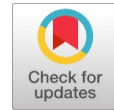


Performance of Hybrid Ensemble Classification Techniques for Prevalence of Heart Disease Prediction

Sachin Kamley



Abstract: In medical science, heart disease is being considered as fatal problem and in every seconds most of the people dies due to this problem. In heart disease, typically heart stops blood supply to other parts of the body. Hence, proper functioning of body stopped and affected. In this way, timely and accurate prediction of heart disease is an important concern in medical science domain. Diagnosing of heart patients with previous medical history is not being considered as reliable in many aspects. However, machine learning techniques have mystery to classify heart disease data efficiently and effectively and provide reliable solutions. In the past, prediction of heart disease problem various machine learning tools and techniques have been adopted. In this study, hybrid ensemble classification techniques like bagging, boosting, Random Subspace Method (RSM) and Random Under Sampling (RUS) boost are proposed and performance is compared with simple base classification techniques like decision tree, logistic regression, Naive Bays, Support Vector Machine, k-Nearest Neighbor (KNN), Bays Net (BN) and Multi Layer Perceptron (MLP). The heart disease dataset from Kaggle data source containing 305 samples and Matlab R2017a machine learning tool are considered for performance evaluation. Finally, the experimental results stated that hybrid ensemble classification methods outperforms than simple base classification methods in terms of accuracy

Index Terms: Ensemble Classification, Heart Disease, Kaggle, Machine Learning, Matlab R2017a.

I. INTRODUCTION

In these days, heart disease is being known as one of the complex and life deadest human disease in all over the world. Heart is an important body organ in human body. If this organ gets affected than it also affect the other important body parts as a result heart failure occurs. According to World Health Organization (WHO) figures, one in every four people dies of heart disease [1]. Mostly in men heart disease possibility greater than women. In US, heart failure rate is higher than other countries. Approximately, 610,000 Americans loss their lives annually due to heart problem [2]. There are various symptoms of heart diseases like feet swollen, breadth shortness, chest pain and physical body weakness etc. and factors that increases the risk of heart diseases like smoking, family history, high blood pressure, obesity, physical inactivity, poor diet and high blood cholesterol etc. However, these factors are utilized to examine the heart disease [3].

For the last couple of decades, machine learning sub field of data mining has been widely used in solving problems in medical

domains such as Cardiology, Urology, Gynecology etc. [4] [5]. The main focus of these techniques is to build an intelligent diagnosis system that can be helpful for medical stack holders in order to perform a expert diagnosis as well as accurately predicting a heart disease in effective manner i.e. particular patient has heart disease or not.

Machine learning (ML) techniques have capability to explore hidden patterns from huge databases as well as adjusting power and utilizing some disease related patterns in order to heart disease diagnosis. This knowledge can be helpful for designing and build expert system that will be helpful for health-care sectors for diagnosing and predicting heart disease for patients [6].

In order to improve the performance of individual classifier, the idea of combining classifier is introduced. Therefore, these classifiers are based on various classifier techniques and could significantly get a different rate of correctly classified instances. Thus, these classifiers are known as hybrid ensemble classifier. Moreover, these classifiers have capability to improve generalize performance by combining several base or weak classifiers by providing training data samples [7]. For the last couple of years, hybrid ensemble classification techniques have been extensively used for heart disease prediction and achieved better performance as compared to base classification techniques [8] [9]. In the past, very few authors have investigated the application of hybrid ensemble classification techniques for heart disease prediction. This study presents a five most important base classification techniques like Support Vector Machine (SVM), Logistic Regression (LR), Naive Bays (NB), K-Nearest Neighbors (KNN), Decision Tree (DT), Bayes Net, Multi layer Perceptron (MLP) and bagging, boosting, Random Subspace Method (RSM) and Random Under Sampling (RUS) boost as hybrid ensemble classification techniques. The Kaggle data source is used for performance evaluation of proposed models. The remaining sections are described as follows. Section 2 describes the brief literature review of significant researchers. Section 3 describes about data pre-processing process of data. Discussion of proposed methodologies discussed by Section 4. Section 5 and Section 6 discuss about various experimental results generated and briefly discussion of experimental results respectively. At last conclusion and future scopes of the study described by Section 7.

