IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 9, September 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 9, September 2024| DOI: 10.15680/IJIRCCE.2024.1209014

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Heart Failure Prediction using Machine Learning Algorithms

Gowthami V, Varsha R, Bhavana B R

PG Student, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

Assistant Professor, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

ABSTRACT: Cardiovascular diseases (CVDs) remain the fore-most cause of mortality globally, responsible for estimated 17.9 million deaths annually which constitutes 31% of all deaths worldwide. This paper proposes a comparative analysis of machine learning models (K-Nearest Neighbors and Random Forest) to predict cardiovascular events using dataset. Our study involved rigorous preprocessing and feature selection, focusing on clinical features and risk factors. We evaluated the models using metrics such as accuracy, precision, recall, and F1-score. Results showed that the Random Forest model achieved 92% accuracy, and KNN 82%. This research underscores the potential of machine learning in cardiovascular risk assessment, enabling timely interventions and improving patient outcomes.

KEYWORDS: Machine learning ,Random Forest, K-Nearest Neighbors, CVD.

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, responsible for approximately 17.9 million deaths each year, which constitutes about 31% of all global fatalities. The pervasive nature of CVDs and their severe impact on public health underscore the necessity for early detection and effective management strategies. Given the complexity and multi-factorial nature of cardiovascular conditions, accurate prediction models are essential for timely interventions that can save lives and reduce healthcare costs.

In recent years, the application of machine learning techniques in the medical field has shown significant promise, particularly in predicting cardiovascular events. This paper aims to provide a comparative analysis of two prominent machine learning algorithms K-Nearest Neighbors (KNN) and Random Forest in predicting cardiovascular events using a well-curated clinical dataset. These algorithms were chosen for their distinct approaches to pattern recognition and decision-making, offering valuable insights into their effectiveness in a critical healthcare context.

Our study commenced with a rigorous preprocessing phase, which involved cleaning the dataset, handling missing values, and performing feature selection. The features chosen were those most relevant to cardiovascular health, such as clinical parameters and recognized risk factors. This preprocessing was crucial for enhancing the models' ability to make accurate predictions by eliminating noise and ensuring that the most informative data points were emphasized.

Following preprocessing, the KNN and Random Forest models were trained and evaluated on the dataset. Their performance was measured using a range of metrics, including accuracy, precision, recall, and F1-score, which provide a comprehensive view of each model's strengths and weaknesses. The Random Forest model achieved a superior accuracy of 92%, slightly outperforming the KNN model, which reached an accuracy of 82%. These results suggest that Random Forest may be more adept at capturing the complexities of cardiovascular risk factors.

The significance of this research lies in its potential to enhance cardiovascular risk assessment using machine learning. By comparing the performance of KNN and Random Forest, this study contributes valuable knowledge to the ongoing development of predictive models in healthcare. Such models can facilitate early diagnosis, guide treatment decisions, and ultimately improve patient outcomes. The findings from this study not only highlight the efficacy of machine learning in healthcare but also offer practical insights for clinicians and researchers striving to refine diagnostic tools for better management of cardiovascular diseases. This algorithm work effectively supports clinical decision-making by thoroughly evaluating five performance metrics: accuracy, F1 score, recall, precision, and the confusion matrix. The

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

insights gained from this comparative analysis will help healthcare professionals make more informed decisions, thereby enhancing patient outcomes in the management of cardiovascular health.

