# In-hospital mortality prediction model of heart failure patients using imbalanced registry data: A machine learning approach

H. Sabahi[a], M. Vali[a*], D. Shafie[b]

[a]Department of Biomedical Engineering, K. N. Toosi University of Technology, Tehran, P.O.BOX: 16315-1355, Iran.
[b]Heart Failure Research Center, Cardiovascular Research Institute, Isfahan University of Medical Sciences, Isfahan, Iran.

**Abstract**

Heart failure (HF) is a cardiac dysfunction disease with a high mortality rate that is mostly calculated via registry data. The objective of this work was to predict in-hospital mortality in patients hospitalized with HF utilizing their before-hospitalization registry data. The data include 3968 HF records extracted from Persian Registry Of cardio Vascular diseasE (PROVE)/HF registry. We proposed a method that contains an imbalanced ensemble probabilistic model which using registry data predicts HF patients who die during hospitalization from those who survive. The suggested ensemble model uses machine learning models that several ones, namely Decision Tree, Random Forest, LDA, Logistic Regression, SVM, KNN, and XGBoost were evaluated. We also used feature importance analysis to find the important ones and reduce the complexity. The results illustrated the proposed method can predict in-hospital mortality of HF patients using XGBoost that outperformed all others. Feature importance ranking obtained by XGBoost demonstrated that the proposed method can achieve an acceptable performance with the first 18 important features and XGBoost (accuracy: 76.4%±1.6%, sensitivity: 76.8%±6.9%, specificity: 76.4%±1.8%). Moreover, statistical analysis presented significant predictors of in-hospital mortality (*P-value<0.01*).
In conclusion the proposed method can effectively predict in-hospital mortality of HF patients using the imbalanced data.

**Keywords**

Heart failure, In-hospital mortality, Registry data, Imbalanced data, Machine learning.

## 1. Introduction

* Corresponding author. Tel: +98 (21)-84062405, Fax**:** +98 (21)-8462066.
*E-mail addresses*: sabahihadi@email.kntu.ac.ir (H. Sabahi); mansour.vali@eetd.kntu.ac.ir (M. Vali); d.shafie87@gmail.com (D. Shafie).

Heart failure (HF) is one of the prevalent causes of hospitalization and mortality worldwide. Despite advances in diagnosis and treatment, HF still has a high mortality rate resulting in a growing burden on health providers [1]. Evidence indicates that 40% of hospitalized patients with HF die or are hospitalized again within one year [2]. The increasing mortality of HF has changed it into a life-threatening disease. Therefore, identifying the prevalent predictors and prediction of mortality has been the main focus of current studies [3].

Early recognition of risk factors of the disease can improve prognosis and will be used as predictors of mortality to help in decision-making. Several clinical predictors such as age and depression are associated with increased HF mortality [2]. There is little research on risk prediction models for elderly patients; however, age is an independent predictor in HF patients [4].

Accurately predicting the mortality allows for effective risk classification and provides more appropriate medical care. Calculating the mortality rate of HF around the world is usually based on registry systems [3]. Miro et al. predicted mortality of Acute Heart Failure (AHF) patients using the Epidemiology of AHF in Emergency department registry data [5].

The usage of machine learning techniques for predicting in-hospital mortality of hospitalized patients has been considered a helpful solution in recent years [6]. Fonarow et al. proposed a regression tree using ADHERE Registry data to predict in-hospital mortality probability in patients hospitalized with HF [7]. Konig et al. developed a reliable algorithm to calculate expected in-hospital mortality in HF cohorts based on routine administrative data by comparing regression analysis with four machine learning models [8]. Luo et al. constructed a risk stratification method using an extreme gradient boosting algorithm and available clinical data to predict the in-hospital mortality of hospitalized HF patients in intensive care units (ICUs) [9].

Although registry systems provide informative data regarding diseases in society, they are usually divided up into imbalanced classes that often result in a low sensitivity when ordinary machine learning algorithms are applied. There are many approaches that address imbalanced classification problems. The most common methods consist of oversampling and undersampling, which are relatively able to improve the classification performance. The undersampling and ensembling approach was proved to be advantageous for imbalanced classification [10], in which some classifiers are trained by the minority class and undersampled majority class. Then, they are combined into an ensemble model. Therefore, undersampling and ensembling approach could overcome imbalanced classification problems, but their performance is not still suitable for all imbalanced datasets and they can be improved on a registry dataset for mortality prediction.

Because of the existence of irrelevant and correlated features to target in actual data, feature importance analysis is usually employed to address dimensionality challenges and to improve the system generalization [11]. Alizadehsani et al. ranked all features of coronary artery disease datasets based on their clinical importance to select the best ones with the machine learning techniques [12].

In this research we want to predict in-hospital mortality of HF patients using their before-hospitalization imbalanced registry data. To address this issue, we proposed an imbalanced ensemble probabilistic model to predict in-hospital mortality of HF patients using imbalanced registry data. We showed that the proposed model with Extreme Gradient Boosting (XGBoost) can identically classify both minority and majority classes with a higher performance in comparison with conventional classifiers.

In this research, we investigated the importance of the features to find the influential ones and to reduce the complexity of the proposed model. The usage of the found important features will reduce the cost and time of the registration of HF patients to predict in-hospital mortality. Furthermore, we also found significant predictors resulting from the statistical analysis of the before-hospitalization registry data that are helpful for health providers to forecast mortality and better manage resources. In addition, we have used a Decision Tree algorithm to extract special rules from a subset of data.

In the remainder of this paper, we will present material and methods, then obtained results and relevant discussion, and finally, the conclusion will be stated.

## 2. Material and methods

### 2.1 Data description

The data of this work were included records of patients hospitalized with decompensated or acute HF from March 2015 until October 2018 using data extracted from the Persian Registry Of cardio Vascular diseasE (PROVE). This is the first registry program for cardiovascular diseases that was launched as a pilot study in Isfahan (Iran) in 2014. PROVE registry was for patients with stroke, acute coronary syndrome, atrial fibrillation, ST elevation myocardial infarction, HF, percutaneous coronary intervention, congenital heart disease, familial hypercholesterolemia, and chronic ischemic cardiovascular disease [13-14]. In this study, informed consent forms were obtained from all patients [14].

PROVE/HF is part of the PROVE registry that registers hospitalized HF patients. The collected data consisted of demographic data, underlying diseases, comorbidities, signs and symptoms, physical examination results, diagnoses, paraclinical tests, treatments, and medications. All the gathered data were related to before and during hospitalization as well as the discharge time of the patients. The PROVE/HF registry was followed at 3, 6, and 12 months after the first admission as needed.

The PROVE/HF registry data included 3968 records belonging to 2918 patients (male: 60.52%, age: 68.97 ± 13.26 years, female: 39.48%, age: 73.27 ± 11.66 years); some patients had more than one admission at different times. Totally, 606 features related to before and during hospitalization, discharge time, and three consecutive follow-ups were registered for each patient.

Given the aim of this study, before-hospitalization features were only used to predict in-hospital mortality of HF patients.

### 2.2 Preprocessing of data

After data acquisition, preprocessing plays a vital role in data mining that transforms raw data into appropriate forms for subsequent uses. Figure 1 depicted all preprocessing steps of raw PROVE/HF registry data. As mentioned before, only before-hospitalization features were used in this study to predict in-hospital mortality of patients. Therefore, these features should firstly be extracted from the registry. There are many unnecessary features that should be removed such as dates of procedures. In addition, we removed some features that were the same for all patients and had no variance. The data features are two types; the first type is categorical which describes categories or groups such as the "cigar status" of the patient. The second type is numerical which takes numerical values and represents a measurement such as the "weight" of the patient. Since some categorical features have a lot of missing values, we removed those features that had missing values more than an arbitrary threshold depending upon the importance of the features. Since we wanted to assess the effect of more categorical features on the prediction results, we removed only the features with more than 80% missing values [15]. We have filled the remaining categorical features after consultation with cardiologists and specialists. There were some drugs in the data belonged to the same type, and we merged them as a feature. We also removed some numerical features with many missing values that did not exist in patients' medical records. Finally, each sample was labeled according to the mortality status of the patient. If a patient dies during hospitalization, his records are labeled as '1', otherwise as '0'.

After all preprocessing steps, the HF registry data comprise 3252 samples (class '0' = 3070, class '1' = 182) and 42 features (categorical = 36, numerical = 6). The features between patients who died in the hospital and those who survived were compared using the $X^2$ test and t-test for categorical and numerical features, respectively. A P-value less than 0.01 was statistically considered significant. Table 1 shows all remaining features after preprocessing that includes 8 different groups: Demographic, Aetiology, Medical History, Vital Sign, Physical Examination, Procedures, Medications and Biomarker. Numerical features are presented as $mean \pm SD$ (Standard Deviation), and categorical features are shown as n (%) (Number (Percentage)). Most of the categorical features have two states. For instance, 453 patients of the class '0' had COPD, out of 3070 patients (14.8%), and others did not. Figures 2 and 3 show bar plot of the categorical features and error bar of the numerical features, respectively.

### 2.3 Method

In this section, we describe the proposed method to predict in-hospital mortality of HF patients using their before-hospitalization features of the preprocessed data. The structure of the proposed method is shown in Figure 4.

As reflected in Figure 4, the obtained features should be normalized after preprocessing the raw data. Since there are two types of categorical and numerical features in our data, we used two different methods for each one. Categorical and numerical features are normalized using the Min-Max scaling and standard scaling methods, respectively [16]:

$$\text{Min-Max scaling: } X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

$$\text{Standard scaling: } X_n = \frac{X - X_{mean}}{SD} \tag{2}$$

This work aimed to predict in-hospital mortality of HF patients with a low in-hospital mortality rate of 5.6%. Therefore, we encounter an imbalanced classification problem with two classes in which survived patients during hospitalization are the majority class. Class imbalance problems are frequently happened in the field of medical data processing [17]. Class imbalance causes most algorithms to assign all samples of both classes into the majority one to achieve a high accuracy [18]. We proposed an imbalanced ensemble probabilistic classifier model to distinguish HF patients who die during hospitalization from those who survive using the imbalanced data. Figure 5 illustrates the structure of the imbalanced ensemble probabilistic model.

The proposed method uses the "undersampling and ensembling" strategy to undersample the majority class samples, together with the minority class samples, to train some classifiers [10]. In this method, the majority class samples are undersampled the same size as the minority class ones. The undersampled data of the majority class are not put back again. Whenever the number of majority class samples is not enough to undersample, the needed number of samples is randomly selected from the majority class sample subsets. Hence, the structure of the imbalanced ensemble probabilistic model is created using each undersampled subset of the majority class in conjunction with all samples of the minority class. The total number of created training subsets is $N = \left[\frac{n_j}{n_m}\right] + 1$, where $n_j$ and $n_m$ indicate the number of majority and minority class samples in the training set, respectively. Each training subset is used to train a classifier, and each classifier has its own hyperparameters. Most machine learning models have parameters known as hyperparameters [19] that need to be fixed before training [20]. As Figure 5 shows (Clf tuners), this work tunes the hyperparameters of each classifier based on the corresponding training subset and then trains them using the best found hyperparameters and the training subset. To find the best hyperparameters, we used the basic grid search technique in which all possible permutations of the hyperparameters of a model are applied to build the models. The best model is selected after the evaluation of the performance of each one. To evaluate the built model, the grid search technique uses the 5-fold cross-validation method, which is a resampling method used for evaluating machine learning models [21]. In addition, we designated 'Accuracy' as an evaluation strategy of the performance of the cross-validated model on the test set. Figure 6 summarizes the grid search to find the best hyperparameters of the models.

