PERFORMANCE EVALUATION OF HEALTH CARE SYSTEM USING MACHINE LEARNING ENSEMBLE MODEL

Vikas Kulshreshtha[1], N.K.Garg[2], Jai Kumar Maherchandani[3], Sachinpal Singh Yadav[4], Amritpal Singh Yadav[5], Gaurav Gupta[6] and Nitin Kumar Suyan[7]

[1]Department of Information Technology, Engineering College Jhalawar, Rajasthan, India- 326001

[2]Department of Electrical Engineering, Engineering College Jhalawar, Rajasthan, India- 326001

[3]Department of Electrical Engineering, College of Technology and Engineering, Udaipur, India- 313001

[4]Department of Computer Science & Engineering, Engineering College Jhalawar, Rajasthan India

[5,6]Department of Computer Engineering, Women Engineering College Ajmer, Rajasthan, India

[7]Department of Electronics & Communication Engineering, Engineering College Jhalawar, Rajasthan, India

ABSTRACT

Today healthcare system has a significant role to give accurate predications in less time. Heart failure is common disease and medical field has advanced technology to diagnose the issues regarding heart. Researchers are continuously working to improve the system. This paper proposes machine learning based a new model taking the symptoms of heart failure. This paper proposes the various ensemble learning methods which helps to convert weak learner into strong learners. It also depicts the correlation between the predictors and the estimator. This paper compares and evaluates performance of the ensemble learning methods which further determine the estimated mortality rate (EMR) , false mortality rate (FMR) and an accuracy of the model. This paper will be useful for further research in medical science.

Index Terms: Ensemble Methods, Machine Learning, Boosting Algorithms, Bagging Algorithms, Imbalanced data

DOI Number: 10.14704/nq.2022.20.9.NQ44172 **Neuro Quantology 2022; 20(9):1491-1498**

INTRODUCTION

A CCORDING to the report of world health organization (WHO), cardiovascular diseases (CVDs) are the dominant cause responsible to surge the mortality rate globally in 2019. Approximately 17.9 million people died due to CVDs in 2019, out of which the cause of 85 percent death is due to heart failure. This characterize 32 percent of all global death [1].The main reason for heart failure are diabetes , high blood pressure, smoking and serum sodium [2]. This paper applied the various ensemble methods used in machine learning model in order to increase the performance and recommends the XGBoost ensemble method which is an effective tool for predicting the persistence of patient which depicts the heart failure symptoms [3][4] and to discover the clinical feature (predictors) which may cause the heart failure [5][6]. Machine learning blended with ensemble methods are prove like a boon for clinical estimation [7] [8] as well as for feature selection [9].The Table I below shows the description of predictor variables used in the population. Author has used the historical data [10] which is the clinical data of the heart failure. There are 11 predictors which are processed in order to determine the estimator variable. This paper considered death event as the estimator variable.

Table I: Feature description used in the data set

Predictor variable	Description	Nature (Boolean/continues)
p1	Anemia	Boolean
p 2	Creatininephosphokinase	Continues
p 3	Diabetes	Boolean
p 4	Ejection fraction	Continues
p 5	High blood pressure	Boolean
p 6	Platelets	Continues
p 7	Serum creatinine	Continues
p 8	Sex	Boolean
p 9	Smoking	Boolean
p 10	Time	continues
p11	Serum sodium	continues

Fig.1 explains the proper flow of the information which is used in this paper for the performance analysis of the ensemble methods. The study done by Bella G.A. et. al shows the cardiac motion analysis for human survival using deep learning [11]. When the survival outcomes are coupled with additional targets like hospitalization model shows improvements [12]. Even if scientists have spotted a wide set of predictors and indicators there are no shared harmony on their relative impact on survival prediction [13]. As indicated by Sakamoto and colleague circumstances are due to shortfall of reproducibility, which intercept drawing conclusive about the importance of detected factors. This result strongly affects the model performance [14]. Ahmad and colleagues [15] employed traditional biostatistics time-dependent models (such as Cox regression [16] and Kaplan–Meier survival plots [17]) to predict mortality and identify the key features of 299 Pakistan patients having heart failure, from their medical records. Previously various classifiers from machine learning areas were used for prediction of the mortality rate. The classifier includes linear regression [18], random forest [19], one rule [20], decision tree [21], artificial neural network [22], and support vector machine [23].

