority vote. To forecast heart illness in a patient, it incorporated logistic regression, multilayer perceptron, and naive Bayes. A classification accuracy of 88.88% was attained, which was outperforming all base classifiers.

2.2 Comparative analysis and summary

This review section showed that previous works have tried to predict heart failure and heart disease on different datasets, table 1 holds an overall summary. For comparison of ML methods [13] [15] have used multiple ML methodologies. However, they needed to improve accuracy with better evaluation for usefulness in clinical practice. Other works [8] [16] purpose to determine the most significant features to classify HF. However, applying two or more different feature selections can help to select the most significant features by comparing their results. Some hybrid classifiers were proposed in [3] [7] [17] [19] by combining the general classifiers with ensemble methods. They combined the general classifier with only one ensemble classifier and compared their results. Where combining two or more ensemble classifiers in the proper order might enhance the multiple advantages and makes the efficient classifier. Some well-known traditional classifiers DT [3] [7] [8] [13] [17], GB [3] [7] [8] [17], and SVM [8] [13] [16] have been used mostly. However, we have implemented these classifiers along with several different ML approaches. The hyper parameter tuning is utilized to improve the performing outcome such as accuracy, precision, recall, f1-score, and AUC score.

TABLE 1: Summary of the existing researches.

Year & Reference	Data Collection	Number of instances	Classes	Reduce Imbalance Issues	Performing Classifiers	Best Performing Classifier	Performance Results
2022 [14]	Faisalabad Institute of Cardiology	299	2	SMOTE	SVM, DT, RF, LGBM	SVM	ACC = 83.33%, PRE = 86.36%
2021 [8]	UCI	299	2	SMOTE	DT, RF, ETC, SVM, GB	ETC	ACC = 0.9262%
2021 [17]	Physionet databases	--	3	--	DT, SVM Gaussian,	SVM Gaussian	ACC = 88.79%, AUC = 94.41%
2021 [13]	UCI	299	2	--	DT, SVM, KNN, RF, GWO-LMP	GWO-MLP	ACC = 87%
2021 [18]	(UCD) Ireland and University Hospital of Ioannina	487	3	SMOTE Undersampling	DT, RF, KNN, SVM, LMT, ROT	ROT	ACC = 91.23%, REC = 93.83%
2020 [11]	Faisalabad Institute of Cardiology	299	2	--	RF, DT, SVM, KNN, GB	RF	ACC = 74%, AUC = 80%
2020 [3]	Z-Alizadeh Sani, Statlog, Cleveland, Hungarian	303, 261, 303, 294	2	--	RF, DT, PSO, GB	PSO	ACC = 98, 93, 85, 91%

2019 [23]	UCI	270	2	--	LR, NB, MLP, Voting	Voting	ACC = 88.88%
2019 [21]	UCI	303	2	--	DT, RF, SVM, GB, HRFLM	HRFLM	ACC = 88.7%

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

In this paper, our main aim is to incorporate different ML methods with significant features which can serve as warnings at the early stages. This research is across several machine learning approaches like data collection, data preprocessing, feature selection, ML classifiers, and explainable AI. Figure 1 holds the working methodology of our study. Some necessary libraries are utilized for this study like Python 3.5, NumPy, Pandas, GridsearchCV, etc.

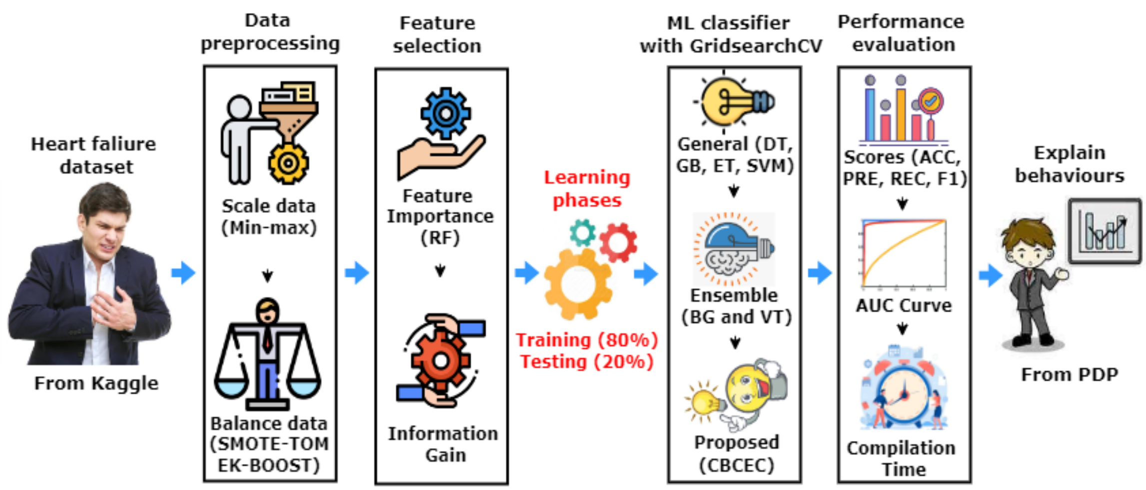

Figure 1: Flow of the working procedure.

3.2 Data Collection

For this research, a heart failure clinical dataset is driven from the Kaggle data repository [20]. This dataset contains 299 medical records with 13 clinical features, these are collected during the follow-up period. The last feature name DEATH_EVENT is the target class, 1 is for dead and 0 is for alive. Where 203 dead cases and 96 alive cases are reported. The full overview of this dataset is provided in table 2.

TABLE 2: Dataset explanations.

Feature Name	Explanation	Measurement	Range
Age	Patient age	Years	40 - 95
Anaemia	Decrease of red blood cells or hemoglobin	Boolean	0 (no), 1 (yes)
High blood pressure (H_b_p)	If the patient has blood pressure	Boolean	0 (no), 1 (yes)
Creatinine phosphokinase (Cr_ph)	Level of the CPK enzyme in the blood	Mgc/L	23-7861
Diabetes	If the patient has diabetes	Boolean	0 (no), 1(yes)
Ejection fraction (Ej_fr)	Blood leaving percentage	Percentage	14-80
Sex	Man or woman	Binary	0 (woman), 1 (man)
Platelets	Platelets in the blood	kiloplatelets/mL	25.01 - 850.00
Serum creatinine (Se_cr)	Level of creatinine in the blood	mg/dL	0.50 - 9.40
Serum sodium (Se_so)	Level of sodium in the blood	mg/dL	114 - 148
Smoking	If patients smoke	Boolean	0 (no), 1(yes)
Time	Follow-up period	Days	4 - 285
DEATH_EVENT (target)	If the patient died in the follow-up period	Boolean	0 (alive), 1(dead)

3.3 Data Preprocessing

The selected dataset for this study is almost clean and preprocessed. There are no missing values are containing in this dataset. However, Creatinine phosphokinase and Platelets features have huge differences between one value from another. It may delay decision-making, and overcome this issue by min-max scaling. Which converts the feature values into a range, quickly learns an algorithm, and is essential for improving results.

