Improving Classification Using SMOTE on Imbalanced Heart Failure Data

Nahary Jannath[1*], Md Abdul Masud[2], Mohammad Jamal Hossain[3], Md Samsuzzaman[4]

[1,2&3]Department of Computer Science and Information Technology (CSIT), Patuakhali Science and Technology University. nahary6201@cse.pstu.ac.bd, masud@pstu.ac.bd, jamalpstu07@gmail.com

[4]Department of Computer Science and Communication Engineering (CCE), Patuakhali Science and Technology University, Bangladesh. sobuz@pstu.ac.bd.

[*]Correspondence: nahary6201@pstu.ac.bd

Abstract: Heart Failure is the leading cause of death globally. As heart disease classification is a very challenging task, this research transforms a huge amount of raw data into useful information and provides an automated system that helps in making valuable decision of infected but alive heart failure patients with reduced costs. This research analyses the patients of heart failure from Kaggle website with 299 records and aims to predict the heart failure survivor with high accuracy. We have presented this research with seven base classifiers, Support Vector Machine (SVM), Gaussian Naïve Bayesian (GNB), K Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost) to investigate the effective class identification performance. Moreover, the given dataset has an imbalanced positive and negative class distribution problem which reduces the overall performance. Random Forest with Synthetic minority oversampling technique (SMOTE) is implemented to overcome the majority and minority class imbalance problems. Therefore, the employed seven machine learning classifiers are trained with the highest ranked selected significant features by RF. The results are compared with base learner algorithms using a full set of features. The experimental results report that among all classifiers SMOTE based Random Forest classifiers outperforms all other models on imbalanced heart failure data.

Keywords: Heart Disease, Feature selection, SMOTE, F-Score, Classification performance.

Introduction

Heart disease is a complex disease where coronary arteries are blocked and the heart can't supply enough oxygenated blood to heart muscles to meet the body's demands. There as on why many people are affected by heart failure is that there are a wide variety of factors like blood pressure, creatine, high cholesterol, diabetics, smoking habit and limited physical exercise.

Nowadays, clinical test results are made based on doctors' personal experience rather than on the information stored in large databases. So, the results lead to irreparable errors, biases, and huge costs which affect the patient's quality of service. Accurate medical diagnosis is an important step for complex disease result analysis. However, class imbalance arises in different medical records because the number of patients suffering from a complex disease is very smaller than the number of healthy people.

These imbalanced datasets heavily affect the performance of the classifier (Luque A., Carrasco A., Martin A, 2019). Class imbalance is the main obstacle to complex medical diagnosis as the number of suffering patients of deadly diseases is very fewer than the number of healthy people. The same occurrence has appeared in heart failure dataset with a clinical history of patients that is bound to be imbalanced which is being used in this research. The heart failure diagnosis is disposed tobias as the dataset is not balanced and the number of dead patients is very fewer than the number of infected patients who are alive (Fernandez, Garcia, Herrera, & Chawla, 2018). As in the heart failure dataset the target variables are not distributed in an equal proportion, incorrectly classifying a heart failure affected patient as a healthy person with no disease can cost the life in future. Therefore, for handing class imbalance problem we have applied sampling method Synthetic minority over sampling technique (SMOTE) (Fernandez A. et al. 2018) on the heart failure data set.

This is the main focusing point that has motivated our attempts to help the heart failure patients as well as healthcare professionals by developing the machine learning automated techniques and accurately prediction of heart disease for saving millions of lives.

Furthermore, in our research Random Forest based Synthetic minority oversampling technique (RF-SMOTE) is proposed to overcome the majority and minority class imbalance problems. Synthetic minority oversampling techniques generate new synthetic samples of target features and the overall classification performance is increased.

In modern era, machine learning is widely used in medicine, education, agriculture, business and engineering fields. Machine learning algorithm strans forma huge amount of raw data into useful information that provides an automated system that helps in making valuable decisions with reduced costs.

Jiang Y. et al. (2021) used real heart disease dataset of China for predicting cardiovascular disease risk by examining feasibility of seven widely used Machine learning models. There is the influence of imbalanced data in their research. But the authors did not use under sampling or over sampling or any sampling methods to deal with imbalanced data.

D. Velusamy (2021), I. D. Mienye (2020), S. Shah (2020), L. Ali (2019), S. Nalluri (2019), S. Mohan (2019), A. Rahaman (2019), J. Li (2020) and A. Baccouche (2020) also developed many machine learning algorithms for heart disease prediction purposes. D. Velusamy used three base classifiers kNN, SVM and RF for correct prediction of cardiovascular disease. The outcome of three classifiers were combined by majority voting using average-voting, majority-voting, and weighted-average voting. F.J Shaikh and D.S Rao (2021) predicted cancer by machine learning approaches. Lu, Men (2021) and A. K. Pathak (2020) used LSTM, fuzzy logic and Neural network for multi disease analysis. I. D. Mienye (2020) also published a paper for heart disease risk predictions. The above discussed papers conducted with classification performance but class imbalanced issue is not addressed well. Though some paper is focused on sampling techniques, the prediction results is not good enough. Azad, C. et al. (2021)applied SMOTE to oversample the minority class in its pre-processing stage and the outcome achieved by the proposed system in terms of CA, CE, precision, sensitivity, FM and AUROC is 82.1256%, 17.8744%, 0.8070%, 0.8598, 0.8326 and 0.8511, respectively.

The Statistics Committee of The American Heart Association displays and analyses heart disease and stroke related records. The statistical report presents the latest heart disease condition along with quality of heart failure patients care and total medical costs. L. Ali (2019) introduced an expert system using the SVM model and SVM was used as a predictive model. A hybrid grid search algorithm was used with SVM. In this paper, the class imbalance problem is not considered and the minority class was not resembled. S. Mohan et al. predicted heart disease using hybrid machine learning technique in 2020 but the prediction accuracy is 0.87 that is not good enough.