II. LITERATURE REVIEW

Since many decades, classification algorithms have extensively used for heart disease prediction and detection purpose. However, performance comparison of different machine learning classification algorithms has been evaluated in several research studies. In this section, brief literature review of some significant researchers is discussed here.

Manuscript published on 30 August 2019.

*Correspondence Author(s)

Sachin Kamley, Department. of Computer Applications, S.A.T.I., Vidisha (M.P.), India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Acharya et al. (2015) [10] have investigated different machine learning classification methods such as Decision Tree (DT), Naive Bays (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Using these methods, they have measured the heart rate variability signals based on Discrete Wavelet Transform method. The 92.02% classification accuracy had secured by DT method with 10-fold cross validation technique, which was the highest among all.

Diwedi (2017) [11] has utilized the performance of different machine learning techniques like Naive Bayes (NB), Classification Tree, K-Nearest Neighbor (KNN), Logistic Regression (LR), Support Vector Machine (SVM) and Artificial Neural Network (ANN). However, the accuracy achieved by proposed algorithms are 83%, 77%, 80%, 85%, 82%, 82% and 84% respectively. The LR method gives the highest classification accuracy 85% among all. Hashi (2017) [12] has designed an expert clinical decision support system based on classification techniques. However, the techniques used for prediction purpose are C4.5 Decision Tree and K-Nearest Neighbor (KNN). The performance of proposed algorithms were evaluated on WEKA data mining software and accuracy recorded by proposed algorithms were 90.43% and 76.99% respectively.

Shahi and Gurm (2013) [13] have used WEKA data mining software for automatic diagnosis of heart disease prediction. For this task, they have used various data mining techniques like Association Rule, Naive Bays (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) and Decision Tree (DT) respectively. Based on the experimental results, SVM method had secured the highest classification accuracy 85% among all.

Baitharu and Pani (2016) [14] have utilized the Liver Disorder dataset for performance evaluation of various data mining techniques in order to design of health care decision support system. They have investigated various techniques for study purposes like Zero R, J48, Multi Layer Perceptron (MLP), Naive Bays (NB), K-Nearest Neighbor (KNN) and Voting Frequency Intervals (VFI). However, the MLP method had recorded the highest classification accuracy 77.59% among all.

Das and Sengur (2009) [15] have applied Neural Network (NN) method using Statistical Analysis Software (SAS) for prediction of heart disease. However, the proposed method had recorded the 89.01% accuracy.

Palaniappan et al. (2008) [16] have investigated the application of various classification techniques like Neural Network (NN), Decision Tree (DT) and Naive Bays (NB) and using these techniques they have designed intelligent heart disease prediction system. The proposed system was implemented using .Net technology and their experimental results had stated that NB method outperformed than others and secured 86.12% classification accuracy.

Pouriyeh et al. (2017) [17] have recorded the performance of various machine learning techniques such as Naive Bays (NB), K-Nearest Neighbor (KNN), Multi Layer Perceptron (MLP), Radial Basis Function (RBF), Single Conjunctive Rule Learner, Decision Tree (DT) and Support Vector Machine (SVM). in their study, they have experimented on different k values (1,3,9,15) for better performance. Finally, SVM method outperformed than others and recorded 90.45% classification accuracy.

Subha et al. (2018) [18] have discussed an ensemble based Extreme Learning Machine (ELM) for Cardiovascular disease prediction. However, their experimental results had shown that proposed model outperformed well and provides better classification accuracy than base classification models.

Kamley and Thakur (2019) [19] have presented the comparative study of three important machine learning classification techniques like Support Vector Machine (SVM), Naive Bays (NB) and K-Nearest Neighbor (KNN) for heart disease prediction. The Kaggle data source is obtained and Matlab R2017a machine learning tool are used for study purpose. Finally, their experimental results had stated that SVM had recorded the highest classification accuracy 86.12% than other methods.

