International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 13, No. 2, August 2024, pp. 297~305
ISSN: 2252-8776, DOI: 10.11591/ijict.v13i2.pp297-305

Comparative analysis of heart failure prediction using machine learning models

Srinivas Kanakala, Vempaty Prashanthi
Department of CSE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

Article Info

Article history:

Received Feb 4, 2024
Revised May 1, 2024
Accepted May 12, 2024

Keywords:

Classification
Confusion matrix
Decision tree
K-nearest neighbor
Logistic regression
Machine learning
Naive Bayes
prediction
Random forest

ABSTRACT

Heart failure is a critical health problem worldwide, and its prediction is a major challenge in medical science. Machine learning has shown great potential in predicting heart failure by analyzing large amounts of medical data. Heart failure prediction with the help of machine learning classification algorithms involves the use of models such as decision trees, logistic regression, and support vector machines to identify and analyze potential risk factors for heart failure. By analyzing large datasets containing medical and lifestyle-related variables, these models can accurately predict the likelihood of heart failure occurrence in individuals. In our research, the heart failure prediction and comparison are done using logistic regression, K-nearest neighbor (KNN), support vector machines (SVM), decision tree and random forest The accurate identification of high-risk individuals enables early intervention and better management of heart failure, reducing the risk of mortality and morbidity associated with this condition. Overall, machine learning algorithms play a major role in improving the accuracy of heart failure risk assessment, allowing for more personalized and effective prevention and treatment strategies.

Corresponding Author:

Srinivas Kanakala
Department of CSE
VNR Vignana Jyothi Institute of Engineering and Technology
Hyderabad, Telangana, India
Email: srinivaskanakala@gmail.com

1. INTRODUCTION

Heart failure is a widespread and serious condition where the heart struggles to pump an adequate amount of blood to fulfill the body's requirements. It impacts millions globally and stands as a primary reason for hospital admissions and deaths. Recognizing and foreseeing heart failure promptly can significantly enhance patient results and decrease healthcare expenses. Over recent years, the utilization of machine learning (ML) methods has displayed potential in assisting with predicting and detecting heart failure at an early stage. By leveraging large datasets and powerful algorithms, ML algorithms can extract patterns and insights from patient data, facilitating accurate predictions and personalized treatment strategies.

Machine learning algorithms possess the capability to analyze extensive sets of patient data, encompassing medical records, imaging scans, genetic profiles, and lifestyle details. These algorithms can recognize hidden patterns and relations between the data that may not be evident to human observers. By considering multiple variables simultaneously, ML algorithms can create predictive models that estimate an individual's risk of developing heart failure or worsening of their condition over time.

One notable benefit of machine learning in predicting heart failure is its capacity to manage intricate, non-linear connections among various risk factors.. Traditional statistical models often assume

linearity, limiting their effectiveness in capturing the intricate interactions that contribute to heart failure. ML algorithms, on the other hand, can model these intricate relationships, incorporating multiple variables and their interactions to provide more accurate predictions. This capability allows for a more complete understanding of the underlying risk factors and aids in the development of personalized treatment plans.

In their analysis [1], data mining techniques such as K-nearest neighbor, T3 Algorithm, and the C4.5 decision tree were employed using the WEKA tool. Various attributes including heredity, age, gender, race, as well as certain medical conditions like hypertension, high cholesterol levels, cardiovascular disease, diabetes mellitus, obesity, and history of stroke were utilized. The C4.5 decision tree algorithm demonstrated an accuracy of 95.42%, while K-nearest neighbor achieved 94.18% accuracy in predicting stroke. Sudha *et al.* [2] suggests the use of Decision Trees, Naive Bayes, and Neural Networks as predictive classification algorithms to spot the existence of disease, along with Principal Component Analysis (PCA) to reduce the number of attributes. The steps that are followed in this paper are data collection, preprocessing, dimensionality reduction, feature subset selection. After that, a suitable classification algorithm such as decision trees or neural networks can be chosen to predict outcomes based on these features. Finally, the model's performance is analyzed. Maintaining a healthy lifestyle and early detection are crucial measures for preventing heart strokes. Amini *et al.* [3], researchers utilize a heart stroke dataset sourced from Kaggle. Their study aims to forecast the likelihood of heart strokes and categorize patients' risk levels by employing various data mining techniques like K-nearest neighbors (KNN), decision tree, and random forest. Shareefunnisa *et al.* [4] collected data from Kaggle and pre-processed it by label encoding, handling missing values, and splitting it into train and test sets. Using logistic regression, decision tree, random forest, k-nearest neighbors, support vector machine, and naïve Bayes, they calculated accuracies based on precision, recall, and F1 scores. Naïve Bayes outperformed the other algorithms, achieving an accuracy of 82%. Its success demonstrates its effectiveness for classification tasks and suggests its potential for future applications. In [5]-[8] machine learning algorithms were applied for haert disesase prediction, brain hemography and covid 19 spread analysis detection were done. In [9], [10] authors discussed how machine learning algorithms are used for student drop out detection and music catogorization. In [11], [12] authors discussed leaf diseses detection using machine learming. Heart failure decomposition study is done by [13]. Artificial Intelligence is the latest area where reaserchers are now showing intrest where AI based heart diseases detection is implemented in [14]-[16]. In [17]-[21] heart disease detection is done with the implementation of hybrid models in machine learning. Taking real time parameters in to considerartions [22]-[24] has designed models for heart failure detection.