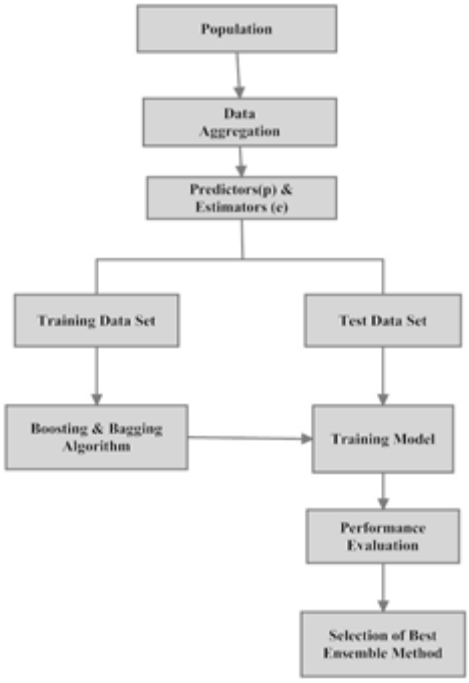

Fig. 1. Data flow diagram.

PROPOSED MODEL

The Machine Learning (ML) is widely used in various real time applications like ecommerce, healthcare, weather forecasting, finance, online fraud detection, self-driving cars, speech recognition, image recognition, product recommendation, malware filtering and many more. The algorithms used for creating the machine learning model are highly acclaimed by the industries. It plays critical role in healthcare domain. Their most propitious approaches contributing in healthcare specially in estimating the mortality rate due to heart failure. This is the encouraging factor for writing the research paper. Typically machine learning models act as a 'black boxes', which shows the poor behavior of the model. In real world, especially in the healthcare sector imbalanced data is commonly used.

Correlation Matrix of Predictor Variables

Correlation is the critical mechanism used for feature engineering. It helps in building up the machine learning model. In simple terms correlation shows how strongly the variables are related to each other. Uncorrelated predictors with the estimator results as the best candidate to trim from the model.While in the case of highly correlated predictors then we can select only one predictor. Pearson product-moment coefficient is widely used on feature engineering. According to it given collection of paired values the coefficient produces value between -1 to 1 which signifies the strength of dependencies between the predictor and estimator. A value of 1 shows positive slope and -1 shows the negative slope, Pearson coefficient value is 0 then it shows that there is no relation between the variables. The Fig.2 shows that the features are very less related to each other (less than0.5). Hence author has to consider all the columns for estimating the mortality rate. In addition, all the input variables are important as it directly or indirect effect the estimator variable. Equation 1 shown below depicts the correlation equation which is fraction of covariance of predictor and estimator data points present in training dataset and the standard deviation of the predictor and estimator which shows how much the data points are close to mean. Correlation equation plays a very important role in feature selection.

$$|f| = \frac{\sum\left(p_i - \bar{p}\right)\left(e_i - \bar{e}\right)}{\sqrt{\sum\left(p_i - \bar{p}\right)^2 \sum\left(e_i - \bar{e}\right)^2}} \quad (1)$$