Finally, this study is enhanced by hybrid ensemble models and evaluates the effectiveness of such models for clinical health care system for heart disease prediction.

III. DATA PRE-PROCESSING

The Heart disease dataset is obtained from Kaggle data source repository for study purpose [20] and can be accessed by various researchers for prediction of heart disease problem. However, the dataset contain sample size of 300 patients, 14 attributes and 5 missing values [20]. During analysis, missing values are filled with mean values. Finally, dataset contain 305 patient records and 13 independent input parameters and 1 class label attribute used for prediction of heart disease [20].

In order to effective representation of data and efficient performance of machine learning classifier, data pre-processing process is used. Therefore, data pre-processing methods removing missing values, incorrect values and outliers from dataset in order to prepare dataset for prediction task [21]. Table I shows the brief description of heart disease dataset.

Table I. Brief Description of Heart Disease Dataset [20]

S.No.	Attributes	Brief Description
1	Age	Age in Years
2	Sex	Male and Female
3	CPT	Chest Pain Type
4	RBP	Resting Blood Pressure
5	FBS	Fasting Blood Sugar Level
6	SC (mg/dl)	Serum Cholesterol
7	ECG	Electro Cardio Gram
8	MHR	Maximum Heart Rate Achieved
9	EIA	Exercise Included Angina
10	Old_Peak	ST Depression Induced by Exercise Relative to Rest
11	SPE	Slope of Peak Exercise ST
12	NMV	No. of Major Vessels Values
13	Thal	Thalassemia;
14	Disease (Class)	Present Disease or Not

IV. PROPOSED METHODOLOGIES

The proposed system is designed with the aim of presence of heart disease as well as improving classification accuracy of machine learning classifier. In this study, 7 popular different machine learning classifiers Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bays (NB), Multi Layer Perceptron (MLP), Decision Tree (DT) and Bays Net (BN) and hybrid ensemble classification techniques bagging, boosting, Random Subspace Method (RSM) and RUS boost are used and their performance is compared with base classification techniques. In this section, ensemble classification and ensemble learning techniques are discussed below.

A. Ensemble Classification: A machine learning classifier that incorporates a no of classifier that work together in order to identify the class label for unlabeled instances. In order to improve the accuracy of classifier, there are various methods are used of association of ensemble classification. A well known machine learning paradigm is called ensemble learning where multiple model are combined to solve the main problem [22]. Hence, ensemble classifier is built with ensemble learning techniques like ensemble bagged tree, ensemble boosted tree, ensemble subspace discriminate etc. Typically, these techniques outperforms better than base classification techniques. Fig. 1 shows the basic diagram of ensemble classification modeling.

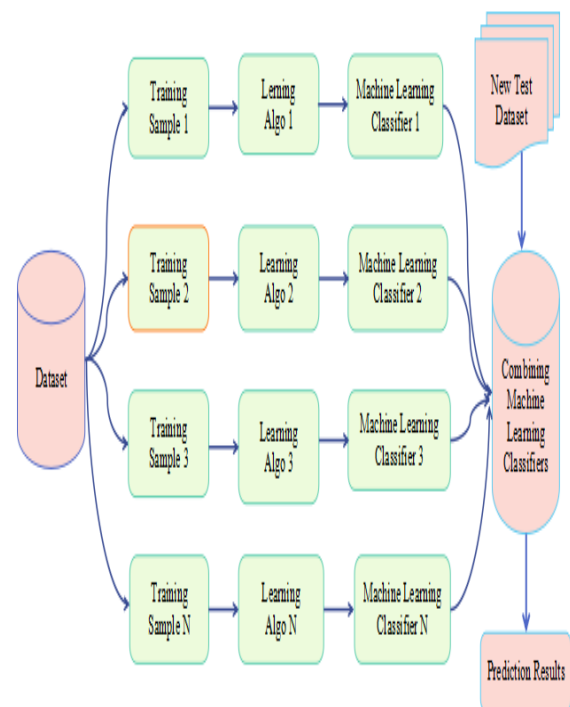


Fig. 1. Basic Diagram of Ensemble Classification Methodology [3] [5] [22]

B. Hybrid Ensemble Classification Techniques: There are some most popular and widely used hybrid ensemble classifications techniques are boosting, bagging, Random Subspace Method (RSM) and RUS boost. In the next section, these techniques are described in brief.