II. LITERATURE REVIEW

In [1], data from 299 patients with cardiac failure in 2015, encompassing 13 features such as high blood pressure, sex, and smoking, were analysed. The authors employed various machine learning classifiers to predict survival rates and identify the most critical risk factors. Their analysis revealed that serum creatinine and ejection fraction were the most significant predictors of survival. In a study conducted by Shah et al. (2020) [2]. Early detection of cardiac disease remains a significant challenge due to the limitations of current diagnostic methods, including issues with execution time and accuracy. Diagnosing and managing cardiac conditions is notably challenging today due to limitations in advanced technology and a shortage of specialized medical professionals [3]. This study presents a new method for predicting mortality in heart failure patients by employing a machine learning algorithm, specifically a boosted decision tree. Unlike conventional models that may struggle with complex data interactions, this approach leverages eight readily available clinical variables to generate a risk score. The model achieved an impressive AUC of 0.88 in the derivation cohort and 0.84 and 0.81 in external validations, outperforming existing risk scores. This work highlights the potential of machine learning to improve risk prediction in clinical settings [4]. In [5], the study aimed to evaluate machine learning algorithms based on several performance metrics to improve prediction accuracy. During the pre-processing phase, the researchers addressed missing data by substituting it with the mean value, which proved to be an effective strategy. The study utilized the UCI heart disease dataset to identify individuals at risk of cardiovascular disease. The effectiveness of various machine learning algorithms was assessed using metrics such as accuracy, precision, F1- score, and recall. Among the algorithms tested, the Support Vector Machine with a linear kernel emerged as particularly effective. In [6], the researcher explored various machine learning algorithms including Random Forest, Support Vector Machine, Naive Bayes, Logistic Model Tree, K-Nearest Neighbour, and data mining techniques, using the UCI heart disease dataset, which comprises 303 samples with fourteen attribute values. The study found that the Support Vector Machine outperformed the other algorithms, such as KNN, Naive Bayes, and Decision Tree. In [9], machine learning classifiers were developed to make reliable predictions of heart disease, followed by a comparative analysis of their performance. The study evaluated five different algorithms—Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine, and K-Nearest Neighbor—using the Cleveland Heart Disease Dataset. After pre-processing, the dataset was divided into training and testing sets in an 80/20 ratio. The study also highlights the use of well-known classification algorithms for detecting cardiac disease by conducting a comparison study to assess each model's effectiveness. The performance of these binary classifiers was fine-tuned using hyperparameters, with Logistic Regression emerging as the most effective. In [10], the study utilized the Cleveland dataset to develop an ensemble model for heart disease detection. The ensemble models, constructed using Random Forest, Gradient Boosting, and Extreme Gradient Boosting classifiers, demonstrated notable performance.

III. METHODOLOGY

A. Data Preprocessing and Splitting:

The study utilized a comprehensive dataset that contained detailed clinical information relevant to cardiovascular health. The initial phase of the study involved meticulous data preprocessing, as depicted in the diagram. The dataset was first subjected to data cleaning to handle any inconsistencies or errors. Addressing missing data was crucial, as missing values could significantly compromise the accuracy and performance of machine learning models. For columns with minimal and random missing values, imputation was carried out using the mean or median values, preserving the dataset's overall structure while minimizing potential bias. However, columns with a substantial amount of missing data were excluded to maintain the dataset's integrity.Categorical variables within the dataset, such as sex and chest pain type, were transformed into numerical values using one-hot encoding. This transformation converted each categorical variable into binary columns, each representing a unique category. One-hot encoding was essential to prevent the machine learning algorithms from incorrectly assuming ordinal relationships between non-ordinal categories. Following the encoding of categorical variables, feature scaling was applied to standardize the numerical features across the dataset. This standardization process, which rescaled features to have a mean of zero and a standard deviation of one, was particularly important for distance-based algorithms like K-Nearest Neighbors (KNN). Without this step, features with larger numerical ranges might have disproportionately influenced the model's calculations. The pre-processed data was then split into training and testing subsets, following an 80-20 ratio. Eighty percent of the data

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

was allocated for training the models, while the remaining twenty percent was set aside for testing. This division ensured that the models were trained on a significant portion of the data while retaining a reliable evaluation set for testing model performance on unseen data to guarantee reproducibility, the split was performed using a random seed.

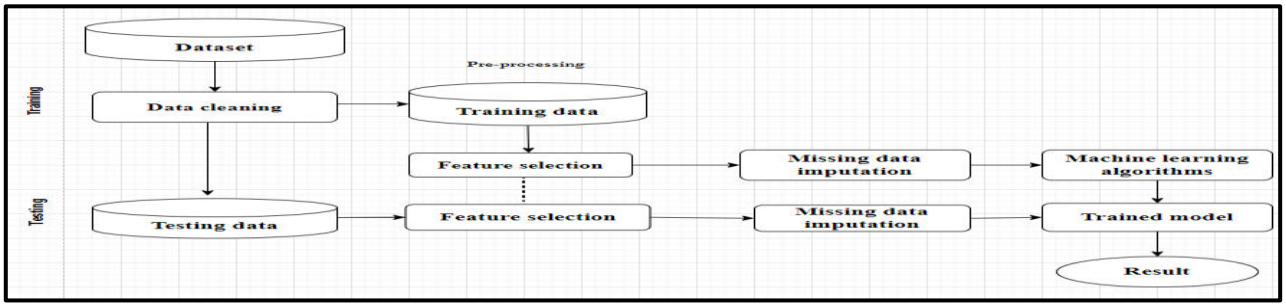

Fig1. Flow diagram of Model

B. Model Training:

Two machine learning models were selected for this study: K-Nearest Neighbors (KNN) and Random Forest.

