

# An early warning system of heart failure mortality with combined machine learning methods

Ananda Sutradhar, Mustahsin Al Rafi, Mohammad Jahangir Alam, Saiful Islam

Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

## Article Info

### Article history:

Received May 24, 2023

Revised Aug 3, 2023

Accepted Aug 18, 2023

### Keywords:

Ensemble method

Feature selection

Heart failure

Imperial boost-stacked

SMOTE-ENN

## ABSTRACT

Heart failure (HF) is currently the leading cause of morbidity and mortality worldwide. Identifying the risk of mortality at the early stages is crucial to reducing the mortality rate. However, the traditional methods for exploring the signs of mortality are difficult and time-consuming. Whereas, machine learning (ML) methods are superior in reducing HF's mortality rate by providing early warnings. This study presents a novel ML classifier called imperial boost-stacked (IBS) that can serve as an effective early warning system for predicting HF mortality. Initially, we performed an efficient data balancing technique named synthetic minority oversampling technique with edited nearest neighbors (SMOTE-ENN) to mitigate the imbalance problem. Next, two well-known feature selection techniques, the extra tree (ET) and information gain (IG), are applied to reduce the data dimensions and select the most significant features. Following that, the prepared feature sets are trained with our proposed IBS classifier. Simultaneously leveraging the advantages of boosting, stacking, and multiple robust methods, it significantly correlates with the intricate patterns of clinical data of HF patients. Finally, the robust outcomes of 92.75% accuracy over existing studies reveal that our proposed study can effectively warn the HF mortality at early stages and reduce the burden on the healthcare sector.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Corresponding Author:

Ananda Sutradhar

Department of Computer Science and Engineering, Daffodil International University

Daffodil Road, Asulia, Dhaka 1341, Bangladesh

Email: anandasutradhar819@gmail.com

## 1. INTRODUCTION

Heart failure (HF) is a chronic condition in which the heart is unable to pump enough blood to meet the body's needs. It has just risen to the top spot as the global cause of death. Sedentary behavior, excessive alcohol use, smoking, obesity, germs, influenza, chest radiation, hypertension, coronary artery disease, and dyslipidemia. are the most typical risk factors for HF [1]. Several non-lifestyle risk factors should also be taken into account, including age, gender, family history, and high levels of fibrinogen [2]. HF is classified as a clinical syndrome, including major symptoms (such as shortness of breath, ankle edema, and fatigue) and one common psychological symptom (anxiety). Hospitalizations for HF are more common in women, and older people have a larger chance of acquiring HF than younger ones [3]. HF deaths have surged 35 times for men and 48 times for women in Bangladesh over the past ten years. According to the Bangladesh Bureau of Statistics, 1,80,408 people have died due to HF [4]. The death rate for individuals with HF after discharge from the hospital is 10.4% at 30 days, 22% at one year, and 42.3% at five years [5].

To lower the fatality rate, it is essential to examine the indicators of mortality as soon as possible and to start providing counseling and medication. Numerous clinical indicators, such as ejection fraction (testing how well the heart pumps blood), B-type natriuretic peptide (a hormone secreted by the heart in response to

HF), renal function (poor kidney function), are being investigated to determine the risk of HF mortality. This manual method is complicated, time-consuming, expensive, and might not always be adequate.

Therefore, researchers are concentrating on machine learning (ML) methods to detect HF, thereby saving time, money, resources, and many lives. However, multiple noteworthy similar concerns have been observed from numerous antecedent studies on this topic. Likewise, the study [6], [7] conducted their study with only the specific feature of the HF dataset. Chicco and Jurman [7] stated that serum creatinine and ejection fraction are sufficient for an effective prediction. Nevertheless, these specific features alone are not appropriate for clinical purposes. Using one or two feature selection techniques can able to rectify the most relevant features for identifying the death cases. Then, the authors [8]–[10] utilized a famous data balancing technique named synthetic minority oversampling technique (SMOTE) to overcome data imbalance techniques and obtained prediction accuracy of 92.6%, 91.23%, and 83.33%, respectively. But SMOTE potential to generate noisy and uninformative samples [11]. As a result, different ML techniques must be used to significantly lessen the drawbacks of SMOTE. Afterward, the researches [12]–[14] introduced the proposed hybrid classifier by combining the baseline classifier with individual ensemble methods and performed generalized accuracies. However, individual ensemble classifiers might produce biased results and constrain decision boundaries [15]. Resulting in doesn't work well with all kinds of data and easily occurs overfitting or underfitting. In contrast, combining several classifiers can help to get beyond these restrictions by utilizing the advantages of several models.

