IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning-Based Prediction And Risk Assessment Of Heart Attacks Using Patient Health Data

[1]Dr.B.Meena Preethi, Associate Professor, Sri Krishna Arts and Science College, Coimbatore.

[2]K.Paramasivam, M.Sc., Computer Science, Sri Krishna Arts and Science College, Coimbatore.

[3]Umadevi.R, M.Sc., Computer Science, Sri Krishna Arts and science college, Coimbatore.

[4]Hari Shankar L B, M.Sc., Computer Science, Sri Krishna Arts and Science College, Coimbatore.

ABSTRACT

Cardiovascular diseases (CVDs), particularly myocardial infarctions (heart attacks), remain one of the leading causes of mortality worldwide. A primary contributing factor to this high mortality rate is delayed diagnosis and inadequate implementation of preventive strategies. In response, this study presents a machine learning-based predictive framework for assessing the risk of heart attacks using a combination of patient health indicators and lifestyle factors. The dataset was subjected to comprehensive preprocessing, including label encoding of categorical attributes and normalization of numerical features. A comparative analysis was conducted among several classification algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Gradient Boosting, and Artificial Neural Networks. This work illustrates the potential of artificial intelligence to support early detection, enhance preventive healthcare interventions, and contribute to reducing the global burden of cardiovascular diseases.

Keywords: Cardiovascular diseases, Machine learning, Heart attacks, Classification

I INTRODUCTION

Cardiovascular diseases (CVDs), especially heart attacks (myocardial infarctions), remain one of the most significant global health challenges, accounting for millions of deaths annually [1], [4]. Despite ongoing advancements in medical science and technology, early detection of heart attack risk continues to be a complex problem due to the multifactorial nature of CVDs. Factors such as age, blood pressure, cholesterol levels, diabetes, smoking habits, and family history all contribute to the onset of heart conditions [6], [11]. However, conventional diagnostic methods often fail to integrate these diverse parameters into a comprehensive and timely risk assessment [12], [15].

In recent years, the emergence of machine learning (ML) has opened new avenues for developing intelligent, data-driven models capable of uncovering hidden patterns in large volumes of patient health data [3], [7], [13]. By learning from historical medical records and patient profiles, ML algorithms can support clinical decision-making by predicting disease onset with greater accuracy and speed [1], [5]. These technologies not only enhance diagnostic precision but also hold promise in enabling early intervention and personalized treatment strategies [8], [14].

This study aims to leverage various supervised ML algorithms to build a robust, automated system that can accurately predict the likelihood and severity of heart attack risk. By analyzing a rich dataset of 3,000 patient records, the study compares multiple classification models—including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine, Gradient Boosting, and Artificial Neural Networks—to identify the most effective approach [2], [4], [9]. The goal is to create a practical tool that supports healthcare providers in proactive decision-making and ultimately reduces the burden of cardiovascular diseases [10].

II LITERATURE SURVEY

In the study by Baban U. Rindhe et al. [1], machine learning algorithms such as Artificial Neural Networks (ANN), Random Forest, and Support Vector Machine (SVM) were used to predict heart disease by analyzing large volumes of patient data. The research highlights the urgent need for early detection systems due to the increasing global death rate caused by cardiovascular diseases. Their findings emphasize that machine learning can support healthcare professionals in accurate and timely diagnosis, improving treatment decisions and potentially saving lives.

Avvaru R. V. Naga Suneetha and Dr. T. Mahalngam [2] proposed a method for predicting cardiovascular disease using both Machine Learning (ML) and Deep Learning (DL) techniques. The study employed SMOTE-ENN to balance the training data and evaluated several classifiers, including KNN, Naive Bayes, Decision Tree, SVM, XGBoost, ANN, and CNN. Their results demonstrated that Artificial Neural Networks achieved the highest prediction accuracy of 95.7%, highlighting the effectiveness of deep learning approaches in early diagnosis of heart disease.

Senthilkumar Mohan et al. [4] proposed a hybrid machine learning approach for effective heart disease prediction using clinical data. The study aimed to enhance prediction accuracy by selecting significant features and applying various classification algorithms. The proposed Hybrid Random Forest with Linear Model (HRFLM) achieved an accuracy of 88.7%, demonstrating improved performance compared to traditional models. Their research highlights the potential of hybrid ML techniques in effectively analyzing healthcare data for early diagnosis of cardiovascular diseases.