Boosting: one of the well known hybrid ensemble classification algorithms also known as model averaging method. However, it was actually designed to solve classification problems. Further, it was extended to solve also regression problems.

In boosting process, weights of an observation are adjusted incrementally based on the last classification performance [7] [23]. If an observation was classified incorrectly then process tries to increase the weights of this observation and vice-versa. In order to create a final model, all succeeding models are weighted according to their success and outputs and they are combined using voting (for classification problem) and averaging (for regression problem) [23].

Bagging: The other most popular algorithm in ensemble classification family is known as Boot-step aggregation or Bagging. In boosting process, multiple models can be created based on randomly drawn sample of the data [7] [8] [23]. In order to label un-label instances, each classifier model has assigned an equal vote. Finally, single model can be generated by combining all model output by using a vote majority. The bagging algorithm is more powerful than boosting i.e. in terms of model over-fitting problem. Hence, the bagging method provides more effective results for unstable non-linear model such as decision tree i.e. minor change in the training set causes major change in the learner classifier [24].

Random Subspace Method (RSM): The other popular method is proposed in ensemble family by combining models known as RSM. In RSM process, machines are trained using randomly chosen sub-spaces of the original feature set. Finally, classifier results are generated using those individual classifier performances by majority voting [23] [24]. The RSM classifier consisting of various classifiers in a sub-space of data feature space such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), linear classifier and many more [24].

Random Under Sampling (RUS) Boost: The RUS Boost is another simple hybrid data sampling method that is used to remove the data distribution imbalances between classes and improves the classification performance of weak classifiers [5] [7] [24]. In order to achieve the intended balanced class distribution, RUS Boosted method randomly eliminates the data from the training samples. So it is also named as random data extraction.

K-Fold Cross-Validation: In K-fold cross-validation technique, dataset is divided into k equal size of parts where k-1 groups are used to train the model and rest part is used to test the model performance in each step. In 10-fold validation, for each fold the process was repeated 10 times. As a result of each fold, training and testing groups were divided over the original dataset prior to selection training and testing new sets for the new cycle [8] [9] [21] [25]. In the end, 10 fold process averages of all performance metrics are evaluated.

C. Performance Evaluation Matrices: In order to evaluate the performance of all base classifiers and hybrid ensemble classifiers various performance evaluation matrices are used.

Classification Accuracy (CA): It indicates the correctly classified instances of classification model. Formula 1 is used to calculate the classification accuracy [21] [26].

$$\text{Classification Accuracy (CA)} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

Classification Error (CE): It indicates the overall incorrectly classified instances of the classification model [21] [26].

$$\text{Classification Error (CE)} = \frac{FP+FN}{TP+TN+FP+FN} * 100\% \quad (2)$$

Sensitivity/Recall/True Positive Rate: Sensitivity denotes the ratio of correctly classified instances to the total no. of heart patients. In other words, it confirms that if a diagnostic test is positive and the patient has the heart disease. The formula for sensitivity is given by equation (3) [21] [26].

$$\text{Sensitivity} = \frac{TP}{TP+FN} * 100\% \quad (3)$$

Specificity: Specificity denotes in the context of a diagnostic test is negative and patient has heart disease. The formula for sensitivity is given by equation no. (4) [21] [26].

$$\text{Specificity} = \frac{TN}{TN+FP} * 100\% \quad (4)$$

Precision: The formula for precision is given by equation no. (5) [21] [26].

