www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

Optimizing Mortality Prediction in Cardiac Patients Using Genetic Algorithm and Random Forest with Class Imbalance Handling

Asim Javed
Department of Computer Science, Times Institute, Multan, 60000, Punjab, Pakistan

Sadaqat Ali Ramay
Department of Computer Science, Times Institute, Multan, 60000, Punjab, Pakistan

Tahir Abbas (Corresponding Author)
Department of Computer Science, Times Institute, Multan, 60000, Punjab, Pakistan. Email: drtahirabbas@t.edu.pk

Ashir Javeed
Blekinge Institute of Technology, Karlskrona, 37179, Sweden

Shahbaz Saeed
Department of Computer Science, Times Institute, Multan, 60000, Punjab, Pakistan

Wasif Akbar
Department of Computer Science, Emerson University, Multan, 60000, Punjab, Pakistan

Abstract
This study presented a decision support system based on data mining and machine learning techniques for accurate prediction of cardiac patients' deaths because resources in the hospitals are limited; therefore, appropriate allocation of resources will improve the survivability of cardiac patients. Data mining techniques are widely used by researchers to uncover hidden information and patterns that could potentially save or prolong patient lives. Age, gender, high blood pressure, cholesterol, and irregular heartbeat rates are some of the variables considered for this study. This study used medical health records data from 368 observations with 55 unique features. This work presents a machine learning-based approach for predicting cardiac patient death by employing an electronic health record (EHR) dataset. The constructed model is based on a genetic algorithm (GA) for selecting important features from the dataset and a Random Forest (RF) model for classifying mortality in cardiac patients. The hyperparameters of RF models were optimized using a grid search algorithm for improved performance of RF. A public dataset was obtained in order to evaluate the efficacy of the constructed GA_RF model. One of the problems that we encountered during this study was imbalance classes in the collected dataset. The machine learning models tend to bias toward the majority class in the dataset. To overcome this problem, the Random Under Sampling (RUS) method was

Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

employed. The performance of the constructed GA_RF model was tested on several evaluation metrics, and results validate the effectiveness of the proposed GA_RF model for mortality prediction in cardiac patients.
Keywords: machine learning; cardiac morality; feature ranking; random forest; imbalance classes.

Introduction

The World Health Organization (WHO) projects that heart disease will be the primary cause of 23.6 million deaths worldwide by 2030; therefore, creative approaches to avoid complications will be needed, even in the face of challenging tasks and false presumptions. This rapid increase in cardiac patients will put a significant burden on the healthcare system. The resources in the healthcare system are limited; therefore, appropriate allocation of resources will improve the survivability of the cardiac patients [1]. Furthermore, the healthcare industry generates a significant amount of data, including repetitive information, numerous assignments, incomplete data, and a warm relationship with time [2]. However, this electronic data is frequently stored by healthcare departments for future use. The researchers are using medical health data for several diseases' prediction, such as dementia, heart failure, stroke, and cancer. The data mining techniques are also employed by the researchers for the identification of risk factors that are associated with diseases. Furthermore, improvements made in the field of artificial intelligence also motivated the researcher to build decision support systems that would help medical practitioners in real-world scenarios make the best decision.

According to the World Health Organization, 17.9 million deaths worldwide in 2019 due to cardiovascular diseases, and 85% of these deaths were caused by heart attacks and strokes. Although identifying diseases based only on symptoms is difficult, early disease detection is essential for survival. To help doctors diagnose patients accurately, computational algorithms have been developed. This study combines support vector machines and neural networks for optimal results in biomedical-focused decision support systems that use classification models to evaluate and compare data [3].

Numerous systems for clinical decision-making have been developed by researchers [4], machine learning and data mining-based automated diagnostic systems for many different kinds of medical ailments, such as hepatitis [5], dementia [6],[7], and heart disease [8], [9]. Proposed an intelligent learning system based on RF and a random search algorithm for the early detection of heart disease. Other researchers created distinct methods as well. Their proposed model's accuracy was 93.33%, the highest level that was feasible [10]. Anxiety, tension, and restlessness are on the rise due to the fast-paced lifestyles of today. The leading cause of illness in younger generations is heart diseases, which include arrhythmia, mitral regurgitation, hypertrophic cardiomyopathy, dilated cardiomyopathy, coronary artery disease, congenital heart diseases, heart attacks, heart failure, and myocardial infarction [11]. The study focused on 13 traits and risk variables while examining data mining techniques for predicting fatal cardiac conditions. Decision trees, naive bayes, and neural networks were used to compare the accuracy rates. The study found that neural networks outperformed

Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

decision trees and naive bayes with the highest accuracy rate of 100% in predicting heart disease and death [12].