Liaqat Ali et al. [5] developed an optimized expert system for heart failure prediction using a stacked Support Vector Machine (SVM) model. The system combines a linear SVM with L1 regularization for feature selection and an L2-regularized SVM for prediction. A Hybrid Grid Search Algorithm (HGSA) was proposed to simultaneously optimize both models. The method outperformed conventional SVM and other ensemble models, achieving up to 3.3% better accuracy and outperforming ten prior models with accuracies ranging from 57.85% to 91.83%. This highlights the efficiency of stacked SVMs and hybrid optimization in improving predictive accuracy for heart failure diagnosis.

Rajesh N. et al. [6] explored the use of various machine learning algorithms for predicting heart disease, focusing on attributes associated with common risk factors. The study evaluated Naive Bayes and Decision Tree classifiers, along with a combination of algorithms, to determine the most suitable method for accurate prediction. Their results indicated that Naive Bayes performs better with smaller datasets, while Decision Trees yield higher accuracy when applied to larger datasets. This research emphasizes the importance of selecting the appropriate algorithm based on dataset size for effective heart disease prediction.

III PROBLEM STATEMENT

Cardiovascular diseases, especially heart attacks, continue to pose a major threat to global health, contributing significantly to premature mortality and long-term disability. Despite advancements in medical diagnostics, early identification of individuals at risk of heart attacks remains a challenge due to the multifactorial nature of the disease and the subtle interplay of health, lifestyle, and genetic factors. Traditional clinical approaches often rely on subjective assessment or isolated risk indicators, which may not capture the comprehensive risk profile of a patient. Moreover, there is a lack of automated, data-driven systems capable of processing large-scale patient data to provide accurate and timely heart attack risk predictions. Therefore, there is a critical need for an intelligent, machine learning-based solution that can analyze diverse patient health parameters and reliably classify the likelihood and severity of heart attack risk, thereby aiding clinicians in making informed decisions and improving preventive care strategies.

IV PROPOSED METHOD

The proposed method presents a machine learning-based approach to predict and assess the risk of heart attacks using structured patient health data. The process begins with data preprocessing, where a dataset comprising 3,000 patient records with 26 features is cleaned and prepared. Categorical variables such as gender, diabetes status, and lifestyle factors are transformed into numerical form using label encoding, while continuous numerical features are standardized using a standard scaler to ensure uniform input for all machine learning algorithms. Irrelevant features like geographical identifiers (country, continent, hemisphere) are removed to focus on medically relevant attributes.

Multiple supervised machine learning models are then trained and evaluated, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Gradient Boosting (XGBoost), and an Artificial Neural Network (ANN). Each model is assessed based on classification accuracy, precision, recall, and F1-score to determine its effectiveness in predicting heart attack risk levels. This system accepts new patient input data, automatically applies the necessary transformations, and provides a precise prediction of heart attack risk, labeled as one of four categories. This end-to-end method ensures accurate, real-time prediction capabilities, offering valuable support for early diagnosis and preventive cardiac care.