Interpreting medical data manually is prone to human error and mistakes, which can significantly impact patient outcomes. Additionally, the process is time-consuming, resulting in delays in diagnosis and treatment initiation. The limited capacity of healthcare professionals to process and analyze huge volumes of data poses a risk of overlooking critical patterns or trends that could inform accurate diagnoses. Furthermore, subjectivity and biases inherent in the human interpretation of medical information can introduce inconsistencies and inaccuracies in the diagnostic process. Consequently, there is a need for advanced technologies and tools to assist healthcare providers in data analysis and interpretation, thereby minimizing errors, improving efficiency, and enhancing patient care.

2. PROPOSED METHOD

In the context of heart failure prediction project, data plays a critical role in identifying patterns that can be utilized to forecast outcomes. To train our model, we provide a CSV file containing a dataset comprising 9,921 rows and 12 columns. Data cleaning is a vital step in making data for analysis. It involves getting rid of irrelevant or incorrect information within a dataset. By removing unnecessary data and reducing duplicates, the dataset becomes more streamlined. Inconsistencies are identified and resolved by validating against known factors. It is important to uphold strict data quality measures during the data import process. Typos can be corrected, and missing regions can be filled in using precise algorithms. Data cleaning ensures high-quality data, which ultimately leads to more accurate analysis and improved performance of models. In the context of classification, a machine learning model refers to a mathematical representation derived from the training process. It involves studying various algorithms that can automatically improve and build the model based on past data and experiences. Like computer software, a classification machine learning model is designed to identify patterns or behaviors using prior data. The learning algorithm analyzes the training data to uncover patterns, which are then encapsulated within the ML model. This model can now be utilized to classify and predict the outcomes of new data depending upon the learned patterns. In our reserach we compared various machine learning models to identify the best approach for our analysis. We evaluated the working of different models, including Naive Bayes, logistic regression, K-nearest neighbors (KNN), support vector machines (SVM), Decision Tree, and Random Forest. Each model offers unique characteristics and

algorithms for learning patterns and making predictions. By comparing these models, we aimed to determine which one achieved the highest accuracy and provided the most reliable results for our specific project requirements. Naive Bayes: This algorithm is depended on Bayes' theorem where it considers all features as independent once. It's majorly utilized for the tasks such as classification which is simple and efficient. One important feature of Naive Bayes is that it works well for large datasets.

Logistic regression: Logistic regression is a extensively utilized for classification algorithms. Inthis algorithm logistic function is used where the relationship among input variables and the possibility of a specific outcome is calculated. It works well for dependent variable when it is binary or categorical.

K-nearest neighbors (KNN): KNN is parametric free model which is utilized for both classification as well as regression tasks. A new data point is given to a class depending upon the larger vote of its k nearest neighbors in feature space. KNN is simple to understand and implement but can be expensive for large datasets in terms of execution.

Support vector machines (SVM): SVM is a powerful model utilized for both classification and regression. It creates a hyperplane in feature space which maximally divides distinct classes. SVM is effective in operating complex datasets and can manage high-dimensional spaces.

Decision tree: Decision trees iteratively separate the dataset depending upon on different features, forming a tree-like structure for decision-making. Each internal node indicates a test on a feature, while the leaves indicates class labels. Decision trees are interpretable, at ease to understand, and can operate both classification and regression tasks.

Random forest: Random Forest follows an ensemble learning methodology where more than one decision tress are combined for creating predictions. It creates a set of diverse decision trees and groups their predictions to achieve better accuracy and robustness. Random Forest handles high-dimensional data well and is less prone to overfitting.

The system lets health care experts to input comprehensive patient data, which include demographic information (age, gender), medical history (previous heart conditions, hypertension, diabetes), lifestyle factors (smoking, alcohol consumption, physical activity), and clinical measurements (blood pressure, cholesterol levels, ejection fraction). This data forms the basis for accurate risk assessment. The system employs sophisticated algorithms and machine learning models to evaluate the patient's data and calculate their risk of heart failure. The risk assessment process considers a wide range of factors and their interactions to generate a personalized risk score. These factors may include age, family history, biomarker levels, comorbidities, and lifestyle habits. The system serves as a effective decision support tool for health care experts. It provides evidence-based insights, guidelines, and clinical best practices to aid in making informed decisions regarding heart failure prevention, diagnosis, and treatment. Overall, the Heart Failure Prediction System empowers healthcare providers with comprehensive tools to identify high-risk patients, implement preventive measures, and deliver personalized care, ultimately improving heart failure outcomes and patient well-being. Maintaining large datasets becomes easier with the use of advanced data management techniques. These techniques allow for efficient storage, organization, and retrieval of data, making it easier to handle and update large volumes of information. Compared to traditional methods, these techniques often require less time and effort, as automation and optimized algorithms streamline data processes. With reduced manual intervention, human effort is minimized, enabling more efficient handling of large datasets in a timely manner.