But in our research the main highlighting point is to resample the minority class with SMOTE for better heart prediction. The most significant features were identified using Random Forest in our research so that the heart failure risk also can be monitored and presented.

Materials and Methods

SMOTE converts imbalanced data into balanced data. Figure 1 shows the working procedure of SMOTE based heart failure prediction. For heart failure prediction, the research process starts with an important feature selection technique using Random Forest after analyzing the data set. Then Synthetic Minority Oversampling Technique (SMOTE) is implemented for the data sampling process in order to handle class imbalance problems. Then balanced data in converted into training and test section. 70% data are trained and 30% data are tested for result analysis. Thus, we fit a balanced dataset with seven machine learning algorithms including Support Vector machine (SVM), Gaussian Naive Bayes (GNB), K Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost) and finally prediction result is obtained by trained classifiers.

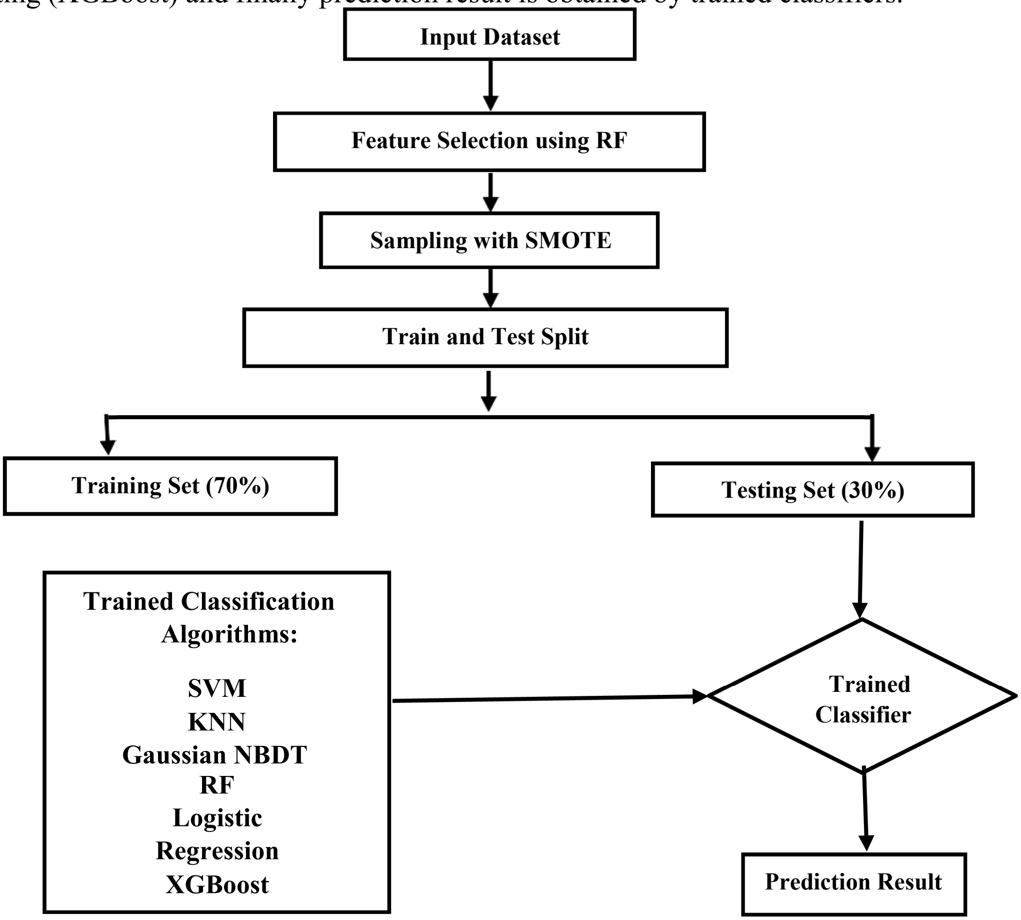

Figure1. Flow chart of SMOTE based heart failure prediction

Feature Selection Technique using Random Forest

Figure 2 shows the feature selection technique which improves predictive performance by choosing best features (Dubey, 2018). Here, we have used Random Forest (RF) for feature ranking. Random forest evaluates all features and selects the significant features according to 100 decision trees per random forest evaluation matrix with cross-validation. Depending on how well a feature can improve the purity of the leaves, Random Forest determine how important the feature is. The significance of the feature increases with the increase in leaf purity. This is carried out for each tree individually, average dal trees and then normalized to 1. So, the sum of the importance scores calculated by a Random Forest is 1. RF selects the features by exploring the training set to avoid over fitting. Thus, RF selects the best significant attributes that provide effective predicting outcomes. RF is used for reducing unnecessary features, reducing the training time, reducing under fitting and over fitting, and for increasing the predictive performance.

