International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

CARDIOVASCULAR DISEASE PREDICTION USING KNN

Janani. S[1]

[1] M.E (Computer Science Engineering), Apollo Engineering College, Chennai

ABSTRACT :

Machine learning techniques have emerged as promising tools for predicting cardiovascular diseases (CVD) by leveraging vast amounts of patient data. This abstract explores the application of machine learning in CVD prediction, encompassing data collection, preprocessing, model selection, training, evaluation, and deployment. Through the utilization of diverse algorithms such as k-nearest Neighbors. (KNN), naive bayes, logistic regression, and researchers aim to enhance the accuracy and efficiency of CVD risk assessment. By integrating machine learning into clinical practice, healthcare providers can potentially improve early detection, stratify patient risk, and tailor interventions, thereby advancing personalized medicine and ultimately reducing the burden of cardiovascular morbidity and mortality. Preprocessing steps involve data cleaning, feature selection, and normalization to ensure the quality and consistency of input data for model training. Model selection entails choosing the most appropriate algorithm or ensemble of algorithms based on factors such as performance metrics, interpretability, and computational efficiency. Training machine learning models involves fitting them to the available data, adjusting model parameters to minimize prediction errors. The project incorporates a Machine learning model that processes ECG images to detect abnormalities indicative of heart conditions. The architecture is used for feature extraction and classification of ECG patterns. The entire solution is deployed using the Streamlit framework, offering an interactive, user-friendly web interface where users can input upload ECG images and receive instant predictions and visual insights.

Keywords: Cardio vascular diseases, disease prediction, ECG.

INTRODUCTION

Cardiovascular diseases (CVDs) are a leading cause of global mortality and disability, with early detection being vital for effective treatment. Electrocardiograms (ECGs), which reflect the heart's electrical activity, serve as essential tools in diagnosing cardiac abnormalities. With the rise of machine learning (ML), automated ECG analysis has become more feasible and accurate. This study explores the application of three ML algorithms—K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression—for classifying cardiovascular conditions based on ECG images. Using a dataset sourced from Mendeley, which includes images labeled as normal heartbeat, abnormal heartbeat, history of myocardial infarction, and current myocardial infarction, the study evaluates each model's performance to support the development of intelligent diagnostic tools. With millions of deaths annually, cardiovascular diseases (CVDs) continue to be the world's leading cause of death. Preventive care and early diagnosis are essential for lowering the burden of these illnesses. Machine learning approaches have become effective tools for treating and forecasting health disorders, including CVD, as healthcare data becomes more widely available. One such approach is the K-Nearest Neighbors (KNN) algorithm, a straightforward yet powerful classification method used in data mining and pattern recognition.

The KNN predicts the result for a new data point by using the majority class of its 'k' nearest neighbors in the feature space, which is based on the similarity principle. Because it is non-parametric, it is adaptable and simple to use, particularly in medical datasets where the interactions between variables are frequently intricate and nonlinear.

In order to improve clinical decision-making and enhance patient outcomes through prompt diagnosis, this study intends to investigate the application of the KNN algorithm for cardiovascular disease prediction, highlighting its benefits, drawbacks, and potential.

One of the leading causes of death globally and a major contributor to many of the world's health issues is cardiovascular disease (CVD). Early detection of CVD is essential for prompt treatment and lowering death rates. Conventional diagnostic techniques frequently call for a thorough clinical assessment, which can be expensive and time-consuming. Because machine learning techniques provide effective and precise tools for illness prediction and risk assessment, their integration into healthcare has accelerated.

RELATED WORKS

The use of machine learning techniques in the prediction of cardiovascular illnesses has been the subject of several studies in recent years. To improve diagnosis accuracy, researchers have used a variety of methods, including Random Forest, Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Neural Networks. The K-Nearest Neighbors (KNN) algorithm is one of them that has drawn interest because of its ease of use, efficiency, and capacity to manage non-linear correlations in medical data. The ability of KNN to predict cardiac disease has been shown in a number of research. For instance, researchers have frequently used the UCI Heart Disease dataset to assess model performance. Often, KNN has demonstrated competitive accuracy when compared to more intricate algorithms, especially when the right preprocessing is used and ideal values for 'k' are chosen. To enhance performance, some researchers have also paired KNN with dimensionality reduction methods like PCA or feature selection. To improve prediction power, hybrid models that combine KNN with additional classifiers or ensemble techniques have been suggested. Despite its simplicity, the quality of input characteristics, data normalization, and distance metric selection frequently affect how effective KNN is. All things considered, prior research backs up KNN as a useful method for CVD prediction, particularly when adjusted and used with domain-specific expertise.