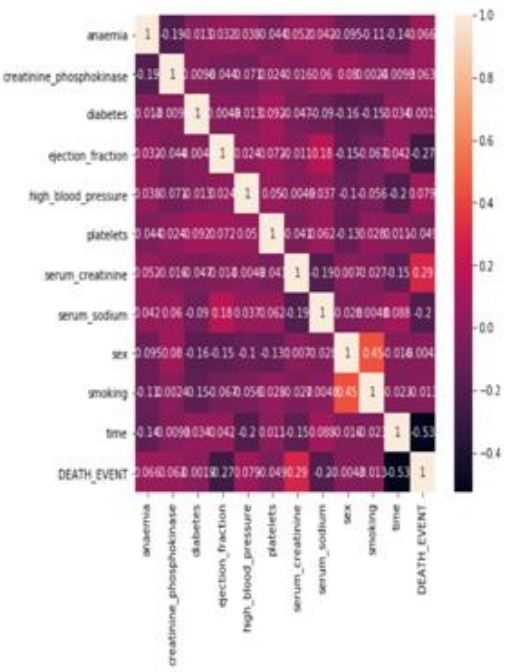

Fig. 2. Correlation matrix of predictors and estimator

Relationship between the Predictor Variables and Estimator

The Fig.3 explains the distribution between the age (predictor) and death event (estimator) which states that peak density (.03) of the population survival varies between 60 to 70 years. Survival of the patient increases from 45 to 75, the density for the maximum survived patient ranges from .008 to 0.03. Patient of 80 years and above has the slimmest density of survival which is below .005 level. Fig.4 reveals the allocation between the ejection fraction (predictor) and death event (estimator) that survival density notably drops when the ejection fraction is below 20 percent and above 70 percent. At around 40 percent of ejection fraction shows the highest density of survival patient. Distribution between the creatinine phosphokinase (predictor) and death event (estimator) is shown in the above Fig.5 which in details tells that the level of the predictor is ranges from 0 mcg/L to 1000mcg/Lencounters rise in the survival density and decreases when the predictor level is higher than 2000 mcg/L. Fig. 6 shows the distribution between the platelets which act as the predictor and the death event which act as the estimator variable which shows that platelets are not much effecting the death rate in the context of the heart failure.