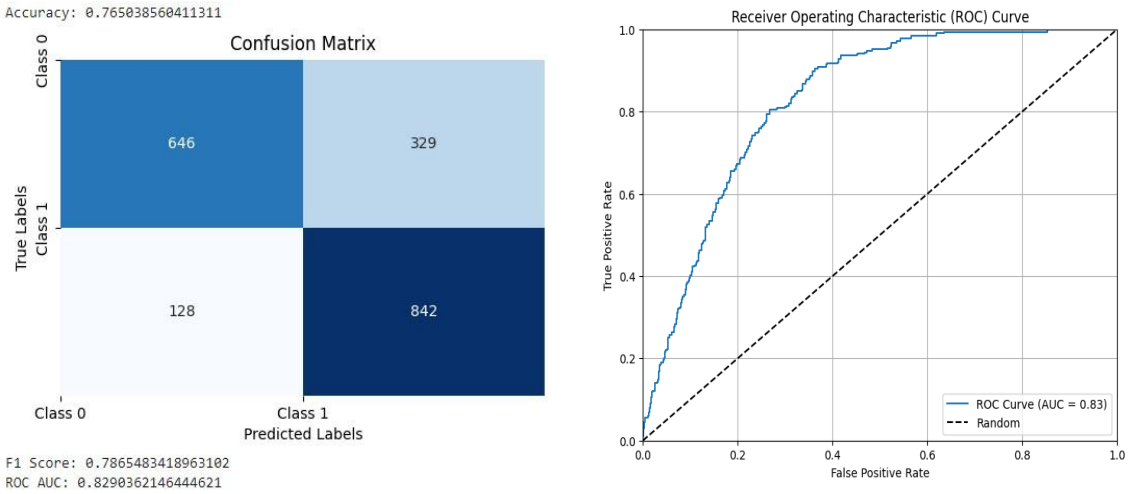
3. RESEARCH METHOD

The heart failure dataset is a inclusive collection of clinical and demographic features of patients, aimed at predicting the likelihood of heart failure occurrence. We have data consisting of 9722 rows and 12 columns, each column has its own significance to classify the patients if they are or not on the verge of suffering from heart failure. Initially we imported all libraries which are required like pandas, matplotlib seaborn etc. Data set is in the CSV format. Data consists of both categorical and numerical data. Some columns have null values which as to be preprocessed Next, we are separating the stroke column from the data frame and storing it in a variable called label. Further Data preprocessing is done on the copy of data that was created in the previous step. Missing values of numerical variables are computed using imputation and categorical variables are converted into numer variables utilising one hot encoding. Missing values are imputed ("Unknown" is replaced with "Never Smoked" which is the mode of the column). Initially, the dataset was highly imbalanced which was balanced using oversampling. Next The data is split into two–training data and test data. These two are used to train the models in the subsequent steps. Next it is applied on different classification models as shown below:

3.1. Naive Bayes

Training the Naive Bayes model on balanced dataset and calculating the correctness. Plotting heatmap for confusion matrix , calculating F1 score and ROC AUC is shown in Figure 1. The confusion matrix above in Figure 1(a) shows a pictorial representation of Naive Bayes' performance. The four quadrants display counts

for true positives (842), false positives (329), true negatives (646), and false negatives (128). The ROC AUC curve in Figure 1(b) measures a Naive Bayes model's capability to discriminate between positive and negative classes, showing the area beneath the curve i.e 0.83 units.