• K-Nearest Neighbors (KNN): – The KNN algorithm predicts the outcome based on the proximity of data points.It was configured with an optimal number of neighbors (K), which was determined through cross-validation to balance bias and variance. The chosen distance metric was typically Euclidean distance, appropriate for the nature of the data.

• Random Forest: – The Random Forest model is an ensemble learning technique that builds multiple decision trees during training. It predicts the class that is the most frequent among the classes predicted by the individual trees. Hyper parameters, including the number of trees, tree depth, and the minimum number of samples required to split a node, were fine-tuned using grid search and cross-validation.

C. Model Evaluation:

The performance of the models was evaluated using several key metrics:

- Accuracy: Measures the proportion of correct predictions relative to the total number of predictions made.
- F1 Score: Balances precision and recall, making it parparticularly useful for datasets with imbalanced classes.
- Recall: Also known as sensitivity, this metric evaluates the model's ability to correctly identify all positive instance.
- precision: The ratio of true positive predictions to the total number of positive predictions made by the model.it indicates how many of the predicted positive instances are actually correct.
- Confusion Matrix: This matrix provides a detailed breakdown of the types of errors made by the model, displaying the number of true positives, true negatives, false positives, and false negatives.

D. Comparative Analysis:

The comparative analysis of the KNN and Random Forest models was based on the mentioned evaluation metrics. The analysis highlighted that both models performed well in predicting cardiovascular events, but the Random Forest model demonstrated a slight edge in accuracy and robustness. The confusion matrix for each model revealed the distribution of prediction errors, providing insights into their respective strengths and weaknesses. The Random Forest model's ensemble approach contributed to its superior performance by reducing variance and avoiding over-fitting, which is often a challenge with single decision trees.Overall, the detailed preprocessing, splitting, and model training procedures ensured that the dataset was in an optimal state for machine learning. This rigorous approach led to accurate and reliable predictions, contributing significantly to the study's findings and underscoring the effectiveness of advanced machine learning techniques in enhancing cardiovascular risk assessments.

IV. IMPLEMENTATION

A comprehensive analysis to predict heart disease using two machine learning algorithms: K-Nearest Neighbors (KNN) and Random Forest (RF). The study utilized a heart disease dataset, which includes 11 clinical features such as age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, resting ECG results, maximum heart rate, exerciseinduced angina, ST depression induced by exercise, and the slope of the peak exercise ST segment. The

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

target variable, heart disease, indicates whether a patient has heart disease. Preprocessed the data by handling missing values and encoding categorical variables. We then split the dataset into training and testing sets. Both KNN and RF models were trained using the training set, and their performance was evaluated on the testing set. Key metrics such as accuracy, recall, F1-score, and confusion matrix were calculated to assess the models' effectiveness. Additionally, we have visualized chestpain type vs heartdisease to further analyze their performance.

A. Performance Metrics Analysis:
The performance of K-Nearest Neighbors (KNN) and Random Forest models was evaluated using several metrics. The results are summarized in Table.

Table	KNN	RF
Accuracy	0.82775	0.92503
Recall	0.85959	0.93696
Precision	0.83565	0.92897
F1 Score	0.84745	0.93295

TABLE PERFORMANCE OF KNN AND RF

Table I shows the performance metrics for the K-Nearest Neighbors (KNN) and Random Forest models. The KNN model recorded an accuracy score of 0.82775, representing the proportion of instances correctly classified. Its precision score was 0.83565, indicating the proportion of true positive predictions out of all positive predictions made. The F1 score was 0.84745, which balances precision and recall, while the recall score was 0.85859, reflecting the proportion of actual positive cases that were accurately identified and the Random Forest model demonstrated an accuracy score of 0.92503, which is higher than that of KNN, suggesting better overall classification performance. The precision score for Random Forest was 0.92897, the F1 score was 0.93295, and the recall score was 0.93696. These metrics indicate that Random Forest performed better in terms of identifying and classifying positive instances accurately, providing a complete robust classification compared to KNN.