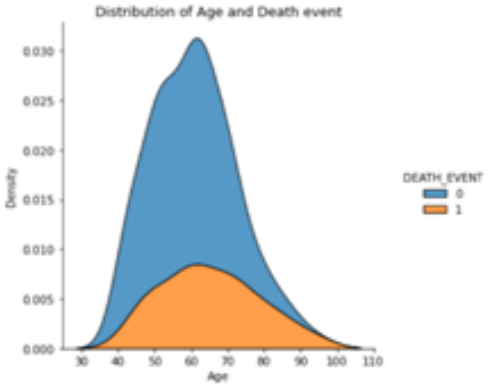

Fig. 3. Distribution of age and death event

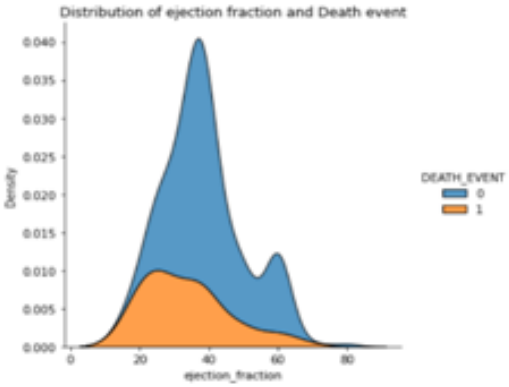

Fig. 4. Distribution of p4 and death event

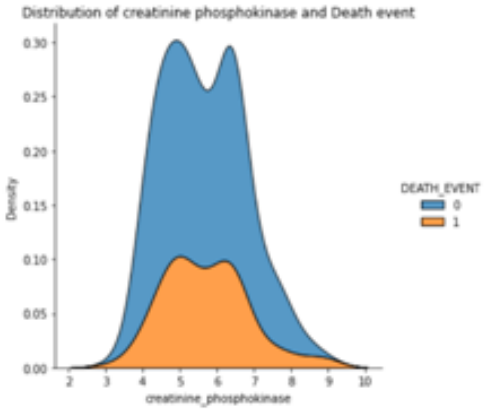

Fig. 5. Distribution of p2 and death event

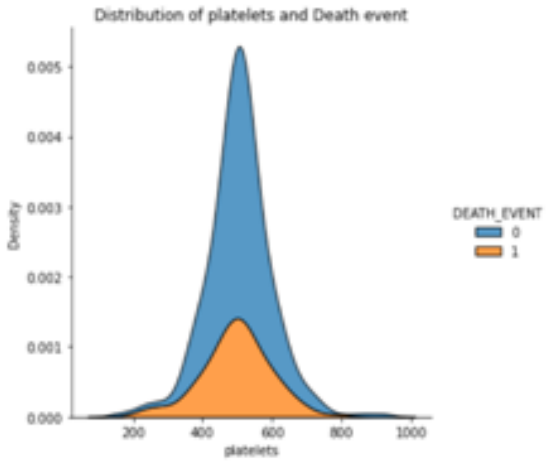

Fig. 6. Distribution of p6 and death event

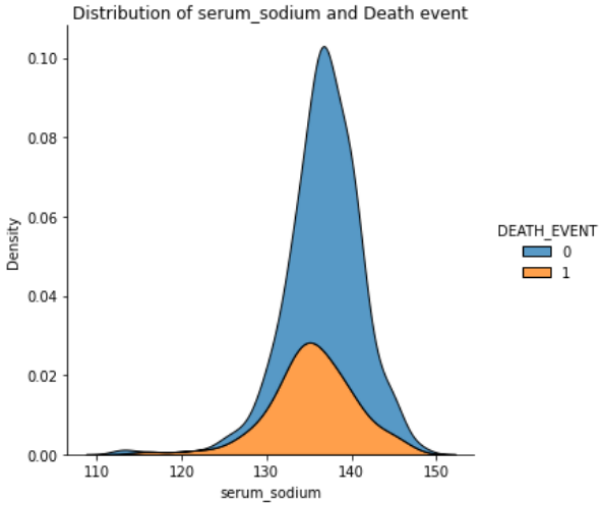

Fig. 7. Distribution of p11and death event

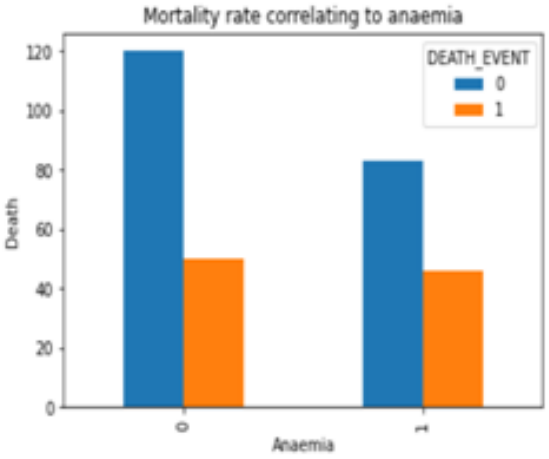

Fig. 8. Mortality rate correlating to anemia

It also shows that the mean value of platelets in the blood of human body is 260 kilo platelets/mL. The Fig.7 shows the significant role of the predictor (serum sodium) in the outcome (death event). The peak survival density is in between 135 to 140 mEq/L. The density of the death event is rising at around 130 mEq/L and declining at around 148 mEq/L.The bar chat (Fig.8) shows the correlation between the anemia (predictor) and the mortality rate. The reveals that patient who were not suffered from anemia has higher mortality rate as compared to that patient who had anemia which was 50 and 40 respectively. The relationship between mortality rate and diabetes is depict in the above Fig. 9 which says that patient who were diabetic has more prone to death as compared to the person who doesn't have diabetes.