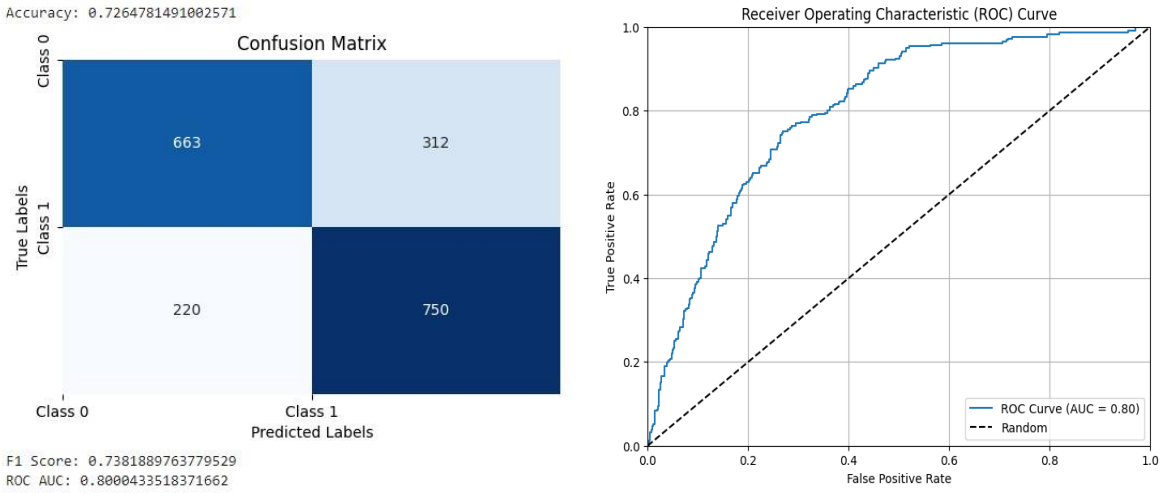


(a) (b)

Figure 1. Navie Bayes (a) Heatmap of confusion matrix and (b) AUC

3.2. Logistic regression

Training the logistic regression model on balanced dataset and calculating the accuracy are shown in Figure 2. The confusion matrix above in Figure 2(a) shows pictorial form of Logistic regression's performance. The four quadrants display the counts for true positive (750), false positive (312), true negative (663), and false negative (220). The ROC AUC curve Figure 2(b) measures a logistic regression model's capability to discriminate between positive and negative classes, showing the area beneath the curve i.e 0.80 units.



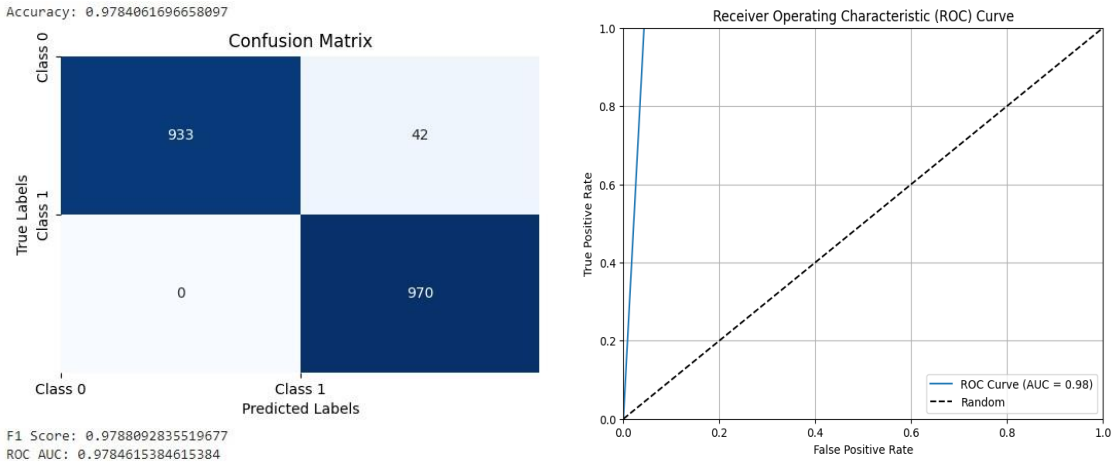
(a) (b)

Figure 2. Logistic regression (a) Heatmap of confusion matrix and (b) AUC

3.3. K-nearest neighbor

Training the K-nearest neighbor model on balanced dataset and calculating the accuracy are shown in Figure 3. The confusion matrix above in Figure 3(a) shows a pictorial representation of KNN's performance. The four quadrants display the counts for true positive (970), false positives (42), true negatives

(933), and false negatives (0). The ROC AUC curve above in Figure 3(b) measures a KNN model's capability to discriminate between positive and negative classes, showing the area beneath the curve i.e 0.98 units.

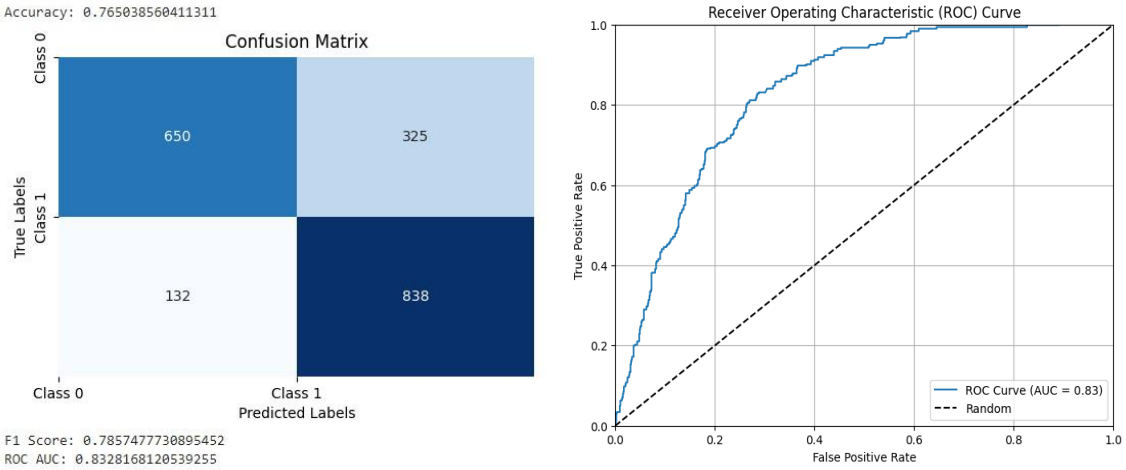


(a) (b)