One more issue we need to handle is the imbalance of the dataset. The synthetic minority oversampling technique (SMOTE) is one of the famous approach for balancing data and researchers mostly use it [8] [14]. But SMOTE has the potential to produce noisy and useless samples. SMOTE-Tomek is a combination of over and under-sampling techniques for dealing with imbalance issues and overcoming the drawbacks of SMOTE [30]. Which combines SMOTE to produce synthetic data for the minority class and Tomek connections to eliminate the data that the majority class has designated as Tomek links. It is more effective when combining over and under-sampling techniques with an ensemble classifier [21]. Since data from the minority class is frequently misclassified, more weight is added to this class with each iteration, boosting algorithms particularly beneficial for this problem [22]. Hence applied the boosting procedure AdaBoost (AB) at the same time with SMOTE-Tomek, which injects SMOTE-Tomek at each boosting iteration. The advantage of this method is that SMOTE-Tomek provides more samples of the minority class at each boosting stage while boosting offers equal weights to all misclassified data. Fig 2 illustrates the process of SMOTETOMEK-Boost.

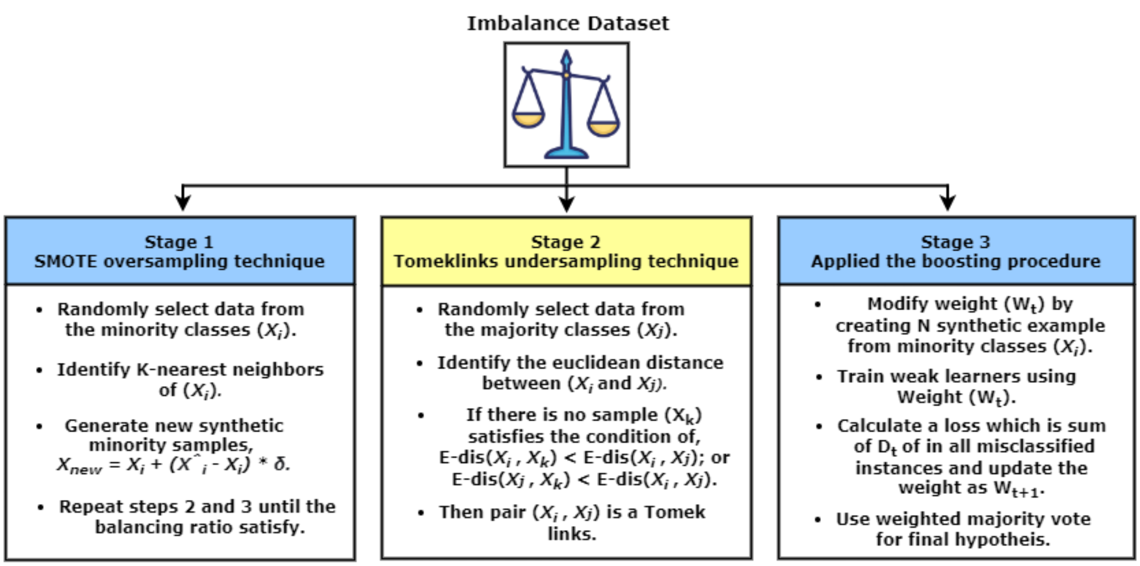

Figure 2: Process of SMOTETOMEK-BOOST.

3.4 Feature Selection

By choosing the most crucial variables and removing irrelevant features, feature selection enhances machine learning and boosts the prediction power of machine learning algorithms. Here, feature importance and information gain are employed to choose the significant features.

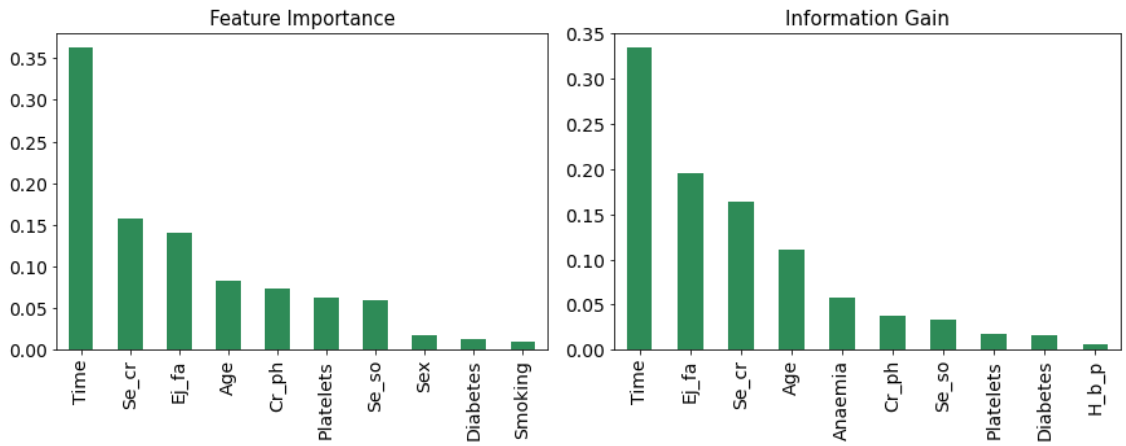

Figure 3: Significant features from two feature selection methods.

The selection method known as feature importance (FI) assigns a score to input features depending on how essential they are to the prediction of the outcome. In essence, it demonstrates the extent to which a particular variable is beneficial for a given model and prediction. As a technique for ML model interpretability, FI is also frequently employed. Random forest is fit with the FI method to evaluate the feature ranking. Information gain (IG) is mostly used to compute each variable's gain of the target variable. It is mainly an entropy-based feature selection method. The major factor used to determine IG is how much of a phrase may be used to categorize information. We have selected the top 10 features from these feature selections based on their importance rank. Figure 3 has demonstrated these features. Then, the processed dataset and the reduction of feature sets are split into 80% for training and 20% for testing.

3.5 ML Classifiers Description

For detecting heart failure, we have employed four traditional classifiers such as decision tree, gradient boost, support vector machine, and extra tree classifier. Along with presenting a combinational ensemble classifier named CBCEC.

3.5.1 Decision Tree

Decision Tree (DT) is thought the most well-known technique for representing classifiers in data classification. It is mostly used to handle non-linear data sets efficiently and is helpful for machine learning since they divide complex data into easier-to-handle components [23]. DT is one of the effective techniques frequently employed in a variety of domains, including pattern recognition, image processing, and machine learning [24].

3.5.2 Gradient Boost

Gradient boost (GB) is a method that stands out for many weak classifiers working together to create a strong classifier. It works based on the concept of the decision tree. The decision criteria of XGBoost, a scalable ensemble method based on gradient boosting, are identical to those of decision trees [25]. GB can strong enough to uncover nonlinear relationships between any model target and features [26].

3.5.3 Support Vector Machine

Support vector machine (SVM) is a potent yet adaptable supervised machine learning technique that, in essence, represents several classes in a multidimensional space using a hyperplane. Models based on SVM are particular difficulties that can benefit from measurements that can easily be utilized to refine the remedy [27]. Normal SVM is not appropriate for classifying huge data sets, despite its strong theoretical underpinnings and high classification accuracy.