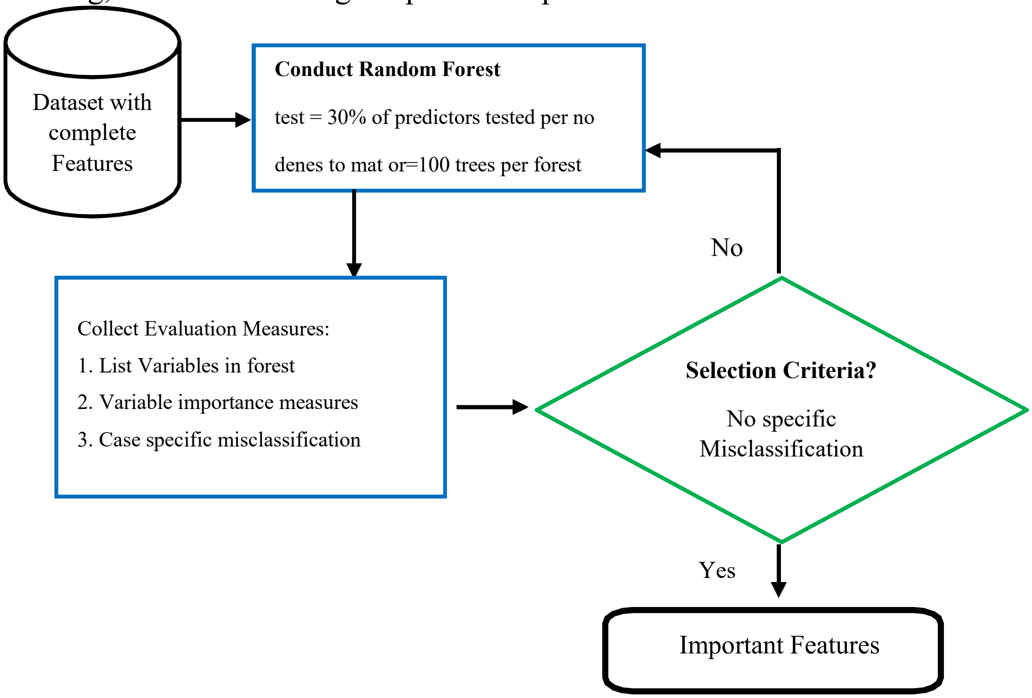

Figure 2. Feature Selection Procedure by Random Forest

Sampling with Synthetic Minority Over sampling Techniques (SMOTE)
While this research is dealing with classification problems, the percentage of the target classes in the total sample plays an important role. In a data set, when the represents the minority class then it is called an imbalanced dataset.

In this case, the machine learning models try to fit the dominating majority class and provide a biased and in correct prediction. At the same time, it gives a false intellect of accuracy. In this situation, even if we classify all records as alive, we will achieve good accuracy but that's not our objective. Now the question is how we can deal with such a scenario. One of the ways is to resample the data set by either decreasing the majority class or by increasing the minority class observations and we can also do that randomly. In case of simple under sampling, we simply select as mall subset of the majority class and keep the minority class as it is. Definitely it helps us balance the dataset but discarded observations can have some valuable information and such an approach could lead us to bias. Similarly in case of random oversampling, we randomly add more observations by copying some or all of those observations by replicating them multiple times. That is why SMOTE technique is used as a sampling technique and also used for handling these problems.

SMOTE creates new synthetic observations or new data points. SMOTE process for creating synthetic dataset is given in Figure 3 by (Fernandez, Garcia, Herrera, & Chawla, 2018).

- First, every data point is plotted then identifies the feature vector and its nearest neighbors and it takes the difference between two.
- Then simply multiply the difference with a random number *gap* between zero and one.
- The next step is to identify a new data point on the line segment by adding the random number to the feature vector.
- Finally, repeat the same process for identified feature vectors. Each*j*-thfeature, *j=1 p*, of the synthetic data sets,

$$X_{synthetic} = \left[X_{synthetic}^{(1)} + X_{synthetic}^{(2)} \dots \dots . X_{synthetic}^{(p)}\right] \quad (1)$$

Equation (1) can be express as,

$$X_{synthetic}^{(j)} = X^{(j)} + gap^{(j)} * [X_{neighbor}^{(j)} - X^{(j)}] \qquad (2)$$

Where,

$X^{(j)}$is the *j*-th feature of the chosen instance X

$Xneighbor^{(j)}$ is the *j*-th feature of a randomly chosen neighbor data point of instance $Xgap^{(j)}$is the uniformly distributed random variable from (0, 1) for the *j-th* feature

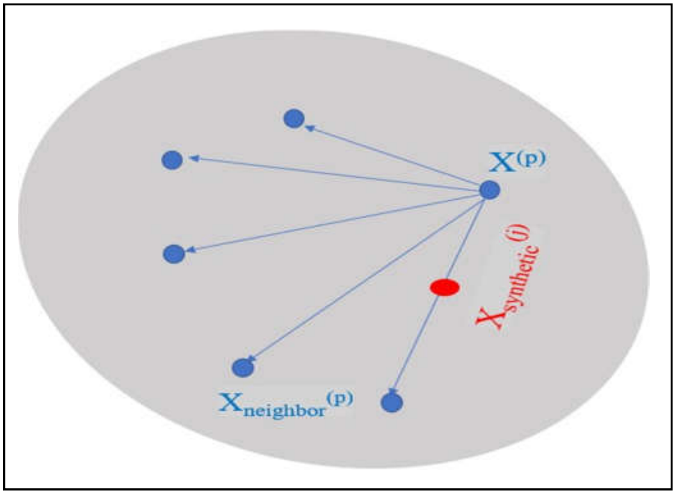

Figure 3. Generating synthetic data with SMOTE

Machine Learning Classifiers

For heart failure prediction we have used seven supervised machine learning algorithms, Support Vector machine (SVM), Gaussian Naive Bayes (GNB), K Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Extreme Gradient Boosting (XGBoost). Random Forest select the best attributes of the dataset. Class level of selected features is evaluated. The working procedure of the classifiers is described.

Support Vector Machine (SVM)

Support Vector Machine takes the labeled heart failure dataset, train the data points and outputs the hyperplane that best separates the classes. As heart failure is two dimensional, the hyperplane is simply a line (Jaiswal, java Tpoint, 2011). This line is the decision boundary: anything that falls to one side of it SVM classifies as dead and anything that falls to the other as infected but alive. The main function of the SVM is to find the best hyperplane having alarge distance from the hyperplane to both classes so that in future if a new data point comes then it can be classified easily.