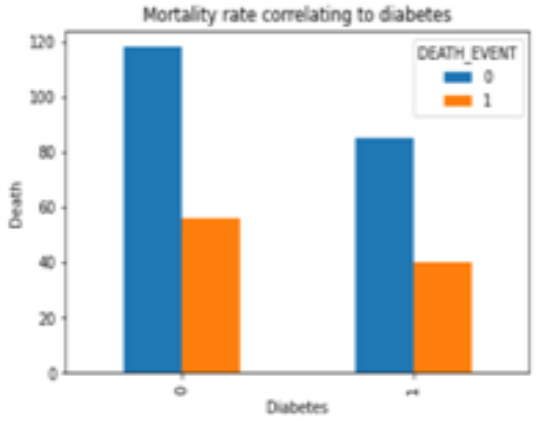

Fig. 9. Mortality rate correlating to diabetes

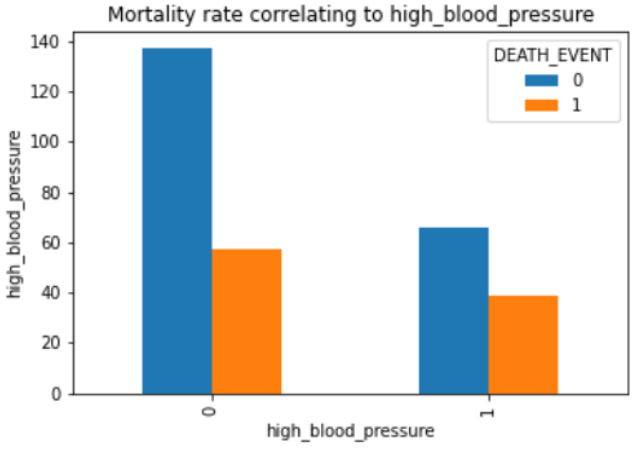

Fig. 10. Mortality rate correlating to blood pressure

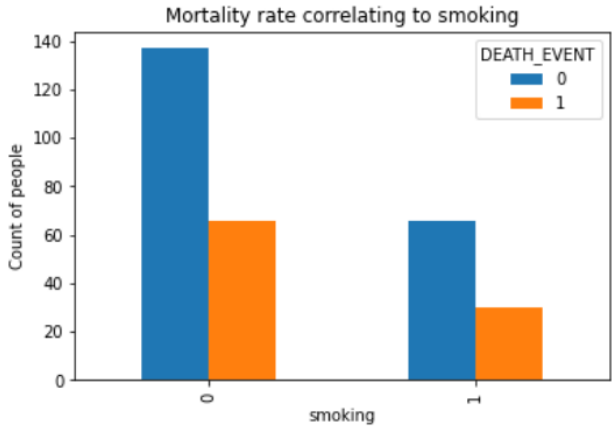

Fig. 11. Mortality rate correlating to smoking

The above bar chart (Fig.10) exhibits the correlation of mortality rate with high blood pressure, it states that mortality rate of non-hypertension patient is high as compared to hypertension patient. In the case of mortality rate correlation with smoking patient (Fig.11) reveals that survival as well as mortality rate is higher in the case of non-smoking patients. Mortality rate correlation with the gender shown in the Fig.12 below exhibit that mortality as well as survival rate is higher in male as compared to females' patients. The Fig.13above shows that count of population with respect to death event (estimator variable), it reveals that imbalance between the dead patients and the survived patients which is nearly 200 and 100 respectively. It signifies that there is a presence of imbalance among the two value present in the estimator variable of the given population. This paper explains the performance of different kind of ensemble techniques viz. bagging and boosting algorithm. Finally the proposed algorithm is best it is capable for handling the imbalance data with the high accuracy.