3.5.4 Extra Tree

Extra tree (ET) classifier fit randomized decision trees on various subsamples of the dataset that were primarily based on decision trees. The initial training sample is used to build each decision tree in the ET forest. It utilizes the idea of averaging to improve accuracy as well as control over data fitting [28].

Algorithms 1: Illustrates the procedure of our proposed classifiers.

Input: Number of base classifiers, $BC = BC_1\ to\ BC_4$. Number of bootstrap samples = B.

Training data, $\mathrm{D_{train}} = \sum_{i=1}^{n}(\mathrm{a_i}, \mathrm{b_i})$.

Output: Combined classifier $CBCEC$ to classify having risk of heart failure or not.

START:

Step1: Compute the best performing classifier from traditional classifiers.

$\boldsymbol{for}\ i = 1;\ i <= 4;\ i++\ \boldsymbol{do}$

$\mathrm{B-PC} = \mathrm{Train}\{\mathrm{DT(D_{train}),\ GB(D_{train}),\ SVM(D_{train}),\ ET(D_{train})}\}$

$\boldsymbol{end\ for}$

$\mathrm{B-PC} = \mathrm{Max_{acc}}\{\mathrm{DT(D_{train}),\ GB(D_{train}),\ SVM(D_{train}),\ ET(D_{train})}\}$

Step2: Apply the best performing classifier as a base estimator on Bagging.

$\boldsymbol{for}\ j = 1;\ j <= B;\ j++\ \boldsymbol{do}$

$D_j, \ldots \ldots D_B = Boostrap(D_{train})$

$\boldsymbol{end\ for}$

$\boldsymbol{for}\ b = 1;\ b <= B;\ b++\ \boldsymbol{do}$

$\mathrm{B-BG} = \mathrm{aggregate}\{\mathrm{B-PC(D_b)}, \ldots \ldots, \mathrm{B-PC(D_B)}\}$

$\boldsymbol{end\ for}$

Step3: Combine the best performing classifier and the integrate classifiers of Bagging with soft Voting.

$\boldsymbol{for}\ m = 1; m <= 2;\ m++\ \boldsymbol{do}$

$\mathrm{CBCEC} = \mathrm{agrmax}\{\mathrm{B-PC(D_{train}), B-BG(D_{train})}\}$

$\boldsymbol{end\ for}$

$\boldsymbol{Return}\ CBCEC$

$\boldsymbol{END}$

3.5.5 Combining Best Classifier with Two Ensemble Classifiers

Nowadays researchers would like to train data with hybrid or combined classifiers to get multiple benefits [3] [8] [17], while individual classifiers are sometimes does not able to reach the desired level [43]. Hence proposed CBCEC classifier, which is a combination of one general and two ensemble classifiers BG and VT. The ensemble classifiers would be grateful to reduce overfitting and underfitting issues [44]. BG mainly works on bootstrapping (creating some bootstrap data samples from the data) and aggregating (aggregating the individual prediction from each bootstrap sample). VT works on training multiple models together and combining the predictions.

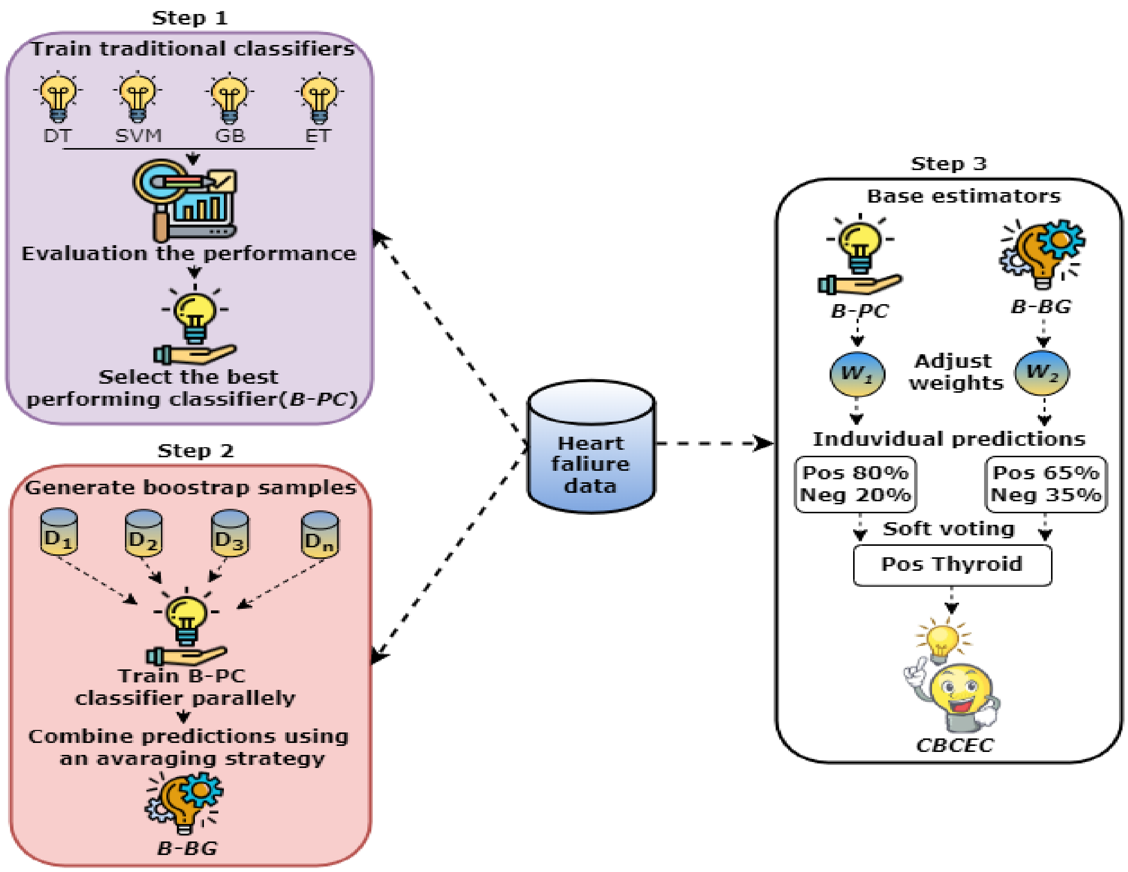

Figure 4: Working diagram of our proposed model.

First of all, we determine the best-performing classifier by comparing the results of our four general classifiers as $B-PC$. Then set the $B-PC$ as a base estimator and parallelly fit from the generated bootstrap samples of BG, let as B-BG.

$$B-BG = \{B-PC(D_b), \dots \dots, B-PC(D_B)\}$$

Here, D_b and D_B are the first and last bootstrap samples respectively. This method could be superior in reducing variance [29]. Another ensemble method VT can perform well when two or more base classifiers are integrated together [31]. Hence finally integrates best-performing classifiers (B-PC) and hybrid boosting classifier (B-BG) with the help of the soft voting. This type of voting works with all classifiers and generates the average probability score for all classes, finally the highest average prediction is selected to create the final prediction. Evaluate as,

$$CBCEC = agrmax\{B - PC(D_{train}), B - BG(D_{train})\}$$

Here, D_{train} is the traning instances, algorithm 1 holds the procedure of $CBCEC$ classifier. For better understanding a working diagram of our proposed algorithm are provided in figure 4.