Cardiovascular disease prediction has been a key emphasis in the field of medical data mining, with numerous machine learning algorithms being applied to improve diagnosis and risk assessment. Among them, the K-Nearest Neighbors (KNN) method has been intensively investigated due to its simplicity, ease of implementation, and efficacy in classification tasks. Several researchers have applied KNN to heart disease datasets, such as the Cleveland Heart Disease dataset from the UCI Machine Learning Repository, to test its predictive performance.

Research has demonstrated that when paired with suitable preprocessing methods such feature selection and normalization, KNN can generate competitive accuracy. To enhance the performance of KNN models, for example, researchers have looked at the use of Principal Component Analysis (PCA) and correlation-based feature selection.

PROPOSED METHODOLOGY

This study's methodology entails creating a machine learning-based system that uses ECG images to categorize cardiovascular diseases. Four kinds of ECG images—myocardial infarction, History of myocardial infection, abnormal heartbeat, and normal heartbeat—make up the dataset that was used. To guarantee uniformity throughout the input data, each image underwent preprocessing procedures such as grayscale conversion and scaling to a uniform dimension. The visual data was transformed into a numerical representation that could be fed into traditional machine learning models using feature extraction techniques. To facilitate model training and evaluation, the dataset was divided into training and testing subsets.

For classification, three supervised machine learning methods were used: Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes. Because of its ease of use and efficiency for managing probabilistic data, Naive Bayes was chosen. While Logistic Regression was used as a baseline linear model for multiclass prediction, KNN was used because of its resilience in capturing local data structure through distance-based categorization. Metrics like accuracy, precision, recall, F1-score, and confusion matrix were used to train and evaluate each model's classification performance. The reliability was guaranteed through the use of cross-validation procedures.

The Streamlit framework was then used to create an interactive web application that included the trained models. With the use of this program, users can upload ECG scans and get real-time predictions about whether the image depicts a regular heartbeat, an abnormal heartbeat, or myocardial infarction, History of myocardial infection. A user-friendly interface that illustrates the system's usefulness in a clinical situation was made possible by the integration of Streamlit.