Heart disease is common in young people; peripheral vascular disease, renal abnormalities, and stroke are the main causes. The Heart Disease Prediction System (HDPS) is crucial for early disease predictions, especially for people in the 30- to 40-year-old age range. Early disease prediction models are developed using data mining techniques such as intelligent support vectors. This technique reduces the cost of treatment by enabling the quicker and more accurate identification of cardiovascular diseases [13]. This study evaluated the predictive power of several data mining techniques for heart diseases. Using a neural network and patient data from 15 different parameters, the most accurate method was identified. When combined with genetic algorithms and feature subset selection methods, decision trees proved to be a helpful strategy [14].

The objective of this study is to leverage from data mining and machine learning approaches for constructing a decision support system based on medical health data for predicting mortality in cardiac patients so that the resources in the hospital should be efficiently used. However, there are a few limitations of machine learning techniques that need to be addressed, such as that machine learning models tend to bias toward the majority class in the dataset. Therefore, proposed decision support should eliminate this problem for predicting mortality. To overcome this problem, the newly developed decision support system utilized random sampling technique to balance the classes in the dataset. Moreover, another problem faced by the machine learning models was overfitting. To avoid this problem, the proposed decision support system employed feature selection techniques so that machine learning should not learn noise or irrelevant information from the dataset. The employed feature selection algorithm also helps us to understand the risk factors that cause mortality in cardiac patients. The following is an overview of the study's primary contributions:

- For predicting cardiac patients' death, a decision support system called GA_RF is proposed;
- Genetic algorithm was employed to overcome the problem of model overfitting;
- The random under sampling (RUS) technique is used to overcome the issue with the bias machine learning model caused by the dataset's imbalanced classes;
- The developed decision support system (GA_RF) improves the performance of the traditional random forest model by 5.0% for the prediction of death in cardiac patients and outperforms other state-of-the-art machine learning models;
- The risk variables that contribute to cardiac patients' mortality are also identified by the proposed GA_RF model. The mortality rate for cardiac patients can be significantly decreased by addressing these risk factors;

Materials and Methods

This section presents a thorough examination of all the techniques used in building the proposed (GA-RF) model as well as the dataset used. The pre-processing of acquired medical health data, the selection of important features

www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

using a genetic algorithm, the application of the random under sampling technique to address the bias problem associated with the proposed model, and the optimization of the random forest model for classification.

2.1. Dataset Description

The study utilized an online publicly available dataset consist of 55 features of cardiac patients with a total sample sized of 368. The demography of observation in the collected dataset were 83 women, 285 men, and 80 cardiac patients as given in Figure 1. The detection of cardiac death and the absence of a connection between heart disease and death were done using binary labeling. Excellent and poor case studies for men and women were both displayed by the statistical data.

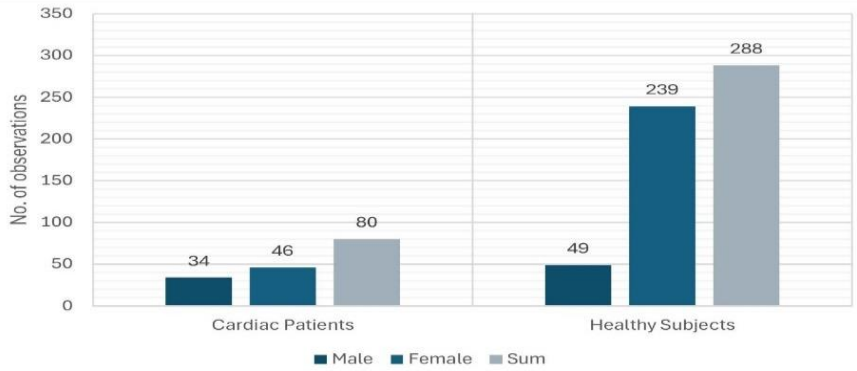

Figure 1. The dataset's sample distribution

Proposed Model

Five stages comprise the proposed work: data collection, preprocessing, feature selection, classifier application, and prediction. The secondary data collected for this study's dataset came from an online source. There are 75 unique attributes in the dataset. The dataset contains 368 samples in total. As a result, data was extracted, and the selection criteria that had the strongest correlation to mortality prediction were identified. Figure 2 presents the workflow of a newly constructed decision support system for predicting mortality in cardiac patients.