Motivated by the limitations of preceding studies, this current study was proposed to address the aforementioned concerns and provide further insight into the topic at hand. In contrast to previous research, we use SMOTE with edited nearest neighbors (SMOTE-ENN) to eliminate noisy data to overcome the SMOTE disadvantage. This combined method offers to efficiently generate synthetic instances and reduce irrelevant samples at the same time. Then instead of considering the specific features, we rectify the most potential features using the extra tree (ET) and information gain (IG) feature selection techniques. From these techniques, we prepared three different feature sets (e.g., ALL, ET, and IG) and determined the most potential one using extensive experiments of multiple performance metrics. To train the following feature sets, two robust ensemble methods (e.g., boosting and stacking) were taken into account. Inspired by the advantages of these methods, we have combined them with baseline classifiers and proposed a novel imperial boost-stacked (IBS) classifier to mitigate their individual disadvantage during the classification phase. By gaining several advantages during training, our suggested IBS classifier can effectively improve classification performance by lowering bias and variance and offering insights into the underlying data patterns and relationships. Moreover, to validate the significance of the proposed IBS classifier, employed three well-known ML classifiers named decision tree (DT), random forest (RF), and gradient boost (GB). Finally, an extensive experiment demonstrated that the proposed IBS classifier can obtain superior outcomes in detecting HF mortality.

## 2. PROPOSED IMPERIAL BOOST-STACKED CLASSIFIER

Researchers nowadays aim to achieve multiple benefits by training data with hybrid or combined classifiers, as individual classifiers may sometimes fall short of the expected level [16]. A combination of multiple classifiers can able to get multiple advantages at the same time and mitigate the individual weakness. Hence, we proposed a robust ML classifier named the IBS classifier to detect the signs of mortality due to HF. The proposed IBS was developed by combining the advantage of employed conventional classifier with two robust methods named boosting (BS) and stacking (ST). The fundamental function of BS is assigning the weights of each instance in the training data and attempting to classify the examples correctly in each iteration by changing the weights based on the errors made by previous weak classifiers. Whereas ST works in two stages, initially, training the set of classifiers and making predictions from the base models. Then the prediction sets are used as input for the second-level classifier. Which map to learn these inputs and generate the final outcome.

To build a strong model, we first trained the dataset by BS. It operates by iteratively reweighting the training data and adding weak learners to the ensemble based on prior weak learners' performance. Initially, we initialize each training example's weight to  $1/n$ , where  $n$  is the total number of examples. For  $t$  number of iterations, train a weak learner (WL) on the weighted training set and calculate the error rate ER. Based on the ER, new weights (NW) were calculated using the formula of  $\ln\{(1 - ER) / ER\} / 2$ . The NW was updated on each iteration, expressed in (1).

$$w(i, t + 1) = w(i, t) * \exp\{-\alpha t * y(i) * ot(y_i)\} / N_F \quad (1)$$

Where  $w(i, t)$  refers the weight of example  $i$  at iteration  $t$ , where  $y(i)$  refers the true level of  $i$ ,  $ot(y_i)$  refers the output of weak classifier  $ot$  on the example of  $x_i$ , and  $N_F$  is a normalization factor. Finally, a weighted

combination of the  $ot$  was conducted as the final classification, represented in (2). Where the OSC refers to the outcome of BS and the  $sign()$  function returns  $+1$  if the output of OSC is dead and  $-1$  for the surviving class.

$$OSC = sign\{\sum_{t=1}^T \alpha t * ot(x_i)\} \tag{2}$$

Next, we used ST to improve the result greatly by gathering information from both base and meta-level classifiers. Initially, we aim to effectively construct the training dataset for the second-level classifier. Hence, setting the three baselines (e.g., DT, RF, and GB) and robust OSC as the base estimator of ST to get the advantage of multiple ML methods, expressed as (3). To construct a new input set ( $xp$ ), we classify the training instances  $x_i$  applying ISLC. The logistic regression classifier (LGC) was utilized as a second-level learner to train  $xp$ . LGC can be extensively used in binary classifications and effectively generate the final predictions of the proposed IBS classifier (4). Where  $IBS_{p(Y=1|X)}$  is the probability of the binary response variable of our proposed IBS classifier. The  $\partial^0$  is the intercept and  $\partial^1$  to  $\partial^n$  is the coefficient of the generated input features  $xp_i$  to  $xp_n$ . Finally, algorithm 1 has represented the whole procedure of the proposed classifier.