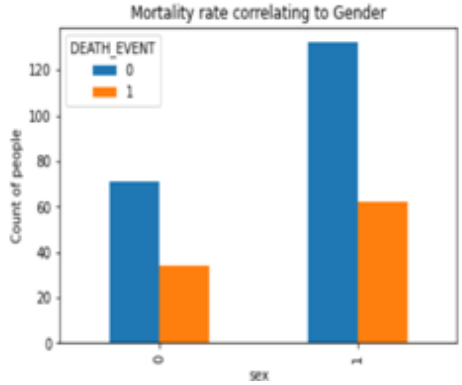

Fig. 12. Mortality rate correlating to anemia

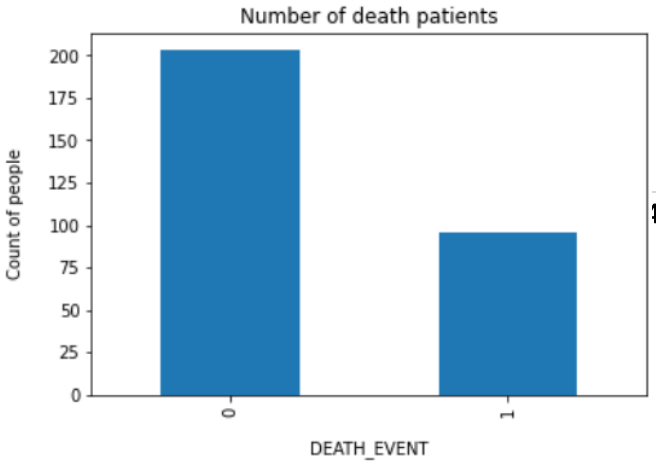

Fig. 13. Count plot of the estimator variable

Principles of Ensemble Methods

Performance is the critical aspect for the machine learning model. It mainly depends upon the accuracy. In data science sample analysis is a pivotal task, it is used as a main building block for creating the model. But, in real world machine learning model faces overfitting problem due to which the model gives the inconsistent outcome even though the accuracy of the model is magnificent. It is caused due to imbalanced data present in the population. Therefore, while dealing with imbalanced data there are other aspects besides accuracy which helps in evaluating the performance of the model. The model should not be suffering from overfitting (high variance low bias) or under fitting (low variance and high bias). Further, the estimator variable of the model is highly depending upon the correlation between the predictor variable. Hence there should be no correlation among the predictor variables. This paper explains the various ensemble methods for estimating the mortality rate of the patient due to the heart failure. It shows the equation for the error in ensemble method. Ensemble methods in data science used collection of learning algorithm instead of single algorithm which enhances performance of the model. Basically ensemble learning depends upon the bias variance trade off. When the average of the estimated values is different from the observed values, it shows the bias error, this results in the under performance of the model. When estimation on the training data is high and shows estimation on the test data badly, this problem is overfitting. This paper evaluates the performance based on bagging algorithm

(random tree forest classifier) and boosting algorithm (Ada boost classifier, gradient boost classifier and XGboost classifier).

Table II shows the general confusion matrix in which death event has considered as the estimator which captures the estimated and the actual values and plug in the values in their respective blocks.

$$EMR = \frac{TP}{TP + FP} \quad (2)$$

$$FMR = \frac{TP}{TP + FN} \quad (3)$$

The above two equation 2 and 3 are used to compute the estimated mortality rate (EMR) and false mortality rate (FMR) used in various ensemble methods. In EMR false positive value are very critical. In this paper imbalanced data medical specific use case is considered, hence in order to make the better performance of the model FP value is supposed to be cost effective and FN value should be minimal.

Table II: GENERAL CONFUSION MATRIX

		Predicted	
		Death Event =Yes	Death Event =Yes
Actual	Death Event =Yes	TP	FN
	Death Event =No	FP	TN

Equation for error in ensemble learning:

$$e(p) = \left(E\left[\hat{f}(p) - f(p) \right] \right)^2 + E\left[\hat{f}(p) - \left[f(p) \right] \right]^2 + r^2 \quad (4)$$

The above equation 4 shows the ensemble equation which is used to calculate the error by taking the summation of expected square of bias, variance and residual error.

In random tree forest multiple decision tress are built. Random tree forest works on the bagging or bootstrap ensemble method which is based in parallel learning process. The dataset consists 299 rows and 13 columns out of which p3, p5 and p9 are the dominating feature which suppresses the remaining predictors. The precision value calculated is 0.79 which is very low. The recall value determined by the random tree forest is .55 which is quiet higher. The F1 measure is .65 and the overall accuracy of the model is 75 percent.

$$\begin{bmatrix} 51 & 6 \\ 19 & 23 \end{bmatrix}$$

The confusion matrix is generated random tree forest ensemble method as shown above.