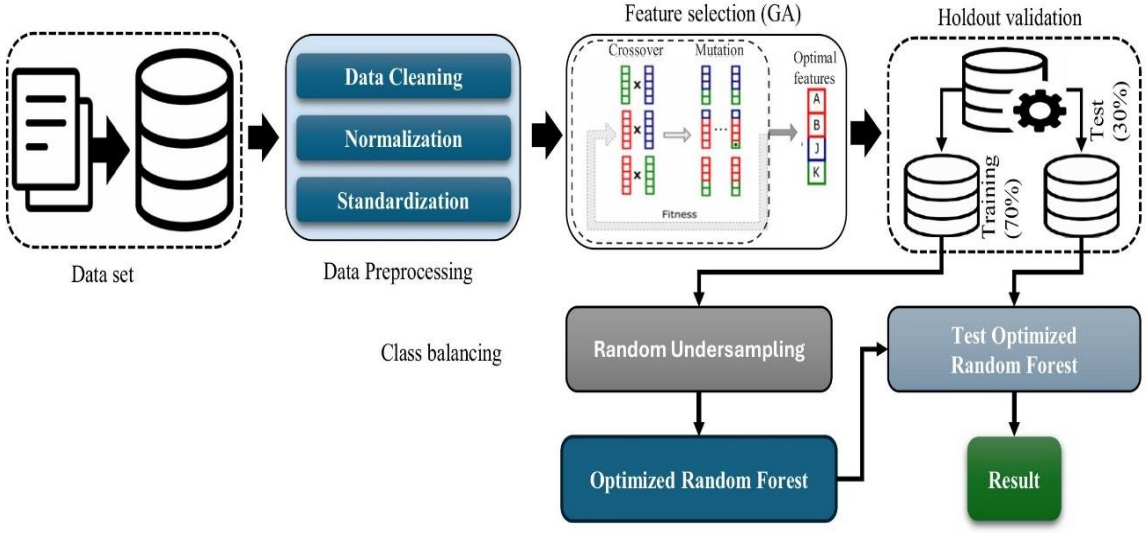

Figure 2. GA-RF model for cardiac mortality prediction.

After loading the dataset, the first of the proposed models is data preprocessing. The machine learning algorithms work well when data is in refined form.

www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

Therefore, several operations were carried out to process the data, such as data cleaning, normalization, and data standardization. In the data cleaning operation, the missing values were handled, while in the normalization operation, the data types of each feature in the dataset were normalized. Data was standardized by employing the z-score standardization method on the whole feature space. Data standardization is a preprocessing technique that rescales the features of your data, so they have a mean of 0 and a standard deviation of 1. This is particularly useful in machine learning when you have features with different scales, as it helps many algorithms perform better. The mathematical formula for z-score standardization for feature (F) and standardized value Standardization is given as:

$$F_{standarization} = \frac{F - \lambda}{\delta}$$

Where λthe mean of the feature is values and δ is the mean of the feature values. After standardization, the features have a mean of approximately 0 and a standard deviation of 1.

Following the data standardization, a feature selection module was deployed to select the most relevant features from the data. The proposed model employed a genetic algorithm (GA) for this task. GA is a successful search approach for investigating complex spaces using natural selection and population genetics concepts. It allows for easy input property selection for neural networks, with each feature having a binary chromosome. Fitter individuals are more likely to pass on their genes to the next generation, resulting in a 'Survival of the Fittest'. The future generation is produced through replication, fitness rating, crossover points, and mutation.

Following feature selection from a GA, data partitioning was performed using a holdout validation strategy for the goal of training and testing the newly built model. Unfortunately, the data contains imbalanced classes, which may result in biased results. Machine learning (ML) models trained on imbalanced data are prone to bias, favoring the majority class while disregarding the minority class. Because minority class cases are taught seldom throughout the training phase, minority class prediction is unusual, ignored, and unreported. In the literature, several approaches to handling imbalanced data have been put forth. The technique that is most frequently used is resampling. Two techniques make up this strategy: under-sampling and oversampling. Oversampling equalizes the size of each class in training data by duplicating minority class samples. In order to balance the size of each class, some majority class samples are eliminated during the training phase. Therefore, a model is supposed to behave impartially when trained with balanced data. Numerous methods for under-sampling have been put out in the literature. However, it has been discovered that random under-sampling is the easiest method and performs similarly to other methods. As a result, the proposed model used the random under-sampling strategy to enhance the training process and minimize bias from the models created. The random under sampling approach selects participants at random from the larger class in each iteration/fold of a cross-validation trial until the training data is balanced. Hence, the training process is improved or balanced. It's important to remember that the resampling methods are only applied to the training data during each cross-validation cycle, not the entire dataset before cross-validation. Figure 3

www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

presents the original distribution of data along with balanced data after applying a random under sampling technique.