Following training each classifier with the best found hyperparameters and the training subset, the trained classifier predicts the probability of test set samples. Afterward, the mean probability of the output of all classifiers is

computed for each test set sample. Then, a decision threshold of 0.5 is considered to calculate the predicted label for each test set sample, where probability values less than 0.5 are assigned to class '0', otherwise to class '1'.
To assess the performance of the model, some metrics are figured out as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$F1\ score = \frac{2TP}{2TP + FP + FN} \tag{6}$$

$$Matthews\ Correlation\ Coefficient\left(MCC\right) = \frac{TP \times TN - FP \times FN}{\sqrt{\left(TP + FP\right)\left(TP + FN\right)\left(TN + FP\right)\left(TN + FN\right)}} \tag{7}$$

Where True Negative (*TN*) and True Positive (*TP*) respectively show the numbers of negative and positive samples that are correctly diagnosed. Moreover, False Negative (*FN*) and False Positive (*FP*) respectively represent the numbers of negative and positive samples that are not correctly diagnosed. In addition, we used two other valuable metrics, which include ROC_auc that calculates the area under the receiver operating characteristic (ROC) curve, and PR_auc that calculates the area under the precision-recall curve from predicted labels [22].
According to our proposed model in Figure 4, we used the 5-fold cross-validation method with 10 times repetition to evaluate our model. The final results are computed using the average of all different metrics for model evaluation. The most crucial advantage of the 5-fold cross-validation method is its lower variance than the single hold-out set method. Therefore, its sensitivity to any partitioning bias on the dataset is less than the single hold-out set method. Besides, 5-fold cross-validation is a more robust method than the single hold-out method that randomly splits data into training and testing sets [23].
In practice, machine learning models have different performances for various datasets because their characteristics are different. Therefore, in this study we evaluated and compared several models to select the one that has the best performance for our purpose. In particular, the seven different evaluated classification models are Decision Tree, Random Forest, Linear Discriminant Analysis (LDA), Logistic Regression, Support Vector Machine (SVM), k-nearest neighbor (KNN), and XGBoost [6, 8, 15].
The model was fitted with the best found hyperparameters using the described method above. The technical hyperparameters of the considered models to find the best ones using the basic grid search technique are listed in Table 2. Additionally, we accomplished hierarchical clustering analysis over models based on the *FN* and *FP* values.

### 2.4 Feature importance analysis

In this study, we used feature importance analysis to reduce the number of features and complexity of the proposed model. Feature importance includes techniques that designate a score to each input feature based on how they are effective at the classification performance of a target variable [24]. There are many types of importance scores such as statistical correlation coefficient scores, decision trees, and permutation scores. Decision tree algorithms suggest importance scores based on the decrease in the criterion of the split points, like Entropy or Gini. This approach can

be applied for ensembles of trees such as the Random Forest and XGBoost algorithms. The models Decision Tree, Random Forest, and XGBoost allowed the importance of features to be derived during the training of the models. Linear machine learning algorithms such as logistic regression calculate coefficient statistics of each feature and target variable in order to apply in a weighted sum to make a prediction. These coefficients can be utilized directly as a feature importance score. The models LDA, SVM, and KNN use permutation importance scores where in the feature importance is difficult to extract, and they do not support native importance scores. Briefly, the model is trained on the data, then it is applied to classify the data while the values of a feature have been scrambled. This is repeated for every feature, and the whole process is repeated several times. The result would be a mean importance score for every feature. The accuracy of the model is considered as a basis for the importance score. Obviously, the effect of scrambling the feature values is small for unimportant features but is considerable for important ones that will reduce the model's accuracy.

### 2.5 Rule Extraction

Finally, in this work we extracted some significant rules using the Decision Tree. Due to the imbalanced data, the majority class samples are undersampled the same as the minority class ones. The Decision Tree model is fitted on the constructed data that both classes have the same size, and then all possible rules are extracted. In order to extract the confident rules, we chose the rules with the accuracy of 100% that are supported by at least ten samples.
All of the experiments of the current study are accomplished using *sklearn* [25] machine learning library in Python (version 3.7.0) and SPSS Statistics for Windows, version 15 (SPSS Inc., Chicago, III., USA).

## 3. Results

At first, we illustrate the results of the statistical analysis using the $X^2$ test for categorical features and t-test for numerical features in patients who died in hospital (class '1') and ones who did not (class '0'). Table 1 declares all used features with their P-value which less than 0.01 is considered statistically significant. Numerical features are presented as *mean*±*SD*, and categorical features are shown as numbers and percentages. All features with *P-value<0.01* in Table 1 are statistically significant and can be considered as predictors of in-hospital mortality for HF patients.
After the statistical analysis of the features, we present the obtained results of the proposed method to predict the in-hospital mortality of HF patients using preprocessed imbalanced registry data. As mentioned before, we used the 5-fold cross-validation method with 10 times repetition to evaluate our model while each set includes the same percentage of each target class as the complete dataset. Then, each set is given to the imbalanced ensemble probabilistic model as the input for training and testing. According to the number of training samples, the total number of created training subsets equals 17, where each training subset contains 292 samples, equally of both classes. Table 3 shows the performance of the proposed model on the test sets for all classifier models used to predict the in-hospital mortality of HF patients.

As Table 3 demonstrates, using the KNN as a classifier in the structure of the imbalanced ensemble probabilistic model of the proposed method gives the best accuracy and specificity (83.7% and 85.5%, respectively). In this case, however, the corresponding sensitivity has the lowest value (50%). The XGBoost achieves the best sensitivity, F1 score, ROC_auc, PR_auc and MCC of 77.3%, 27.1%, 84.7%, 34.6% and 28.2%, respectively. According to Table 3, XGBoost has the highest number of top metrics and, therefore, it outperforms all other classifiers. The average ROC and Precision-Recall (PR) curves with the 5-fold data resampling and 10 times repetition are depicted in Figure 7. Most classifiers have AUC values above 80%, but the value of KNN is lower (78%). We also used the AUC value as the criterion of the PR curve. The lowest and the highest values of AUC of PR curves are for KNN and XGBoost, respectively (22%, 35%).

In this study, the hierarchical clustering analysis was applied to cluster the seven classifiers using the FN and FP values from a random sampling. Figure 8 shows the hierarchical clustering analysis that represents similar classifiers

culminate in similar results; for instance, tree based classifiers Decision Tree, Random Forest, and XGBoost are clustered closely.

As previously mentioned, we used feature importance analysis to reduce the number of features and complexity of the proposed model in the current study. We applied three different types of importance scores, including tree based scores for the Decision Tree, Random Forest, and XGBoost models (Figure 9), statistical correlation coefficient scores for Logistic Regression model (Figure 10), and permutation scores for LDA, SVM, and KNN models (Figure 11).

As can be seen, each classifier model creates its feature importance scores based on the corresponding method. In order to reduce the dimension of the features and complexity of the proposed model, we should use a specific model to compute its feature importance scores. According to Table 3, XGBoost outperforms all other models; therefore, it can be used to find the important features. Then, we apply different numbers of sorted important features to the proposed model with the XGBoost and figure out the ROC_auc metric. The result is presented in Figure 12.

As Figure 12 shows, at first, the ROC_auc increases with the number of the important features. But after 18 features, the ROC_auc does not have significant changes. Therefore, we can consider that at least the first 18 important features are required to achieve acceptable model performance. The list of the first 18 important features can be found from the feature importance diagram of XGBoost in Figure 9. Table 4 shows the result of using the first 18 important features as the input of the proposed model for all classifiers used to predict the in-hospital mortality of HF patients.

According to Table 4, XGBoost achieves the best ROC_auc, PR_auc, and MCC of 84.9%, 34.6%, and 27.7%, respectively. Therefore, XGBoost has the highest number of top metrics and outperforms all other classifier models with the first 18 important features. A comparison of the results of the proposed model with the XGBoost in both cases (using all of the features (Table 3) or the first 18 important features (Table 4)) indicates that the model can slightly perform better with the all features, but the improvement is not significant. Therefore, in order to reduce the complexity of the model, we can only use the first 18 important features that are extracted by XGBoost.
Finally, we describe the significant extracted rules. Based on the Decision Tree, seven rules, which are illustrated as IF (Antecedent) and Then (Consequent) in Table 5, were generated with the accuracy of 100% and at least ten samples. The presence of high BUN, low heart rate, low Hb, and NIV usage was associated with HF patients who died during hospitalization. On the other hand, the normal range of BUN, Hb, SBP, and not using NIV was associated with those that did not die during hospitalization. However, more investigation with larger data sets and more features is still required.

## 4. Discussion

The present research applied the data of the first Iranian national registry of cardiovascular diseases and, besides the statistical analysis of the significant features, proposed a model to predict the in-hospital mortality of HF patients. As mentioned before, some of the features were statistically significant (*P-value<0.01*). The results of the statistical analysis are in line with previous related studies reporting the predictors of mortality rate and morbidity in HF patients. According to Table 1, anemia and kidney disease have a significant relationship to in-hospital mortality which is consistent with the previous findings in other countries [26]. The effect of anemia on HF mortality has been clarified in several studies that have recommended treatment of anemia as a preventive factor to reduce HF mortality [27]. According to Table 1, the average systolic and diastolic blood pressures (SBP and DBP) of patients who died during hospitalization are lower than those who did not die. This issue induces hypotension (low blood pressure) in

patients with severe HF. The higher mortality rate of HF patients with hypotension is consistent with our result. It is completely a logical finding because the low blood pressure in HF disease is related to cardiogenic shock that shows the pump failure of the heart and a lower left ventricular ejection fraction (LVEF) [26]. On the other hand, if the blood pressure is low, the heart will struggle to deliver enough oxygenated blood to the cells; therefore, the body will increase the heart rate (HR) to push more oxygen-rich blood to the cells [28]. Our finding of the HR of severe HF patients justifies this result again. In spite of advances in technology, physical examination remains essential in the management of HF patients. Edema, JVP, crackle, and CPO are all statistically significant physical examinations of the current study which can be considered as predictors of in-hospital mortality. Patients with any type of procedure in their medical history are mostly at a severe stage of the disease. In these patients, the higher mortality rate is a sensible result demonstrated by a statistically significant relationship to PCI, hemodialysis, and NIV therapy. The investigation of the predictive role of medications presents that losartan as an Angiotensin receptor blocker (ARB) and ASA as an antiplatelet were more often used in the patients who did not die during hospitalization while Hydrochlorothiazide as a diuretic was utterly vice versa. Therefore, losartan and ASA can be considered as preventive factors for mortality. There is also disagreement on the role of diuretics medications on the mortality of HF patients in the results of various researches. Some studies suggested higher mortality rates of HF patients by diuretic use, which is in line with the outcome of the current research [29], while others offered protective [30]. The effect of diuretics on the mortality of HF patients has to be further studied. The crucial role of biomarkers is growingly recognized in HF management, diagnosis, and screening of severe patients [31]. An increased level of serum creatinine during HF hospitalization is associated with worse outcome [32]. This issue is in line with our result about the higher creatinine levels in patients who died in hospital. A persistently high level of BUN is also associated with an increased risk of cardiovascular readmission and death [10] which is close to our result according to Table 1. Another important primary biomarker is cardiac troponin which its level can be elevated in HF patients [33]. This issue is demonstrated in our findings in Table 1 which patients who died in hospital have more positive troponin than others. In the current data, the in-hospital mortality rate of HF patients is 5.6% which is in the range of several published registries (4%-7%) [34-35]. To reduce the mortality of HF patients, on the one hand, healthcare staff and physicians should pay more attention to the predictors and treatment of underlying disease (such as anemia and hypotension). On the other hand, patients should adhere to medications (especially ARB and ASA).