Gaussian Naive Bayes (GNB)

GNB is independent of features that indicate that covariance matrices are diagonal and the classes of target features follow the Gaussian distribution. GNB presents continuous values of attributes according to the Normal Distribution Brownlee, Machine Learning Mastery, 2019).Gaussian Naïve Bayes contains hypothesis, posterior and prior probability of the classes and evaluates maximum likelihood estimation. Gaussian distribution calculates the mean of the featured values and provides a bell-shaped symmetric curve.

K Nearest Neighbor (KNN)

The purpose of KNN is to classify new objects based on training data. Firstly data is normalized from the heart failure data points. Then data is trained. KNN finds k objects from the training data which are very close to the testing data. The optimal value of k that we used in our work is the square root of N, where N is the total number of samples in the dataset. KNN works based on the nearest distance from training data to testing data to determine the k-nearest neighbor. KNN uses Euclidean distance for similarity measures of given training to testing data.

Decision Tree (DT)

DT mainly deploys as a tree-based structure which is used for classifying different items and predictions (Jijo & Abdul azeez, 2021). The implementation of the decision tree starts with a root node and the total functions of the tree is based on the root node. Then interior nodes of the tree handle different types of attributes and represent a test. The brunch node represents the results of the test. Finally, each terminal node identifies the class level.

Random Forest (RF)

Random Forests are built from decision trees. RF creates a boots trapped data set and the size of dataset sis the same as the original and the boots trapped data set randomly selects samples from the dataset. Thus, Decision tree is trained with each of the boots trapped dataset sin dependently but the twist is that all features are not used for training and the subset of features is picked randomly (Jaiswal, java T point, 2015). Here, RF contains 64 to 128 decision trees for a proper balance of running time according to volume of datasets (Dubey, 2018). When a new test data arises, pass the data through each tree and evaluate prediction results. All prediction results are combined and according to majority voting the testing data is classified. we use two random processes: boots trapping and random feature selection. As both row sampling with replacement and feature sampling with replacement is used, high variance is getting low and improves the classification and overall accuracy.

Logistic Regression (LR)

Logistic Regression is used for binary classification and analyses independent variables as well as relations between them and predicts one or more dependent variables. LR is implemented according to the term of probability and it is a statistical algorithm (Jaiswal, javaT point, 2015). LR evaluates and estimates the approximation probability between independent and dependent variables.

Extreme Gradient Boosting (XGBoost)

Though XGBoost follows the gradient boosting principle, it is ten times faster than the gradient boosting and more generalized model used for reducing over fitting problems. But XGBoost is often sensitive to outliers and it doesn't give better performance when the data is unstructured (Brownlee, J. 2021).

Results and Discussion

This section includes Dataset, Experimental Settings and Result Analysis.

Dataset

In this section, we have used the 'Heart failure clinical records dataset' from Kaggle (online community of data scientists and machine learning practitioners). In the dataset, there are 299 patients' records. The dataset contains 13 clinical features. Among 299 records, 105 are women and 194 are men who are above 40 years old.

Table 1. Dataset Details

Features	Ranges	Description
Time	4-285 times	Follow up period
Gender	0,1	Man or Woman
Smoking	0,1	If patients Smoke
Diabetics	0,1	If diabetics have
Blood Pressure	0,1	If Blood Pressure have
Anemia	0,1	Hemoglobin
Age	40-90 years	Age of patients
Ejection Fraction	14-80 percentage	Blood leaving in the heart
Serum Sodium	114-148 mEq/L	Level of sodium in blood
Creatinine	0.5-9.4mg/dL	Creatinine in blood
Platelets	25.01-850.00kilo platelets/mL	Platelets in blood
Creatinine Phosphokinase	23-7861mcg/L	Creatinine Phosphokinase in blood

Experimental Settings

We have used RF to choose the most important features using heart failure dataset. Then SMOTE is used and for implementing 7 supervised machine learning algorithms: SVM, GNB, KNN, DT, RF, LR and XGBoost. Performance of the algorithms are evaluated using four performance metrics. All experiments are implemented in python platform using various libraries on 4 GB Asus Vivo BOOK T30 graphical processing by 2x Intel processor with Xeon CORE i5 operating on

2.2 GHz CPU. Here four different metrics will be calculated namely Accuracy, Precision, Recall and F- Score for examining the best machine learning algorithm. The metrics will be calculated by a confusion matrix that contains four elements: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Evaluation of metrics is represented below: to resample the dataset. We have deployed python code for feature selection

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

Recall = $\frac{TP}{TP+FN}$ (5)

F- Score = 2 * $\frac{precision*Recall}{precision+Recall}$ (6)

In this segment, results of all experiments are presented and described as follows:

- First, all sets of features are used for analysis classification performance of all machine learning models. Here the heart failure dataset contains 13 features. The Death Event feature is used as testing because it is the target class. Dataset Details are presented in table 1. Machine learning models have been trained with 13 features and we examine the model performance Table 2.
- Then, we have used SMOTE to balance the heart failure dataset. Balanced dataset is used for training the machine learning models using all sets of features. The results are described in Table 3 and visualization of resampling data with SMOTE is shown in figure 5.
- Finally, Experimental results are evaluated based on important features. Here 7 important features time, serum creatinine, creatinine phosphokinase, platelets, ejection fraction, age and serum sodium which are selected by Random Forest. Then, Machine learning models are trained with 7 features and SMOTE created balanced dataset is used to fit the models. The overall results are compared and the best fitted model has been selected for heart failure prediction. The results are described in Table 4.