$$ISLC = \{DT(x_{train}), RF(x_{train}), GB(x_{train}), OSC(x_{train})\} \tag{3}$$

$$IBS_{p(Y=1|X)} = 1/\{1 + \exp\{-(\partial^0 + \partial^1(xp_i) + \partial^2(xp_2) + \dots + \partial^n(xp_n))\}\} \tag{4}$$

Algorithm 1. Illustrates the procedure of our proposed imperial boost-stacked classifier

Input: Training data,  $D_{train} = \sum_{i=1}^n(x_i, y_i)$ ; Number of iterations,  $t = 1$  to  $T$ ; Weak Learner =  $WL$ ;  
 1. Calculate the new weights =  $Nw$ ; Updated the weights on training sets =  $w(i, t + 1)$ ; Error rate =  $ER$ ;  
 2. Output of  $WL = ot$ ; Normalization factor =  $NF$ ; Outcome of Boosting =  $OBS$ ; Input for second-level  
 3. classifier =  $xp$ .  
 4. Output: Classify the survival (0) or alive (1) classes.  
 5. for  $t = 1, 2, 3, \dots, T$ :  
 6.     Train  $WL$  on the  $D_{train}$   
 7.     Calculate the  $ER$  of  $WL$  on  $D_{train}$   
 8.      $NW = \ln\{(1 - ER) / ER\} / 2$   
 9.      $w(i, t + 1) = w(i, t) * \exp\{-\alpha t * y(i) * ot(y_i)\} / N_F$   
 10. end for  
 11.  $OSC = sign\{\sum_{t=1}^T \alpha t * ot(x_i)\}$   
 12.  $ISLC = \{DT(x_{train}), RF(x_{train}), GB(x_{train}), OSC(x_{train})\}$   
 13. for  $i = 1, 2, 3, \dots, n$ :  
 14.     Apply  $ISLC$  to classify training instances  $x_i$   
 15.      $xp = ISLC(x_i)$   
 16. end for  
 17.  $IBS_{p(Y=1|X)} = 1/\{1 + \exp\{-(\partial^0 + \partial^1(xp_i) + \partial^2(xp_2) + \dots + \partial^n(xp_n))\}\}$   
 18. Return  $IBS_{p(Y=1|X)}$ : The predicted binary response from our proposed IBS classifier.

### 3. METHOD

This research uses several ML approaches, like data preprocessing, feature selection, performing classifiers, and evaluating the performance. To train the processed data, significant features are chosen by two well-known feature selection techniques. Then, evaluate the performance of our proposed classifier by accuracy, precision, recall, and f1 score. Figure 1 holds the overall workflow of our study.

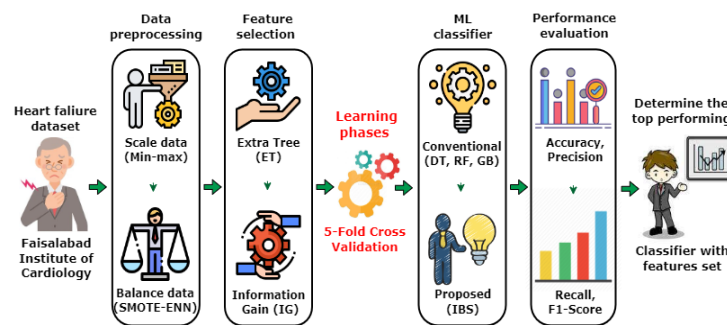


Figure 1. Flow diagram for the working procedure of our study

### 3.1. Data description

This study employed the Faisalabad Institute of Cardiology and Allied Hospital's heart failure clinical records data set, available in [17]. This dataset contains 299 medical records with 13 clinical features, these are collected during the follow-up period. The name of these features is age, anaemia (anmia), high blood pressure (H\_b\_p), creatinine phosphokinase (Cr\_ph), diabetes (dibts), ejection fraction (Ej\_fr), sex, platelets (plts), serum creatinine (Se\_cr), serum sodium (Se\_so), smoking (Smkg), time, and DEATH\_EVENT (the patient survived or not). The last feature name DEATH\_EVENT has been selected as the target class, 1 is for dead and 0 is for alive. Where 96 surviving instances and 203 reported death cases result in an unbalanced dataset. More information about this dataset can be found in the original dataset curators' release [18].