Besides the statistical analysis of the features, we proposed a new model to predict in-hospital mortality of the HF patients using the imbalanced registry data. Most algorithms frequently obtain poor performance with imbalanced datasets because they tend to get high accuracy and assign the most samples to the majority class, which causes low sensitivity.

There are many approaches that address imbalanced classification problems. Oversampling and undersampling are the most commonly used approaches. These methods improve the overall performance of the classification. However, Oversampling may increase the likelihood of occurring overfitting, especially for higher rates of oversampling. Furthermore, it will increase the computational effort and decrease the classifier performance. In undersampling, huge number of data are discarded. This can be very problematic as the elimination of such data may make the decision boundary between majority and minority classes harder to learn, resulting in a high variance and performance loss.

The undersampling and ensembling approach was displayed to be more effective than others for the imbalanced classification [10] that trains several classifiers using the minority and the undersampled majority class samples and then combines the output of classifiers into an ensemble structure. Ensemble methods will reduce the variance of the results by aggregating the prediction performance of the classifiers [18]. Inspired by this approach, we proposed a new ensemble model to predict the in-hospital mortality of HF patients using imbalanced registry data.

In the suggested model, the class probability of each test set sample is calculated after training the classifier. Then, the mean class probability of all classifiers is computed for labeling the sample. In the proposed model, we compute the class probability of each test set sample instead of predicting the class label for each classifier directly. This method will provide a more accurate class probability for each test set sample and, therefore, will reduce the *FP* and *FN* that will cause to increase in the performance of the classification.

In the proposed model, different classifiers are used and each one has its own performance. According to Table 3 and Table 4, although KNN has the highest accuracy and specificity among all used classifiers, its main drawback is the great difference between sensitivity and specificity that shows it cannot equally classify both minority and

majority classes. According to Table 3 XGBoost has the highest sensitivity with all features but according to Table 4 Random Forest has the highest sensitivity with only the first 18 important features. These tables demonstrate that F1 score of all classifiers are close to each other but XGBoost has the highest F1 score when all features are used and Logistic Regression has the highest one when only the first 18 important features are used. According to Table 3 and Table 4, Random Forest and XGBoost have the best ROC_auc, PR_auc, and MCC among all classifiers and their values are so close to each other. However, XGBoost is preferred because its values are a bit better than Random Forest. In XGBoost the variance of ROC_auc and MCC are lower and the mean of PR_auc is higher than the values of Random Forest. Also, sensitivity and specificity of XGBoost are so close to each other that means it can classify both classes identically. According to Table 3 and Table 4, XGBoost has the highest number of top metrics and, therefore, it outperforms all other classifiers.

The number of the training samples is a fundamental determinant of classification accuracy, and they are correlated to each other [36]. In the current study, we assessed the effect of the training sample size on the classification performance for all applied classifier models.

To evaluate the effect of the training size on the model performance, we examined the proposed model for different numbers of training sets that are applied in the structure of the imbalanced ensemble probabilistic model (Figure 5). As Figure 13, Figure 14, and Figure 15 demonstrate, the more training sets, the better performance the proposed model can achieve for all classifier models. A larger training sample size improves the learning process, and classifiers can stratify the new samples with more accuracy. Therefore, if we have more data, we will obtain higher classifications performance. In addition, these figures restate that KNN has the lowest sensitivity for different numbers of training sets despite it has the best accuracy and specificity, therefore, KNN is not a proper classifier for our purpose. According to Figure 13, Figure 14, and Figure 15 SD of sensitivity is more than SD of accuracy and specificity, which it is because our data are highly imbalanced and the number of patients who died in hospital (class '1') is much less than survived ones (class '0').

Feature importance provides insight into the model and data and it is the basis for feature selection which can improve the performance of a predictive model. In short, feature Importance score is used to perform feature selection. Supervised feature selection methods include filter methods, wrapper methods and embedded methods that each one includes different techniques. The filter methods choose the best subset of feature space immediately before feeding it to a learning algorithm. The remaining two approaches, embedded and wrapper, create the optimal subset of features in conjunction with the learning algorithm. Contrary to the other methods, the embedded methods put together the advantages of both the wrapper and filter methods. Dissanayake et al. have conducted an experimental evaluation of the performance of models created for heart disease prediction using various feature selection techniques such as ANOVA, Chi-square, mutual information, Relief algorithm, forward and backward feature selection and so on [37]. Finally, the feature subset achieved by the backward technique that belongs to wrapper methods led to the highest classification accuracy. In this study, we used the feature importance analysis to reduce the number of features and complexity of the proposed model. We applied the classifier models to extract the feature importance scores, and the result of each classifier completely differed over the order of the features. Because the XGBoost had the best performance, we chose its feature importance scores, and only the first 18 important features had a significant effect on the XGBoost performance. According to the feature importance analysis, for the most models, some features, including NIV, Hb, Heart Rate, Bun, SBP, DBP, Creatinine, and Troponin have high importance scores and can affect the model performance considerably. On the other hand, there are some features such as CRT-D, Primary Heart Ischemic, Cold Peripheral Organs (CPO), Captopril, and ICD which showed less importance scores and would not affect the performance significantly. NIV was shown with the highest importance score in all seven models; therefore, it is a notable predictor of in-hospital mortality for HF patients. NIV can help to decrease respiratory effort and will improve gas exchange and cardiac output [38]. Hence, the HF patients who are at severe stage could use NIV to provide relief to HF symptoms, and for this reason in our data the percentage of dead patients who used NIV is more than others. We aim to add more data to decrease the effect of imbalanced data in the future. In addition, using of the registry data to predict mortality of HF patients during 3, 6 and 12 months after discharge can be investigated in future studies. It would be also interesting to develop the proposed model for the imbalanced multiclass classification problems.

## 5. Conclusion

In this work, we proposed a method to predict in-hospital mortality of HF patients using PROVE/HF imbalanced registry data of hospitalized patients with HF. The method contains an imbalanced ensemble probabilistic model that using an undersampling and ensembling approach can distinguish HF patients who die during hospitalization from those who do not. The suggested model uses machine learning models, and among the various models evaluated, XGBoost could outperform all others with higher performance. Feature importance analysis using XGBoost showed the proposed method could achieve acceptable performance with fewer but important features (accuracy: 76.4% ± 1.6%) which can reduce the system complexity considerably. In addition, the statistical analysis of the features suggests predictors that can be used by health providers to determine the required medical resources to reduce the in-hospital mortality of HF patients.

## Declaration of competing interest

The authors have no conflict of interest to disclose.


## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or Not-for-profit sector.


## 6. References


1. Yancy, C. W., Jessup, M., Bozkurt, B., et al. "2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines", *Journal of the American College of Cardiology*, 62(16), pp. 147-239 (2013).
2. Pocock, S. J., Wang, D., Pfeffer, M. A., et al. "Predictors of mortality and morbidity in patients with chronic heart failure", *European Heart Journal*, 27(1), pp. 65-75 (2005).
3. Mahmooda, S. S, Levy, D., Vasan, R. S., et al. "The Framingham Heart Study and the epidemiology of cardiovascular diseases: A historical perspective", *Lancet*, 383(9921), pp. 1933-1945 (2014).
4. Azad, N. and Lemay, G. "Management of chronic heart failure in the older population", *Journal of geriatric cardiology: JGC*, 11(4), pp. 329 (2014).
5. Miró, Ò., Rossello, X., Gil, V., et al. "Predicting 30-day mortality for patients with acute heart failure in the emergency department: a cohort study", *Annals of internal medicine*, 167(10), pp. 698-705 (2017).
6. Seki, T., Kawazoe, Y., and Ohe, K. "Machine learning-based prediction of in-hospital mortality using admission laboratory data: A retrospective, single-site study using electronic health record data", *PloS one*, 16(2), pp. 0246640 (2021).
7. Fonarow, G. C., Adams, K. F., Abraham, W. T., et al. "Committee ASA. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis", *Jama*, 293(5), pp. 572-580 (2005).
8. König, S., Pellissier, V., Hohenstein, S., et al. "Machine learning algorithms for claims data-based prediction of in-hospital mortality in patients with heart failure", *ESC heart failure*, 8(4), pp. 3026-3036 (2021).
9. Luo, C., Zhu, Y., Zhu, Z., et al. "A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure", *Journal of translational medicine*, 20(1), pp. 1-9 (2022).
10. Wallace, B. C., Small, K., Brodley, C. E., et al. "Class imbalance, redux", *Paper presented at: 2011 IEEE 11th international conference on data mining*, (2011).
11. Wojtas, M. and Chen, K. "Feature importance ranking for deep learning", *arXiv. arXiv preprint arXiv*, 201008973, pp. 18 (2020).
12. Alizadehsani, R., Khosravi, A., Roshanzamir, M., et al. "Coronary artery disease detection using artificial intelligence techniques: A survey of trends, geographical differences and diagnostic features 1991–2020", *Computers in Biology and Medicine*, 128, pp. 104095 (2021).
13. Sakata, Y. and Shimokawa, H. "Epidemiology of heart failure in Asia", *Circulation Journal*, 0971, pp. CJ-13 (2013).

14. Givi, M., Sarrafzadegan, N., Garakyaraghi, M., et al. "Persian Registry of cardioVascular diseasE (PROVE): Design and methodology", *ARYA atherosclerosis*, 13(5), pp. 236 (2017).
15. Hachesu, P. R., Ahmadi, M., Alizadeh, S., et al. "Use of data mining techniques to determine and predict length of stay of cardiac patients", *Healthcare informatics research*, 19(2), pp. 121-129 (2013).
16. Freedman, D. "*Statistics: Fourth International Student Edition",* W.W. Norton & Company (2007).
17. Guo, X., Yin, Y., Dong, C., et al. "On the class imbalance problem", *Paper presented at: 2008 Fourth international conference on natural computation* (2008).
18. Wang, Y., Wang, D., Ye, X., et al. "A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction", *Information Sciences*, 474, pp. 106-124 (2019).
19. Probst, P., Boulesteix, A. L., and Bischl B. "Tunability: importance of hyperparameters of machine learning algorithms", *The Journal of Machine Learning Research*, 20(1), pp. 1934-1965 (2019).
20. Weerts, H. J., Mueller, A. C., and Vanschoren, J. "Importance of tuning hyperparameters of machine learning algorithms" *arXiv preprint arXiv*, 200707588 (2020).
21. Daghistani, T. A., Elshawi, R., Sakr, S., et al. "Predictors of in-hospital length of stay among cardiac patients: A machine learning approach", *International journal of cardiology*, 288, pp. 140-147 (2019).
22. Flach, P. and Kull, M. "Precision-recall-gain curves: PR analysis done right", *Advances in neural information processing systems*, 28 (2015).
23. Refaeilzadeh, P., Tang, L., and Liu, H. "Cross-validation", *Encyclopedia of database systems*, 5, pp. 532-538 (2009).
24. Malliaris, M. E. and Pappas, M. "Revenue generation in hospital foundations: neural network versus regression model recommendations", *International Journal of Management & Information Systems (IJMIS)*, 15(1) (2011).
25. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. "Scikit-learn: Machine learning in Python", *the Journal of machine Learning research*, 12, pp. 2825-2830 (2011).
26. Mosterd, A. and Hoes, A.W. "Clinical epidemiology of heart failure", *heart*, 93(9), pp. 1137-1146 (2007).
27. Kaldara-Papatheodorou, E. E., Terrovitis, J.V., and Nanas, J.N. "Anemia in heart failure: should we supplement iron in patients with chronic heart failure?", *Polskie Archiwum Medycyny Wewnetrznej*, 120(9), pp. 354-360 (2010).
28. Norcliffe-Kaufmann, L., Kaufmann, H., Palma, J. A., et al. "Orthostatic heart rate changes in patients with autonomic failure caused by neurodegenerative synucleinopathies" *Annals of neurology*, 83(3), pp. 522-531 (2018).
29. Domanski, M., Tian, X., Haigney, M., et al. "Diuretic use, progressive heart failure, and death in patients in the DIG study", *Journal of cardiac failure*, 12(5), pp. 327-332 (2006).
30. Ahmed, A., Husain, A., Love, T. E., et al. "Heart failure, chronic diuretic use, and increase in mortality and hospitalization: an observational study using propensity score methods", *European heart journal*, 27(12), pp. 1431-1439 (2006).
31. Spoletini, I., Coats, A. J., Senni, M., et al. "Monitoring of biomarkers in heart failure", *European Heart Journal Supplements*, 21(Supplement_M), pp. M5-M8 (2019).
32. Shlipak, M. G., Chertow, G. C., and Massie, B. M. "Beware the rising creatinine level" *Journal of cardiac failure",* 9(1), pp. 26-28 (2003).
33. Wettersten, N. and Maisel, A. "Role of cardiac troponin levels in acute heart failure" *Cardiac failure review*, 1(2), pp. 102 (2015).
34. Kociol, R. D., Hammill, B.G., Fonarow, G.C., et al. "Generalizability and longitudinal outcomes of a national heart failure clinical registry: Comparison of Acute Decompensated Heart Failure National Registry (ADHERE) and non-ADHERE Medicare beneficiaries", *American heart journal*, 160(5), pp. 885-892 (2010).
35. Abraham, W. T., Fonarow, G. C., Albert, N. M., et al. "Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF)" *Journal of the American College of Cardiology*, 52(5), pp. 347-356 (2008).
36. Ramezan, C. A., Warner, T. A., Maxwell, A. E., et al. "Effects of Training Set Size on Supervised Machine-Learning Land-Cover Classification of Large-Area High-Resolution Remotely Sensed Data" *Remote Sensing*, 13(3), pp. 368 (2021).

