International Journal of Advancement in Electronics and Computer Engineering (IJAECE)
Volume 13, Issue 12, December. 2024, pp. 105-109 ISSN 2278 -1412

A Critical Review of Machine Learning Models for Heart Disease Prediction

Bajrangi Kumar Gupta[1], Jeetendra Singh Yadav[2]
[1] Research Scholar, sudhamarit@gmail.com, CSE RKDFCE, Bhopal, India
[2,]Assis. Prof., jeetendra2201@gmail.com, CSE RKDFCE, Bhopal, India

Abstract *– Heart disease remains a leading global health concern, necessitating accurate and early prediction for improved patient outcomes. Machine learning (ML) has emerged as a powerful tool in cardiovascular diagnostics, offering enhanced predictive capabilities over traditional methods. This review critically evaluates various ML models, including logistic regression, decision trees, random forests, support vector machines (SVM), artificial neural networks (ANNs), and deep learning techniques, in terms of accuracy, interpretability, and real-world applicability.*

The study highlights that ensemble learning and deep neural networks achieve high predictive performance but face challenges such as data imbalance, interpretability, and computational demands. Recent advancements in explainable AI (XAI), federated learning, and hybrid ML models aim to enhance model reliability and clinical integration. The findings emphasize the need for a standardized evaluation framework to improve ML adoption in healthcare.

This review provides key insights for researchers and clinicians, underscoring the potential of AI-driven predictive analytics in revolutionizing heart disease diagnosis and personalized treatment.

Keywords: Heart Disease Prediction, Machine Learning, Deep Learning, Cardiovascular Diagnostics, AI in Healthcare, Predictive Analytics

I. INTRODUCTION

Heart disease continues to be a leading cause of morbidity and mortality worldwide, placing a significant burden on healthcare systems. Early and accurate diagnosis is crucial for improving patient outcomes, reducing complications, and enabling timely medical interventions. Traditional diagnostic methods, such as electrocardiograms (ECG), echocardiography, and clinical assessments, while effective, often rely on manual interpretation, making them susceptible to variability, delays, and human error. In response to these challenges, machine learning (ML) models have emerged as powerful tools for predictive analytics in cardiovascular disease detection and risk assessment.

Machine learning algorithms can analyze vast amounts of patient data, recognize patterns, and generate predictive models with high accuracy. By leveraging techniques such as logistic regression, decision trees, random forests, support vector machines (SVM), artificial neural networks (ANNs), and deep learning architectures, ML models can efficiently assess risk factors, detect abnormalities, and assist clinicians in making data-driven decisions. These models not only improve diagnostic precision but also enable early disease prediction and personalized treatment planning, leading to better patient outcomes.

Despite these advancements, several challenges remain in the adoption of ML for heart disease prediction. Model interpretability, data imbalance, feature selection, and generalizability pose significant obstacles to integrating ML into clinical practice. Black-box deep learning models, while highly accurate, often lack explainability, making them difficult for clinicians to trust and adopt. Additionally, the quality and availability of medical datasets, differences in patient demographics, and variations in feature engineering techniques can significantly impact model performance.

This paper presents a comprehensive review of ML models used for heart disease prediction, evaluating their effectiveness, limitations, and real-world applicability. We explore various classification algorithms, feature selection methodologies, and evaluation metrics to provide a critical analysis of their predictive capabilities. Furthermore, recent advancements in explainable AI (XAI), federated learning, and hybrid ML approaches are discussed to address the key challenges associated with ML-based cardiovascular diagnostics.

By examining the efficacy of ML models in heart disease prediction, this review aims to contribute to the development of more reliable, interpretable, and clinically applicable AI-driven solutions. The insights gained from this study will help bridge the gap between computational intelligence and medical decision-making, paving the way for precision medicine and AI-assisted healthcare innovations..

II. BACKGROUND

In the last few decades, medical science has used the technological advancements very well to improve the

quality of healthcare. These advancements in technology have paved ways for accurate diagnosis and prediction of diseases [1]. Various researchers have proposed a number of models for predicting heart diseases using different technologies such as artificial neural networks, machine learning, data mining, etc. This paper analyses the work done by various researchers on the accuracy of heart disease prediction through the different approaches. A detail literature review has been provided in the study. The analysis has also been presented on the basis on technology used.