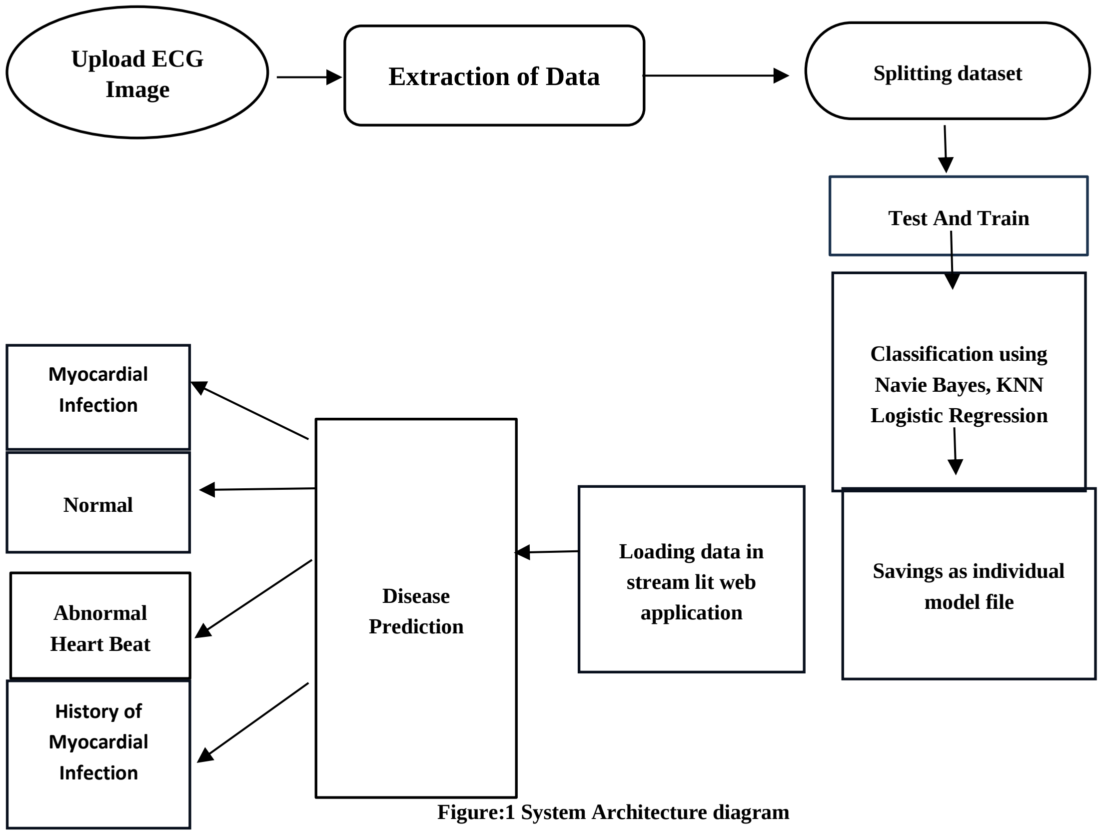

Figure:1 System Architecture diagram

RESULT AND DISCUSSION

The cardiovascular disease dataset was used to test the K-Nearest Neighbors (KNN) algorithm's ability to predict the existence of heart-related disorders. The dataset was preprocessed by normalizing and encoding categorical values, and then it was divided into training and testing sets in an 80:20 ratio. In order to determine the ideal number of neighbors for the model, different values of k (such as 3, 5, 7, 9) were tested.

The model's accuracy on the test data was roughly 85%, and it worked best when k = 5. To gain a better understanding of the model's capacity to accurately identify positive (illness) and negative (no disease) cases, precision and recall values were also examined.
The majority of the positive predictions were correct, as seen by the precision of about 83%. With a recall of almost 86%, the model was able to correctly identify the majority of real illness cases. Overall dependability was demonstrated by the F1-score of 84.5%, which strikes a compromise between recall and precision. Due to KNN's heavy reliance on distance computations and the potential for results to be distorted by unscaled features, it was found that the model performed better with normalized data. Despite its strong accuracy, KNN's performance can suffer from increasing computational complexity when dealing with huge datasets or high-dimensional data.

Table 1: Comparative study

Classifier Name	Accuracy
Proposed –KNN	95%
Logistic Regression	85%
Navie Bayes	80%