Ada boost is the part of boosting ensemble method which overcome the shortcomings of random tree forest. It is based on sequential learning process, which means that outcome of one tree is dependent on the input of other tree. The precision value compute by the model of 299 observations is .84 which higher as compared to random tree forest. The model has calculated the recall value is .50 which is lesser than former model. Like the former model the accuracy of the model is same as 75 percent. The F1 measure is calculated as .63.

$$f(p) = h_0(p) + \beta_1 h_1(p) + \ldots\ldots\ldots\ldots + \beta_n h_n(p) \quad (5)$$

$$f(p) = \text{Limit}_{h_0(p) \to 0} \sum_{i-1}^{n} \beta_i h_i(p) \quad (6)$$

The above equation 5 and 6 are used to calculate the estimator value of ada boost ensemble method. The confusion matrix generated through ada boost classifier is shown below

$$\begin{bmatrix} 53 & 4 \\ 21 & 21 \end{bmatrix}$$

In gradient boost learning process is done by optimizing the loss function. In this type of boosting method residual error plays a very important role. Here we are trying to make the residual error tends to zero. By doing this author has optimizing the loss function. The model has determined the precision value as .85 which is much better, but it needs more improvement. The recall value is calculated as .52 which is higher than the former model but lesser than bagging algorithm. This model shows the high accuracy as 76 percent which is higher than both of the models. The model compute F1 measure as .65 which is higher than the former boosting algorithm. The equations 7 and 8 are shown below used to compute the mean square error (MSE) in order to optimize the cost function.

$$e = ap + b + R \quad (7)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i - e)^2 \quad (8)$$

The confusion matrix is calculated of gradient boost ensemble technique and shown below.

$$\begin{bmatrix} 53 & 4 \\ 20 & 22 \end{bmatrix}$$

As the name suggests it is an extension of the gradient boosting algorithm. The best part of this boosting algorithm is its speed and the accuracy, which mean the kind of outcome it gives are commendable. It is implemented on top of gradient boost. It is designed on the concept of parallelization which is responsible for enhance its speed. The mechanism which makes it different from normal gradient boost is the regularization parameter. It reduces the overfitting problem. It also uses auto pruning, which means it will not allow the tree to grow after certain level. It results in maintaining the bias and variance balance. It will increase the effectiveness of the model by making it robust. It also handles missing values by using sparse feature handling. The recall value is calculated by this model it is lowest (.48) and the prime criteria of the outcome. The XGBoost calculates the precision value as .87 which the highest among all the algorithms. The F1 measure and the accuracy of the model is .62 and 75 percent.

$$SW = \frac{\sum(R)^2}{\sum(P(1-P)+\lambda} \quad (9)$$

The equation 9 is used to calculate the similarity weight (SW) by penalizing it with the lambda value. The confusion matrix calculated by using XGBoost ensemble method is shown below.

$$\begin{bmatrix} 54 & 3 \\ 22 & 20 \end{bmatrix}$$

The Table III explains the EMR and FMR value followed with F1 measure and accuracy of the model. It clearly shows that XGBoosthas the highest EMR value and lowest FMR value which is 0.87 and 0.48 respectively. Hence this paper strongly recommends XGBoost ensemble technique used as minority data points plays a crucial role.