Heart disease is a leading cause of death worldwide. However, it remains difficult for clinicians to predict heart disease as it is a complex and costly task. Hence, [2] they proposed a clinical support system for predicting heart disease to help clinicians with diagnostic and make better decisions. Machine learning algorithms such as Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Random Forest, and Decision Tree are applied in this study for predicting Heart Disease using risk factors data retrieved from medical files. Several experiments have been conducted to predict HD using the UCI data set, and the outcome reveals that Naïve Bayes outperforms using both cross-validation and train-test split techniques with an accuracy of 82.17%, 84.28%, respectively. The second conclusion is that the accuracy of all algorithm decrease after applying the cross-validation technique. Finally, we suggested multi validation techniques in prospectively collected data towards the approval of the proposed approach.

Detection of heart disease through early-stage symptoms is a great challenge in the current world scenario. If not diagnosed timely then this may become the cause of death. In developing countries where heart specialist doctors are not available in remote, semi-urban, and rural areas; an accurate decision support system can play a vital role in early-stage detection of heart disease. In this paper [3], the authors have proposed a hybrid decision support system that can assist in the early detection of heart disease based on the clinical parameters of the patient. Authors have used multivariate imputation by chained equations algorithm to handle the missing values. A hybridized feature selection algorithm combining the Genetic Algorithm (GA) and recursive feature elimination has been used for the selection of suitable features from the available dataset. Further for pre-processing of data, SMOTE (Synthetic Minority Oversampling Technique) and standard scalar methods have been used. In the last step of the development of the proposed hybrid system, authors have used support vector machine, naive bayes, logistic regression, random forest, and adaboost classifiers. It has been found that the system has given the most accurate results with random forest classifier. The proposed hybrid system was tested in the simulation environment developed using Python.

Heart disease is one of the significant reason of death and disability. The shortage of Doctors, experts and ignoring patient symptoms lead to big challenge that may cause death, disability to the patient. Therefore, we need expert system that serve as an analysis tool to discover hidden information and patterns in heart disease medical data. Data mining is a cognitive procedure of discovering the hidden approach patterns from large data set. The available massive data can used to extract useful information and relate all attributes to make a decision. Various techniques listed and tested here to understand the accuracy level of each. In previous studies, researchers expressed their effort on finding best prediction model. This paper [4] proposes new heart disease prediction system that combine all techniques into one single algorithm, it called hybridization. The result confirm that accurate diagnose can be taken by using a combined model from all techniques.

Heart disease is one of the most common causes of death around the world nowadays. Often, the enormous amount of information is gathered to detect diseases in medical science. All of the information is not useful but vital in taking the correct decision. Thus, it is not always easy to detect the heart disease because it requires skilled knowledge or experiences about heart failure symptoms for an early prediction. Most of the medical dataset are dispersed, widespread and assorted. However, data mining is a robust technique for extracting invisible, predictive and actionable information from the extensive databases. In this paper [5], by using info gain feature selection technique and removing unnecessary features, different classification techniques such that KNN, Decision Tree (ID3), Gaussian Naïve Bayes, Logistic Regression and Random Forest are used on heart disease dataset for better prediction. Different performance measurement factors such as accuracy, ROC curve, precision, recall, sensitivity, specificity, and F1-score are considered to determine the performance of the classification techniques. Among them, Logistic Regression performed better, and the classification accuracy is 92.76%.

Health care industries are one among the top in generating data. To mine the complex data advance algorithms and techniques are needed. The data extraction techniques are used to convert these raw facts as meaningful information. One of the popular data extraction techniques is data mining and machine learning. With the patient data Health care industries are now focusing on optimizing the efficiency and quality of the treatment using various data analytical tools. Data mining and Machine learning has been used by many industries, however they are the proven methodology in health care. Non communicable disease such a heart disease, diabetics and cancer are major reason for the death around the world. Heart disease is one among the top reason for death. In this research paper [6] they have implemented popular data mining algorithms viz.,

Support vector machine and decision tree with the relevant heart disease data set using Python.

Nowadays, machine learning algorithms have become very important in the medical sector, especially for diagnosing disease from the medical database. Many companies using these techniques for the early prediction of diseases and enhance medical diagnostics. The motivation of this paper [7] is to give an overview of the machine learning algorithms that are applied for the identification and prediction of many diseases such as Naïve Bayes, logistic regression, support vector machine, K-nearest neighbor, K-means clustering, decision tree, and random forest. In this work, many previous studies were reviewed that used machine learning algorithms for detecting various diseases in the medical area in the last three years.