$$\text{Precision} = \frac{TP}{TP+FP} * 100\% \quad (5)$$

Matthews Correlation Coefficient (MCC): MCC value [-1, 1] denotes the prediction ability of the classifier. Here +1 value denotes that ideal classification prediction and -1 value denotes that completely wrong prediction. On the other side, the value near to 0 denotes random predictions generated by classifier. The MCC formula is given by equation no. (6) [21] [26].

$$\text{MCC} = \frac{TP*TN-FP*FN}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} * 100\% \quad (6)$$

Receiver Optimistic Curve (ROC): In order to analyze the prediction ability of the machine learning classifier, Receiver Optimistic Curve (ROC) is used. However, it represents the graphical curve between True Positive Rate (TPR) and False Positive Rate (FPR) of machine learning classification algorithms [21] [26].

V. EXPERIMENTAL RESULTS

In this study, discussion of different machine learning classification models and their outcomes from different perspective is carried out. The Matlab R2017a machine learning tool is used to experimental purpose. All the computations are performed on Intel (R) Core (TM) i3-2350M CPU @ 2.30 GHz machine. However, the experimental results are conducted in six parts and discussed below one by one.

1st Experiment: In the first experiment, six well known classification approaches are considered and experimental results are generated of applying as single base classifiers. No validation is used in the experiment. Table II shows performance of all base classification techniques respectively.

Table II. Performance Comparison of Base Classification Techniques

Approach	Sensitivity	Specificity	Precision	MCC	ROC	Classification Accuracy (CA) in %	Classification Error (CE) in %
DT	77.98%	77.39%	78.98%	78.2%	79.6%	77.71%	22.29%
NB	85.53%	80.82%	83.71%	89.3%	89.2%	83.27%	16.72%
LR	87.42%	80.13%	85.43%	91.5%	89.7%	83.93%	16.06%
KNN	79.87%	65.75%	75.29%	80.5%	81.1%	73.11%	26.88%
SVM	83.94%	83.19%	84.23%	84.2%	89.7%	83.93%	16.06%
BN	82.43%	78.76%	82.32%	81.3%	90.4%	81.96%	18.03%
MLP	81.13%	81.32%	81.42%	82.7%	86.5%	81.31%	18.68%

2nd Experiment: In second experiment, performance of hybrid ensemble classification techniques is evaluated without any validation. Table III shows performance

comparison of hybrid ensemble classification techniques without any validation.

Table III. Performance Comparison of Hybrid Ensemble Classification Techniques without Validation

Approach	Sensitivity	Specificity	Precision	MCC	ROC	Classification Accuracy (CA) in %	Classification Error (CE) in %
Boosting	83.01%	80.13%	78.10%	84.42%	88.12%	79%	21%
Bagging	86.79%	80.13%	86.79%	87.71%	89.85%	83.6%	16.4%
RSM	82.52%	73.28%	77.19%	82.19%	87.47%	78.4%	21.6%
RUS Boost	89.30%	78.08%	81.60%	86.32%	92.32%	84.1%	15.9%

3rd Experiment: In third experiment, performances of base classification techniques with 10-fold cross validation technique are evaluated. Table IV shows performance of base

classification techniques with 10-fold cross-validation technique.

Table IV. Performance Comparison of Base Classification Techniques with 10-Fold Cross-validation Technique

Approach	Sensitivity	Specificity	Precision	MCC	ROC	Classification Accuracy (CA) in %	Classification Error (CE) in %
Boosting	97.20%	98%	98.14%	97.65%	98.89%	97.20%	2.80%
Bagging	97.60%	99.31%	99.37%	99.60%	99.56%	99.30%	0.70%
RSM	88.50%	91.82%	88.60%	92.24%	95.80%	88.52%	11.47%
RUS Boost	96.85%	87.67%	89.60%	95.63%	97.60%	92.50%	7.50%

4th Experiment: In fourth experiment, performances of hybrid ensemble classification techniques with 10-fold cross validation techniques are evaluated. Table V shows performance of hybrid ensemble classification techniques with 10-fold cross-validation technique.