TableIII: PERFORMANCE EVALUATION OF THE ENSEMBLE LEARNING METHODS

	EMR	FMR	F1	Accuracy
Ada Boost	0.84	0.50	0.63	0.75
GDB	0.85	0.52	0.65	0.76
XG Boost	0.87	0.48	0.62	0.75
RTF	0.79	0.55	0.65	0.75

CONCLUSION

This paper describes on minority class of imbalanced data of healthcare system. The proposed model has good accuracy as well as performance for diagnosis of heart failure. This paper uses various ensemble methods which are consist of bagging and boosting algorithm. These methods are used to reduce the variance and helps in building the balance model. In imbalanced applications performance of the model is more important than the accuracy. The results show that EMR (precision metric) is high and FMR is found less which shows that proposed model is competitive to these type of imbalanced applications. Consequently, this paper recommends XGBoost as the best ensemble method.. Further the performance can be improved by tuning the hyper parameter of XGBoost boosting algorithm. The proposed model can be used for betterment for diagnosis of heart failure and further research

REFERENCES

[1] World Health Organization https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] The Guardian. UK heart disease fatalities on the rise for first time in 50years.https://www.theguardian.com/society/2019/may/13/heartcirculatory-disease-fatalities-on-rise-in-uk.Accessed 25 Oct 2019.

[3] S. J. Al'Aref et al., "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging," European Heart Journal. 2019, doi: 10.1093/eurheartj/ehy404.

[4] S. J. Al'Aref et al., "Determinants of In-Hospital Mortality After Percutaneous Coronary Intervention: A Machine Learning Approach," J. Am. Heart Assoc., 2019, doi: 10.1161/JAHA.118.011160.

[5] W. B. Dunn et al., "Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate," Metabolomics, 2007, doi: 10.1007/s11306-007-0063-5.

[6] J. Gallagher et al., "A systematic review of clinical prediction rules for the diagnosis of chronic heart failure," ESC Hear. Fail., 2019, doi: 10.1002/ehf2.12426.

[7] S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis: Evaluation for cardiovascular diseases," Expert Syst. Appl., 2013, doi: 10.1016/j.eswa.2013.01.032.

[8] M. Panahiazar, V. Taslimitehrani, N. Pereira, and J. Pathak, "Using EHRs and Machine Learning for Heart Failure Survival Analysis," 2015, doi: 10.3233/978-1-61499-564-7-40.

[9] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, "Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques," Computational and Structural Biotechnology Journal. 2017, doi: 10.1016/j.csbj.2016.11.001.

[10] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Med. Inform. Decis. Mak., 2020, doi: 10.1186/s12911-020-1023-5.

[11] [G. A. Bello et al., "Deep-learning cardiac motion analysis for human survival prediction," Nat. Mach. Intell., 2019, doi: 10.1038/s42256-019-0019-2.

[12] [A. A. Voors et al., "Development and validation of multivariable models to predict mortality and hospitalization in patients with heart failure," Eur. J. Heart Fail., 2017, doi: 10.1002/ejhf.785.

[13] W. C. Levy et al., "The Seattle Heart Failure Model: Prediction of survival in heart failure," Circulation, 2006,doi:10.1161/CIRCULATIONAHA.105.584102.

[14] M. Sakamoto et al., "The impact of creating mathematical formula to predict cardiovascular events in patients with heart failure," Sci. Rep., 2018, doi: 10.1038/s41598-018-22347-0.

[15] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," PLoS One, 2017, doi: 10.1371/journal.pone.0181001.

[16] A. Fitrianto and R. L. T. Jiin, "Several types of residuals in cox regression model: An empirical study," Int. J. Math. Anal., 2013, doi: 10.12988/ijma.2013.38193.

[17] [Kleinbaum DG, Klein M. Kaplan–Meier survival curves and the log-ranktest. Heidelberg, Germany: Springer; 2012, pp. 55–96

[18] Seber GA, Lee AJ. Linear Regression Analysis, Wiley Series in Probabilityand Statistics. vol. 329. Hoboken: John Wiley and Sons; 2012.

[19] Breiman L. Random forests. Mach Learn. 2001;45:5–32.

[20] R. C. Holte, "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets," Mach. Learn., 1993, doi: 10.1023/A:1022631118932.

[21] W. Y. Loh, "Classification and regression trees," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 2011, doi: 10.1002/widm.8.

[22] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436.

[23] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," Neural Networks, 1999, doi: 10.1016/S0893-6080(99)00032-5.