3.6 Hyper parameter Tuning

Substantially, we utilized the hyperparameter technique to control the learning process. It significantly reduces the loss of function and improves the performance of the ML model [32]. In the tuning process, we used the GridSearchCV method, which can search through a model's optimal parameter values from the given grid of parameters. Table 3 illustrates the best parameter of our employed traditional classifiers for different feature sets. Afterward, the best performing classifier is integrated into some ensemble classifiers with the same parameters and set as the default parameter of all ensemble classifiers during the proposed model.

TABLE 3: Utilized parameters for different feature sets.

	DT	**GB**	**SVM**	**ET**	**CBCEC**
All features	max_depth = 5, random_state = 5, criterion = ‘entropy’, splitter = ‘best’	max_depth = 1, random_state = 5, loss = ‘log_loss’, learning_rate = 0.2	degree = 1, random_state = 5, kernel = ‘linear’	n_estimator = 10, random_state = 10, criterion = ‘entropy’, max_features = ‘sqrt’	BG = [base_estimator = ‘B-PC’, random_state=10] VT = [base_estimator = ‘B-PC, B-BG’, type = ‘soft’]

FI features	max_depth = 8, random_state = 15, criterion = 'gini', splitter = 'best'	max_depth = 3, random_state = 20, loss = 'log_loss', learning_rate = 0.1	degree = 1, random_state = 5, kernel = 'rbf'	n_estimator = 8, random_state = 15, criterion = 'gini', max_features = 'sqrt'	BG = [base_estimator = 'B-PC', random_state=10] VT = [base_estimator = 'B-PC, B-BG', type = 'soft']
IG features	max_depth = 5, random_state = 20, criterion = 'gini', splitter = 'best'	max_depth = 4, random_state = 20, loss = 'log_loss', learning_rate = 0.1	degree = 1, random_state = 10, kernel = 'rbf'	n_estimator = 7, random_state = 10, criterion = 'gini', max_features = 'sqrt'	BG = [base_estimator = 'B-PC', random_state=10] VT = [base_estimator = 'B-PC, B-BG', type = 'soft']

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we discussed all the experimental results for our proposed work. For proper evaluation, several classification results are measured for all features, FI and IG-based selected features, after applying SMOTETOMEK-BOOST.

4.1 Comparison of all the performed results

We compare all the classifier's results on different three feature sets based on accuracy, precision, recall, f1-score, AUC score, and computational time.

4.1.1 Accuracy

The accuracy of a machine learning model is a measurement used to assess which model performs the best at finding relationships and trends between variables in a dataset based on their training data. Fig 4 shows the accuracy of different classifiers like DT, GB, SVM, ET, and CBCEC respectively.

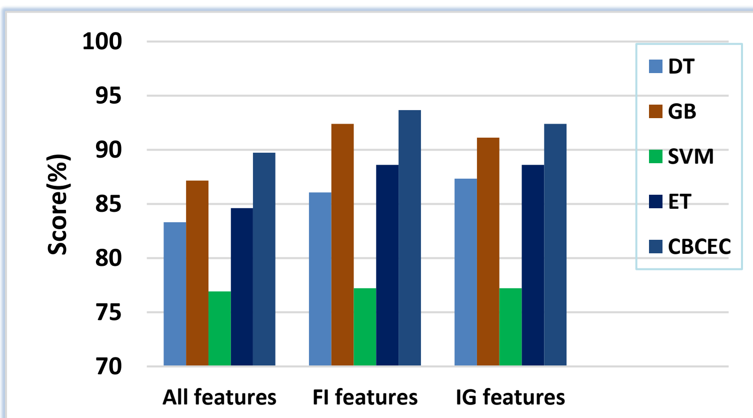

Figure 5: Accuracy on the different feature sets.

Considering all features, the best accuracy is obtained by CBCEC which is 89.74%, whereas ET, GB, and DT gain 84.61%, 87.17%, and 83.33% respectively. Then when considering FI features DT and ET gained almost the same result of 86.07% and 88.60%.

However, results are seriously better for the proposed classifier CBCEC with 93.67%. We get 87.34%, 88.60%, and 77.21% accuracy for DT, ET, and SVM classifiers respectively with the IG features. Comparing those features on different classifiers based on the accuracy we found our proposed classifier CBCEC obtained outperformed with FI-based features.

4.1.2 Precision

Precision refers is how good the model is at predicting a specific category. It is determined by dividing the total number of correctly anticipated positive examples by the ratio of correctly predicted positive examples.

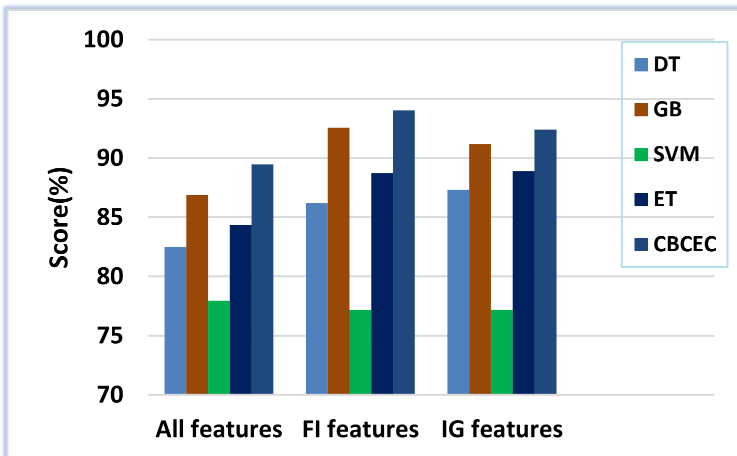

Figure 6: Precision on the different feature sets.

Considering all features, the proposed CBCEC classifiers performed the highest precision score as 89.47%. When applied to the FI features, GB and CBCEC gained good accuracy of 92.57% and 94.02%. For the IG features, DT and ET achieved almost the same result. Here, SVM gained the lowest result between 77 to 78% in all different feature sets.

4.1.3 Recall

The recall gauges how well the model can identify Positive samples. The more positive samples that are identified, the larger the recall. Fig 6 shows the recall scores for the different algorithms and feature sets.