B. Confusion Matrix for KNN model:
The confusion matrix graph provides a clear view of how well the model classified instances of" Heart Disease" versus" No Heart Disease." It shows that the model correctly identified 219 cases of heart disease (true positives) and 300 cases of no heart disease (true negatives). However, it also misclassified 59 cases as positive when they were negative (false positives) and missed 49 cases that were actually positive (false negatives). This indicates that while the model has a high overall accuracy, it may be overfitting to the training data, meaning it could struggle to make accurate predictions on new data. To improve its generalization, techniques like regularization and cross-validation could be beneficial. By looking closely at the confusion matrix and other performance metrics, we can better understand the model's strengths and weaknesses, ultimately working toward a more reliable tool for diagnosing heart disease and improving patient care.

C. Confusion Matrix for Random Forest model:
The confusion matrix for the random forest model indicates the following performance metrics: it correctly classified 253 instances of heart disease (true positives) and 327 instances of no heart disease (true negatives). However, there were also 25 instances misclassified as positive when they were actually negative (false positives), and 22 instances missed as positive when they were indeed positive (false negatives). These results suggest that the model is performing well overall, with a significant number of correct classifications. However, the presence of false positives and false negatives highlights areas where the model could improve. Specifically, the false positives might lead to unnecessary anxiety or further testing for patients incorrectly identified as having heart disease, while false negatives could result in missed diagnoses, potentially endangering patients' health. To enhance the model's accuracy, further analysis could be conducted to understand the causes of these misclassifications. Techniques such as adjusting the classification threshold or further tuning model parameters may help in reducing these errors. Overall, the confusion matrix provides critical insights into the model's strengths and weaknesses, guiding efforts for improvement.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