III. METHOD

This study will systematically evaluate the efficacy of various machine learning (ML) models in predicting heart disease by following a structured methodology that includes data collection, preprocessing, feature selection, model selection, training, evaluation, and performance comparison. The research will utilize benchmark heart disease datasets such as the Cleveland Heart Disease Dataset, Framingham Heart Study, and UCI Machine Learning Repository, ensuring comprehensive analysis across diverse patient demographics.

In the first phase, raw data will be collected, consisting of patient attributes such as age, blood pressure, cholesterol levels, heart rate, electrocardiogram (ECG) readings, and lifestyle factors. Since medical datasets often contain missing values and inconsistencies, a preprocessing step will be applied, including data normalization, missing value imputation, and outlier detection to improve data quality. Additionally, categorical features such as gender and medical history will be encoded to ensure compatibility with ML models.

The feature selection process will employ statistical techniques such as correlation analysis, recursive feature elimination (RFE), and principal component analysis (PCA) to identify the most relevant predictors of heart disease. Selecting the most informative features will enhance model interpretability and improve computational efficiency.

For model selection, various ML algorithms will be implemented and compared, including logistic regression, decision trees, random forests, support vector machines (SVM), k-nearest neighbors (KNN), artificial neural networks (ANNs), and deep learning architectures such as long short-term memory (LSTM) networks and convolutional neural networks (CNNs). Each model will be trained using a training dataset and optimized through hyperparameter tuning to maximize performance.

The evaluation phase will involve assessing each model based on key performance metrics, including accuracy, precision, recall, F1-score, area under the curve (AUC-ROC), and confusion matrix analysis. Cross-validation techniques such as k-fold cross-validation will be employed to ensure model robustness and generalizability across different datasets.

IV. CONCLUSION

This study will aim to enhance heart disease prediction by evaluating the efficacy of various machine learning (ML) models in analyzing patient health data. By implementing a structured methodology that includes data preprocessing, feature selection, model training, and performance evaluation, this research will provide valuable insights into the predictive capabilities of different ML techniques. The expected results will demonstrate that advanced ML models, particularly deep learning architectures, can significantly improve diagnostic accuracy and early detection of heart disease.

Future findings will highlight the importance of feature selection techniques in optimizing model performance and reducing computational complexity. Additionally, the study will show that integrating explainable AI (XAI) approaches will enhance model transparency, enabling better trust and adoption of AI-driven diagnostics in clinical practice. Moreover, the research will explore the potential of federated learning as a privacy-preserving approach for distributed medical data analysis, ensuring data security while maintaining predictive accuracy.

REFERENCES

[1] Sakshi Goel, Abhinav Deep, Shilpa Srivastava, “Comparative Analysis of various Techniques for Heart Disease Prediction”, IEEE, International Conference on Information Systems and Computer Networks, 2019, pp. 88-94.

[2] Halima El Hamdaoui, Saïd Boujraf, Nour El Houda Chaoui, Mustapha Maaroufi, “A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques”, International Conference on Advanced Technologies For Signal and Image Processing, 2020, pp. 1-5.

[3] Pooja Rani, Rajneesh Kumar, Nada M. O. Sid Ahmed, Anurag Jain, “A decision support system for heart disease prediction based upon machine learning”, Journal of Reliable Intelligent Environments, Springer, 2021, pp. 1-14.

[4] Monther Tarawneh, Ossama Embarak, “Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques”, ACTA scientific nutritional health, 2019, pp. 1-6.

[5] S. M. M. Hasan, M. A. Mamun, M. P. Uddin, M. A. Hossain, “Comparative Analysis of Classification

Approaches for Heart Disease Prediction", 2020, PP. 1-4.

[6] S.Poonguzhali, P.Sujatha, P.Sripriya, V.Deepa, K. Mahalakshmi, "Performance Evaluation of Classification Methods For Predicting Heart Disease", 2019, pp. 1-9.

[7] Ibrahim Mahmood Ibrahim, Adnan Mohsin Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases", Journal of Applied Science And Technology Trends, 2021, pp. 1-10.

[8] Manoj Diwakar, Amrendra Tripathi, Kapil Joshi, Minakshi Memoria, Prabhishek Singh, Neeraj kumar, "Latest trends on heart disease prediction using machine learning and image fusion", Materials Today: Proceedings, 2020, pp. 1-7.

[9] Pronab Ghosh, Sami Azam, Mirjam Jonkman, Asif Karim, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques", IEEE Access, 2021, pp. 19304-19326.