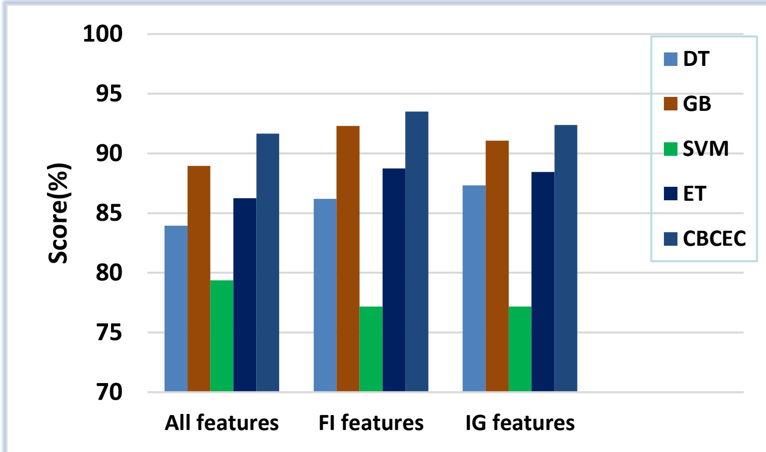

Figure 7: Recall on the different feature sets.

4.1.4 F1-Score

One of the most essential evaluation metrics in machine learning is the f1-score. It is the harmonic mean of precision and recall. The outcomes of the f1-Score are displayed in Fig 7.

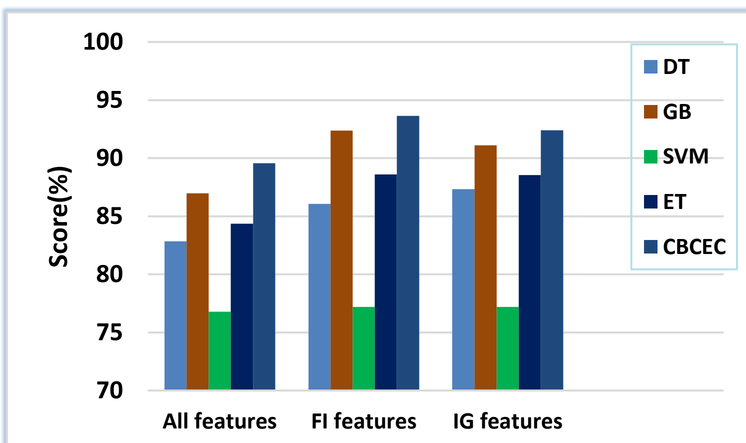

Figure 8: F1-score on the different feature sets.

For All features, DT, GB, ET, and CBCEC are performed mostly 80% to 90%. Considering IG features the highest f1-score is 92.39% achieved with the CBCEC and the lowest F1 Score is 77.18% generated with SVM, the same as like in the other two feature sets.

4.1.5 AUC score

Another important evaluation metric is AUC, which is computed by adjusting the value in the matrix. In fig 8 the AUC plot shows for three different feature sets with the score. Where CBCEC classifiers generated the overall highest AUC score of 98% with FI-based selected features.

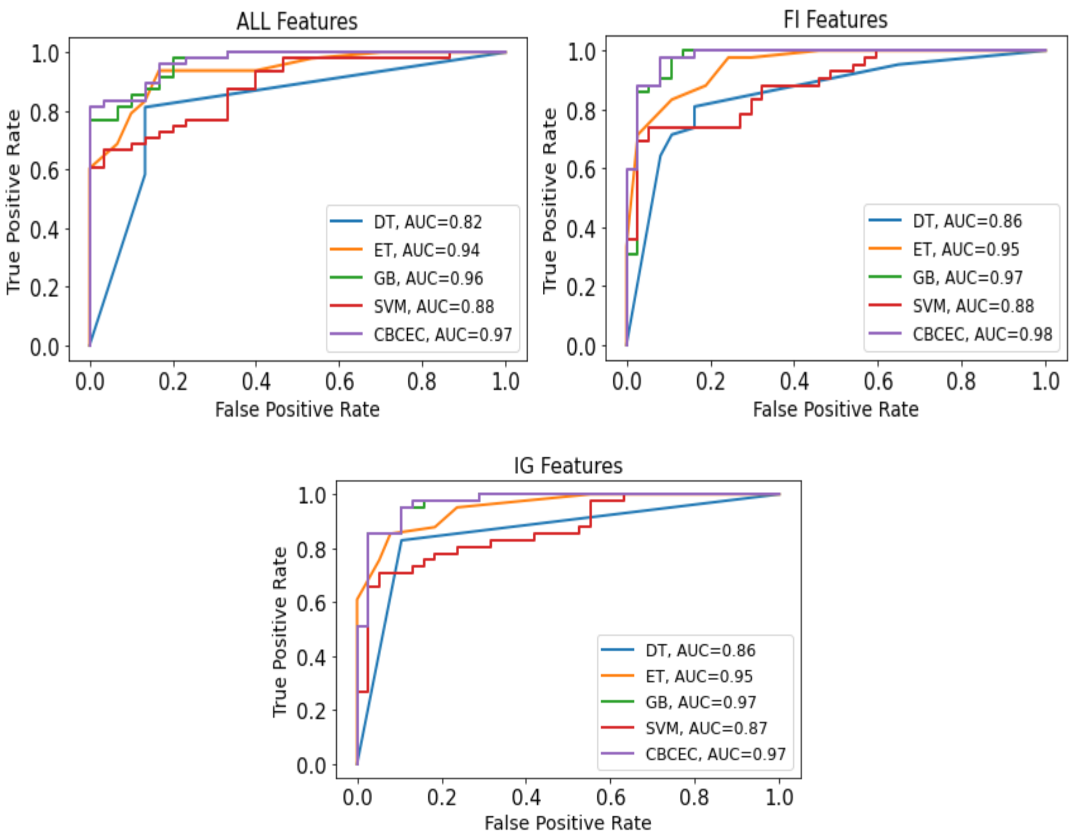

Figure 9: AUC scores on the different feature sets.

4.1.6 Compilation Time

We have measured the computational time for our performing classifiers. Table 4 shows the computational time of all classifiers with different features, which are measured in milliseconds. According to all different feature sets, CBCEC has the highest runtime for compilation of 1351, 957, and 754ms on all, FI, and IG-based features respectively. On the other hand, DT has the lowest runtime of 15.3, 12.2, and 11.8ms in those features.

TABLE 4: Compilation times on the different feature sets measured in milliseconds.

	DT	GB	SVM	ET	CBCEC
All features	15.3	106	82.8	53.2	1351
FI features	12.2	105	77.1	26.3	957
IG features	11.8	82.6	53.6	24.5	754

According to the experimental results, we can assert that supervised machine learning models can serve to warn efficiently of heart failure patients. From the general classifiers GB performed the highest results than the other three classifiers, hence we were determining GB as the best-performing classifier and proposed our intended classifier by combining with it. Furthermore, FI-based feature selection produced better results than all and IG features. Therefore, FI-selected features have a greater impact than others.