### 3.2. Data preprocessing

The dataset that was used for this investigation is nearly uncontaminated and preprocessed; it contains no missing values. In the case of creatinine phosphokinase and platelet characteristics, values vary significantly from one to the next. It might put off making a choice, allowing min-max scaling to solve the problem. It turns the feature values into a range and is crucial for enhancing outcomes. The imbalance of the dataset is another problem that has emerged. The SMOTE is one of the famous approaches to deal with this issue and researchers mostly use it. But SMOTE has the potential to produce noisy and useless samples. Hence, we apply the SMOTE-ENN technique to balance the dataset. At first, SMOTE chose the nearest neighbor in the process of synthesizing a new sample  $X_n$  from sample  $X$ , then the new sample is built according to (5). Balancing the distribution of the classes, it avoids the classifier from being biased toward the minority class.

$$X_{new} = X + rand(0,1) * (X_n - X) \quad (5)$$

The second strategy, known as ENN, uses under-sampling to eliminate samples whose k-nearest neighbors incorrectly classify them. ENN assists in reducing redundancy and noise from the SMOTE-generated dataset. In SMOTE-ENN, the minority class is first oversampled with SMOTE, and then noisy and superfluous samples are eliminated with ENN. It can enhance the performance of classifiers on unbalanced datasets by boosting the proportion of minority class samples and lowering noise and redundancy in the data [19]. After applying SOMTE-ENN, 201 cases also belonged to the dead class and 199 cases to the surviving class. Figure 2 describes the working procedure of the SMOTE-ENN.

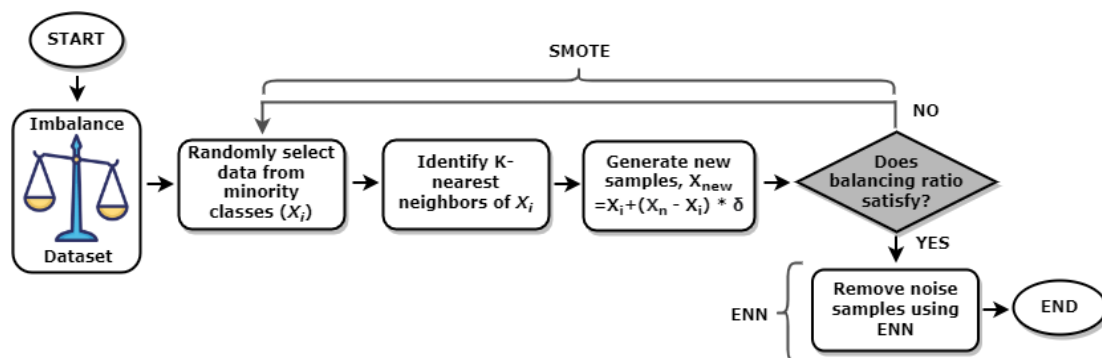


Figure 2. Working diagram of SMOTE-ENN

### 3.3. Feature selection

Feature selection improves machine learning and increases the predictive capability of ML algorithms by selecting the most important variables. Here, the significant characteristics are chosen using ET and IG. ET builds a lot of decision trees on different dataset subsets during the feature selection process and chooses the most discriminatory features according to the impurity decrease criterion. On the other hand, the primary function of IG is to calculate each variable's gain relative to the target variable. We selected the top 10 features from these methods, Figure 3 lists these features. Where Figure 3(a) displays the ET-based selected features and Figure 3(b) for the IG-based features. Then, the processed dataset and reduced feature set were used to evaluate the performance of models by 5-fold cross-validation, this aids in providing a more precise estimate of the model's performance on the dataset [20].