37. Dissanayake, K. and Md Johar, G. M. "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms" *Applied Computational Intelligence and Soft Computing*, 2021, pp.1-17 (2021).
38. Baram, M. "BiPAP and the relief of CHF symptoms" *Critical Care*, 1(1), pp. 1-3 (2001).

**Figure Captions**

- Figure 1. Preprocessing of heart failure (HF) registry data (Fs: Features, Cat: Categorical, Num: Numerical).
- Figure 2. Categorical Features.
- Figure 3. Numerical Features, SD: Standard Deviation.
- Figure 4. The overall structure of the proposed method, SD: Standard Deviation.
- Figure 5. The structure of the imbalanced ensemble probabilistic model (Clf: Classifier).
- Figure 6. The grid search for the hyperparameter tuning with 5-fold cross-validation.
- Figure 7. ROC curve (Left) and Precision-Recall curve (Right) of the proposed model.
- Figure 8. Hierarchical clustering analysis of the classifier models.
- Figure 9. Feature importance bar chart of the Decision Tree, Random Forest, and XGBoost models.
- Figure 10. Feature importance bar chart of the Logistic Regression model.
- Figure 11. Feature importance bar chart of the SVM, KNN, and LDA models.
- Figure 12. ROC_auc metric of the proposed model with XGBoost for different numbers of the important features.
- Figure 13. Accuracy of the proposed model and the number of the training subsets for different classifiers. Dashed lines: Linear Regression, Vertical lines: Standard Deviation.
- Figure 14. Sensitivity of the proposed model and the number of the training subsets for different classifiers. Dashed lines: Linear Regression, Vertical lines: Standard Deviation.
- Figure 15. Specificity of the proposed model and the number of the training subsets for different classifiers. Dashed lines: Linear Regression, Vertical lines: Standard Deviation.

**Table Captions**

- Table 1. Features of the preprocessed HF registry data including categorical (n (%)) and numerical (*mean* ± *SD*) features in 8 different groups. *P-value<0.01* is considered significant statistically. Index: Myocardial Infarction (MI) of a patient, COPD: Chronic Obstructive Pulmonary Disease, SBP: Systolic Blood Pressure, DBP: Diastolic Blood Pressure, JVP: Jugular Venous Pulse, CPO: Cold Peripheral Organs, CABG: Coronary Artery Bypass Grafting, PCI: Percutaneous Coronary Intervention, CRT-D: Cardiac Resynchronization Therapy-Defibrillator, ICD: Implantable Cardioverter Defibrillator, NIV: Non-Invasive Ventilation, Hb: Hemoglobin, BUN: Blood Urea Nitrogen.
- Table 2. Technical hyperparameters of the classification models to find the best ones using the basic grid search technique.
- Table 3. The performance of the proposed model on the test sets for various classifiers used to predict the in-hospital mortality of HF patients, *mean*±*SD* (%) (SD: Standard Deviation).
- Table 4. The performance of the proposed model on the test sets with the first 18 important features for various classifiers used to predict the in-hospital mortality of HF patients, *mean*±*SD* (%)(SD: Standard Deviation).
- Table 5. Significant extracted rules with the accuracy of 100% and at least ten samples using the Decision Tree.

**List of Figures**

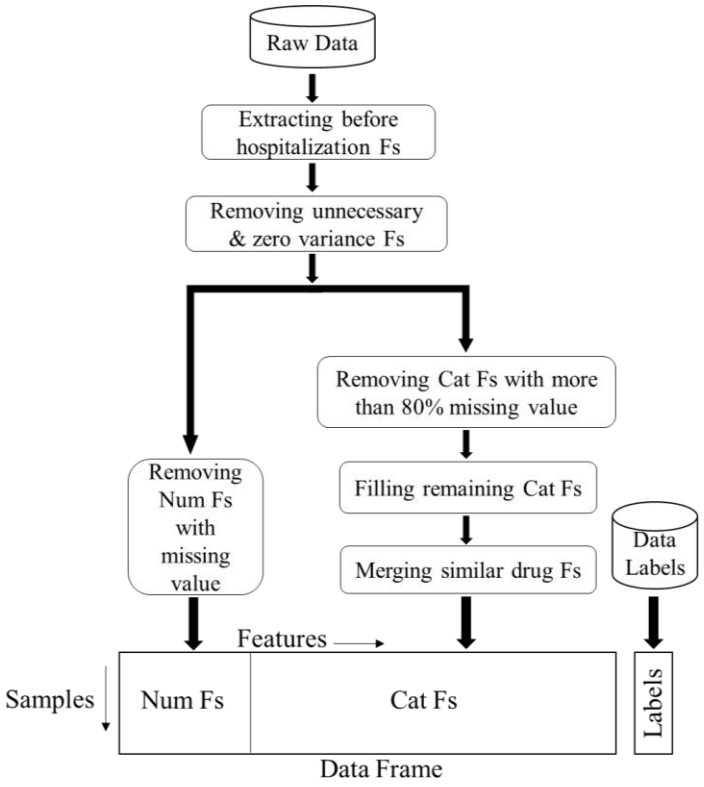



Figure 16.

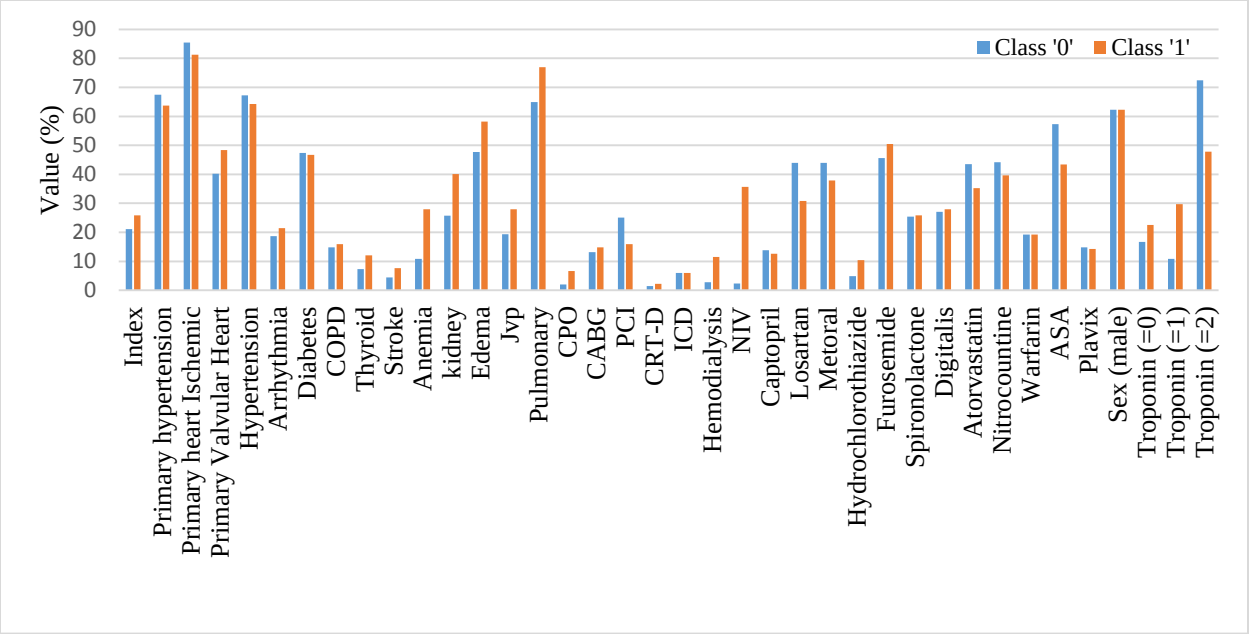



Figure 2.

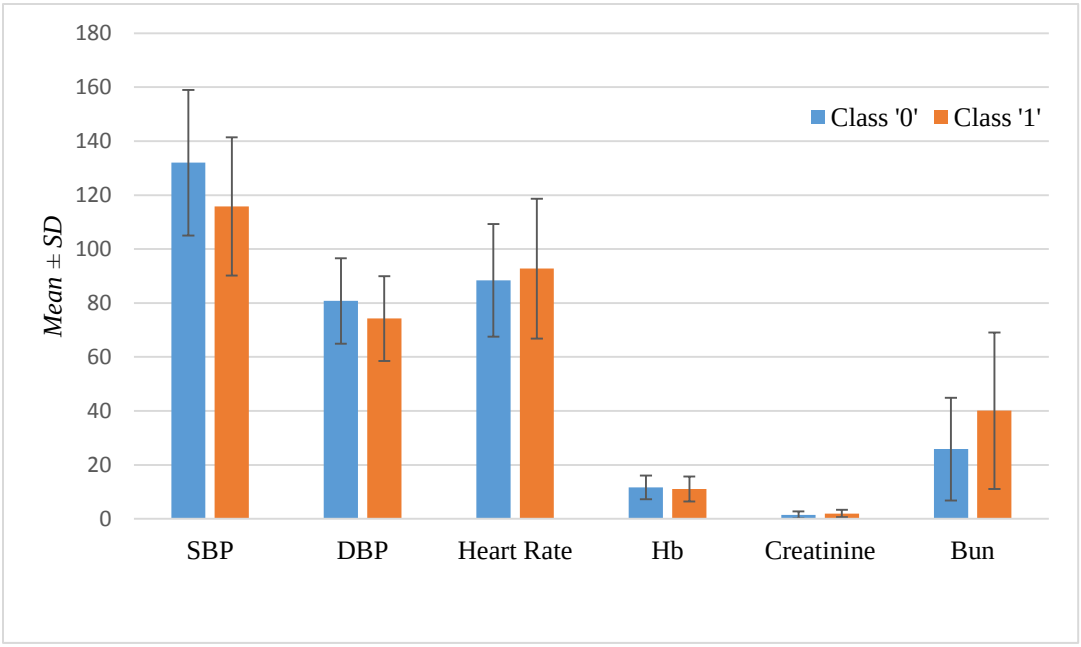



Figure 3.