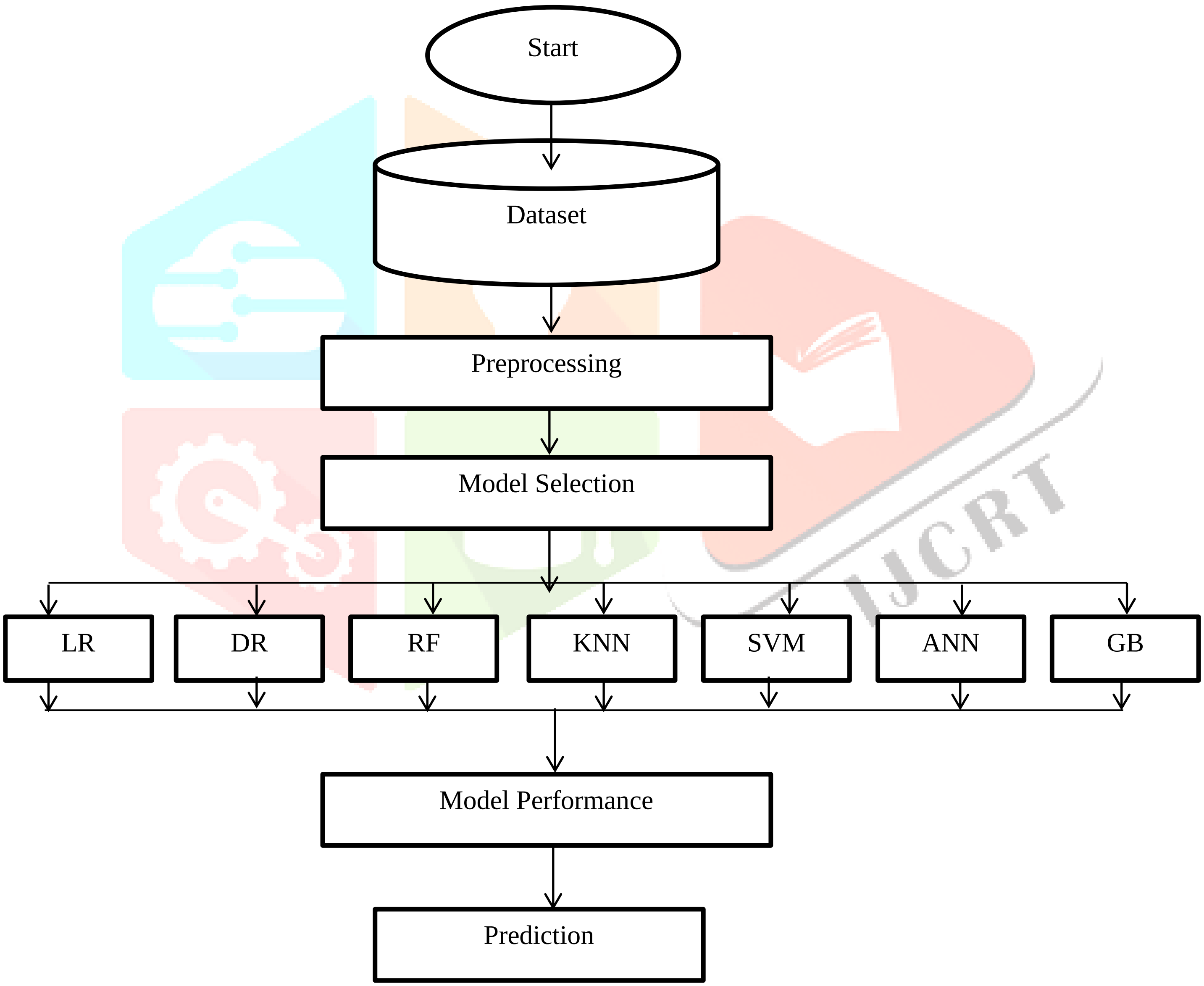

Figure.1 Proposed model

Figure.1 System Diagram

V RESULTS AND DISCUSSION

This study implemented and evaluated several machine learning algorithms to predict heart attack risk levels based on demographic, lifestyle, and medical history data. The dataset comprised 3,000 records with 26 features, including target classes: “No Attack Detected,” “Low Risk,” “Moderate Risk,” and “High Risk.” The performance of the models was assessed using accuracy, precision, recall, and F1-score.

5.1 Data Preprocessing

Label Encoding (Categorical to Numeric Conversion)

Label Encoding transforms categorical variables into numeric codes.

For a categorical feature $x \in \{A,B,C\}$, label encoder maps:

Encoded (x) ={0 if x=A, 1 if x=B, 2 if x=C

5.2 Standardization (Feature Scaling)

Standardizes features by removing the mean and scaling to unit variance.

Formula:

For each feature x:

$$x_{scaled} = \frac{x - \mu}{\sigma}$$

where:

μ = mean of the feature

σ = standard deviation of the feature

5.3 Gradient Boosting

Boosting combines weak learners (usually decision trees) in sequence.

$$y^{t} = y^{(t-1)} + n f_t(x)$$

Where:

n is the learning rate

f_t is the decision tree at iteration t

5.4 Model Performance

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	92.83	0.93	0.93	0.93
Decision Tree	88.67	0.89	0.89	0.88
Random Forest	90.00	0.92	0.90	0.87
K-Nearest Neighbors	85.50	0.84	0.85	0.84
Support Vector Machine	91.17	0.91	0.91	0.90
Gradient Boosting	96.17	0.97	0.96	0.96
Artificial Neural Network	90.83	0.91	0.91	0.91