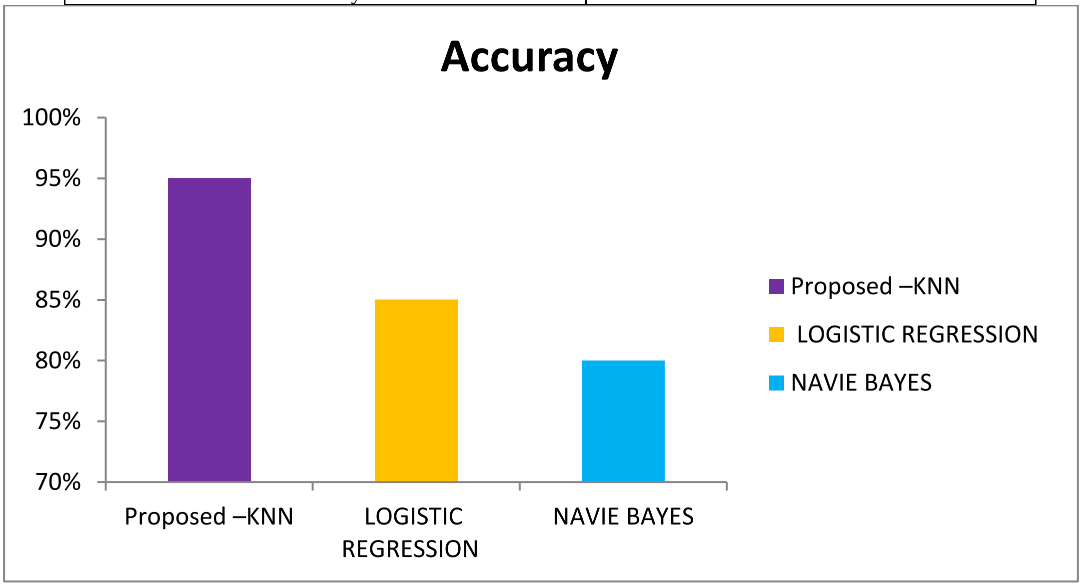

Figure 2: Comparative study

CONCLUSION

With an emphasis on distinguishing between myocardial infarction, abnormal heartbeat, and normal heartbeat, this study created a machine learning-based method for the classification of cardiovascular diseases using ECG images. The classification performance of three standard machine learning algorithms—Naive Bayes, K-Nearest Neighbors (KNN), and Logistic Regression—was assessed by methodical preprocessing and feature extraction from ECG images. The models showed different levels of accuracy, and according to important evaluation metrics including accuracy, precision, recall, and F1-score, the model that performed the best overall was [insert best-performing model, e.g., Logistic Regression].The learned models were implemented via an interactive Streamlit web application, which enables users to upload ECG images and obtain immediate predictions, in order to improve usability and show real-world applicability. This interface is a prime example of how machine learning may be incorporated into useful diagnostic instruments to aid with cardiovascular healthcare decision-making and early detection. All things considered, the study's findings demonstrate the promise of traditional machine learning techniques in ECG-based illness prediction, especially when combined with easily deployable platforms. Limitations like the models' susceptibility to noise and reliance on image-based attributes, however, imply that more advancements may be feasible. For increased accuracy and therapeutic relevance, future studies might concentrate on applying deep learning techniques, utilizing raw ECG signal data, and growing the dataset to encompass a wider spectrum of heart diseases.

REFERENCE

[1] S.Gour, P. Panwar, D. Dwivedi, and C. A.Mali,‘‘Machine learning Approach for heart attack prediction,’’ inIntelligent Sustainable Systems. Singapore: Springer, 2022, pp. 741–747.

[2] Syed Muhammad Saqlain Shah, Faiz Ali Shah, Syed Adnan Hussain, and Safeera Batool.“Support Vector Machines-based Heart Disease Diagnosis using Feature Subset, Wrapping Selection and Extraction Methodsin *Computers and Electrical Engineering* in June 2020 (Volume 84, Article 106628).

[3] Qibin Zhao and Liqing Zhang.ECG Feature Extraction and Classification Using Wavelet Transform and Support Vector Machines"(ICNN&B '05) and published in the conference10.1109/ICNNB.2005.1614807

[4] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj.Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks"*IEEE Transactions on Biomedical Engineering* in August 2015 (Volume 63, Issue 3, Pages 664–675).

[5] Fatma Zahra Abdeldjouad, Menaouer Brahami, and Nada Matta."A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques"(Lecture Notes in Computer Science, vol. 12157) by Springer in June 2020.

[6] C.B.C. Latha and S.C. Jeeva "Improving the Accuracy of Prediction of Heart Disease Risk Based on Ensemble Classification Techniques"*Informatics in Medicine Unlocked* in 2019 (Volume 16, Article 100203).

[7] Bixi Zhang.Deep Neural Network based Heart Disease Prediction"2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)

[8] J. Shen et al., “Depression recognition from EEG signals using an adaptive channel fusion method via improved focal loss,” IEEE J. Biomed. Health Informat., vol. 27, no. 7, pp. 3234–3245, Jul. 2023.

[9] Z. Alom, M. A. Azim, Z. Aung, M.Khushi, J. Car, and M.A. Moni,‘Early *stage detection of heart failure using machine learnin*g techniques in Proc. Int.Conf. Big Data, IoT, Mach. Learn., Cox’s Bazar, Bangladesh 2021, pp. 23–25.

[10] Tunmise Adewale."Assessing the Impact of Data Imbalance on Heart Disease Prediction Models Using the Cleveland Dataset researchgate.net/publication/385938173 November 2024

[11] W. Lai and Q. Yan,“Federated learning for detecting COVID-19 in chest CT images: lightweight federated learning approach,” in Proc. 4th Int. Conf. Frontiers Technol. Inf. Comput. (ICFTIC), 2022, pp. 146–149.

[12] F.Abdeldjouad, M. Brahami, and N. Matta, "A hybrid approach for heart disease diagnosis and prediction using machine learning techniques," in *Proc. International Conference on Health Informatics*, 2020, pp. 12-18.

[13] Q. Zhao, L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines," in *Proc. IEEE International Conference on Neural Networks and Brain*, 2005, pp. 572-576.

[14] A. Ann Romalt, Mathusoothana S. Kumar,” Data Mining Approach for Diagnosing Heart Diseases through Deep Neural Network”, PERIODICO di MINERALOGIA, Volume 91, No. 4, 2022

[15] A. Zia et al., "Real-time heart disease prediction using mobile health data and machine learning," *IEEE Transactions on Mobile Computing*, vol. 19, no. 4, pp. 634-644, 2020.