Table V. Performance Comparison of Hybrid Ensemble Classification Techniques with 10-Fold Cross-validation Technique

Approach	Sensitivity	Specificity	Precision	MCC	ROC	Classification Accuracy (CA) in %	Classification Error (CE) in %
DT	91.80%	77.98%	91.90%	97.20%	94.40%	91.80%	8.19%
NB	85.60%	88.67%	85.60%	95.40%	91%	85.57%	14.42%
LR	84.90%	88.67%	85%	91.50%	92.10%	84.91%	15.08%
KNN	76.70%	79.75%	76.80%	84.70%	89.80%	76.72%	23.27%
SVM	84.60%	87.19%	84.23%	90.40%	90.70%	84.60%	15.04%
BN	82.60%	84.91%	82.60%	89.60%	91.70%	82.62%	17.37%
MLP	95.70%	96.85%	95.80%	98.80%	98.10%	95.73%	4.26%

5th Experiment: In this experiment, performance is compared with other heart disease prediction models. The proposed model has achieved acceptable results performance by using hybrid ensemble classification methods but another important result is to compare the performance of proposed

models with other previous studies. Table VI shows performance comparison of proposed model with previous studies.

Table VI. Performance Comparison of Proposed Model with Previous Studies

Year	Author	Method	Classification Accuracy (CA)
2013	[13]	SVM, NB, Association Rule, KNN, ANN and DT	SVM 85%
2009	[15]	Neural Network Ensembles	89.01%
2018	[23]	Hybrid Ensemble Models (Adaboost, Logit boost, MLP, Random Forest)	MLP 96%
2009	[26]	Hybrid Model (GA+SVM)	84.07%
2008	[16]	NN, DT and NB	NB 86.12%
2017	[17]	NB, KNN, MLP, Radial Basis Function (RBF), Single Conjunctive Rule Learner, DT and SVM	SVM 90.45%
2013	[24]	RIPPER, Decision tree, ANNs, SVM	SVM 84.12%
2017	[22]	RFRS	92.59%
2010	[25]	GA + Naïve Bayes	85.87%
2019	Proposed Model	Bagging, Boosting, RSM, RUS Boost	Bagging 99.3%

Table VI clearly states that proposed model provides excellent performance over other previous studies.

VI. DISCUSSION

The heart disease dataset is considered for study purpose and different machine learning techniques are applied to predict heart disease. The main concern of this study is to compare the performance of different machine learning classification models and present the best one. Based on all the results, different algorithms are performed better according to situation (with or without validation). However, every algorithm has intrinsic power to outperform than other algorithm depending on situation.

In this way, performance of base classification techniques are compared (Table II). However, Support Vector Machine (SVM) and Logistic Regression (LR) methods have highest and equal performance accuracy i.e. 83.93%. The high value

of sensitivity i.e. 87.42% shows that probability that a diagnostic test is positive and model is accurate in detecting and making an accurate prediction of heart disease if new patient comes to health care centers with undiagnosed heart disease. The 80.13% specificity value shows that 80.13% of probability that diagnostic test is negative and patient does not have heart disease. Similarly, 85.43% high precision value indicates that 85.43% positive identifications were actually correct. The ROC values 91.5% show the LR model covered 91.5% area which was greater as compared to other models. Similarly, MCC values 89.7% shows model has achieved accurate prediction performance.

In Table III, hybrid ensemble technique RUS Boost (without validation) had achieved the highest classification accuracy i.e. 84.1% which is the highest among as base classification techniques.

The sensitivity value 89.3% denotes that 89.3% of probability of patient has heart disease. On the other side, specificity value 78.08% denotes that diagnostic test is negative and 78.08% probability of patient does not have the heart disease. The precision value 81.6% denotes that 81.6% positive identifications were actually correct. The MCC value 86.32% shows the good classification predictions and ROC value 92.32% denotes that model has covered 92.32% area which was the highest among all classification models.