Tabel.1 Model Performance

Table 1 presents a comparative analysis of various machine learning models used for heart disease prediction based on four key performance metrics: Accuracy, Precision, Recall, and F1-Score. Among all the models evaluated, the Gradient Boosting (GB) algorithm outperformed the others with the highest accuracy of 96.17%, a precision of 0.97, recall of 0.96, and F1-score of 0.96. This demonstrates its strong predictive capability and robustness in handling imbalanced data. Among all models evaluated, the Gradient Boosting Classifier achieved the highest accuracy of 96.17%, demonstrating its superior ability to learn complex decision boundaries from the feature set. The high F1-score and recall values further confirm its reliability in correctly identifying various heart attack risk levels, including moderate and high-risk cases.

5.5 Comparison chart

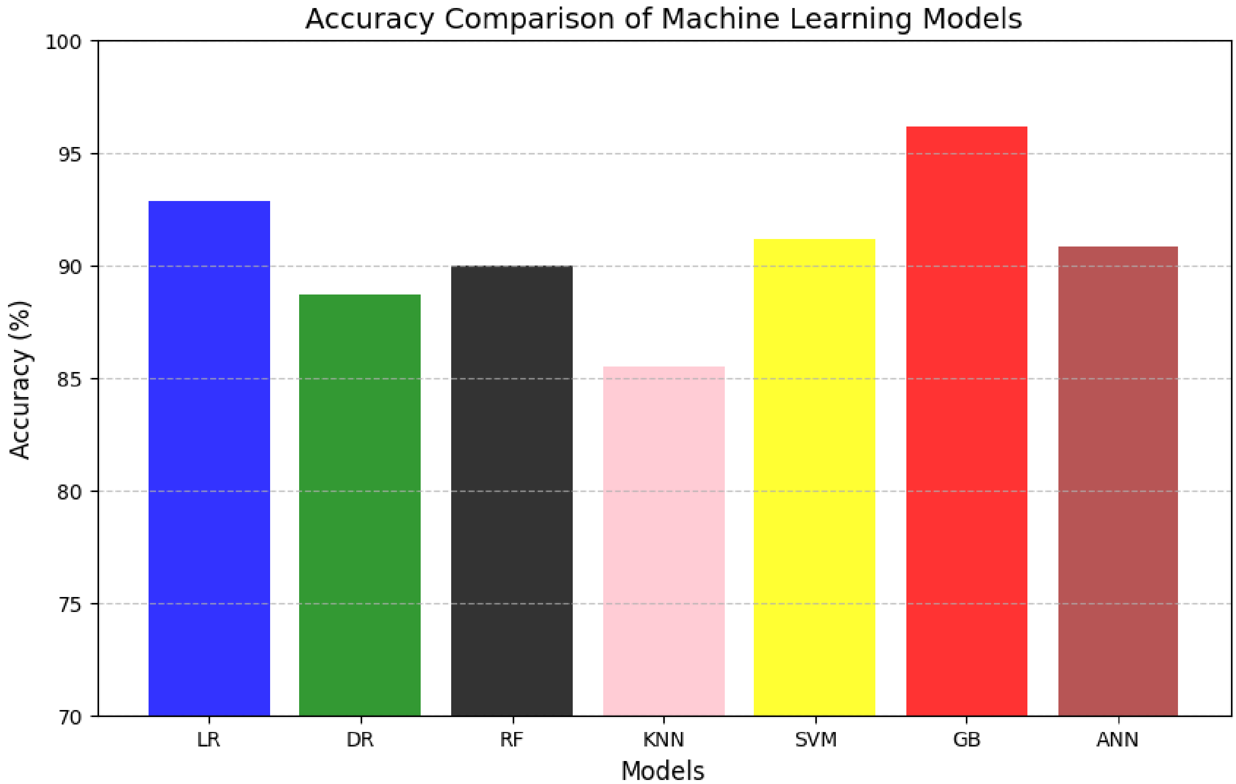

Figure.2 Comparison Graph

Figure 2 presents a bar chart that compares the accuracy of various machine learning models used in the study. The models include Logistic Regression (LR), Decision Tree (DR), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Gradient Boosting (GB), and Artificial Neural Network (ANN). Among all the models, Gradient Boosting (GB) achieved the highest accuracy. This graphical representation highlights that ensemble-based methods like GB and RF tend to outperform simpler algorithms in terms of prediction accuracy for the given dataset.