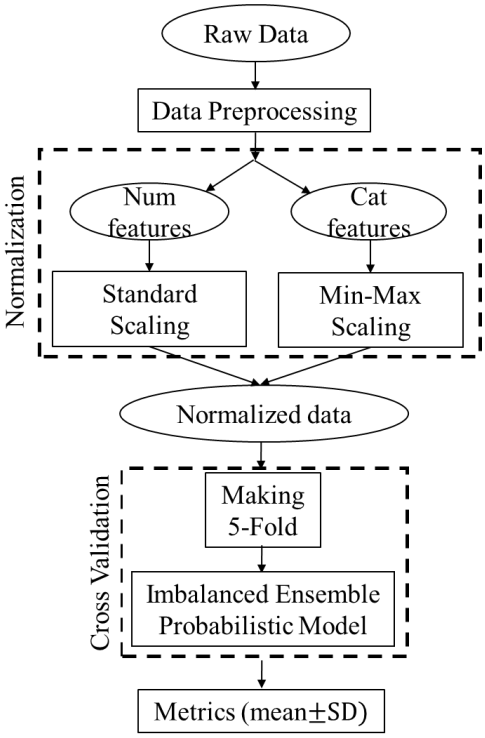



Figure 4.

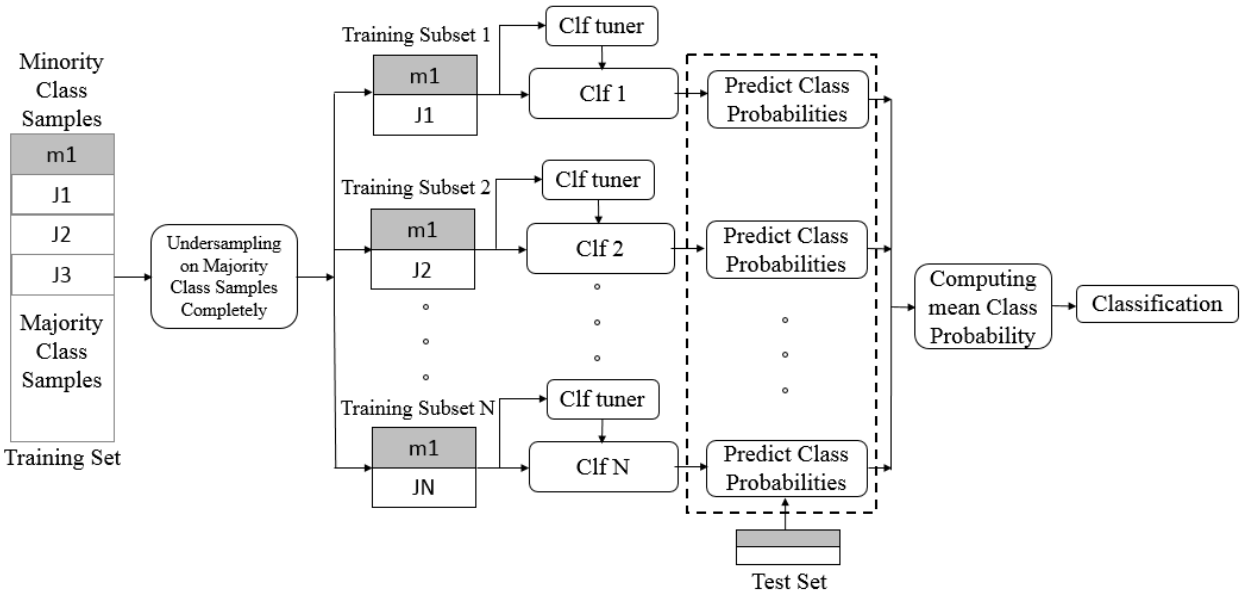



Figure 5.

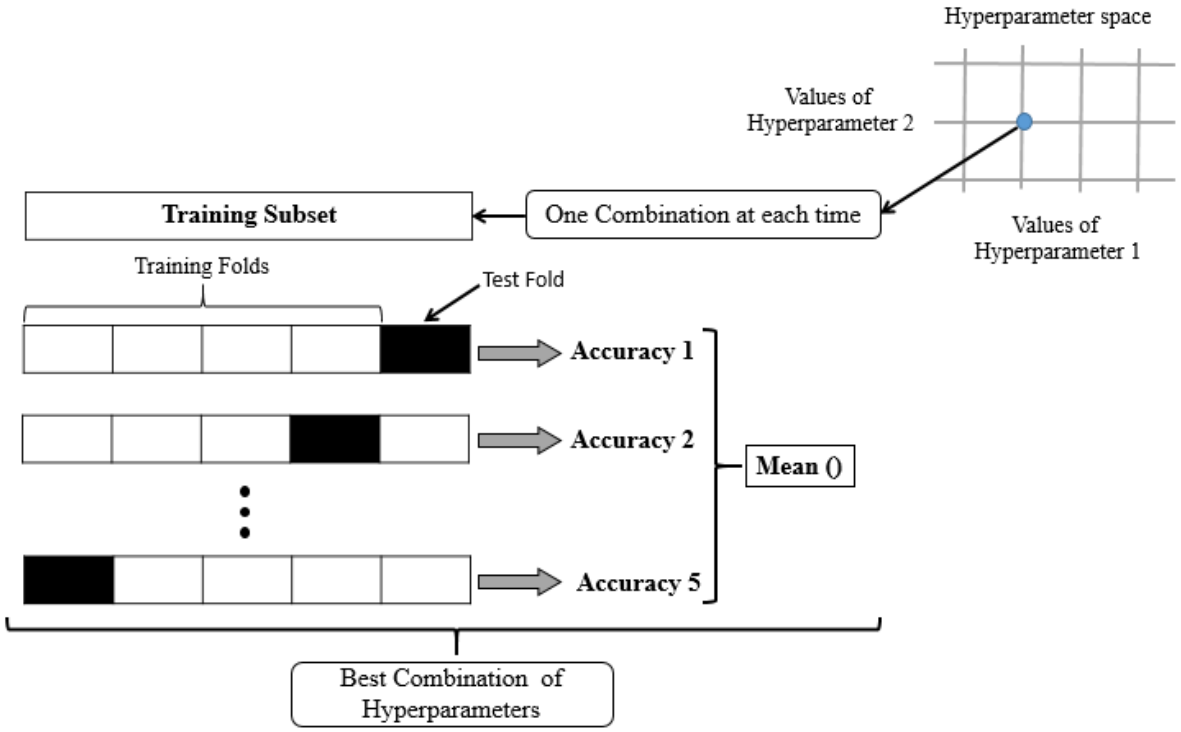



Figure 6.

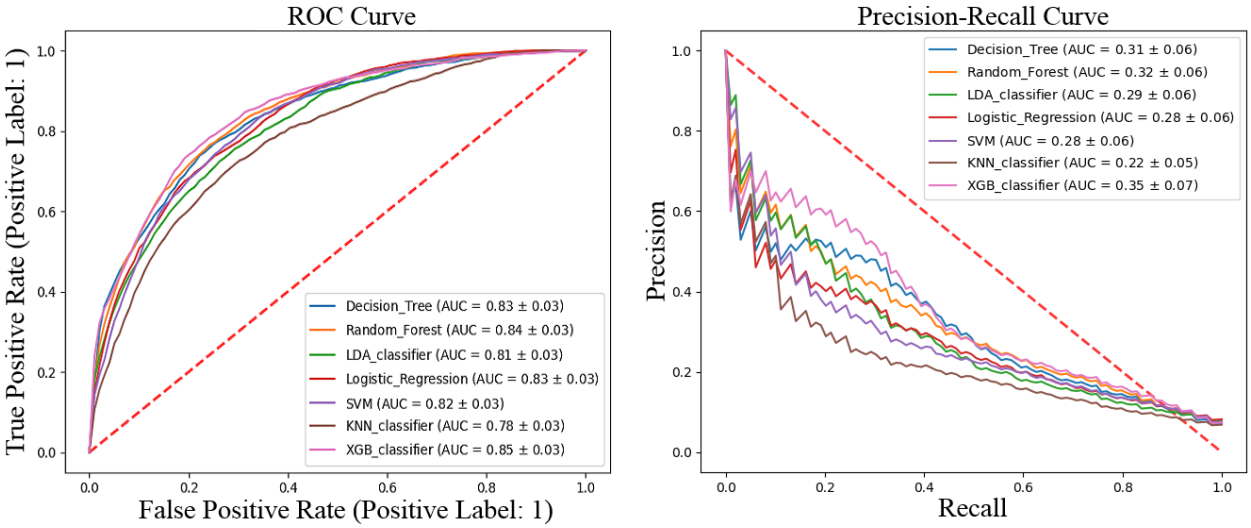



Figure 7.

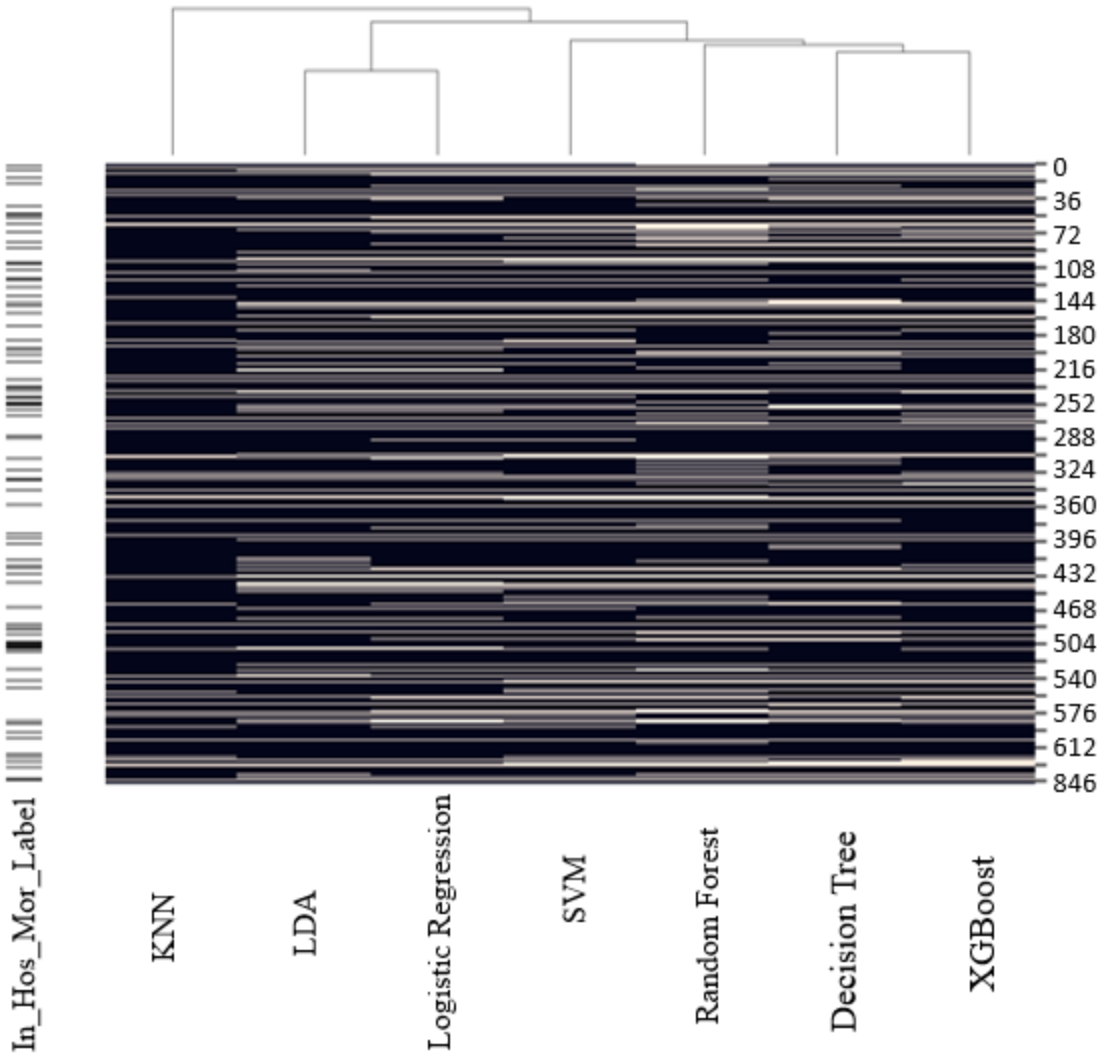

0
36
72
108
144
180
216
252
288
324
360
396
432
468
504
540
576
612
846
In_Hos_Mor_Label
KNN
LDA
Logistic Regression
SVM
Random Forest
Decision Tree
XGBoost

Figure 8.