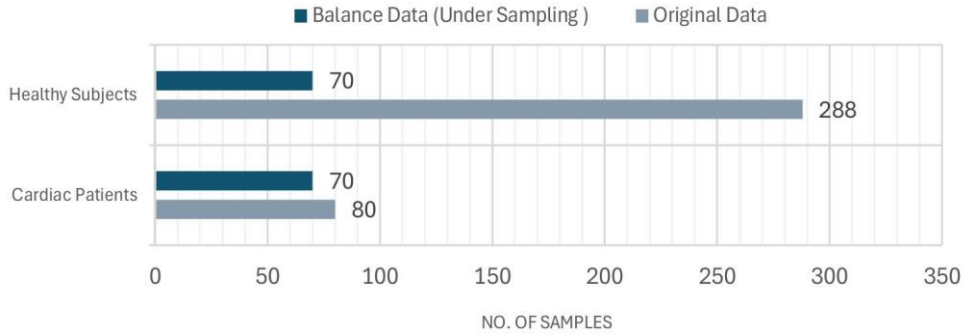

Figure 3. Data distribution overview before and following under sampling.

After balancing the training dataset, the selected features were fed into a random forest model for the classification of the mortality prediction. The hyperparameters of the RF model were optimized using the grid search algorithm. The optimized RF model was trained on the balanced data. In the next step, the test data was used for the performance evaluation of the optimized trained RF model. The results were reported on the test data.

Experiments and Results

3.1. Validation and Evaluation Metrics

Researchers test the efficacy of the machine learning models through a variety of validation techniques [15]–[16][17]. In this study, we employed a holdout-validation scheme. In the holdout validation scheme, the collected dataset is partitioned into sections such as 70% of the dataset is used for the training purpose of the proposed model while 30% dataset is used for the testing of the newly constructed model [18].

The effectiveness of the constructed model has been assessed using several evaluation metrics, such as accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), and Mattew's correlation coefficient (MCC). MCC was selected for the statistical analysis of the proposed model. A prediction with an MCC value of 1 is considered the best; one with a value of 1 is the worst. The evaluation metrics that were selected for the evaluation of the proposed model are given as follows:

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

$$Sensitivity = \frac{TP}{T\,P + F\,P}$$

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

$$\text{F1 Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

$$MCC = \frac{TP * TN - FP * F\,N}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Where TP stands for true positive, FP stands for false positive, TN denotes true negative, and finally, FN presents false negative.

3.2. Experimental Configuration

www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

This section provides information regarding all four experiments that were set up to analyze the impact of imbalance classes in the dataset as well as the efficiency comparison of the proposed GA_RF models along with other state-of-the-art machine learning models using all features in the dataset and top selected features by the genetic algorithm. All the experiments were done using a Windows 10 PC with 32 GB of RAM and an Intel Core TM i7-8250U CPU running at 2 GHz. The Python 3.7.0 software package and the Anaconda Spyder integrated development environment (IDE).

3.2.1. Experiment 1: ML Models Performance Using All Features on an imbalanced dataset

We employed state-of-the-art ML models such as decision tree (DT), random forest (RF), naïve Bayes (NB), logistic regression (LR), k-nearest neighbors (KNN), and support vector machines (SVM) with various kernels (polynomial, RBF, linear, and sigmoid functions). The performance of the developed ML models was assessed using all features of the dataset. It is significant to remember that the classes in the dataset differ from one another. We thus preserved the dataset in its entirety and original form for this experiment. Table 1 displays the sensitivity, specificity, accuracy, and MCC of the various machine-learning algorithms for cardiac disease-related death prediction. Regarding cardiac mortality prediction, the linear kernel support vector machine (SVM) showed the highest accuracy of 85.24%. The specificity values of each machine learning model in Table 1 are less than the sensitivity values, suggesting that the ML models in the dataset tend to favor the majority class. Therefore, we must address this problem in the upcoming experiment. Moreover, non-constant accuracy tests and train values are observed in some machine learning models, like SVM, suggesting an overfitting phenomenon. Therefore, we also need to address this issue in the next experiments.

Table 1: Performance of ML models on an imbalanced data set using all features.

MODEL	ACC._TRAIN	ACC._TEST	SENSITIVITY	SPECIFICITY	F1 SCORE	MCC
NB	68.06	73.87	99.50	68.32	74.00	0.5230
LR	89.85	81.55	92.70	59.37	84.70	0.5557
DT	86.35	77.70	85.70	61.50	86.70	0.4354
RF	93.70	83.77	78.78	63.70	89.70	0.4658
KNN	83.75	84.38	86.51	59.30	87.70	0.4757
AdaBoost	80.29	82.88	82.70	64.70	82.70	0.5828
SVM_RBF	57.50	81.70	81.98	60.50	79.90	0.3991
SVM_sigmoid	66.21	67.56	80.33	66.34	68.70	0.3367

Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

SVM_linear	81.92	85.24	80.70	68.00	91.20	0.4190
SVM_poly	83.32	82.78	82.70	71.71	83.70	0.3539

3.2.2. Experiment 2: Performance of ML Models Using All Features on Balance Dataset

The first experiment clearly shows that ML models typically favor the class that makes up most of a dataset. First, to avoid this prejudice problem, we balance the training data. We used the random under sampling (RUS) approach to ensure that each class's size in the training data is balanced. Any attempt to balance the testing set will lead to overfitting of the model; the training set is the only set of data that is balanced. Table 2 shows that the ML models perform better when the data is balanced, suggesting that bias has no impact on the models' functionality. The SVM with a linear kernel had the best accuracy, at 90.44%. When comparing Table 2 to Table 1, it is observed that both sensitivity and specificity have increased, suggesting that the biased problem is no longer affecting our machine learning models.