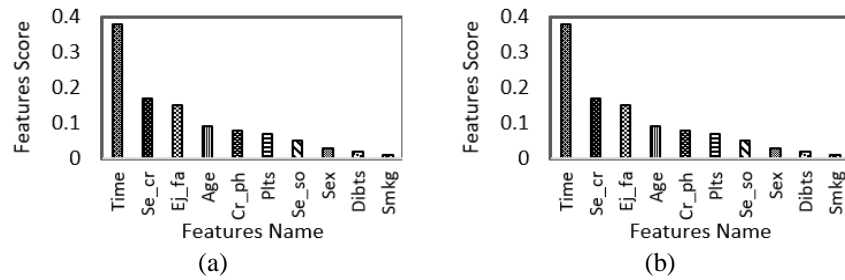


Figure 3. The selected features from two feature selection methods; (a) extra tree and (b) information gain

**3.4. Performance metrics**

In order to validate the effectiveness of our study, we have evaluated four classification matrices named accuracy, precision, recall, and f1-score [21]. The accuracy is used to assess the correctly classified instances and trends between variables in a dataset based on their training data. Precision refers to how good the model is at predicting a specific category. The recall gauges how well the model can identify positive samples. Where the f1-score is a key measure by evaluating the harmonic mean of precision and recall.

**4. RESULTS AND DISCUSSION**

**4.1. Analysis of the performed results**

In this section, we discussed all the experimental results for our proposed work. To determine the significance of our proposed IBS classifier, we have trained three well-known traditional classifiers (e.g., DT, RF, and GB) and compared the outcomes with the IBS classifier. Table 1 shows the accuracy of performed classifiers on 5 different folds of each feature set. Considering all features, the best average accuracy was obtained by the IBS classifier which is 90%, whereas DT, RF, and GB produced 86.50%, 87.75%, and 87% accuracy respectively. The ET-based selected features significantly produced the highest accuracy with the proposed IBS classifier. Additionally, with the IG-based selected features, the proposed classifier was able to obtain 91.50% accuracy. The conventional ML classifiers showed an accuracy range between 86-90% for all different feature sets.

Table 2 shows the recall scores for the performing classifiers on each feature set. A robust recall score of 94.30% has been generated from the proposed IBS classifier, while DT obtained the competitively lowest recall score of 88.25%. The proposed IBS classifier generated another identical score of 92.79% from the IG-based features set. These robust recall scores demonstrated that the proposed IBS classifier has the ability to accurately identify a high percentage of positive instances, out of all genuine positive examples in a dataset.

Table 1. The obtained accuracy (%) for the performed classifiers on multiple feature sets

Folds	All features				ET features				IG features			
	DT	RF	GB	IBSC	DT	RF	GB	IBSC	DT	RF	GB	IBSC
1 <sup>st</sup>	86.25	88.75	85.00	88.75	86.25	88.75	87.50	92.50	86.25	87.50	88.75	91.25
2 <sup>nd</sup>	87.50	87.50	86.25	90.00	88.75	91.25	88.75	93.75	86.25	88.75	88.75	91.25
3 <sup>rd</sup>	86.25	86.25	88.75	88.75	87.50	88.75	88.75	91.25	88.75	88.75	87.50	88.75
4 <sup>th</sup>	85.00	88.75	88.75	91.25	90.00	91.25	91.25	92.50	87.50	91.25	92.50	93.75
5 <sup>th</sup>	87.50	87.50	86.25	91.25	88.75	91.25	90.00	93.75	88.75	88.75	88.75	92.50
Avg	86.50	87.75	87.00	90.00	88.25	90.25	89.25	92.75	87.50	89.00	89.25	91.50

Table 2. The obtained recall (%) for the performed classifiers on multiple feature sets

Folds	All features				ET features				IG features			
	DT	RF	GB	IBSC	DT	RF	GB	IBSC	DT	RF	GB	IBSC
1 <sup>st</sup>	87.17	89.74	86.84	89.74	87.17	89.74	89.47	94.73	87.17	89.47	89.74	92.30
2 <sup>nd</sup>	89.47	89.47	87.17	90.00	89.74	92.30	89.74	94.87	87.17	89.74	89.74	92.30
3 <sup>rd</sup>	87.17	87.17	89.74	89.74	89.47	89.74	89.74	92.30	89.74	89.74	89.47	89.74
4 <sup>th</sup>	86.84	89.74	89.74	92.30	90.00	92.30	92.30	94.73	89.47	92.30	94.73	94.87
5 <sup>th</sup>	89.47	89.47	87.17	92.30	89.74	92.30	90.00	94.87	89.74	89.74	89.74	94.73
Avg	88.02	89.19	88.13	90.82	89.22	91.27	90.25	94.30	88.66	90.20	90.68	92.79

The precision scores of the different classifiers have comparatively represented in Table 3. Considering all features, the proposed IBS classifier performed a precision score of 89%. When considering ET features in the account, the performed precision scores comparatively gained out of the other feature sets.