Classification Results Using Full Set of Features

In our heart failure prediction dataset, the dataset specifies two class problems. The instance of the target class (Death Event) is Death and Alive. The target class specifies two class problems, majority class and minority class where 0 is for alive and 1 is for dead. Alive (0) indicates majority class and dead (1) indicates minority class. The number of dead people is very few compared to the number of people alive. Out of 100% records, almost 68% data belong to the majority class and 32% data identify minority class. Figure 4 represents the target class. So, it is clear that target class specifies two class problems, majority class and minority class where 0 is for alive and 1 is for dead. Alive (0) indicates majority class and dead (1) indicates minority class. Figure 4 shows the actual heart failure dataset highlighting target features.

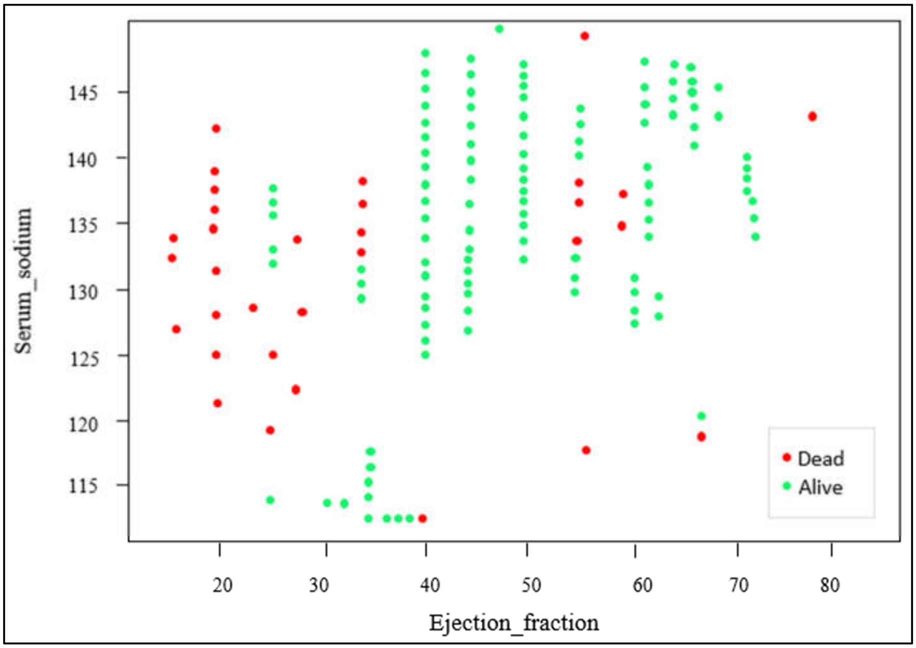

Figure 4. Target class visualization of Heart Failure Data

Table 2 highlights the performance evaluation metrics on the heart failure dataset for seven machine learning supervised models. Here, we have applied statical based, regression, tree and ensemble based classifiers. SVM and GNB are statical based models. DT, RF and XGBoost models are tree based ensemble models. LR is a regression based classifier.

Table 2. Classification results with all set of features and without SMOTE

Model	Class	Overall Accuracy	Precision	Recall	F- Score
SVM	0 (Majority)	0.78	0.71	0.88	0.79
	1 (Minority)		0.48	0.27	0.43
GNB	0	0.73	0.80	0.65	0.72
	1		0.40	0.59	0.50
KNN	0	0.61	0.67	0.57	0.62
	1		0.30	0.21	0.42
DT	0	0.78	0.77	0.93	0.85
	1		0.66	0.35	0.46
RF	0	0.85	0.86	0.84	0.86
	1		0.69	0.82	0.77
LR	0	0.82	0.84	0.72	0.77
	1		0.48	0.65	0.60
XGBoost	0	0.79	0.85	0.61	0.71
	1		0.62	0.71	0.66

Table 2 highlights the classification performance. Here, some models perform good and some models perform poor. We have used all features used for analysing the results and the class imbalance problem is not solved. From the table we can see that, for minority class (1), the result of precision, recall and f-score is very low compared to majority class (0). RF outperforms seven base classifiers and obtains good results with 0.85 accuracy. RF provides

0.86 recall for majority class and 0.69 recall for minority class. Though LR is in second best classifier's position with 0.82 accuracy, it achieves only 0.48 precision, 0.65 recall and 0.60 f-score. The worst model for heart failure prediction is KNN and it also provides 0.62 f-score for the majority class and 0.42 f-score for minority class. It is clear that the accuracy of all models is based on the majority class. Majority class Alive, is dominating in performance metrics. The machine learning models machine learning models fit the dominating majority class during dataset training. As a result, in spite of being the low results in the minority class, the overall accuracy of each model is comparatively high.

Classification results with all set of features using SMOTE

In the heart failure dataset among 299 records 203 records are for alive and 96 records are for dead. We used 70% data for training the model and 30% data is for testing. So, when the dataset is not balanced then among 299 instances 209 data is trained and 90 instances are used for testing the performance of models. The machine learning models train randomly and the number of alive records is greater than dead records. So, the number of alive records trained by Machine learning model is always very much higher than dead records. So, there is a possibility that among 209 training records 204 possible alive records can be trained and machine learning models achieve good results because of the majority class. Here, no dead records are trained. So, complex disease like heart failure prediction results will be wrong and completely biased and the situation will be harmful for our patients. So, SMOTE is used for solving this worst situation. SMOTE balances the dataset. SMOTE resample the dataset by synthetically creating instances. Figure 5 shows the visualization of resampling data by SMOTE.