4.2 Global Behaviors of Most Impactful Features

The explainable AI would greatly facilitate the implementation of AI/ML in the medical domain, notably through fostering transparency and trust. In the healthcare sector, it is crucial to understand which factors are most likely affected by the disease. Hence, we generate the global explanations of the most ten significant features (from FI) by using the Partial Dependence Plot (PDP). The PDP generates the dependence between the target feature and the set of input features. Earlier in the preprocessing stages, we scaled Cr_ph and Platelets feature. Here in the PDP, without scaling data are fitted to provide the right value range of features. Fig 8 displays the PDP plot for FI-based features. The y-axis holds the partial dependence of the feature and the x-axis holds the value of the feature. The minor ticks on the x-axis represent the diverse values of the features.

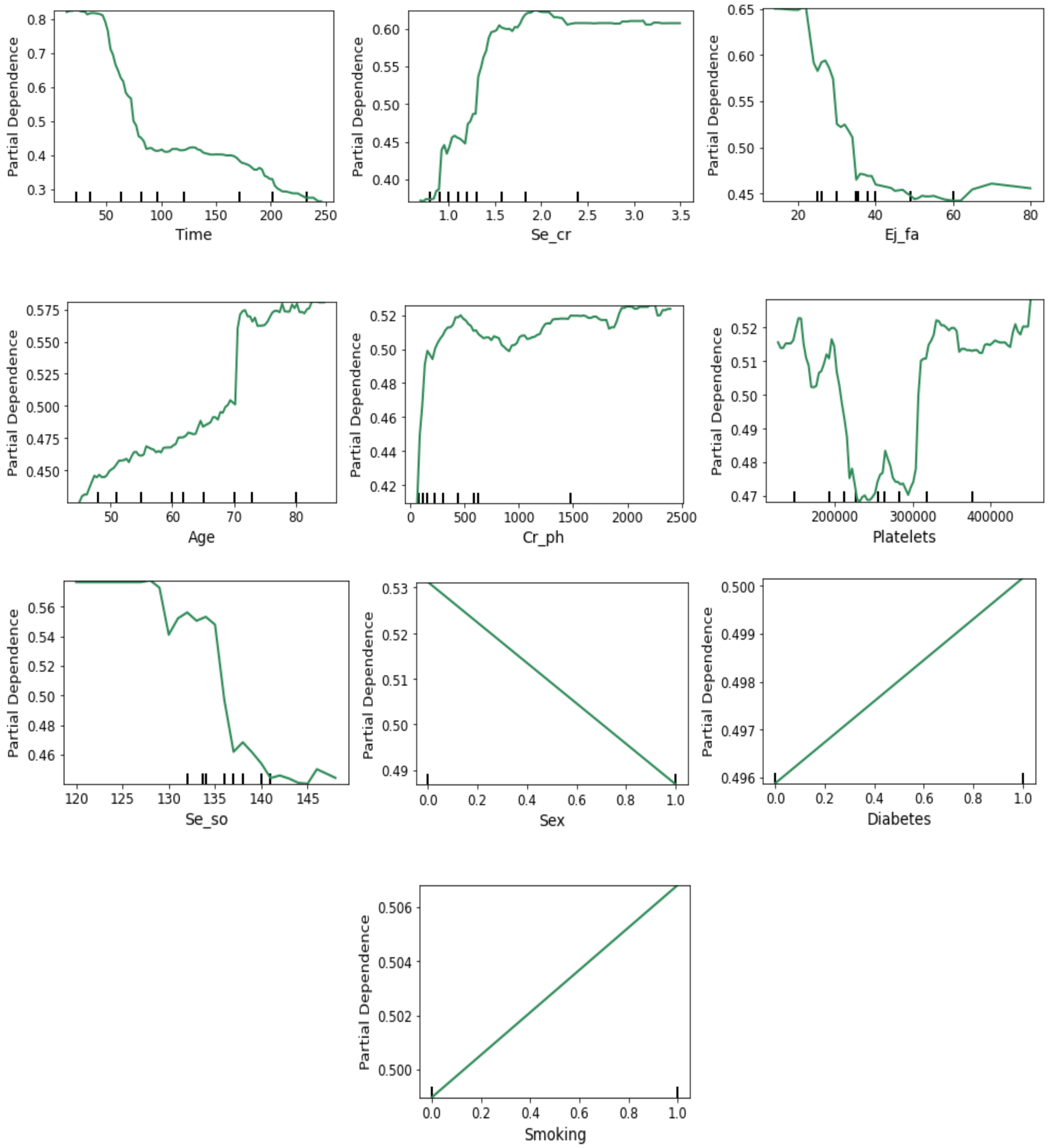

Figure 10: Displays the PDP plot for the most impactful features.

From these plots, we determine the riskiest value or classes of each feature, so that the stakeholders/patients can be aware of their cases. Along with we attach existing justification, which provides more clarification of our study. Table 5 describes the riskiest value ranges or classes from the PDP plots with the existing justifications.

TABLE 5: Determine the riskiest value of the most impactful features.

Feature	**Riskiest value range or classes**	**Existing Justification**
Time	Within 4 - 40 follow-up days	Recommended follow-up within 14 days. [33]
Se_cr	Within 1.5 – 3.5 mg/dl	A higher Se_cr value can increase mortality. [34]
Ej_fa	Within 14 – 20 percent	Below 30% is severely abnormal Ej_fa. [35]
Age	Within 70 – 95 years	HF mostly occurs in older people. [36]
Cr_ph	Within 200 – 2500 mcg/L	10 – 120 mcg/L is normal, otherwise abnormal. [37]
Platelets	<100000 and >350000 per uL	Moderate to severe platelets <100000 per uL. [38]
Se_so	Within 114 – 130 mEq/L	<135 mEq/L is the prevalence value of Se_so in HF. [39]
Sex	Women	Women are more likely to be affected by HF than men. [40]
Diabetics	Having diabetics	People with diabetes are more susceptible to HF. [41]
Smoking	If smoke	Smoking can cause HF. [42]

4.3 Comparison with Existing Studies

Comparing research results with prior studies provides a fresh perspective on the topic that may be useful for any future investigations in the area. The proposed CBCEC classifier obtains robust outcomes for all different feature sets with the SMOTETOMEK-Boost method. However, the FI-based selected feature outperformed others with the CBCEC classifier. Earlier we mention that FI clearly identifies Time, Se_cr, Ej_fa, Age, Cr_ph, Platelets, Se_so, Sex, Diabetics, and Smoking. So, our aspect is (SMOTETOMEK-Boost, FI-based features, CBCEC classifier) useful in patient care and reduces the mortality rate by warning at early stages. These methodologies also outperform existing studies, table 6 holds it.

TABLE 6: Comparison of our aspect with existing studies.