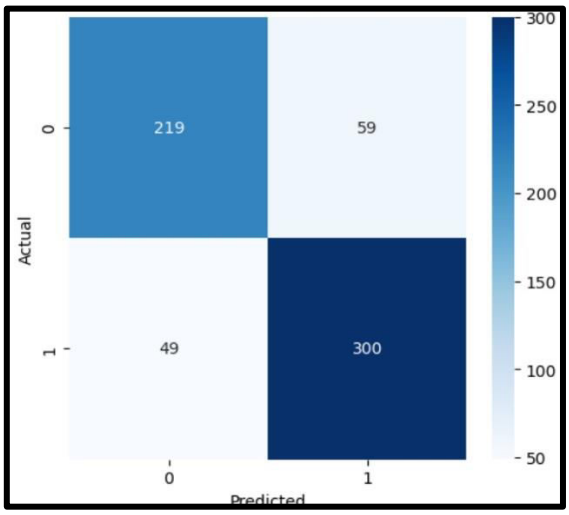

Fig 2. KNN Confusion Matrix

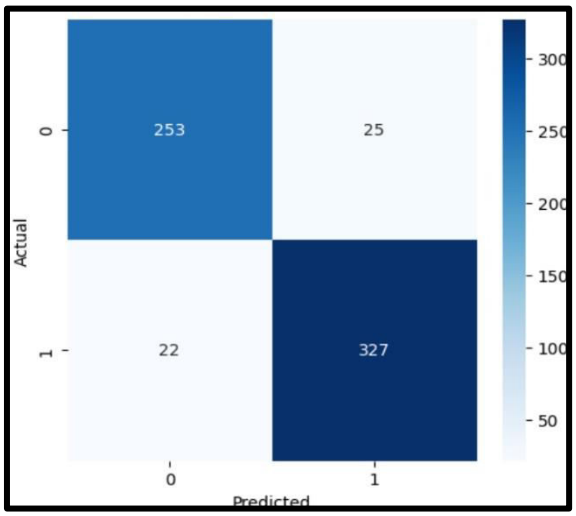

Fig 3. RF Confusion Matrix

D. Chestpain types vs heart disease:

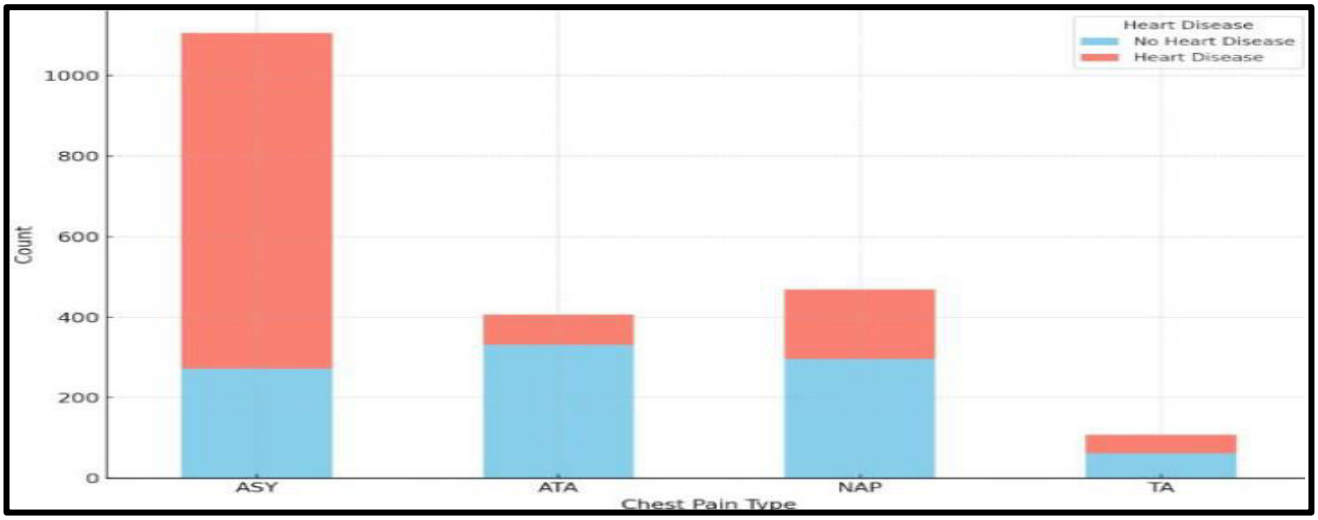

Fig 4. Stacked bar plot of Chestpain types vs heart disease

This stacked bar plot visualizes the relationship between chest pain type and the presence or absence of heart disease. The x-axis represents the four different types of chest pain: ASY (asymptomatic), ATA (typical angina), NAP (nonanginal pain), and TA (atypical angina). The y-axis represents the count of individuals within each category. The bars are stacked, with the blue portion representing individuals without heart disease and the orange portion representing individuals with heart disease. The plot reveals a clear trend: individuals with ASY chest pain have a significantly higher prevalence of heart disease compared to the other chest pain types. Conversely, individuals with TA chest pain have the lowest prevalence of heart disease. This suggests that ASY chest pain may be a more indicative symptom of heart disease than other chest pain types, while TA chest pain might be a less reliable indicator. The plot also highlights that individuals with ATA and NAP chest pain types exhibit an intermediate prevalence of heart disease, suggesting that these pain types may be less strongly associated with heart disease compared to ASY but more strongly associated than TA. This information could be valuable for medical professionals in assessing the risk of heart disease

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

based on chest pain type. Further analysis could explore the underlying reasons for these observed associations and investigate if specific subgroups within each chest pain type exhibit different heart disease prevalence.

V. COMPARATIVE ANALYSIS OF MODEL PERFORMANCE

The comparative analysis of model performance between the K-Nearest Neighbors (KNN) and Random Forest models for heart disease classification highlights significant distinctions in their effectiveness. The KNN model achieved an accuracy of 0.82775, indicating that approximately 82.8% of instances were correctly classified. Its precision score of 0.83565 suggests that about 83.6% of positive predictions were true positives, while the recall score of 0.85859 reflects that roughly 85.9% of actual positive cases were accurately identified. The F1 score of 0.84745 demonstrates a reasonable balance between precision and recall, but the model's confusion matrix reveals concerning misclassifications. specifically, the KNN model correctly identified 219 cases of heart disease (true positives) and 300 instances of no heart disease (true negatives). However, it misclassified 59 cases as positive when they were negative (false positives) and missed 49 cases that were genuinely positive (false negatives). This suggests that KNN may be prone to overfitting, where it performs well on training data but struggles with generalization to unseen instances. The implications of false positives could lead to unnecessary anxiety for patients, while false negatives pose a significant risk of missing actual cases of heart disease. In contrast, the Random Forest model demonstrated superior performance, achieving an accuracy score of 0.92503, which indicates a more robust classification capability. Its precision score of 0.92897 means that approximately 92.9% of positive predictions were accurate. The recall score of 0.93696 indicates that about 93.7% of actual positive cases were identified, while the F1 score of 0.93295 reflects excellent harmony between precision and recall. The confusion matrix for the Random Forest model shows it correctly classified 253 instances of heart disease and 327 instances of no heart disease. The model reported only 25 false positives and 22 false negatives. Although misclassifications still exist, the lower counts compared to KNN signify greater reliability in accurately identifying true cases of heart disease. This aspect is particularly critical in healthcare, where precise diagnosis is vital for effective patient management.