VI CONCLUSION

In this study, a comparative analysis of various machine learning algorithms was conducted to predict heart attack risk levels using a comprehensive dataset comprising demographic, lifestyle, and clinical attributes. The research aimed to develop an automated system capable of accurately identifying individuals at different levels of heart attack risk—ranging from "No Attack Detected" to "Low", "Moderate", and "High" risk. Seven popular machine learning models were implemented, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine, Gradient Boosting (XGBoost), and Artificial Neural Network. After extensive training, evaluation, and validation, Gradient Boosting (XGBoost) emerged as the best-performing model, achieving 96.17% accuracy, along with high precision, recall, and F1-score. This indicates its superior ability to capture complex, nonlinear patterns and relationships in the dataset.

The findings validate the applicability of machine learning in the healthcare domain, especially for early detection and preventive care of cardiovascular diseases. The trained models, particularly XGBoost, can be integrated into clinical decision support systems to assist medical professionals in identifying high-risk patients quickly and accurately. Overall, the study highlights the effectiveness of data-driven techniques in enhancing diagnostic accuracy and supports the continued integration of artificial intelligence into healthcare systems for proactive and personalized care delivery.

REFERENCES

[1] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*(6), 345.

[2] Suneetha, A. R. V. N., & Mahalngam, T. (2022). 15. cardiovascular disease prediction using ml and dl approaches. International Journal on Recent and Innovation Trends in Computing and Communication.

[3] Beyene, C., & Kamat, P. (2018). Survey on prediction and analysis the occurrence of heart disease using data mining techniques. International Journal of Pure and Applied Mathematics, 118(8), 165-174.

[4] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, 81542-81554.

[5] Ali, L., Niamat, A., Khan, J. A., Golilarz, N. A., Xingzhong, X., Noor, A., ... & Bukhari, S. A. C. (2019). An optimized stacked support vector machines based expert system for the effective prediction of heart failure. IEEE access, 7, 54007-54014.

[6] Nikhar, S., & Karandikar, A. M. (2016). Prediction of heart disease using machine learning algorithms. International Journal of Advanced Engineering, Management and Science, 2(6), 239484.

[7] Keya, M. S., Shamsojjaman, M., Hossain, F., Akter, F., Islam, F., & Emon, M. U. (2021, March). Measuring the heart attack possibility using different types of machine learning algorithms. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 74-78). IEEE.

[8] Mostafa, N., Azim, M. A., Kabir, M. R., & Ajwad, R. (2020, June). Identifying the risk of cardiovascular diseases from the analysis of physiological attributes. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1014-1017). IEEE.

[9] Ullah, T., Ullah, S. I., Ullah, K., Ishaq, M., Khan, A., Ghadi, Y. Y., & Algarni, A. (2024). Machine learning-based cardiovascular disease detection using optimal feature selection. IEEE Access, 12, 16431-16446.

[10] Rajput, K., Chetty, G., & Davey, R. (2019, November). Risk factors identification for heart disease in unstructured dataset using deep learning approach. In 2019 international conference on data mining workshops (ICDMW) (pp. 1056-1059). IEEE.

[11] Verma, L., Srivastava, S., & Negi, P. C. (2016). A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. Journal of medical systems, 40(7), 178.

[12] Sultana, M., Haider, A., & Uddin, M. S. (2016, September). Analysis of data mining techniques for heart disease prediction. In 2016 3rd international conference on electrical engineering and information communication technology (ICEEICT) (pp. 1-5). IEEE.

[13] Shadid, S. H., Shafkat, A., Yasmeen, M. F., Sibli, S. A., & Rafi, M. R. (2019). Prediction of heart disease using data mining techniques: A Case study. Prediction of heart disease using data mining techniques: A Case study, 41(1), 11-11.

[14] Methaila, A., Kansal, P., Arya, H., & Kumar, P. (2014). Early heart disease prediction using data mining techniques. Computer Science & Information Technology Journal, 24, 53-59.

[15] Dangare, C., & Apte, S. (2012). A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering and Technology (IJCET), 3(3).