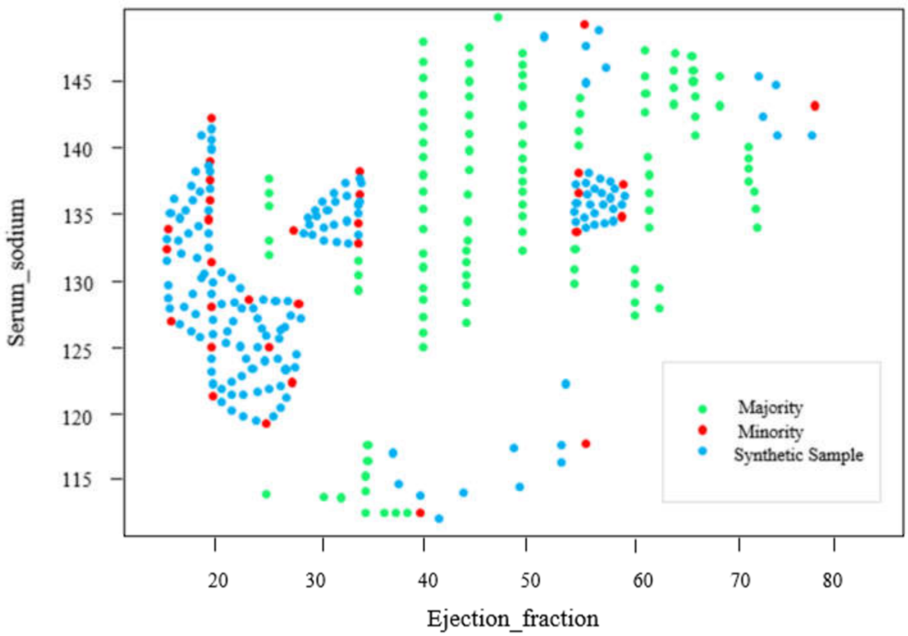

Figure 5. Data resampling with SMOTE

When data points are resampled with SMOTE, then the same number of dead and alive records is trained. So, the machine learning models perform better as they are fitted with balanced records alive and dead.The result is represented in Table 3. We clearly see that, Classification result using SMOTE is better than those of without SMOTE. Tree based ensemble classifiers improve the results with SMOTE. Statistical based models SVM and GNB increase the performance with SMOTE. DT, RF and XGBoost also improve the results. Only regression based model Logistic Regression decreases the results when data is resampled with SMOTE. SVM increases 2% accuracy with SMOTE. RF produces the highest results among all classifiers with 90% accuracy. In RF the minority class precision has been improved from 0.69 to 0.83 and f-score improved from 0.77 to 0.83. We also see that all model performance has been increased for minority class and overall accuracy has also been increased except LR. In LR 10% accuracy is decreased with SMOTE. So, it is also proved from the experiments that SMOTE performs well on classification models but badly performs on regression based classifiers. Moreover, the class imbalance problem is also reduced with SMOTE.

Table 3. Classification result using all features with SMOTE

Model with SMOTE	Class	Overall Accuracy	Precision	Recall	F- Score
SVM	0 (Majority)	0.80	0.81	0.80	0.80
	1 (Minority)		0.74	0.77	0.75
GNB	0	0.76	0.80	0.75	0.77
	1		0.60	0.71	0.63
KNN	0	0.71	0.67	0.77	0.72
	1		0.52	0.65	0.58
DT	0	0.82	0.77	0.93	0.85
	1		0.73	0.65	0.69
RF	0	0.90	0.92	0.89	0.90
	1		0.83	0.82	0.83
LR	0	0.72	0.76	0.68	0.72
	1		0.68	0.75	0.70
XGBoos	0	0.84	0.85	0.71	0.75
	1		0.72	0.77	0.71

Important Feature selected by RF

Figure 6 highlight that RF clearly selects seven important features time, serum creatinine, creatinine phosphokinase, platelets, ejection fraction, age and serum sodium as the most important features.

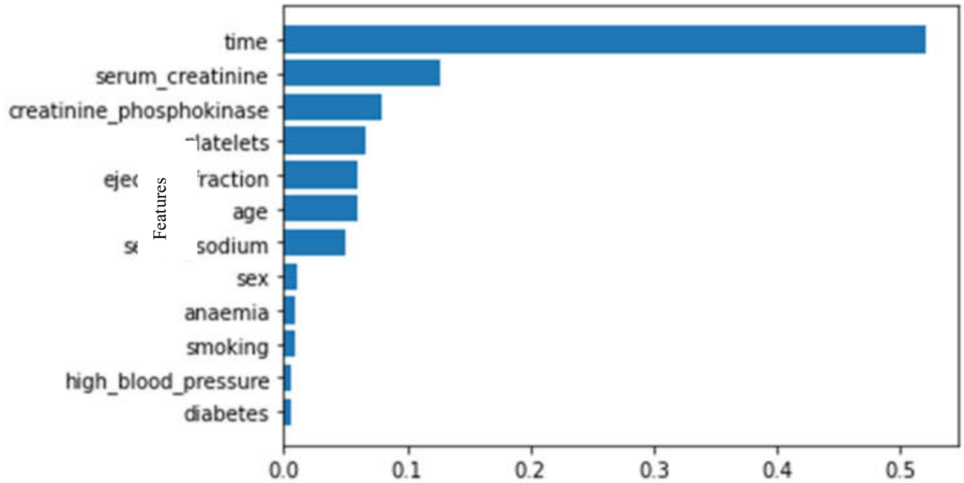

Figure 6. Important features selected by RF

Classification Accuracywith Important Features and SMOTE

RF performs well with seven important features with 0.92 accuracy. In other models, accuracy of all models is also improved using seven best features which is highlighted in Table 4. RF performs well with seven important features with 0.92 accuracy. In other models, accuracy is also improved using seven best features.