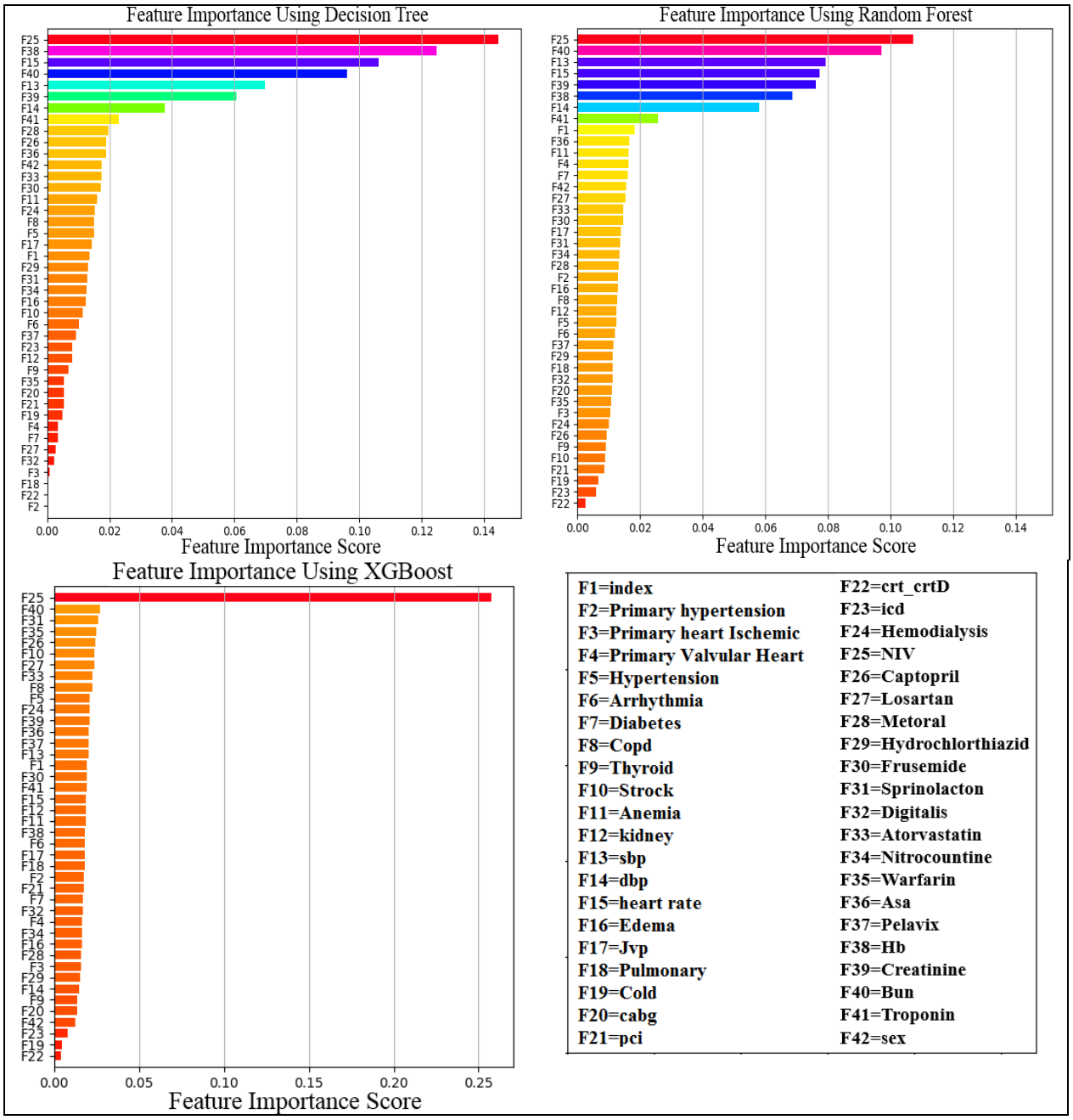
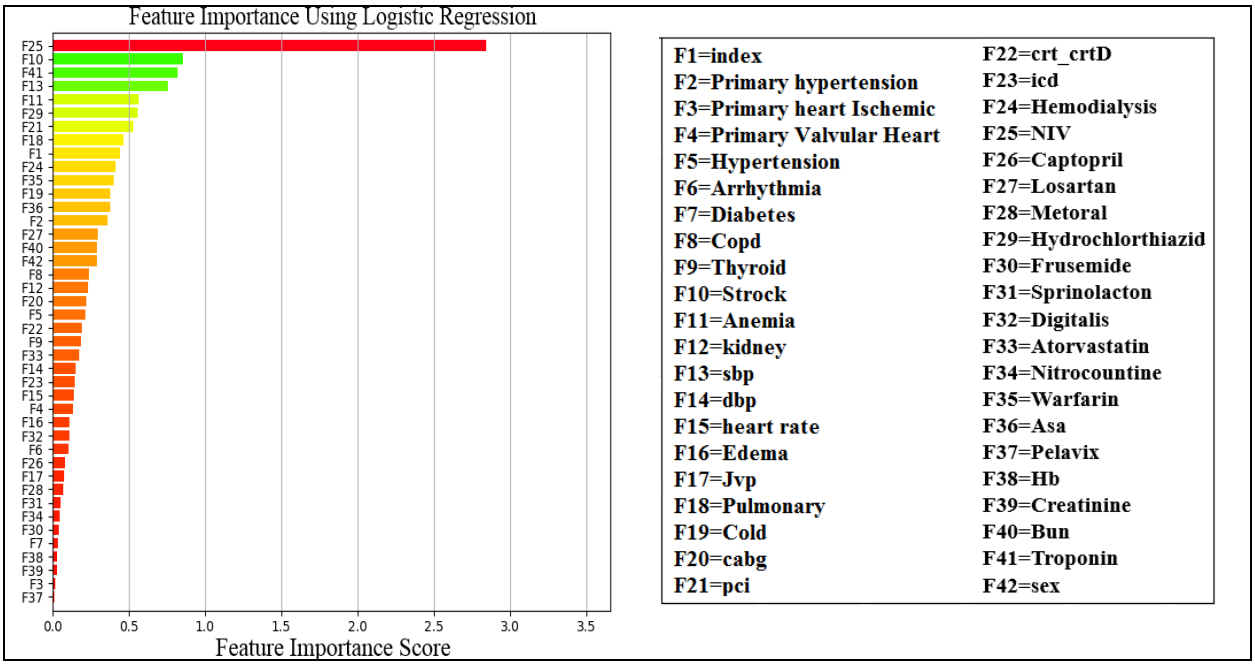
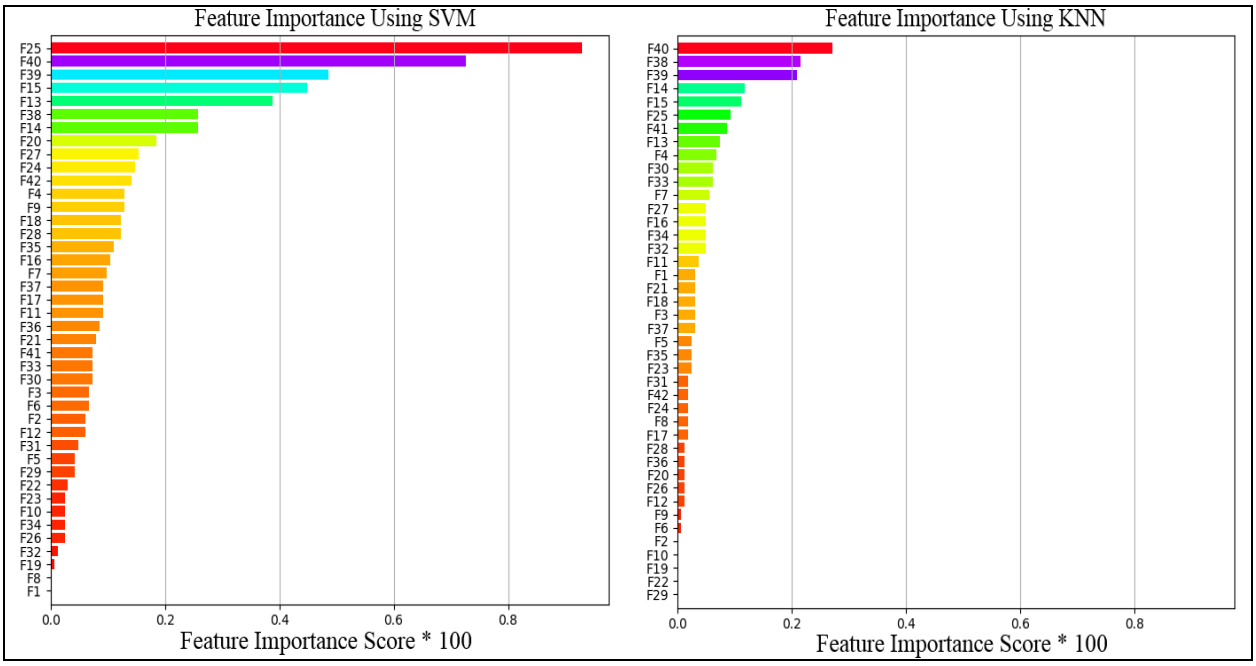

Feature Importance Using Decision Tree
Feature Importance Score
Feature Importance Using Random Forest
Feature Importance Score
Feature Importance Using XGBoost
Feature Importance Score
F1=index
F2=Primary hypertension
F3=Primary heart Ischemic
F4=Primary Valvular Heart
F5=Hypertension
F6=Arrhythmia
F7=Diabetes
F8=Copd
F9=Thyroid
F10=Strock
F11=Anemia
F12=kidney
F13=sbp
F14=dbp
F15=heart rate
F16=Edema
F17=Jvp
F18=Pulmonary
F19=Cold
F20=cabg
F21=pci
F22=crt_crtD
F23=icd
F24=Hemodialysis
F25=NIV
F26=Captopril
F27=Losartan
F28=Metoral
F29=Hydrochlorthiazid
F30=Frusemide
F31=Sprinolacton
F32=Digitalis
F33=Atorvastatin
F34=Nitrocountine
F35=Warfarin
F36=Asa
F37=Pelavix
F38=Hb
F39=Creatinine
F40=Bun
F41=Troponin
F42=sex

Figure 9.