Author & year	**Maximum accuracy**	**Performed classifier**	**Time**
Abid & 2021 [8]	92.62%	Extra Tree	-
Minh & 2021 [13]	85%	Random Forest	-
Saurav & 2022 [14]	83.33%	SVM	-
Lorenzoni & 2019 [15]	81.2%	GLMN	-
Hussain & 2021 [16]	88.79%	SVM	-
Mohan & 2019 [17]	88.7%	Hybrid (HRFLM)	-
Dafni & 2021 [18]	91.23%	Rotation Forest	-
Reza & 2019 [19]	88.88%	Voting (Logistic + Naïve)	-
Our study	**93.67%**	**CBCEC**	**957ms**

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on society

Our suggested approach has numerous advantages, both economically and socially. Our model, which was created to analyze and identify the crucial elements or characteristics of a heart failure patient, is based on real-world data. The ability to inform people about the prevalence of heart failure and the available preventative methods is advantageous to society. Because of the accurate diagnosis and frequent examinations, we are able to recommend early therapy. Because they are more likely to be aware of diseases and be able to predict whether they will be afflicted or not. Our approach requires fewer compilations and is faster. This makes predicting illnesses simple and accurate. Using advanced diagnosis approaches, we have studied the data in our model to determine the underlying reason for heart failure. We hope that our suggested course of action will be accepted and carried out on a societal level.

5.2 Impact on environment

Due to the streamlined diagnosis techniques, our suggested paradigm is particularly successful in remote places. Using the device model, we can cut down on complexity and time. We can be sure that our method will improve the environment because it is simple and doesn't have any negative side effects. People don't need to go to metropolitan areas to find out if they need to risk developing heart failure. The patient's diagnostic report may be simply supplemented by the prediction model, which also forecasts likely outcomes. The price of local therapy or the low cost of diagnosing won't worry patients. It is simpler and may be used by everyone, regardless of level.

It is possible to determine whether or not a patient is at risk for heart failure using our suggested model. Our suggested model will improve the political and social climate. We are confident that if our suggested approach is implemented, the state of medical scientific technology will significantly advance.

5.3 Sustainability Plan

We can assure that our proposed model can be accepted by worldwide research and heart failure technologies. We are confident our proposed model can be useful among the victim who can easily predict their ratio of getting affected by heart failure. If we get proper utilities and scope to implement, we can be motivated and we will be ready to implement in real life to help the rural areas. We hope our proposed model will be sustainable and beneficial.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Conclusion

The mortality rate of HF will be reduced through the processing of raw health data of heart information using machine learning algorithms. In this study, we aim to provide a machine learning-based early warning procedure for the efficient detection of HF. Several ML classifiers are employed to detect HF and overcome the data imbalance problem by SMOTETOMEK-Boost. Significant improvement in the result section has been noticed when reducing the number of selected features by FI and IG feature selection methods. Furthermore, our proposed CBCEC classifier performs the overall highest results compared to others with FI-based selected features. These experimental results demonstrated that the proposed CBCEC classifier can achieve the highest outcomes with SMOTETOME-Boost and FI-based feature selection.

This work has the potential to advance the medical field and help doctors anticipate heart failure patients' chances of survival. It also helps to understand the riskiest value ranges or classes with the riskiest or impactful features of HF. Thus, patients may readily identify the characteristics that affect HF and take medication per basis.

6.2 Future Work

The utilized dataset in this work is quite small, whereas ML algorithms are likely to yield an optimal outcome when it is trained with a much larger dataset. Therefore, future work is indeed to work with a larger dataset to identify HF. Specially, we have proposed two combined hybrid method SMOTETOMEK-BOOST and CBCEC. Substantially, in future we would like to test our suggested aspects using numerous datasets to increase the validity and provide further rationale.

References

[1] WHO. The Top 10 Causes of Death. Accessed: Dec. 30, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10- causes-of-death.

[2] Maculewicz, E., Pabin, A., Kowalczuk, K., Dziuda, Ł. and Białek, A., 2022. Endogenous Risk Factors of Cardiovascular Diseases (CVDs) in Military Professionals with a Special Emphasis on Military Pilots. Journal of Clinical Medicine, 11(15), p.4314.

[3] Tama, B.A., Im, S. and Lee, S., 2020. Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. BioMed Research International, 2020.

[4] Lawson, C.A., Lam, C., Jaarsma, T., Kadam, U., Strömberg, A., Ali, M., Tay, W.T., Clayton, L., Khunti, K. and Squire, I., 2022. Developing a core outcome set for patient-reported symptom monitoring to reduce hospital admissions for patients with heart failure. European Journal of Cardiovascular Nursing.

[5] Peters, S.A., Colantonio, L.D., Dai, Y., Zhao, H., Bittner, V., Farkouh, M.E., Dluzniewski, P., Poudel, B., Muntner, P. and Woodward, M., 2021. Trends in recurrent coronary heart disease after myocardial infarction among US women and men between 2008 and 2017. Circulation, 143(7), pp.650-660.

[6] Tromp, J., Paniagua, S.M., Lau, E.S., Allen, N.B., Blaha, M.J., Gansevoort, R.T., Hillege, H.L., Lee, D.E., Levy, D., Vasan, R.S. and Van Der Harst, P., 2021. Age dependent associations of risk factors with heart failure: pooled population based cohort study. bmj, 372.

[7] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F.J.M., Ignatious, E., Shultana, S., Beeravolu, A.R. and De Boer, F., 2021. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access, 9, pp.19304-19326.

[8] Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V. and Nappi, M., 2021. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE access, 9, pp.39707-39716

[9] Yasmin, F., Shah, S.M.I., Naeem, A., Shujauddin, S.M., Jabeen, A., Kazmi, S., Siddiqui, S.A., Kumar, P., Salman, S., Hassan, S.A. and Dasari, C., 2021. Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. Reviews in Cardiovascular Medicine, 22(4), pp.1095-1113.

[10] Jiang, Z., Pan, T., Zhang, C. and Yang, J., 2021. A new oversampling method based on the classification contribution degree. Symmetry, 13(2), p.194.

[11] F. M. Zahid, S. Ramzan, S. Faisal, and I. Hussain, ''Gender based survival prediction models for heart failure patients: A case study in Pakistan,'' PLoS ONE, vol. 14, no. 2, Feb. 2019, Art. no. e0210602.

[12] D. Chicco and G. Jurman, ''Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone,'' BMC Med. Informat. Decis. Making, vol. 20, no. 1, p. 16, Dec. 2020

[13] Le, M.T., Vo, M.T., Pham, N.T. and Dao, S.V., 2021. Predicting heart failure using a wrapper-based feature selection. Indonesian Journal of Electrical Engineering and Computer Science, 21(3), pp.1530-1539.

[14] Mishra, S., 2022. A Comparative Study for Time-to-Event Analysis and Survival Prediction for Heart Failure Condition using Machine Learning Techniques. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 4(3), pp.115-134.

[15] Lorenzoni, G., Sabato, S.S., Lanera, C., Bottigliengo, D., Minto, C., Ocagli, H., De Paolis, P., Gregori, D., Iliceto, S. and Pisanò, F., 2019. Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. Journal of clinical medicine, 8(9), p.1298.

[16] Hussain, L., Aziz, W., Khan, I.R., Alkinani, M.H. and Alowibdi, J.S., 2021. Machine learning based congestive heart failure detection using feature importance ranking of multimodal features. Math Biosci Eng, 18(1), pp.69-91.

[17] Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, pp.81542-81554.