Figure 3. K-nearest neighbor (a) Heatmap of confusion matrix and (b) AUC

3.4. Support vector machine (Linear)

Training the SVM (linear) model on balanced dataset and calculating the accuracy are shown in Figure 4. The confusion matrix shown below in Figure 4(a) gives a representation of SVM's performance. The four quadrants display the counts for true positive (838), false positives (325), true negatives (650), and false negatives (132). The ROC AUC curve shown below in Figure 4(b) measures a SVM model's capability to discriminate between positive and negative classes, showing the area beneath the curve i.e 0.83 units.

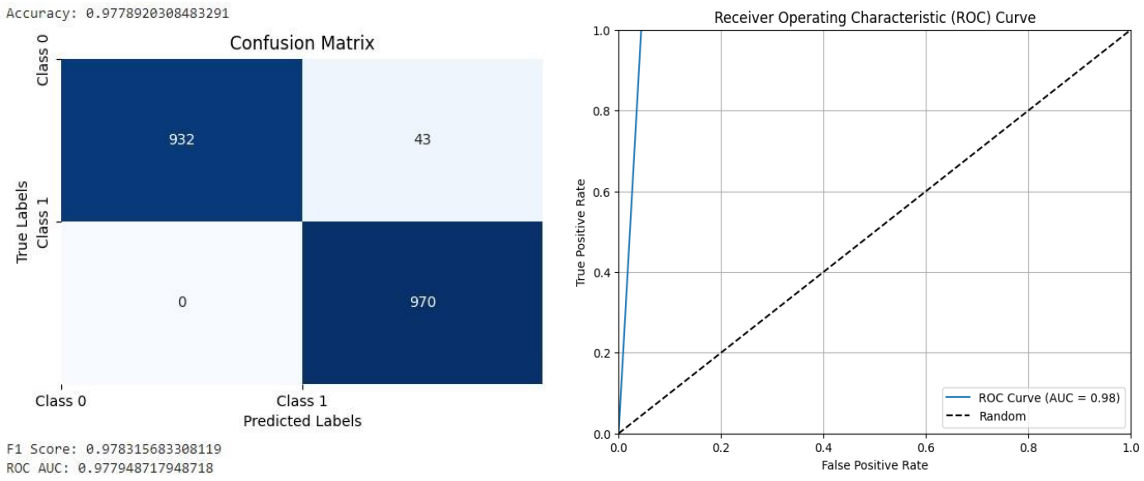


(a) (b)

Figure 4. Support vector machine (a) Heatmap of confusion matrix and (b) AUC

3.5. Decision tree

Training dataset on decision Tree model and their results are shown in Figure 5. The confusion matrix above in Figure 5(a) shows representation of Decision Tree's performance. The four quadrants display the counts for true positive (970), false positive (43), true negatives (932), and false negatives (0). The ROC AUC curve shown in Figure 5(b) measures a Decision Tree model's capaability to discriminate between positive and negative classes, showing the area beneath the curve i.e 0.98 units.

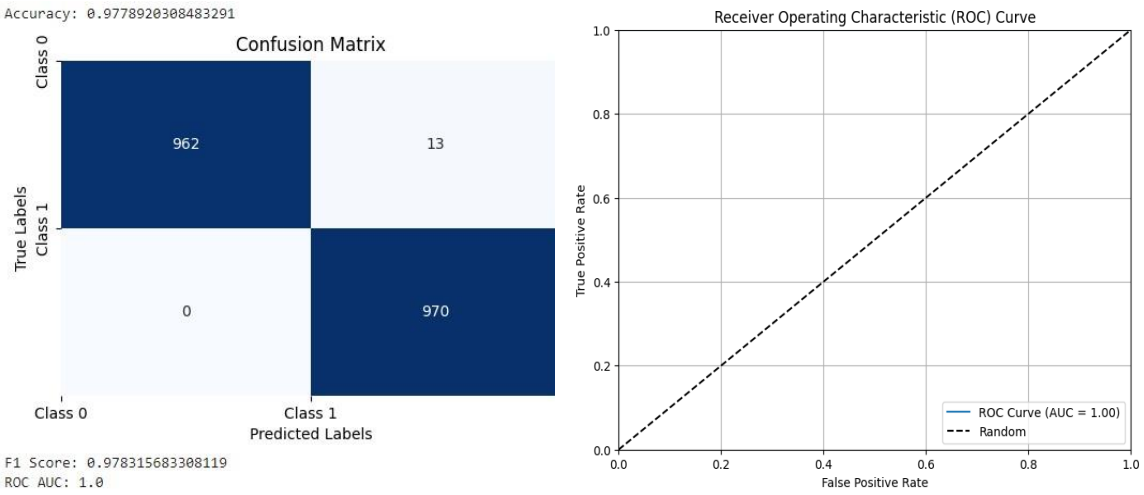


(a) (b)