578 Figure 10.



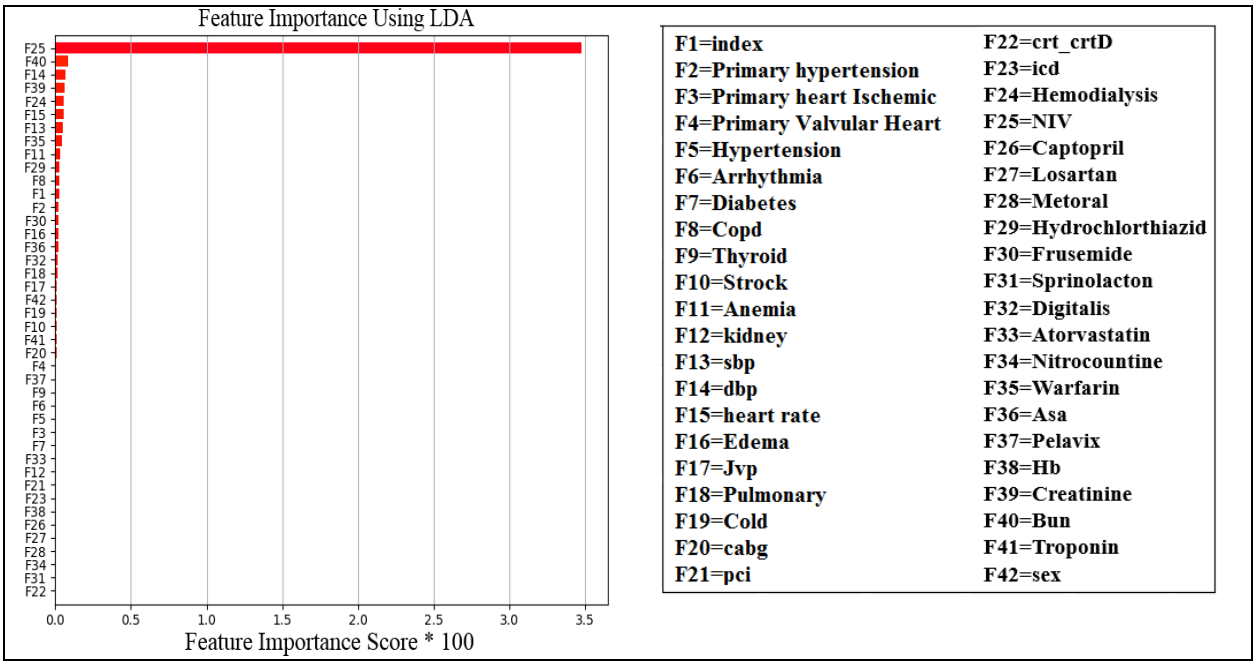



Figure 11.

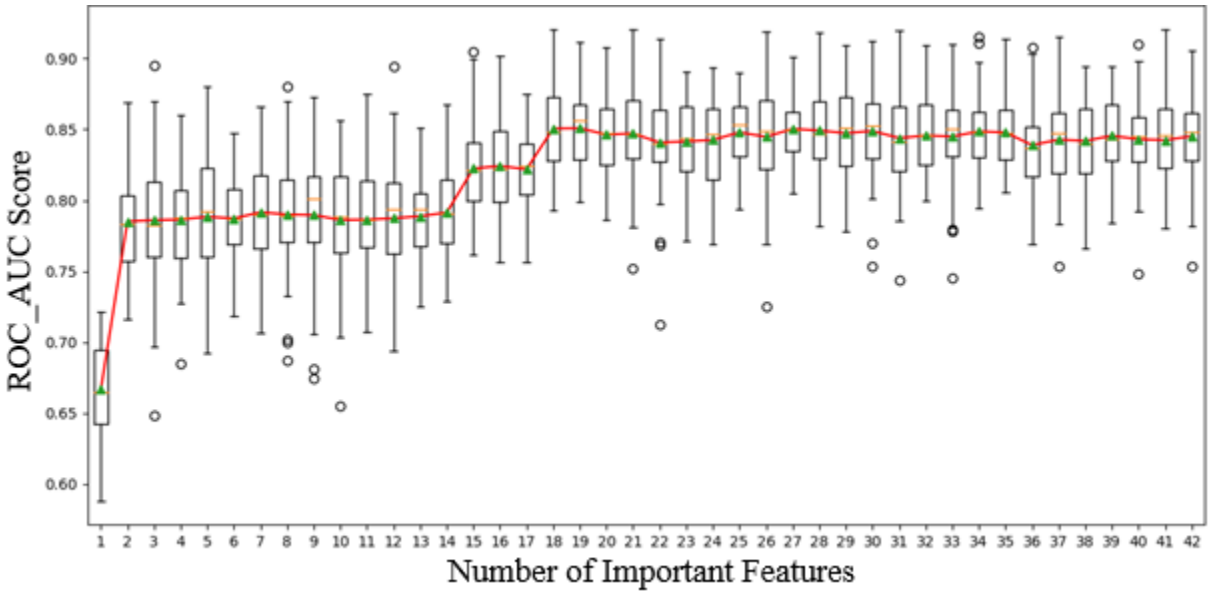



Figure 12.

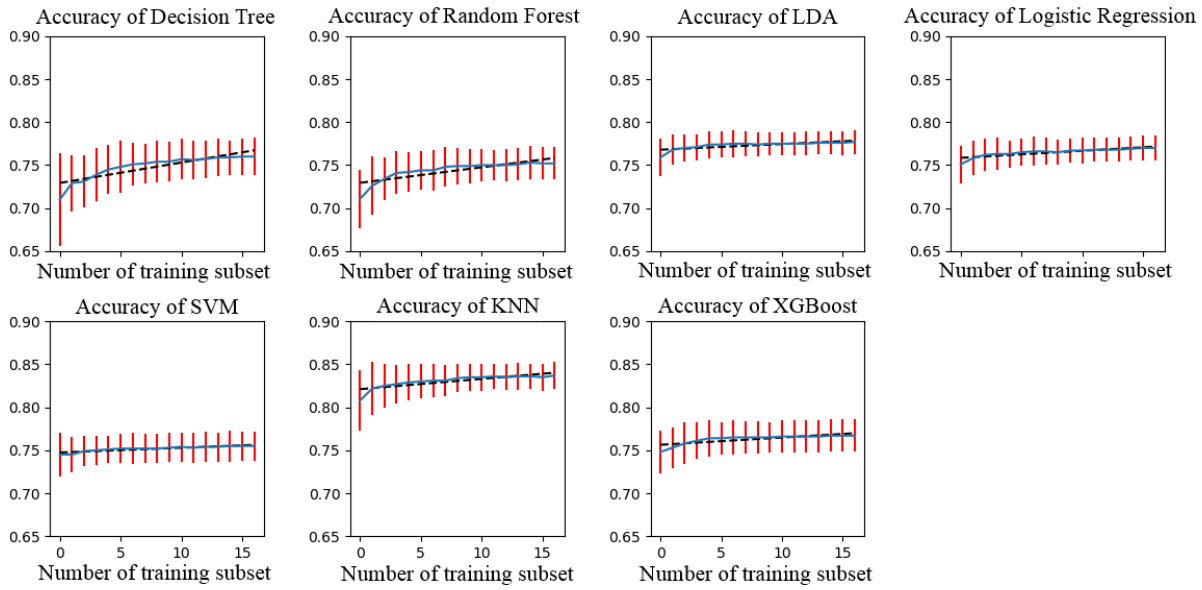



Figure 13.

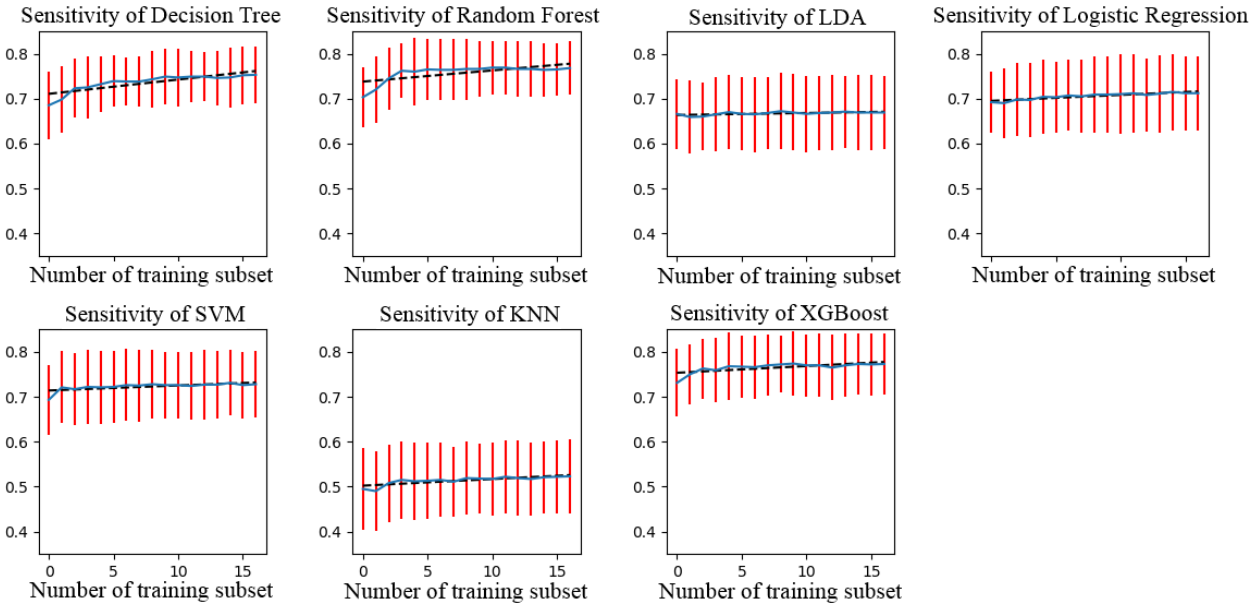



Figure 14.

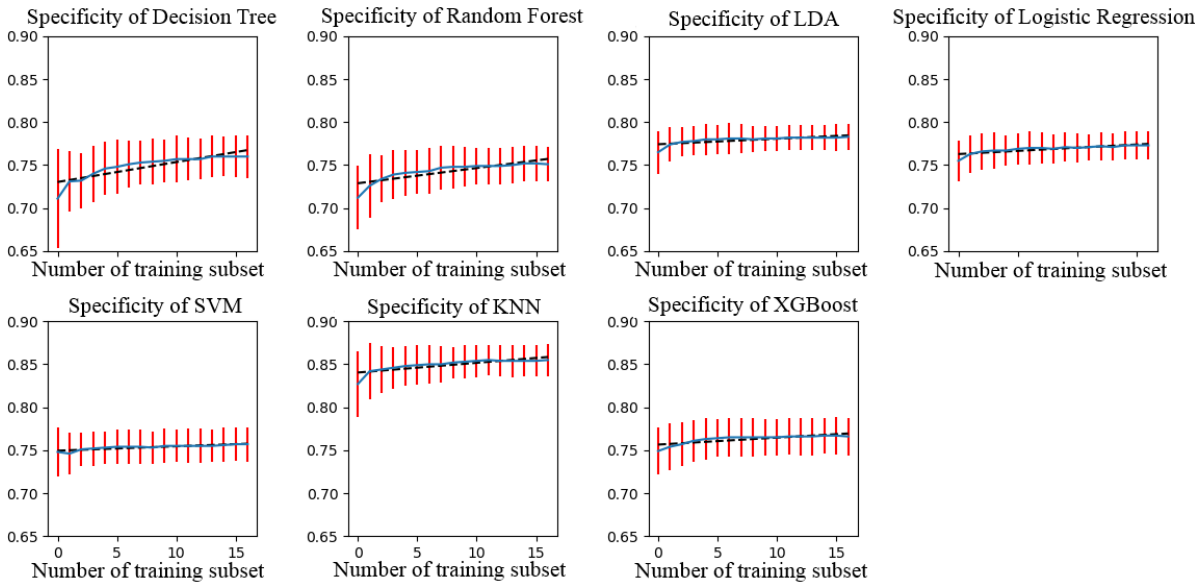

Specificity of Decision Tree
Specificity of Random Forest
Specificity of LDA
Specificity of Logistic Regression
Specificity of SVM
Specificity of KNN
Specificity of XGBoost
Number of training subset

Figure 15.