The performing conventional classifiers have generated approximately identical scores for different feature sets. However, the overall recall scores demonstrated that the proposed IBS classifier has shown the highest score of 91% with ET-based features, compared to the baseline classifiers.

Table 3. The obtained precision (%) for the performed classifiers on multiple feature sets

Folds	All features				ET features				IG features			
	DT	RF	GB	IBSC	DT	RF	GB	IBSC	DT	RF	GB	IBSC
1 <sup>st</sup>	85.00	87.50	82.50	87.50	85.00	87.50	85.00	90.00	85.00	85.00	87.50	90.00
2 <sup>nd</sup>	85.00	85.00	85.00	90.00	87.50	90.00	87.50	92.50	85.00	87.50	87.50	90.00
3 <sup>rd</sup>	85.00	85.00	87.50	87.50	85.00	87.50	87.50	90.00	87.50	87.50	85.00	87.50
4 <sup>th</sup>	82.50	87.50	87.50	90.00	90.00	90.00	90.00	90.00	85.00	90.00	90.00	92.50
5 <sup>th</sup>	85.00	85.00	85.00	90.00	87.50	90.00	90.00	92.50	87.50	87.50	87.50	90.00
Avg	84.50	86.00	85.50	89.00	87.00	89.00	88.00	91.00	86.00	87.50	87.50	90.00

The outcomes of the f1-Score from the classifiers are displayed in Table 4. Initially, for all processed features, 86-87% scores were performed with the conventional classifiers, which reached 89.93% in the case of performing the IBS classifier. Again, the overall generalized score was obtained from the ET-based selected features, which is 92.61% in the case of the IBS classifier. The proposed IBS classifiers consistently perform efficient results compared to baseline classifiers in the case of all different classification metrics.

Table 4. The obtained f1-score for the performed classifiers on multiple feature sets

Folds	All features				ET features				IG features			
	DT	RF	GB	IBSC	DT	RF	GB	IBSC	DT	RF	GB	IBSC
1 <sup>st</sup>	86.04	88.70	84.36	88.70	86.04	88.70	87.35	92.31	86.04	87.35	88.70	91.13
2 <sup>nd</sup>	87.35	87.35	86.04	90.00	88.70	91.13	88.70	93.66	86.04	88.70	88.70	91.13
3 <sup>rd</sup>	86.04	86.04	88.70	88.70	87.35	88.70	88.70	91.13	88.70	88.70	87.35	88.70
4 <sup>th</sup>	84.36	88.70	88.70	91.13	90.00	91.13	91.13	92.31	87.35	91.13	92.31	93.66
5 <sup>th</sup>	87.35	87.35	86.04	91.13	88.70	91.13	90.00	93.66	88.70	88.70	88.70	92.31
Avg	86.23	87.69	86.77	89.93	87.90	90.16	89.18	92.61	87.37	88.92	89.15	91.39

#### 4.2. Discussion

To reduce the mortality of HF, it is crucial to detect the signs and take treatment as soon as possible by counseling the medician. Whereas, the ML methods have significantly emerged as the potential tool for detecting the sign mortality of HF at early stages. Therefore, this study proposed an efficient ML system to detect mortality. Initially, the raw heart patient's dataset was balanced by a robust combined method named SMOTE-ENN and the most significant features are selected by two feature selection techniques ET and IG. Then these different feature sets were trained by the proposed IBS classifier. In order to mitigate the overfitting and underfitting we adjusted the scale of features that had large differences between the data point. It can stop the model from emphasizing features with larger magnitudes, resulting in the model's capacity for generalization can be enhanced by training to weigh each feature equally. Additionally, k-fold cross-validation can help to limit the risk of overfitting or underfitting to a specific subset of the data by using many folds for training and validation [22]. This can help to provide a more accurate assessment of the model's true performance on unseen data. Furthermore, the proposed IBS classifier is developed by employing two robust ensemble methods. These methods would be helpful to reduce overfitting and underfitting issues [23].