Table 2: ML model performance used all features on a balanced dataset.

MODEL	**ACC._TRAIN**	**ACC._TEST**	**SENSITIVITY**	**SPECIFICITY**	**F1 SCORE**	**MCC**
NB	69.16	74.87	99.50	74.52	73.50	0.5290
LR	90.35	85.55	92.70	89.17	84.50	0.5547
DT	86.35	86.70	85.70	77.70	86.70	0.4354
RF	93.70	89.57	78.38	73.70	91.50	0.4658
KNN	82.85	87.38	86.51	72.30	87.70	0.4749
AdaBoost	81.39	82.88	82.70	89.10	82.70	0.5838
SVM_RBF	47.70	81.70	81.98	88.90	81.70	0.3937
SVM_sigmoid	68.31	67.56	80.23	76.14	67.70	0.3323
SVM_linear	92.12	90.44	80.70	85.20	92.70	0.4190
SVM_poly	80.52	81.98	81.70	77.91	81.10	0.3539

3.2.3. Experiment 3: Proposed Model GA_RF Performance on balanced Data

In this experiment, we used the GA module to confirm the effectiveness of a recently proposed method (GA_RF) for extracting the most relevant features from the dataset. In the meantime, categorization was done using the optimized RF. A lattice research method was used to optimize the RF model's hyperparameters. Table 3 demonstrates the performance of the proposed model

Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

on the number of selected features (SF) based on a genetic algorithm. F1 score, ACC._TRAIN, ACC._TEST, sensibility, specificity, and the Matthews correlation coefficient (MCC) are the metrics used to measure training accuracy. Using only 10 of the dataset's characteristics, the proposed model GA_RF obtained the highest test accuracy (ACC._TEST) of 95.09% from Table 3, along with a training accuracy (ACC._TRAIN) of 91.22%, sensitivity of 98.50, specificity of 81.27, F1 score of 94.50%, and MCC of 0.8463, with the best RF hyperparameter values provided at a tree's maximum depth (D = 5). By comparing Tables 2 and Table 3 results, it is demonstrated that the suggested model GA_RF is about 4% superior to the conventional RF model, which makes use of all the dataset's features. Nonetheless, the suggested model GA_RF only employs ten features, minimizing the temporal complexity of the suggested approach.

Table 3: On balance dataset; performance of the proposed GA_RF model.

SF	HYPER.	ACC._TRAIN	ACC._TEST	SENSITIVITY	SPECIFICITY	F1 SCORE	MCC
08	E:10, D:10	92.41	88.78	92.81	70.50	88.50	0.6164
09	E:100, D:5	91.15	84.28	55.34	95.50	84.50	0.5759
07	E:100, D:5	95.46	88.88	63.83	98.03	88.50	0.6951
11	E:10, D:5	96.71	91.49	68.24	99.12	91.50	0.7809
11	E:200, D:5	96.20	92.39	70.50	97.46	92.50	0.7987
12	E:200, D:2	96.27	89.68	64.14	96.85	89.50	0.7476
13	E:100, D:5	96.27	93.29	99.50	72.91	93.50	0.8172
14	E:200, D:5	97.72	92.39	98.15	73.58	92.50	0.7697
10	E:100, D:10	91.22	95.09	98.50	81.27	94.50	0.8463
	E:10, D:2						0.8148
12	E:100,	94.06	91.39	95.76	86.50	91.50	0.81
13	D:10	95.06	93.49	96.65	77.81	93.50	80
14	E:200,	97.21	93.32	97.02	71.50	93.50	0.82
14	D:2	97.72	94.19	99.50	75.50	94.50	81
	E:100, D:5						0.8366

Additionally, a thorough evaluation of the newly proposed method (GA_RF) is carried out by utilizing the receiver operator characteristic (ROC) curve. The ROC curve is a visual representation used to assess the effectiveness of a binary classification model by depicting the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR). Furthermore, the ROC curve calculates the area

Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

under the curve (AUC). AUC is a single measure that summarizes the ROC curve. It goes from 0 to 1, with 1 indicating a perfect model, 0.5 representing random guessing, and anything less than 0.5 indicating worse-than-random performance. Hence, a graph with a large area under the curve is deemed more accurate. AUC is useful because it gives an overall picture of performance across all categorization levels. Thus, the performance of the newly constructed model (GA_RF) is compared with the conventional RF model. It can be depicted from Figure 5 that the proposed model achieved an AUC of 95% in comparison to the conventional RF model, whereas in Figure 6, the conventional RF model obtained an AUC of 90%.