Table 4. Accuracy of all models using 7 important features and with SMOTE

Models	Accuracy	Precision	Recall	F-Score
SVM	0.82	0.80	0.81	0.80
GNB	0.75	0.76	0.74	0.75
KNN	0.72	0.72	0.71	0.72
DT	0.84	0.86	0.82	0.84
RF	0.92	0.92	0.92	0.92
LR	0.85	0.85	0.86	0.85
XGBoost	0.88	0.88	0.88	0.88

The comparison of model performance between SMOTE and without SMOTE is shown in Figure 7. When SMOTE technique is applied, the result of SVM is increased with 2% accuracy (from 0.78 to 0.80), GNB is increased with 3% accuracy (from 0.73 to 0.76), KNN is increased with 10% accuracy (from 0.61 to 0.71), DT is increased with 4% accuracy (from 0.78 to 0.82) and RF is increased with 5% accuracy (0.85 to 0.90) and XGBoost is also increased with 5% accuracy (from 0.79 to 0.84. LR decreases with 10% accuracy (from 0.82 to 0.72) after applying SMOTE. So, it is also proved from the experiments that SMOTE performs well on classification models but badly performs on regression based classifiers. Moreover, the class imbalance problem is also reduced with SMOTE.

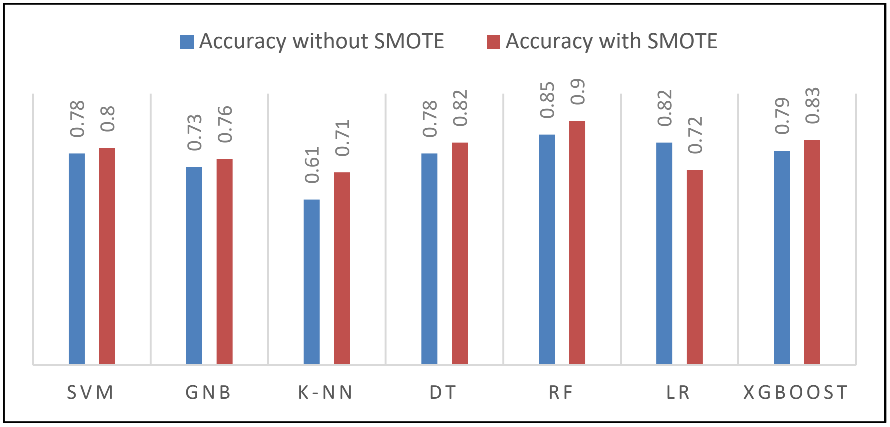

Figure 7. Comparison of accuracy with SMOTE and without SMOTE

Comparison of accuracy with SMOTE between all features and seven features is represented in Figure 8.By using seven important features the classification results are improved. When we have used all features for metrics performance evaluation then accuracy of SVM was 0.8. After applying seven features SVM achieved 0.82 accuracy. KNN has improved 1% accuracy (from 0.71 to 0.72). DT are increased with 2% accuracy (from 0.82 to 0.84) and XGBoost are increased with 4% accuracy (from 0.84 to 0.88) and RF is increased with 2% (from 0.90 to 0.92) accuracywhen SMOTE is applied with best seven features RF and DT improved with 2% accuracy.

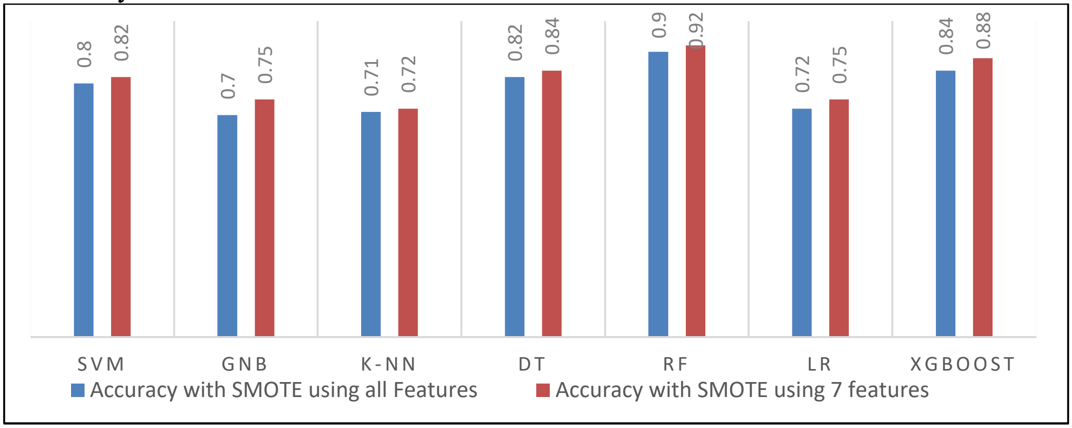

Figure 8. Comparison of accuracy with SMOTE between all features and 7 features.

It is clear from the overall discussion part that after using SMOTE with feature selection technique the prediction performance is increased. RF performs well with seven important features with 0.92 accuracy. For predicting heart failure researcher S. Mohan et al. in 2020 used hybrid machine learning technique but the prediction accuracy is 0.87 that is not good enough. Waqar, Muhammad, et al. in 2021 usedan efficient SMOTE-based deep learning model for heart attack prediction and in their work, they obtained only 0.85 accuracy using Random Forest with SMOTE. Compared with them our work provides best accuracy with random Forest. So, after analysing all the experimental results, it is clear to us that Random Forest provides the best accuracy among all machine learning classifiers. So, heart failure can be predicted using seven significant features and with Random Forest classifiers.

Conclusion

In this research, we have investigated the prediction of heart failure using significant features. RF is used for selecting the most significant features. SMOTE technique is used for solving the class imbalance problem. Balanced datasets created by SMOTE are fitted with different supervised machine learning algorithms. BY investigating the overall experiment, we see that SMOTE enhances the classification performance when SMOTE is fitted with the RF model. RF with SMOTE using 7 significant features provide highest results in all evaluation metrics and gain 0.92 accuracy. So, it is clear that by using SMOTE based Random Forest classifiers, prediction of heart failure will help medical practitioners and they will be able to focus on major heart failure risk factors. The future work for this research can be developed better feature selection techniques to improve the performance of the models.