Figure 5. Decision tree (a) Heatmap of confusion matrix and (b) AUC

3.6. Random forest classifier

Training the random forest model on balanced dataset and calculating the accuracy is shown in Figure 6. The confusion matrix below in Figure 6(a) shows a visual representation of Random forest's performance. The four quadrants display the counts for true positives (970), false positives (13), true negatives (962), and false negatives (0). The ROC AUC curve Figure 6(b) measures a Random forest model's capaability to discriminate between positive and negative classes, showing the area under the curve i.e 1 unit.



(a) (b)

Figure 6. Random forest (a) Heatmap of confusion matrix and (b) AUC

4. RESULTS AND DISCUSSION

Accuracy measures the perfection of predictions, representing the percentage of correct classifications made by a model or algorithm. Figure 7 shows plotting of bar graph to compare accuracies of various models. Higher Accuracy values indicate better model performance. F1-score is a metric that combines precision and recall, providing a balanced measure of a model's performance by considering both false positives and false negatives. Higher F1-score values indicate better model performance. Figure 8 shows Plotting bar graph to compare F1 scores of various models. The Area under the ROC Curve (AUC) is a performance measure for classification models, measuring their ability to differentiate between classes.

Higher AUC values indicate good model performance as shown in Figure 9. Based on the evaluation metrics provided, we can draw some conclusions about the performance of different classification models for the heart failure prediction task. Naïve Bayes achieved an accurateness of 0.765, an F1 score of 0.787, and an ROC AUC of 0.829. Logistic regression performed slightly lower with an exactness of 0.726, an F1 score of 0.738, and an ROC AUC of 0.800. KNN demonstrated excellent performance with an accuracy of 0.978, an F1 score of 0.979, and an ROC AUC of 0.978. SVM showed similar results to Naïve Bayes, with an accurateness of 0.765, an F1 score of 0.786, and an ROC AUC of 0.833. Decision Tree and Random Forest both achieved high accuracy and F1 scores, with Decision Tree achieving an ROC AUC of 0.978 and Random Forest achieving a perfect ROC AUC of 1.0. In conclusion, KNN, decision tree, and random forest demonstrated superior performance across the evaluated metrics, exhibiting high accuracy, F1 scores, and ROC AUC values. Naïve Bayes and SVM showed comparable results in accuracy and F1 score, but SVM had a slightly higher ROC AUC. Logistic regression performed slightly lower compared to the other models. These findings suggest that KNN, decision tree, and random forest are promising models for heart implementation in practical healthcare settings.