The proposed IBS classifier has significantly generated robust outcomes compared to the baseline classifiers. To validate the effectiveness compared to the existing studies, a comparison summary of the performed accuracy between our proposed aspects with the existing studies is shown in Table 5. Our proposed aspects outperformed previous studies with an accuracy of 92.75%. Therefore, we can conclude that our aspects are substantially more generalized in-patient care and reduce the mortality rate by warning early.

Table 5. A comparative summary between our study with the existing studies based on the accuracy

Author and Ref	Balanced dataset	Performed classifier	Accuracy
Ishaq <i>et al.</i> [8]	SMOTE	Extra tree	92.6%
Plati <i>et al.</i> [9]	SMOTE	Rotation forest	91.23%
Mishra [10]	SMOTE	Support vector machine	83.33%
Mohan <i>et al.</i> [13]	--	Hybrid (HRFLM)	88.4%
Raza [14]	--	Voting (Logistic+Naïve)	88.88%
Le <i>et al.</i> [24]	--	Random forest	85%
Lorenzoni <i>et al.</i> [25]	--	GLMN	81.2%
Hussain <i>et al.</i> [26]	--	Support vector machine	88.79%
Our study	SMOTE-ENN	Imperial boost-stacked	92.75%




## 5. CONCLUSION

HF's mortality rate will be reduced by processing raw health data about hearts using ML algorithms. In this study, we aim to provide a machine learning-based early warning procedure for efficiently detecting HF. Several ML classifiers are employed to detect HF and overcome the data imbalance problem by SMOTE-ENN. Significant improvement in the result section has been noticed when reducing the number of selected features by ET and IG feature selection methods. Our proposed classifier IBS has outperformed others with ET-based selected features in terms of overall performance. This work has the potential to advance the medical field and help doctors anticipate heart failure patients' chances of survival. However, in the future, a large dataset can be utilized for better performance. Furthermore, there are several risk factors or ill groups included by HF, and those would be the secondary or main cause of death. We would like to include these attributes in the HF and consider other potential ML ensemble methods to get a more accurate outcome.