Overall, the Random Forest model's superior performance metrics higher accuracy, precision, and recall, combined with fewer misclassifications position it as the more effective tool for heart disease classification. This analysis emphasizes the importance of selecting models based on comprehensive performance metrics, especially in healthcare contexts, where accurate and timely diagnoses are crucial for improving patient outcomes. In summary, while both models have merits, the Random Forest's enhanced reliability underscores its potential to significantly improve diagnostic accuracy in heart disease detection.

VI. CONCLUSON AND FUTURE WORK

In conclusion, the comparison between the K-Nearest Neighbors (KNN) and Random Forest (RF) models clearly highlights the superior performance of RF in heart disease prediction tasks. The Random Forest model consistently outperforms KNN across all key metrics, including accuracy, recall, precision, and F1 score. Its confusion matrix indicates fewer misclassifications, underscoring its reliability and robustness in effectively identifying cases of heart disease. While KNN demonstrates reasonable performance, its tendency for overfitting leads to lower effectiveness in diverse scenarios, resulting in a higher number of false positives and false negatives. These challenges reveal limitations in KNN's ability to maintain consistent performance. Overall, the findings affirm that the Random Forest model serves as a more effective and reliable tool for heart disease prediction, offering enhanced accuracy and a better balance between precision and recall. This study emphasizes the critical importance of model selection in clinical applications, where precise and trustworthy predictions are essential. Future research could explore further enhancements to the Random Forest model, focusing on advanced techniques and improving interpretability to provide clear and actionable insights for medical decision-making.