[18] Plati, D.K., Tripoliti, E.E., Bechlioulis, A., Rammos, A., Dimou, I., Lakkas, L., Watson, C., McDonald, K., Ledwidge, M., Pharithi, R. and Gallagher, J., 2021. A machine learning approach for chronic heart failure diagnosis. Diagnostics, 11(10), p.1863.

[19] Raza, K., 2019. Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. In U-Healthcare Monitoring Systems (pp. 179-196). Academic Press.

[20] Kaggle heart failure clinical records, available at <<https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data>>, last accessed on 05.06.2022

[21] Shahabadi, M.S.E., Tabrizchi, H., Rafsanjani, M.K., Gupta, B.B. and Palmieri, F., 2021. A combination of clustering-based under-sampling with ensemble methods for solving imbalanced class problem in intelligent systems. Technological Forecasting and Social Change, 169, p.120796.

[22] Kaur, H., Pannu, H.S. and Malhi, A.K., 2019. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Computing Surveys (CSUR), 52(4), pp.1-36.

[23] B. Kotsiantis, "Decision trees: a recent overview," Artificial Intelli-gence Review, vol. 39, p. 261–283, 2013.

[24] Bahzad Taha Jijo, Adnan Mohsin Abdulazeez ,"Classification Based on Decision Tree Algorithm for Machine Learning",March 15th, 2021.

[25] Ma, Baoshan; Meng, Fanyu; Yan, Ge; Yan, Haowen; Chai, Bingjie; Song, Fengju (2020). Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Computers in Biology and Medicine, 121(), 103761–. doi:10.1016/j.compbiomed.2020.103761

[26] Otchere, D.A., Ganat, T.O.A., Ojero, J.O., Tackie-Otoo, B.N. and Taki, M.Y., 2022. Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. Journal of Petroleum Science and Engineering, 208, p.109244.

[27] Hao Chena, Chao Zhanga, Ninghong Jiac , Ian Duncand , Shenglai Yanga, YongZhi Yang," A Machine Learning Model for Predicting the Minimum Miscibility Pressure of CO2 and Crude Oil System Based on a Support Vector Machine Algorithm Approach", 2020 published by Elsevier.

[28] Lalwani, P., Mishra, M.K., Chadha, J.S. and Sethi, P., 2022. Customer churn prediction system: a machine learning approach. Computing, 104(2), pp.271-294.

[29] A. Moubayed, E. Aqeeli, and A. Shami, “Detecting dns typo-squatting using ensemble-based feature selection classification models,” IEEE Canadian Journal of Electrical and Computer Engineering, vol. 44,no. 4, pp. 456–466, 2021.

[30] Wang, Z.H.E., Wu, C., Zheng, K., Niu, X. and Wang, X., 2019. SMOTETomek-based resampling for personality recognition. IEEE Access, 7, pp.129678-129689.

[31] H. K. T. L. J. Sushruta Mishra, Pradeep K Mallick and G.-S. Chae, “Stacked knn with hard voting predictive approach to assist hiring pro-cess in it organizations,” International Journal of Electrical Engineering and Education, 02 2021

[32] Hertel, L., Collado, J., Sadowski, P., Ott, J. and Baldi, P., 2020. Sherpa: Robust hyperparameter optimization for machine learning. SoftwareX, 12, p.100591.

[33] McAlister, F.A., Youngson, E., Kaul, P. and Ezekowitz, J.A., 2016. Early follow-up after a heart failure exacerbation: the importance of continuity. Circulation: Heart Failure, 9(9), p.e003194.

[34] Metra, M., Cotter, G., Gheorghiade, M., Dei Cas, L. and Voors, A.A., 2012. The role of the kidney in heart failure. European heart journal, 33(17), pp.2135-2142.

[35] Cleveland Clinic, available at <<https://my.clevelandclinic.org/health/articles/16950-ejection-fraction>>, last accessed on: 05.06.2022.

[36] Pandey, A., Kitzman, D. and Reeves, G., 2019. Frailty is intertwined with heart failure: mechanisms, prevalence, prognosis, assessment, and management. JACC: Heart Failure, 7(12), pp.1001-1011.

[37] Andini, S., Sitanggang, R., Wanto, A., Okprana, H., Daengs, G.A. and Aryza, S., 2021, June. Utilization of Rough Sets Method with Optimization Genetic Algorithms in Heart Failure Cases. In Journal of Physics: Conference Series (Vol. 1933, No. 1, p. 012038). IOP Publishing.

[38] Mojadidi MK, Galeas JN, Goodman-Meza D, Eshtehardi P, Msaouel P, Kelesidis I, Zaman MO, Winoker JS, Roberts SC, Christia P, Zolty R. Thrombocytopaenia as a Prognostic Indicator in Heart Failure with Reduced Ejection Fraction. Heart Lung Circ. 2016 Jun;25(6):568-75. doi: 10.1016/j.hlc.2015.11.010. Epub 2016 Jan 18. PMID: 26868831.

[39] Abebe, T.B., Gebreyohannes, E.A., Tefera, Y.G., Bhagavathula, A.S., Erku, D.A., Belachew, S.A., Gebresillassie, B.M. and Abegaz, T.M., 2018. The prognosis of heart failure patients: Does sodium level play a significant role?. PloS one, 13(11), p.e0207242.

[40] Beale, A.L., Meyer, P., Marwick, T.H., Lam, C.S. and Kaye, D.M., 2018. Sex differences in cardiovascular pathophysiology: why women are overrepresented in heart failure with preserved ejection fraction. Circulation, 138(2), pp.198-205.

[41] Liccardo, D., Cannavo, A., Spagnuolo, G., Ferrara, N., Cittadini, A., Rengo, C. and Rengo, G., 2019. Periodontal disease: a risk factor for diabetes and cardiovascular disease. International journal of molecular sciences, 20(6), p.1414.

[42] Aune, D., Schlesinger, S., Norat, T. and Riboli, E., 2019. Tobacco smoking and the risk of heart failure: A systematic review and meta-analysis of prospective studies. European journal of preventive cardiology, 26(3), pp.279-288.

[43] A. Lal and C. R. S. Kumar, "Hybrid classifier for increasing accuracy of fitness data set," *2017 2nd International Conference for Convergence in Technology (I2CT)*, 2017, pp. 1246-1249, doi: 10.1109/I2CT.2017.8226326.

[44] Yang, P., Hwa Yang, Y., B Zhou, B. and Y Zomaya, A., 2010. A review of ensemble methods in bioinformatics. Current Bioinformatics, 5(4), pp.296-308.

Thank You!
Best of Luck

Report plagiarism

ORIGINALITY REPORT

15%	12%	8%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080	Internet Source	2%
2	Submitted to Daffodil International University	Student Paper	2%
3	research-repository.griffith.edu.au	Internet Source	1%
4	doctorpenguin.com	Internet Source	1%
5	Submitted to Heriot-Watt University	Student Paper	1%
6	Submitted to Liverpool John Moores University	Student Paper	1%
7	Submitted to Birkbeck College	Student Paper	1%
8	Submitted to Coventry University	Student Paper	1%
9	www.hindawi.com	Internet Source	1%