**List of Tables**

Table 1.

| # | Group Features | Type | Features | | Class '0', n=3070 | Class '1', n=182 | P-value |
|---|---|---|---|---|---|---|---|
| 1 | Demographic | Categorical | Index | | 650(21.1) | 47(25.8) | 0.516 |
| 2 | Aetiology | Categorical | Primary Hypertension | | 2072(67.5) | 116(63.7) | 0.294 |
| 3 | | Categorical | Primary Heart Ischemic | | 2623(85.4) | 148(81.3) | 0.128 |
| 4 | | Categorical | Primary Valvular Heart | | 1234(40.2) | 88(48.4) | 0.03 |
| 5 | Medical History | Categorical | Hypertension | | 2063(67.2) | 117(64.3) | 0.417 |
| 6 | | Categorical | Arrhythmia | | 574(18.7) | 39(21.4) | 0.36 |
| 7 | | Categorical | Diabetes | | 1454(47.4) | 85(46.7) | 0.863 |
| 8 | | Categorical | COPD | | 453(14.8) | 29(15.9) | 0.664 |
| 9 | | Categorical | Thyroid | | 224(7.3) | 22(12.1) | 0.018 |
| 10 | | Categorical | Stroke | | 137(4.5) | 14(7.7) | 0.044 |
| 11 | | Categorical | Anemia | | 332(10.8) | 51(28) | <0.001 |
| 12 | | Categorical | Kidney Disease | | 789(25.7) | 73(40.1) | <0.001 |
| 13 | Vital Sign | Numerical | SBP | | 131.99±26.95 | 115.81±25.65 | <0.001 |
| 14 | | Numerical | DBP | | 80.77±15.86 | 74.26±15.7 | <0.001 |
| 15 | | Numerical | Heart Rate | | 88.43±20.9 | 92.73±25.92 | 0.008 |
| 16 | Physical Examination | Categorical | Edema | | 1464(47.7) | 106(58.2) | 0.006 |
| 17 | | Categorical | Jvp | | 594(19.3) | 51(28.0) | 0.004 |
| 18 | | Categorical | Crackle | | 1992(64.9) | 140(76.9) | 0.001 |
| 19 | | Categorical | CPO | | 62(2.0) | 12(6.6) | <0.001 |
| 20 | Procedures | Categorical | CABG | | 404(13.2) | 27(14.8) | 0.517 |
| 21 | | Categorical | PCI | | 770(25.1) | 29(15.9) | 0.005 |
| 22 | | Categorical | CRT-D | | 47(1.5) | 4(2.2) | 0.482 |
| 23 | | Categorical | ICD | | 183(6.0) | 11(6.0) | 0.963 |
| 24 | | Categorical | Hemodialysis | | 85(2.8) | 21(11.5) | <0.001 |
| 25 | | Categorical | NIV | | 73(2.4) | 65(35.7) | <0.001 |
| 26 | Medications | Categorical | Captopril | | 424(13.8) | 23(12.6) | 0.655 |
| 27 | | Categorical | Losartan | | 1348(43.9) | 56(30.8) | 0.001 |
| 28 | | Categorical | Metoral | | 1349(43.9) | 69(37.9) | 0.111 |
| 29 | | Categorical | Hydrochlorothiazide | | 150(4.9) | 19(10.4) | 0.001 |
| 30 | | Categorical | Furosemide | | 1399(45.6) | 92(50.5) | 0.190 |
| 31 | | Categorical | Spironolactone | | 779(25.4) | 47(25.8) | 0.892 |
| 32 | | Categorical | Digitalis | | 832(27.1) | 51(28.0) | 0.786 |
| 33 | | Categorical | Atorvastatin | | 1336(43.5) | 64(35.2) | 0.027 |
| 34 | | Categorical | Nitrocountine | | 1357(44.2) | 72(39.6) | 0.220 |
| 35 | | Categorical | Warfarin | | 590(19.2) | 35(19.2) | 0.997 |
| 36 | | Categorical | ASA | | 1759(57.3) | 79(43.4) | <0.001 |
| 37 | | Categorical | Plavix | | 453(14.8) | 26(14.3) | 0.862 |
| 38 | Biomarker | Numerical | Hb | | 11.70±4.39 | 11.11±4.61 | 0.081 |
| 39 | | Numerical | Creatinine | | 1.476±1.26 | 1.963±1.36 | <0.001 |
| 40 | | Numerical | Bun | | 25.85±19.08 | 40.1±29.02 | <0.001 |
| 41 | | Categorical | Troponin | 0=Not done | 514(16.7) | 41(22.5) | <0.001 |
| | | | | 1=Positive | 333(10.8) | 54(29.7) | |
| | | | | 2=Negative | 2223(72.4) | 87(47.8) | |
| 42 | Demographic | Categorical | Sex | Male | 1914(62.3) | 106( 58.2) | 0.268 |
| | | | | Female | 1156(37.7) | 76(41.8) | |

Table 2.

| **Models** | **Hyperparameters** |
|---|---|
| Decision Tree | Criterion (To measure the split quality),<br>Min_samples_split (Minimum number of samples to split an internal node) |
| Random Forest | Min_samples_split,<br>Max_features (The number of features for the best split) |
| LDA | Solver (Algorithm to use in optimization problem) |
| Logistic Regression | Solver,<br>C (Regularization) |
| SVM | Kernel, C (Regularization) |
| KNN | k (Number of neighbors),<br>weights (Weight function),<br>metric (Distance metric) |
| XGBoost | n_estimators (Number of boosting stage),<br>learning_rate,<br>subsample (The fraction of samples to fit the individuals learners),<br>max_depth (The maximum depth of individuals learners) |

Table 3.

| | *Accuracy* | *Sensitivity* | *Specificity* | *F1 score* | *ROC_auc* | *PR_auc* | *MCC* |
|---|---|---|---|---|---|---|---|
| Decision Tree | 76±2.2 | 75.3±6.4 | 76±2.5 | 26±2.2 | 83.4±2.6 | 31.3±6 | 26.7±3.2 |
| Random Forest | 75.2±1.9 | 76.8±6 | 75.1±2 | 25.8±2 | 84.2±2.6 | 32.1±6 | 26.7±3 |
| LDA | 77.7±1.4 | 66.9±8.1 | 78.3±1.5 | 25.1±2.5 | 81.4±3.3 | 29.5±6 | 24.3±3.9 |
| Logistic Regression | 77±1.4 | 71.2±8.2 | 77.3±1.6 | 25.7±2.4 | 82.8±3.1 | 27.7±6 | 25.6±3.9 |
| SVM | 75.5±1.7 | 72.8±7.5 | 75.7±2 | 25±2.1 | 82.5±2.9 | 27.7±6 | 25.1±3.4 |
| KNN | **83.7±1.6** | 52.3±8.2 | **85.5±1.9** | 26.4±3.4 | 77.9±3.1 | 22.5±5 | 23.4±4.4 |
| XGBoost | 76.7±1.9 | **77.3±6.9** | 76.6±2.2 | **27.1±2.1** | **84.7±2.9** | **34.6±6.7** | **28.2±3.1** |

Table 4.

| | *Accuracy* | *Sensitivity* | *Specificity* | *F1 score* | *ROC_auc* | *PR_auc* | *MCC* |
|---|---|---|---|---|---|---|---|
| Decision Tree | 76.5±1.8 | 74.9±7.9 | 76.5±2 | 26.3±2.4 | 84±3.5 | 33.1±6.6 | 26.9±3.8 |
| Random Forest | 75.9±1.9 | **77.6±7.1** | 75.8±2.1 | 26.5±2.1 | 84.9±3.1 | 34.3±6.5 | 27.7±3.3 |
| LDA | 80.2±1.6 | 65.8±8.1 | 81.1±1.9 | 27.1±2.4 | 83.3±3.3 | 31.6±6.5 | 26.2±3.6 |
| Logistic Regression | 78.7±1.5 | 71.6±7.2 | 79.2±1.8 | **27.4±2.1** | 83.8±3.3 | 30.7±6 | 27.5±3.3 |
| SVM | 77.4±1.7 | 74.2±7.3 | 77.6±2 | 26.9±1.9 | 83.9±3 | 28.8±5.7 | 27.4±3.1 |
| KNN | **82.9±1.4** | 57.3±7.3 | **84.4±1.5** | 27.3±3 | 80.7±3.2 | 24.7±5.4 | 25±4 |
| XGBOOST | 76.4±1.6 | 76.8±6.9 | 76.4±1.8 | 26.7±1.9 | **84.9±2.8** | **34.6±7.5** | **27.7±3.1** |

Table 5.

| | Antecedent | Consequent |
|---|---|---|
| **1** | (NIV=No), (Bun<25), (Troponin>1.5), (Creatinine<=1.45), (SBP>110.32), (Hb<=15.39) | class 0 |
| **2** | (NIV=Yes), (DBP>40.44), (10.49< Hb<=18.44), (Bun<84) | class 1 |
| **3** | (NIV=No), (Bun>25), (SBP<=119.81), (3.69<Hb<17), (Troponin<=1.5) | class 1 |
| **4** | (NIV=No), (Troponin<=1.5), (SBP>132.31), (Bun<=13.12), (Heart Rate<=165.61) | class 0 |
| **5** | (NIV=Yes), (DBP>40.44), (Bun<84), (Hb<10.5), (Troponin<=1.5) | class 1 |
| **6** | (NIV=No), (SBP<=99.83), (Heart Rate>52.6), (Bun<=20.52), (Hb<=15.74) | class 1 |
| **7** | (NIV=No), (SBP<=119.8), (Hb<17), (Troponin>1.5), (heart rate<=122.6), (Bun>30.2), (DBP>64.4) | class 1 |

**Biographies**

**Hadi Sabahi** is a PhD Candidate in the field of medical data analysis at the Biomedical Engineering Department,
K.N. Toosi University of Technology under supervision of Professor Mansour Vali. His research interests remain in the fields of large data processing in medical application using artificial intelligence as well as medical image processing. He has focused on analysis of medical registry data of cardiovascular diseases using machine learning approaches to predict mortality and morbidity.

**Mansour Vali** received his PhD degree in 2006 from Amirkabir University of Technology in Biomedical Engineering, Tehran, Iran. He has been working as an Assistant Professor and a faculty member at the Department of Biomedical Engineering, K.N. Toosi University of Technology since 2013. His main research interests are “sound and speech processing in medical and psychological assessments”. He has developed a new course with the same title at the Electrical Engineering Department of K.N. Toosi University. Also, he is working on large data processing in medical applications.

**Davood Shafie** received his Diploma of Medicine in 2002 from Isfahan University of Medical Sciences, Isfahan, Iran. Due to his interest in studying cardiovascular disease, he studied cardiology at the Isfahan University of Medical Sciences. He completed his professional training in the field of HF and transplantation at the Iran University of Medical Sciences. He currently serves as a HF specialist at Shahid Chamran Cardiovascular, Medical & Research Center in Isfahan. He has also been the deputy director of the Heart Failure Research Center since 2014. Over the years, he has also coordinated the HF registry program in Isfahan. He is also a member of the Isfahan Cardiovascular Research Center (WHO Collaborating Center in EMR).