REFERENCES

1.Chicco, Davide, and Giuseppe Jurman.” Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone.” BMC medical informatics and decision making 20 (2020).
2.Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. “Heart disease prediction using machine learning techniques.” SN Computer Science 1.6 (2020): 345.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.Adler, Eric D., et al." Improving risk prediction in heart failure using machine learning." European journal of heart failure 22.1 (2020): 139- 147.
4.Li, Jian Ping, et al." Heart disease identification method using machine learning classification in e-healthcare." IEEE access 8 (2020): 107562- 107582.
5.M. A. Nabaouia Louridi, Bouabid El Ouahidi," Identification of Cardiovascular Diseases Using Machine Learning," presented at the 2019 7th Mediterranean Congress of Telecommunications (CMT), 2019,1.
6.K. V. s. K. Nikhil Kumar Muthyala, K.Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools," IJSRCSEIT, vol. 3, no. 3, p. 13, 15 March 2018 2018, doi: 10.13140/RG.2.2.28488.83203.
7.P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain," A decision support system for heart disease prediction based upon machine learning," Journal of Reliable Intelligent Environments, vol. 7, no. 3, pp. 263-275, 2021, doi: 10.1007/s40860-021-00133-6.
8.J. L. Amin ul Haq, Muhammad Hammad Memon, Muhammad Hunain Memon, Jalaluddin Khan, Syeda Munazza Marium," Heart Disease Prediction System Using Model Of Machine Learning and Sequential Backward Selection Algorithm for Features Selection," presented at the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 29 -31 March 2019, 2019, 1.
9.S. Hiren Kumar Thakkar, Sonali Patil," A Comparative Analysis of Machine Learning Classifiers for Robust Heart Disease Prediction," presented at the 2020 IEEE 17th India Council International Conference (INDICON, 2020, 1.
10.B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," BioMed Res. Int., vol. 2020, pp. 1–10, Apr. 2020.
11.Bhunia, P.K.; Debnath, A.; Mondal, P.; D E, M.; Ganguly, K.; Rakshit, P. Heart Disease Prediction using Machine Learning. Int. J. Eng. Res. Technol. 2021, 9.
12.Hassan, C.A.U.; Iqbal, J.; Irfan, R.; Hussain, S.; Algarni, A.D.; Bukhari, S.S.H.; Alturki, N.; Ullah, S.S. Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers. Sensors 2022, 22, 7227.
13.Waigi, R.; Choudhary, S.; Fulzele, P.; Mishra, G. Predicting the risk of heart disease using advanced machine learning approach. Eur. J. Mol. Clin. Med. 2020, 7, 1638–1645.
14.Drozd˙ z, K.; Nabrdalik, K.; Kwiendacz, H.; Hendel, M.; Olejarz, A.; ˙ Tomasik, A.; Bartman, W.; Nalepa, J.; Gumprecht, J.; Lip, G.Y.H. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol. 2022, 21, 240.
15.K. D. K. Venkatesh, M. Prathyusha, and C. H. Naveen Teja," Identification of Disease Prediction Based on Symptoms Using Machine Learning," JAC: A Journal Of Composition Theory, vol. 14, no. 6, 2021.
16.R. Kumar and P. Rani," Comparative analysis of decision support system for heart disease," Adv. Math. Sci. J., vol. 9, no. 6, pp. 3349–3356, 2020.
17.R. R. Razonable et al., "A framework for outpatient infusion of antispike monoclonal antibodies to high-risk patients with mild-tomoderate coronavirus disease-19: the Mayo Clinic model," in Mayo Clinic Proceedings, vol. 96, no. 5, pp. 1250-1261, 2021.
18.M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review," Artificial Intelligence in Medicine, vol. 128, p. 102289, 2022.
19.P. Singh, N. Singh, K. K. Singh, and A. Singh, "Diagnosing of disease using machine learning," in Machine learning and the internet of medical things in healthcare, pp. 89-111, 2021.
20.V. Chang, V. R. Bhavani, A. Q. Xu, and M. Hossain," An artificial intelligence model for heart disease detection using machine learning algorithms," Healthc. Anal., vol. 2, 2022.
21.M. M. Ahsan and Z. Siddique," Machine learning-based heart disease diagnosis: A systematic literature review," Artif. Intell. Med., vol. 128, p. 102289, 2022.
22.A. Alanazi," Using machine learning for healthcare challenges and opportunities," Inform. Med. Unlocked, vol. 30, p. 100924, 2022.
23.L. Yahaya, N. David Oye, and E. Joshua Garba," A comprehensive review on heart disease prediction using data mining and machine learning techniques," Am. J. Artif. Intell., vol. 4, no. 1, p. 20, 2020.
24.R. M. Aziz," Application of nature-inspired soft computing techniques for gene selection: a novel framework for cancer classification," Soft Comput., vol. 26, no. 22, pp. 12179–96, 2022.
25.A. Yaqoob, R. M. Aziz, N. K. Verma, P. Lalwani, A. Makrariya, and P. Kumar," A review on nature-inspired algorithms for cancer disease prediction and classification," Mathematics, vol. 11, no. 5, p. 1081, 2023.
26.M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid Machine Learning Model," Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021, pp. 1329–1333, 2021.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

27.R. Shang, B. Ara, I. Zada, S. Nazir, Z. Ullah, and S. U. Khan," Analysis of simple K-mean and parallel K-mean clustering for software products and organizational performance using education sector dataset," Sci. Program., vol. 2021, 2021.

28.M. A. Kadhim and A. M. Radhi," Heart disease classification using optimized Machine learning algorithms," Iraqi Journal For Computer Science and Mathematics, vol. 4, no. 2, pp. 31-42, 2023.

29.D. Mpanya, T. Celik, E. Klug, and H. Ntsinjana, "Machine learning and statistical methods for predicting mortality in heart failure," Heart Fail. Rev., vol. 26, no. 3, pp. 545–552, 2021. Diagnostics, vol. 11, no. 9, p. 1714, Sep. 2021.

30.M. F. Ansari, B. AlankarKaur, and H. Kaur, "A prediction of heart disease using machine learning algorithms," Adv. Intell. Syst. Comput., vol. 1200 AISC, pp. 497–504, 2021, doi: 10.1007/978-3-030-51859-24.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details