Funding: This research was funded by ICT Division research Innovation fund, Bangladesh for the year 2022-23.

Acknowledgements: This research work is financially supported by ICT Division Bangladesh so we would like to thank ICT Division Bangladesh.

Conflicts of Interest: The authors declare no conflict of interest.

References

Alber Fernandez, Salvador Garcia, Francisco Herrera, and Nitesh V. Chawla (2018). *SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary,* Journal of Artificial Intelligence Research, vol. 61, pp. 863–905.

Ashikur Rahaman, Md.Milon Islam, Md. Rashedul Islam, Muhammad Sheikh Sadi, and Sheikh Nooruddin (2019). *Developing IoT based smart health monitoring systems: a review,* in International Information and Engineering Technology Association, vol. 33, no. 6, pp. 435–440, DOI: https://doi.org/10.18280/ria.330605

Asim Kumar Pathak, J. Arul Valan (2020). *A predictive model for heart disease diagnosis using fuzzy logic and decision tree,* in Smart Computing Paradigms: New Progresses and Challenges. Advances in Intelligent Systems and Computing, vol 767. pp. 131–140 Springer, Singapore. https://doi.org/10.1007/978-981-13-9680-9_10

Asma Baccouche, Begonya Garcia-Zapirain, Cristian Castillo Olea and Adel Elmaghraby, (2020). *Ensemble deep learning models for heart disease classi_cation: A case study from Mexico*, in Information, vol. 11, no. 4, p. 207, DOI: https://doi.org/10.3390/info11040207

Dubey, A. (2018, December 15). *Towards Data Science*. Retrieved from Towards Data Science in https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f

Durgadevi Velusamy, Karthikeyan Ramasamy (2021). *Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset*, Computer Methods and Programs in Biomedicine, Volume 198, Article 105770.

Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). *SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary*. Journal of Artificial Intelligence Research, 61, 863-905.

F.J Shaikh, D.S Rao (2022). *Predication of cancer disease using machine learning approach,* in ELSEVIER, Science Direct. Volume 50, Part 1, Pages 40-47, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.03.625

Imomoiye Domor Mienye, Yanxia Sun, Zenghui Wang (2020). *An improved ensemble learning approach for the prediction of heart disease risk*, in Informatics in Medicine Unlocked, Elsevier, Science Direct, Volume 20,100402, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2020.100402. (https://www.sciencedirect.com/science/article/pii/S2352914820304184)

Jaiswal, S. (2015). *javaTpoint*. Retrieved from javaTpoint WEb site: https://www.javatpoint.com/machine-learning-random-forest-algorithm

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends, 2*(1), 20-28.

Ju Li, Muhammad Nazir Jan, and Mohammad Faisal (2020). *Big data, scientific programming, and its role in internet of industrial things: a decision support system*, Scientific Programming, vol. 2020, Article ID 8850096.

L. Ali *et al. (2019)*, *An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure*, in *IEEE Access*, vol. 7, pp. 54007-54014, doi: 10.1109/ACCESS.2019.2909969.

Lu, Men, Ilk, Noyan, Tang, Xinlin, Liu, Yuan (2021). *Multi-Disease Prediction Using LSTM Recurrent Neural Networks* Expert Systems with Applications 177. 114905. 10.1016/j.eswa. 2021.114905, 2021.

Mohammad Shafenoor Amin, Yin Kia Chiam, and Kasturi Dewi Varathan (2019). *Identification of significant features and data mining techniques in predicting heart disease,* Telematics and Informatics, vol. 36. Pages 82-93, ISSN 0736-5853, https://doi.org/10.1016/j.tele.2018.11.007. (https://www.sciencedirect.com/science/article/pii/S0736585318308876)

S Mohan, C. Thirumalai, G. Srivastava (2019). *Effective Heart Disease Prediction Using Hybrid Machine Learning Technique*s, in *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

Sravani Nalluri, R. Vijaya Saraswathi, Somula. Ramasubbareddy, K. Govinda, and E. Swetha (2019). *Chronic heart disease prediction using data mining techniques*, in Data Engineering and Communication Technology, vol. 2020, pp. 903–912, Springer, Singapore, Asia.

S. Shah, N. Habib, M. Nadeem, A. Alshdadi, M. Alqarni, and W. Aziz (2020). *Extraction of Dynamical Information and Classification of Heart Rate Variability Signals using Scale Based Permutation Entropy Measures,* in International Information and Engineering Technology Association, vol. 37, no. 3, pp. 355–365, DOI**:** https://doi.org/10.18280/ts.370302

Sravani Nalluri, R. Vijaya Saraswathi, Somula. Ramasubbareddy, K. Govinda, and E. Swetha (2019). *Chronic heart disease prediction using data mining techniques*, in Data Engineering and Communication Technology, vol. 2020, pp. 903–912, Springer, Singapore, Asia.

Waqar, Muhammad, et al. "An efficient SMOTE-based deep learning model for heart attack prediction." *Scientific Programming* 2021 (2021).

Yunxing Jiang, Xianghui Zhang, Rulin Ma, Xinping Wang, Jiaming Liu, Mulatibieke Keerman, Yizhong Yan, Jiaolong Ma, Yanpeng Song, Shuxia Guo, Heng Guo (2021*). Cardiovascular Disease Prediction by Machine Learning Algorithms Based on Cytokines in Kazakhs of China*, Clin Epidemiol CLEP.S313343. eCollection