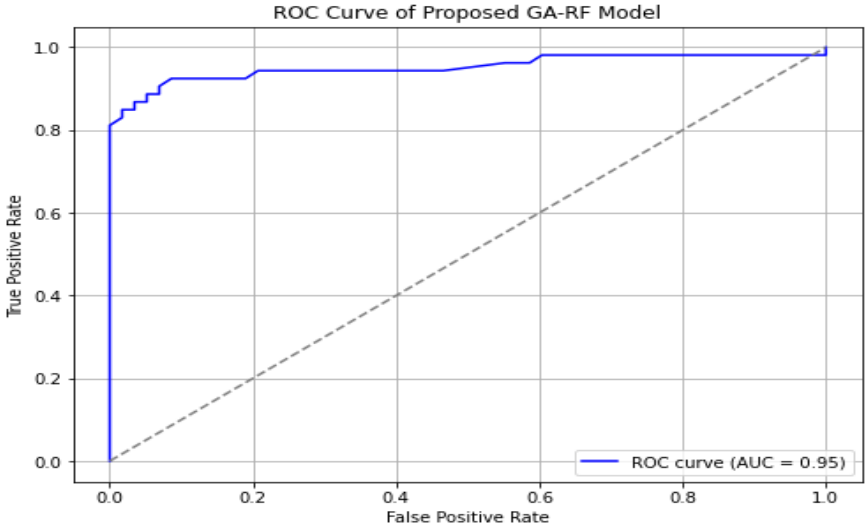

Figure 4. ROC of the newly proposed model's (GA_RF).

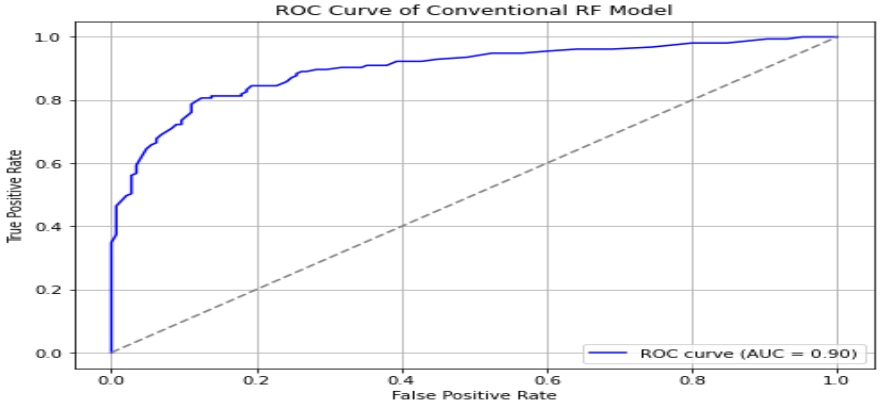

Figure 5. The ROC curve of the traditional RF model.

Furthermore, the proposed model's performance was also assessed using the confusion matrix that was employed. Figure 6 provides an overview of the confusion matrix (GA_RF) for the proposed model.

www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

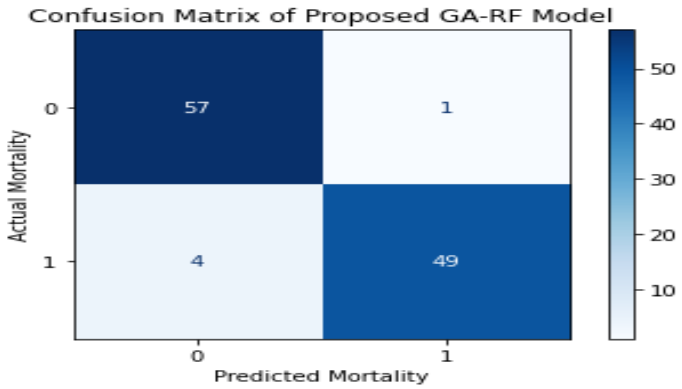

Figure 6. GA_RF proposed the model's confusion matrix.

To validate the proposed model on balanced data, we also evaluated the model's performance on imbalanced data, as Table 1, 3 show. To analyze the performance of the proposed model (GA_RF) on imbalanced data, several assessment indicators were employed. Crucial factors include MCC, F1 score, precision, sensitivity, and specificity. Unbalanced data (Table 4) reveals that the best accuracy of 87.15% was obtained by using only ten carefully chosen features, in contrast to balanced data, where the proposed model achieved the highest level of accuracy (95.09%) by using the same number of features (10).