## REFERENCES

- [1] T. A. McDonagh *et al.*, "2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," *European Heart Journal*, vol. 42, no. 36, pp. 3599–3726, Sep. 2021, doi: 10.1093/eurheartj/ehab368.
- [2] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," *BioMed Research International*, vol. 2020, pp. 1–10, Apr. 2020, doi: 10.1155/2020/9816142.
- [3] J. Tromp *et al.*, "Age dependent associations of risk factors with heart failure: Pooled population based cohort study," *The BMJ*, vol. 372, p. n461, Mar. 2021, doi: 10.1136/bmj.n461.
- [4] O. Rahman *et al.*, "Internet of things based electrocardiogram monitoring system using machine learning algorithm," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3739–3751, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3739-3751.
- [5] F. Yasmin *et al.*, "Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future," *Reviews in Cardiovascular Medicine*, vol. 22, no. 4, pp. 1095–1113, 2021, doi: 10.31083/j.rcm2204121.
- [6] 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, p. e0210602, Feb. 2019, doi: 10.1371/journal.pone.0210602.
- [7] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 16, 2020, doi: 10.1186/s12911-020-1023-5.
- [8] A. Ishaq *et al.*, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021, doi: 10.1109/ACCESS.2021.3064084.
- [9] D. K. Plati *et al.*, "A machine learning approach for chronic heart failure diagnosis," *Diagnostics*, vol. 11, no. 10, p. 1863, Oct. 2021, doi: 10.3390/diagnostics11101863.
- [10] S. Mishra, "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*, vol. 4, no. 3, pp. 115–134, Jul. 2022, doi: 10.35882/jeeemi.v4i3.225.
- [11] Z. Jiang, T. Pan, C. Zhang, and J. Yang, "A new oversampling method based on the classification contribution degree," *Symmetry*, vol. 13, no. 2, pp. 1–13, Jan. 2021, doi: 10.3390/sym13020194.
- [12] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
- [13] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [14] K. Raza, "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule," in *U-Healthcare Monitoring Systems: Volume 1: Design and Applications*, 2018, pp. 179–196, doi: 10.1016/B978-0-12-815370-3.00008-6.
- [15] R. Ahmad, I. Alsmadi, W. Alhamdani, and L. Tawalbeh, "A comprehensive deep learning benchmark for IoT IDS," *Computers and Security*, vol. 114, p. 102588, Mar. 2022, doi: 10.1016/j.cose.2021.102588.
- [16] A. Lal and C. R. S. Kumar, "Hybrid classifier for increasing accuracy of fitness data set," in *2017 2nd International Conference for Convergence in Technology, I2CT 2017*, Apr. 2017, vol. 2017-January, pp. 1246–1249, doi: 10.1109/I2CT.2017.8226326.
- [17] Fedesoriano, "Heart failure dataset," 2021, <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.
- [18] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: a case study," *PLoS ONE*, vol. 12, no. 7, p. e0181001, Jul. 2017, doi: 10.1371/journal.pone.0181001.
- [19] Z. Xu, D. Shen, T. Nie, and Y. Kou, "A hybrid sampling algorithm combining M-SMOTE and ENN based on random forest for medical imbalanced data," *Journal of Biomedical Informatics*, vol. 107, p. 103465, Jul. 2020, doi: 10.1016/j.jbi.2020.103465.
- [20] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, Feb. 2016, pp. 78–83, doi: 10.1109/IACC.2016.25.
- [21] A. Sutradhar, S. Tajmen, A. A. Dhaly, F. M. J. M. Shamrat, M. S. R. Talukder, and A. Khater, "Skin cancer classification and early detection on cell images using multiple convolution neural network architectures," in *3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings*, Oct. 2022, pp. 1089–1094, doi: 10.1109/ICOSEC54921.2022.9952115.
- [22] P. Charilaou and R. Battat, "Machine learning models and over-fitting considerations," *World Journal of Gastroenterology*, vol. 28, no. 5, pp. 605–607, Feb. 2022, doi: 10.3748/wjg.v28.i5.605.
- [23] P. Yang, Y. H. Yang, B. B. Zhou, and A. Y. Zomaya, "A review of ensemble methods in bioinformatics," *Current Bioinformatics*, vol. 5, no. 4, pp. 296–308, 2010, doi: 10.2174/157489310794072508.
- [24] M. T. Le, M. T. Vo, N. T. Pham, and S. V. T. Dao, "Predicting heart failure using a wrapper-based feature selection," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 21, no. 3, pp. 1530–1539, 2021, doi: 10.11591/ijeecs.v21.i3.pp1530-1539.
- [25] G. Lorenzoni *et al.*, "Comparison of machine learning techniques for prediction of hospitalization in heart failure patients," *Journal of Clinical Medicine*, vol. 8, no. 9, p. 1298, Aug. 2019, doi: 10.3390/jcm8091298.
- [26] L. Hussain, W. Aziz, I. R. Khan, M. H. Alkinani, and J. S. Alowibdi, "Machine learning based congestive heart failure detection using feature importance ranking of multimodal features," *Mathematical Biosciences and Engineering*, vol. 18, no. 1, pp. 69–91, 2020, doi: 10.3934/MBE.2021004.




**BIOGRAPHIES OF AUTHORS**

**Ananda Sutradhar**    received the B.Sc. degree in computer science and engineering from Daffodil International University, Bangladesh, in 2023 respectively. His research interests include machine learning, deep learning, and artificial intelligence. He can be contacted at email: [anandasutradhar819@gmail.com](mailto:anandasutradhar819@gmail.com).






**Mustahsin Al Rafi**    received the B.Sc. degree in computer science and engineering from Daffodil International University, Bangladesh, in 2023 respectively. His research areas are machine learning and deep learning. He can be contacted at email: [mustahsin.rafi@gmail.com](mailto:mustahsin.rafi@gmail.com).



**Mohammad Jahangir Alam**    has been a senior lecturer at Daffodil International University since 2017. He is a consistent research enthusiast interested in embedded systems, IoT, machine learning and artificial intelligence. He has now been part of numerous efficient research publications as author/co-author. Along with research, he has expertise in software application development. He can be contacted at email: [jahangir.cse@diu.edu.bd](mailto:jahangir.cse@diu.edu.bd).



**Saiful Islam**    has been a senior lecturer at Daffodil International University since 2016. He is a consistent research enthusiast interested in embedded systems, IoT, machine learning and artificial intelligence. He has now been part of numerous efficient research publications as author/co-author. Along with research, he has expertise in software application development. He can be contacted at email: [saiful.cse@diu.edu.bd](mailto:saiful.cse@diu.edu.bd).