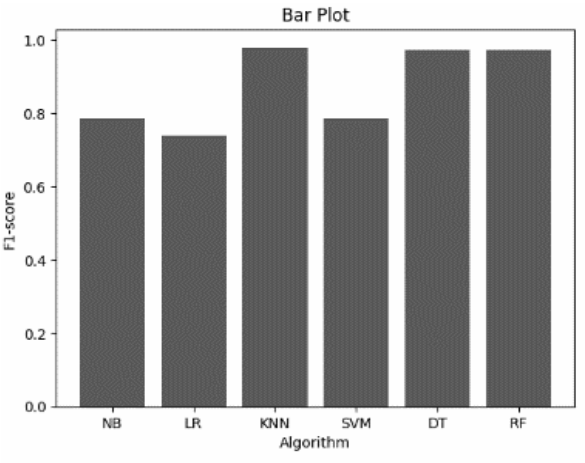


Figure 7. Accuracies with KNN, DT, RF

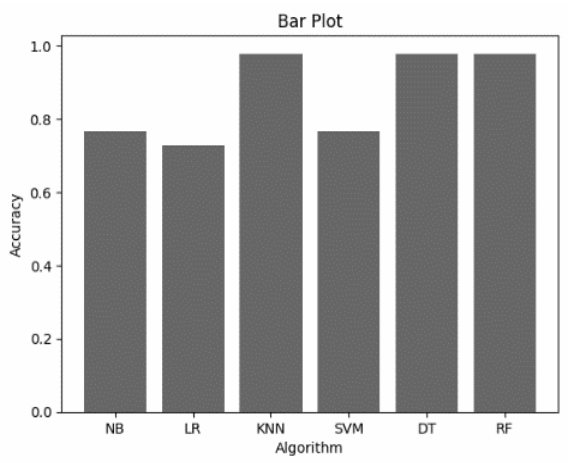


Figure 8. F1 scores with RF

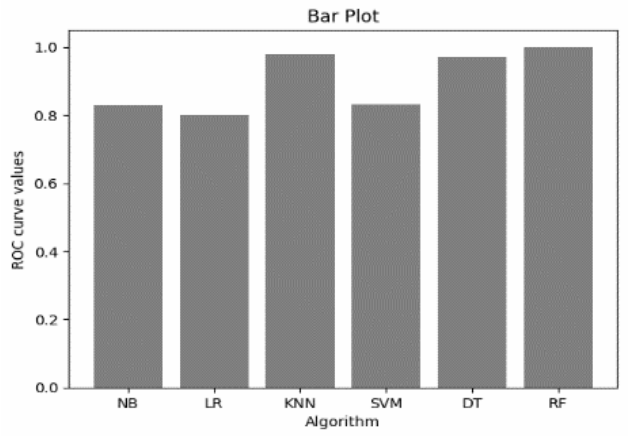


Figure 9. Plot of AUCs with RF

5. CONCLUSION

In conclusion, this machine learning project aimed to predict heart failure using various classification algorithms such as Naive Bayes, logistic regression, K-nearest neighbors (KNN), support vector machines (SVM), Decision Trees, and Random Forest. The execution of each algorithm was evaluated depending on their accuracy, F1 score and area under the Receiver Operating Characteristic (ROC) curve. After careful analysis, it was found that decision trees and random forest outperformed the other algorithms

in predicting heart failure. Random forest demonstrated the highest accuracy and the highest area under the ROC curve, indicating its effectiveness in differentiating the positive and negative instances of heart failure. Overall, this machine learning project underscores the significance of selecting right algorithms for predictive tasks in healthcare. By leveraging the power of random forest, we can improve heart failure prediction accuracy and contribute to early intervention and better patient outcomes.

To enhance the practical application and usability of the heart failure prediction model built using Random Forest, the development of a user interface (UI) can be considered as the next step. A user-friendly interface will allow healthcare professionals and researchers to interact with the model more easily and obtain predictions efficiently. The UI can provide an intuitive input form where users can enter relevant patient data such as age, gender, medical history, and vital signs. Upon submission, the model will process the inputs and generate the prediction results, displaying the probability of heart failure for the given patient.

To further validate and generalize the heart failure prediction model, it is crucial to test its performance on diverse and larger datasets. Obtaining access to real-world clinical data can give valuable insights into the model's applicability and robustness. Collaborations with healthcare institutions and data-sharing initiatives can facilitate the acquisition of comprehensive and diverse datasets, allowing the model to be trained and validated on more representative samples. By expanding the dataset, the model can capture a wider range of patient characteristics, thus enhancing its predictive capabilities and enabling its deployment in real-world scenarios.

REFERENCES

[1] G. Sailasya and G. L. A. Kumari, "Analyzing the performance of stroke prediction using ML classification algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120662.

[2] A. Sudha, P. Gayathri, and N. Jaisankar, "Effective analysis and predictive model of stroke disease using classification methods," *International Journal of Computer Applications*, vol. 43, no. 14, pp. 26–31, 2012, doi: 10.5120/6172-8599.

[3] L. Amini *et al.*, "Prediction and control of stroke by data mining," *International Journal of Preventive Medicine*, vol. 4, pp. S245–S249, 2013.

[4] S. Shareefunnisa, S. N. L. Malluvalasa, T. R. Rajesh, and M. Bhargavi, "Heart stroke prediction using machine learning," *Journal of Pharmaceutical Negative Results*, vol. 13, no. 5 (Special Issue), pp. 2551–2558, 2022, doi: 10.47750/pnr.2022.13.S05.395.

[5] V. Prashanthi, S. Kanakala, D. Borgaonkar, and D. S. Babu, "Rainfall prediction using catboost machine learning algorithm," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, IEEE, Sep. 2023, pp. 1–5. doi: 10.1109/NMITCON58196.2023.10276336.

[6] O. S. Priya, K. Srinivas, and S. Yeruva, "Heart failure prediction using classification methods," in *Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems*, A. B. Reddy, B. V. Kiranmayee, R. R. Mukkamala, and K. Srujan Raju, Eds., in Algorithms for Intelligent Systems. Singapore: Springer Nature Singapore, 2022, pp. 545–553. doi: 10.1007/978-981-16-7389-4.

[7] S. Kanakala and V. Prashanthi, "Covid-19 Spread Analysis". In: Satapathy, S.C., Bhateja, V., Favorskaya, M.N., Adilakshmi, T. (eds) *Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies,* Springer, Singapore vol 224, 2021, doi: 10.1007/978-981-16-1502-3_5

[8] K. V. Sharada, V. Prashanthi, and S. Kanakala, "Detection and classification of intracranial brain hemorrhage," in *Lecture Notes in Networks and Systems*, vol. 244, 2022, pp. 455–464. doi: 10.1007/978-981-16-2919-8_41.

[9] H. Dasi and S. Kanakala, "Student dropout prediction using machine learning techniques," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 4, pp. 408–414, 2022.

[10] V. Prashanthi, S. Kanakala, V. Akila, and A. Harshavardhan, "Music genre categorization using machine learning algorithms," in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, IEEE, Nov. 2021, pp. 1–4. doi: 10.1109/ICCICA52458.2021.9697137.