Table 4: On imbalanced data, the suggested model (GA_RF) performed well.

SF	**ACC._TRAIN**	**ACC._TEST**	**SENSITIVITY**	**SPECIFICITY**	**F1 SCORE**	**MCC**
02	77.54	82.10	81.58	75.75	81.50	0.2405
04	80.48	82.98	84.15	57.64	82.50	0.2581
06	79.48	83.00	83.70	60.50	82.50	0.2545
08	84.74	84.78	91.41	57.02	82.50	0.4958
10	88.82	87.15	91.61	62.40	86.50	0.5351
13	88.50	85.34	90.61	60.50	85.50	0.4957

Out of the total 55 features, the GA selected the ten most significant features, which are 6, 10, 12, 22, 27, 36, 37, 41, 48, and 53. These are the main risk factors that contribute to cardiac mortality.

Table 5: Description of the top 10 selected features from the data by GA.

Sr. No	**Selected Feature**	**Description of feature**
01	06	Lifestyle
02	10	Hyperlipi
03	12	Family History
04	22	S. Sodium

www.journalofeducationalresearch.online
Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

05	27	ESR
06	36	Neutrophil
07	37	Lympho
08	41	Co
09	48	Restecg
10	53	Ca

Experiment 4: ML Model Performance Utilizing the GA Feature
In this experiment, we compared the performance of state-of-the-art machine learning models, such as NB, KNN, RF, DT, LR, and SVM with different kernels, using the GA feature selection module and the recently released model (GA_RF). The hyperparameters of the selected ML models were also adjusted through the grid search method (optimization process). From the dataset, we selected balanced classes to ensure a precise comparison. The outcomes of the selected machine learning model are shown in Table 6, which also includes the proposed model (GA_RF) with the number of selected features (SF) from the data by genetic algorithm. Together with other metrics such as sensitivity, specificity, Matthew's correlation coefficient (MCC), training accuracy (ACC._train), and test accuracy (ACC._test) used to assess the performance of different machine learning models.

Table 6: Performance of ML models utilizing the GA feature selection module on balanced data.

MODEL	SF	ACC._TRAIN	ACC._TEST	SENSITIVITY	SPECIFICITY	MCC
NB	03	65.15	72.57	87.84	67.87	0.2562
LR	13	84.34	78.87	97.64	61.84	0.5408
DT	14	94.69	88.78	95.02	67.16	0.6453
KNN	09	84.59	87.88	90.92	71.08	0.5660
AdaBoost	14	91.64	86.98	95.68	72.21	0.6198
SVM_linear	14	83.88	72.57	95.88	79.63	0.4391
SVM_sigmoid	13	81.05	68.06	97.05	63.34	0.4182
SVM_poly	11	93.17	86.58	97.96	65.87	0.6624
SVM_RBF	11	92.67	90.71	99.52	72.83	0.7022
Proposed Model	10	91.22	95.09	98.50	81.27	0.8463

When comparing the test accuracy of various machine learning models that utilize the same feature selection module (GA), the newly constructed model (GA_RF) achieved the highest test accuracy of 95.09% by utilizing only 10

Dialogue Social Science Review (DSSR)
ISSN Online: 3007-3154
ISSN Print: 3007-3146
Vol. 2 No. 3 (October) (2024)
DIALOGUE SOCIAL SCIENCE REVIEW

features from the dataset. The decision tree is in the second spot with a test accuracy of 94.69% using only 14 features from the dataset, while in third place is the support vector machine with a polykernel, which obtained a test accuracy of 93.17% by using only 11 features from the dataset. Additionally, the effectiveness of the machine learning models is also assessed using the ROC curve. Figure 7 displays the ML models' performance using the AUC criterion on the balanced dataset and the GA feature selection module. SVM has the highest AUC of 91.70% among machine learning models.

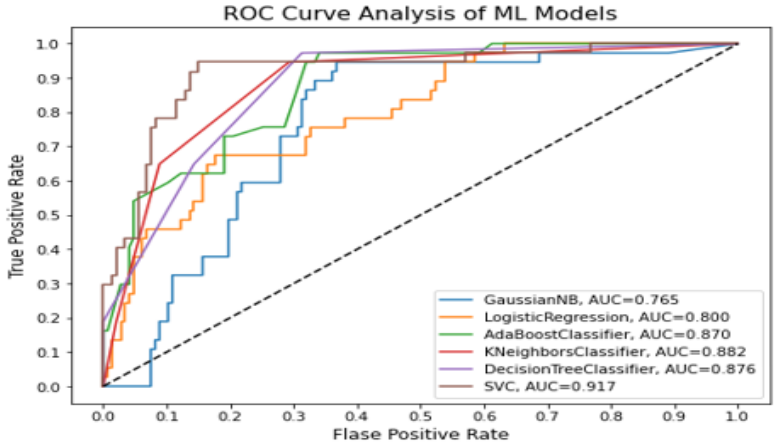

Figure 7. ROC curves analysis of ML models based on GA feature selection methods.