In Table IV, MLP model with 10-fold cross validation technique had achieved the highest classification accuracy i.e. 95.73% which was the highest as compared to hybrid ensemble classification techniques (RUS Boost, 84.1%, Table III) and base classification techniques (83.93%, LR, SVM, Table II). The high values of sensitivity, specificity, precision, MCC and ROC (above than 95%) denote that model had achieved outstanding performance in terms of all accuracy parameters.

In Table V, hybrid ensemble classification techniques bagging and boosting with 10-fold cross validation technique had secured the excellent classification performance i.e. 99.3% and 97.2% respectively. However, the proposed techniques had performed better than base classification techniques (with or without validation) and hybrid ensemble classification techniques (without validation). The sensitivity, specificity, precision, MCC and ROC values were recorded up to 97.6%, 99.31%, 99.37%, 99.6% and 99.56% respectively which was very close to 100. Thus, we can say that hybrid ensemble techniques performed better than base classification techniques in terms of accuracy

VII. CONCLUSION AND FUTURE SCOPES

In this research study, different machine learning techniques like LR, DT, NB, SVM, MLP, KNN and BN have been applied as base classifier and hybrid ensemble classification techniques like Bagging, Boosting, RSM and RUS boost with or without validation have applied on the heart disease dataset. However, the performance of each technique are evaluated based on different evaluation metrics like Classification Accuracy (CA), Classification Error (CE), sensitivity, specificity, precision, Mathew Correlation Coefficient (MCC) and Receiver Optimistic Curve (ROC).

In this direction, we have recorded some improvements. Base classification techniques with validation performed better than base classification techniques without validation. Hybrid ensemble classification techniques without validation performed better than base classification techniques with validation. Finally, hybrid ensemble classification techniques with validation outperformed than base classification techniques (with or without validation) and hybrid ensemble classification techniques without validation respectively. However, the main goal of this study is to compare the performance of different machine learning techniques on a small dataset. In this direction, we have tried to improve the accuracy of fore-mentioned techniques.

In future, other hybrid optimization techniques and feature selection algorithms will be adopted in order to improve the performance of machine learning techniques as well as big data samples and more attributes will be considered for better comparison of performance.