[11] C. A. Sarma, S. Inthiyaz, B. T. P. Madhav, and P. S. Lakshmi, "An Inter digital- poison ivy leaf shaped filtenna with multiple defects in ground for s-band bandwidth applications," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 8, pp. 55–66, Aug. 2022, doi: 10.17762/ijritcc.v10i8.5668.

[12] D. R. Gill, "A study of framework of behavioural driven development: methodologies, advantages, and challenges," *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol. 8, no. 2, pp. 09–12, Jun. 2022, doi: 10.17762/ijfrcsce.v8i2.2068.

[13] J. Henriques, *et al.*, "Prediction of heart failure decompensation events by trend analysis of telemonitoring data," IEEE J. Biomed. Health Informat., vol. 19, no. 5, pp. 1757–1769, Sep. 2014, doi: 10.3390/s21113575

[14] G. Yang *et al.*, "Homecare robotic systems for healthcare 4.0: visions and enabling technologies," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2535–2549, Sep. 2020, doi: 10.1109/JBHI.2020.2990529.

[15] X. Yuan, J. Chen, K. Zhang, Y. Wu, and T. Yang, "A stable AI-based binary and multiple class heart disease prediction model for IoMT," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2032–2040, Mar. 2022, doi: 10.1109/TII.2021.3098306.

[16] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Gheni, "Improving the heart disease detection and patients' survival using supervised infinite feature selection and improved weighted random forest," *IEEE Access*, vol. 10, pp. 67363–67372, 2022, doi: 10.1109/ACCESS.2022.3185129.

[17] K. A. L. Hernandez, T. Rienmüller, D. Baumgartner, and C. Baumgartner, "Deep learning in spatiotemporal cardiac imaging: A review of methodologies and clinical usability," *Computers in Biology and Medicine*, vol. 130, March 2021, doi: 10.1016/j.compbiomed.2020.104200.

[18] S. Narula, K. Shameer, A. M. S. Omar, J. T. Dudley, and P. P. Sengupta, Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. J*ournal of the American College of Cardiology*, vol. 68, no. 21, pp. 2287-2295, 2016, doi: 10.1016/j.jacc.2016.08.062.

[19] Y. Wang, X. Ge, H. Ma, S. Qi, G. Zhang, and Y. Yao, "Deep learning in medical ultrasound image analysis: a review," *IEEE Access*, vol. 9, pp. 54310–54324, 2021, doi: 10.1109/ACCESS.2021.3071301.
[20] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
[21] C.-A. Cheng and H.-W. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, Jul. 2017, pp. 2566–2569. doi: 10.1109/EMBC.2017.8037381.
[22] M. T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, "Heart disease risk prediction using deep learning techniques with feature augmentation," *Multimedia Tools and Applications*, vol. 82, no. 20, pp. 31759–31773, Aug. 2023, doi: 10.1007/s11042-023-14817-z.
[23] K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, and T. Suryawanshi, "Human disease prediction using machine learning techniques and real-life parameters," *International Journal of Engineering, Transactions B: Applications*, vol. 36, no. 6, pp. 1092–1098, 2023, doi: 10.5829/ije.2023.36.06c.07.
[24] N. Sinha, M. A. G. Kumar, A. M. Joshi, and L. R. Cenkeramaddi, "DASMcC: Data augmented SMOTE multi-class classifier for prediction of cardiovascular diseases using time series features," *IEEE Access*, vol. 11, pp. 117643–117655, 2023, doi: 10.1109/ACCESS.2023.3325705.

BIOGRAPHIES OF AUTHORS

Srinivas Kanakala working as Assistant Professor, Department of CSE, VNRVJIET, Hyderabad. He completed Ph.D. from Osmania University, Hyderabad in the year 2018 and has 19 years of academic and research experience. Prior to PhD, he had earned M.Tech in CSE from JNTUH in the year 2008 and B.Tech in computer science and information technology from JNTUH in the year 2004. His research interests include developing algorithms and models for building systems and applications in areas of data mining, machine learning and big-data analytics. He has around 30 publications, in various international journals and conferences. He can be contacted at email: srinivaskanakala@gmail.com.

Vempaty Prashanthi working as Senior Assistant Professor, Department of CSE, VNRVJIET, Hyderabad. She completed Ph.D. from JNTU, Hyderabad in the year 2020 and has 18 years of academic and research experience. Her Ph.D. work was on Network Coding Aware Routing Techniques in Mobile Ad hoc Networks. Prior to PhD, she had earned M.Tech in CSE from JNTUH in the year 2010 and B.Tech in Computer Science and Information Technology from JNTUH in the year 2005.V.Prashanthi's research interests include developing algorithms and models for building systems and applications in areas of data mining, machine learning and big-data analytics. She has around 30 publications, in various journals and conferences. She can be contacted at email: prashuvempaty@gmail.com.