Conclusions

This study describes a machine learning-based system for predicting cardiac patient death by using medical health data. The proposed model is composed of two components: an RF model for mortality classification and a genetic algorithm (GA) for choosing important features. To balance the classes in the dataset, the random under sampling (RUS) method was used on the training data. Using just 10 features from the first module, the proposed model obtained the highest test accuracy of 95.09%. Additionally, the proposed model (GA_RF) performed better than other state-of-the-art machine learning models.

References

[1] Javed, R., Abbas, T., Khan, A.H. *et al.* Deep learning for lungs cancer detection: a review. *Artif Intell Rev* **57**, 197 (2024). https://doi.org/10.1007/s10462-024-10807-1

[2] W. Lipworth, "Real-world data to generate evidence about healthcare interventions: The application of an ethics framework for big data in health and research," *Asian Bioethics Review*, vol. 11, no. 3, pp. 289–298, 2019.

[3] R. Katarya and P. Srinivas, "Predicting heart disease at early stages using machine learning: A survey," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 302–305.

[4] Javeed A, Dallora AL, Sanmartin Berglund J, Ali A, Anderberg P, Ali L. Predicting dementia risk factors based on feature selection and neural networks. Computers, Materials and Continua. 2023;75(2):2491-508.

[5] W. Akbar *et al.*, "Development of hepatitis disease detection system by exploiting sparsity in linear support vector machine to improve strength of

adaboost ensemble model," *Mobile Information Systems*, vol. 2020, no. 1, p. 8870240, 2020.
[6] Ihnaini, B., Khan, M. A., Khan, T. A., Abbas, S., Daoud, M. S., Ahmad, M., & Khan, M. A. (2021). A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning. Computational Intelligence and Neuroscience, 2021(1), 4243700.
[7] A. Javeed *et al.*, "Early prediction of dementia using feature extraction battery (feb) and optimized support vector machine (svm) for classification," *Biomedicines*, vol. 11, no. 2, p. 439, 2023.
[8] A. Javeed, S. S. Rizvi, S. Zhou, R. Riaz, S. U. Khan, and S. J. Kwon, "Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification," *Mobile Information Systems*, vol. 2020, no. 1, p. 8843115, 2020.
[9] Sayyid Kamran Hussain, Sadaqat Ali Ramay, Tahir Abbas, Muhammad Kaleem, & Asif Khanzada. (2024). Ocular Diseases Detection Using Machine Learning, Deep Learning and Artificial Intelligence Based Techniques. *Journal of Computing & Biomedical Informatics*. Retrieved from https://jcbi.org/index.php/Main/article/view/657.
[10] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE access*, vol. 7, pp. 180235–180243, 2019.
[11] S. Sharma and M. Parmar, "Heart diseases prediction using deep learning neural network model," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 3, pp. 2244–2248, 2020.
[12] K. Patel *et al.*, "An Analysis of Epidemiological Factors in Heart Failure Outcomes," *Cureus*, vol. 14, no. 2, 2022.
[13] C. E. Hamo *et al.*, "OUTpatient intravenous LASix Trial in reducing hospitalization for acute decompensated heart failure (OUTLAST)," *Plos one*, vol. 16, no. 6, p. e0253014, 2021.
[14] J. I. Janjua, T. A. Khan and M. Nadeem, "Chest X-Ray Anomalous Object Detection and Classification Framework for Medical Diagnosis," 2022 International Conference on Information Networking (ICOIN), Jeju-si, Korea, Republic of, 2022, pp. 158-163, doi: 10.1109/ICOIN53446.2022.9687110.
[15] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics*, vol. 10, no. 5, p. 593, 2021.
[16] Abbas, T., Khan, A. H., Kanwal, K., Daud, A., Irfan, M., Bukhari, A., & Alharbey, R. (2024). IoMT-Based Healthcare Systems: A Review. Computer Systems Science & Engineering, 48(4)..
[17] M. A. Saleem, N. Thien Le, W. Asdornwised, S. Chaitusaney, A. Javeed, and W. Benjapolakul, "Sooty tern optimization algorithm-based deep learning model for diagnosing NSCLC tumours," *Sensors*, vol. 23, no. 4, p. 2147, 2023.
[18] A. Javeed, J. S. Berglund, A. L. Dallora, M. A. Saleem, and P. Anderberg, "Predictive power of XGBoost_BiLSTM model: a machine-learning approach for accurate sleep apnea detection using electronic health data," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 188, 2023.