REFERENCES

1. World Health Organization (WHO) (2014). Fact Sheet Available Online at "http://www.who.int/mediacentre/factsheets/fs317/en/".
2. K. D. Kochanek, Xu. J. Murphy, S. L. Miniño A.M. and H.C. Kung, "Deaths: Final Data for 2009. National Vital Statistics Reports, 60(3), pp. 1–116, 2011.
3. W. Dai, T.S. Brisimi, W.G.Adams, T. Mela, V. Saligrama, and I.C. Paschalidis (2015). Prediction of Hospitalization due to Heart Diseases by Supervised Learning Methods. International Journal of Medical Informatics, 84(3), 189–197.
4. G. L. Ki Heon, K.N. Yong and Ryu Ho (2007). Mining Biosignal Data: Coronary Artery Disease Diagnosis Using Linear and Nonlinear Features of HRV. LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, 56–66.
5. M. Fatima and M. Pasha (2017). Survey of Machine Learning Algorithms for Disease Diagnostic. Journal of Intelligent Learning Systems and Applications, 9(1), 1–6.
6. K.C. Hian and T. Gerald (2005). Data Mining Applications in Healthcare. Journal of Healthcare Information Management, 19(2), 64–72.
7. E. Alpaydin (2004). Introduction to Machine Learning (Adaptive Computation and Machine Learning). The MIT Press.
8. M.K. Obenshain, "Application of Data Mining Techniques to Healthcare Data. Infection Control & Hospital Epidemiology", 25(8), pp. 690–695, 2004.
9. A. Malav, K. Kadam and P. Kamat (2017). Prediction of Heart Disease Using K-means and Artificial Neural Network as a Hybrid Approach to Improve Accuracy. International Journal of Engineering and Technology, 9(4), 1–7.
10. U. Acharya, K.S. Rajendra, D.N. Vidya Ghista, W.J.E. Lim, F. Molinari and M. Sankaranarayanan, "Computer Aided Diagnosis of Diabetic Subjects by Heart Rate Variability Signals Using Discrete Wavelet Transform Method", Knowledge Based System, 81, pp. 56–64, 2015.
11. A.K. Dwivedi, "Analysis of Computational Intelligence Techniques for Diabetes Mellitus Prediction. Neural Computing and Application", 13 (3), pp. 1–9, 2017.
12. E.K. Hashi, M.S.U. Zaman and M.R. Hasan, "An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques", International Conference on Electrical Computation, Communication, Engineering, pp. 396–400, 2017.
13. M. Shahi and R. Gurm Kaur, "Heart Disease Prediction System Using Data Mining Techniques", Oriental Journal Computer Science Technology, 6(4), pp. 457–466, 2013.
14. T.R. Baitharu and S.K. Pani (2016). Analysis of Data Mining Techniques for Health-care Decision Support System Using Liver Disorder Dataset. Procedia Computer Science, 85, 862–870.
15. R. Das and T. Sengur (2009). Effective Diagnosis of Heart Disease through Network Ensembles. Expert System with Applications, 36, 7675–7680.
16. S. Palaniappan and R. Awang (2008). Intelligent Heart Disease Prediction System Using Data Mining Techniques. International Journal of Computer Science and Network Security, 8(8), 1–6.
17. S. Pouriyeh, S. Vahid, G.D. Pietro, H. Arabnia and J. Gutierrez (2017). A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease. 22nd IEEE Symposium on Computers and Communications (ISCC 2017): Workshops-ICTS4eHealth.
18. R. Subha, K. Anandakumar and A. Bharathi (2018). An Ensemble based Extreme Learning Machine for Cardiovascular Disease Prediction. International Journal of Applied Engineering Research, 13(10), 7903–7912.
19. S. Kamley and R.S. Thakur (2019). Machine Learning Techniques: Performance Analysis for Prevalence of Heart Disease Prediction. International Journal of Engineering and Advanced Technology (IJEAT), 8(4), 188–191.
20. Data Source Repository Downloaded on 19/03/2019 from "http://www.kaggle.com".
21. J. Han and M. Kamber, (2006). Data Mining: Concepts and Techniques. 2nd edition, Morgan Kaufmann Publishers, San Francisco.

22. X. Liu, Q. Su. Wang, M., Zhang, Y.Q. Zhu Wang and Q. Wang (2017). A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method. Computational and Mathematical Methods in Medicine, pp.11-16.
23. E. Nikookar and E. Naderi (2018). Hybrid Ensemble Framework for Heart Disease Detection and Prediction. International Journal of Advanced Computer Science and Applications (IJACSA), 9(5), 243-248.
24. D. Vadicherla, and S. Sonawane (2013). Decision Support System for Heart Disease Based on Sequential Minimal Optimization in Support. International Journal of Engineering Sciences and Emerging Technologies, 4(2), 19–26.
25. A.G. Karegowda, A.S. Manjunath and M.A. Jayaram (2010). Feature Subset Selection Problem Using wrapper Approach in Supervised Learning. International Journal of Computer Applications, 1(7), 13–17.
26. K.C.Tan, E.J. Teoh, Q. Yu and K.C. Goh (2009). A Hybrid Evolutionary Algorithm for Attribute Selection in Data Mining. Expert Systems with Applications, 36(4), 8616-8630.

AUTHORS PROFILE



Sachin Kamley completed his Masters degree from S.A.T.I., Vidisha in 2006 affiliated to Rajiv Gandhi Technological University, Bhopal (M.P.). He is working as an assistant professor in Department of Computer Applications since May 2007 at Samrat Ashok Technological Institute (S.A.T.I), Vidisha and also completed Ph. D from Barkatullah University, Bhopal in the year 2015. He has published more than 18 papers in International journals and conferences and attended many workshops of National repute.