[10] Pronab ghosh, Sami Azam, Asif Karim, "Use of Efficient Machine Learning Techniques in the Identification of Patients with Heart Diseases", ICISDM 2021, pp. 14-20.

[11] P Kalpana, S Shiyam Vignesh, L M P Surya, V Vishnu Prasad, "Prediction of Heart Disease Using Machine Learning", Journal of Physics: Conference Series, 2021, pp. 1-5.

[12] Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain, Preeti Nagrath, "Heart disease prediction using machine learning algorithms", IOP Conf. Series: Materials Science and Engineering, 2021, pp. 1-11.

[13] Abid Ishaq, Saima Sadiq, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques", IEEE Access, 2021, pp. 39707-39716.

[14] Sandeep Chandra Bollepalli, Ashish Kumar Sahani, "An Optimized Machine Learning Model Accurately Predicts In-Hospital Outcomes at Admission to a Cardiac Unit", Diagnostics 2022, pp. 1-14.

[15] Pronab Ghosh Sami Azam, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques", IEEE Access, 2021, pp. 19304-19326.

[16] A Angel Nancy, Dakshanamoorthy Ravindran, "IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning", Electronics 2022, pp. 1-19.

[17] Prasannavenkatesan Theerthagiri, Usha Ruby A, "Diagnosis and Classification of the Diabetes Using Machine Learning Algorithms", 2021, pp. 1-24.

[18] Jasjit S. Suri, Mrinalini Bhagawati, "Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review", Computers in Biology and Medicine, 2022, pp. 1-24.

[19] George Konstantonis, Krishna V. Singh, "Cardiovascular disease detection using machine learning and carotid/femoral arterial imaging frameworks in rheumatoid arthritis patients", Rheumatology International, Springer 2022, pp. 1-25.

[20] Oliver Faust, Wanrong Hong, "Heart rate variability for medical decision support systems: A review", Computers in Biology and Medicine, 2022, pp. 1-36.

[21] Navchetan Awasthi, Laslo van Anrooij, "Bandwidth Improvement in Ultrasound Image Reconstruction Using Deep Learning Techniques", Healthcare 2023, pp. 1-19.

[22] Nurul Absar, Emon Kumar Das, "The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction", Healthcare 2022, pp. 1-19.

[23] Kaouter Karboub, Mohamed Tabaa, "A Machine Learning Based Discharge Prediction of
Cardiovascular Diseases Patients in Intensive Care Units", Healthcare 2022, pp. 1-23.

[24] Evan D. Peet, Dana Schultz, "Variation in the infant health effects of the women, infants, and children program by predicted risk using novel machine learning methods", Health Economics, 2022, pp. 194-217.

[25] Forum Desai, Deepraj Chowdhury, "HealthCloud: A System for Monitoring Health Status of Heart Patients using Machine Learning and Cloud Computing", Elsevier, 2021, pp. 1-20.

[26] Juwairiya Siraj Khan , Manoj Kaushik, "Cardi-Net: A deep neural network for classification of cardiac disease using phonocardiogram signal", Computer Methods and Programs in Biomedicine, 2022, pp. 1-16.

[27] D. Cenitta, R Vijaya Arjunan, Prema K V, "Ischemic Heart Disease Multiple Imputation Technique Using Machine Learning Algorithm", Eng. Sci., 2022, pp. 262-272.

[28] Sebastian Raubitzek, Kevin Mallinger, "Combining Fractional Derivatives and Machine Learning: A Review", Entropy 2023, pp. 1-23.

[29] Ghulab Nabi Ahmad, Hira Fatima, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without Grid Search CV", IEEE Access, 2022, pp. 80151-80173.

[30] Maheriya Arpita, Shailesh Panchal , "Smart Health and Cybersecurity in the Era of Artificial Intelligence ", 2022, pp. 1-9.

[31] Joy Iong Zong Chen, P. Hengjinda, "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method - A Comparative Study", Journal of Artificial Intelligence and Capsule Networks, 2021, pp. 17-33.

[32] Nagaraj M. Lutimath, Neha Sharma , "Prediction of Heart Disease using Biomedical Data through Machine Learning Techniques ", EAI Endorsed Transactions , 2021, pp. 1-6.

[33] Angier Allen, Zohora Iqbal, "Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus", BMJ Open Diab Res Care, 2022, pp. 1-7.

[34] Manoj Diwakar, Amrendra Tripathi, "Latest trends on heart disease prediction using machine learning and image fusion", 2020, pp. 1-6.

[35] Dhyan Chandra Yadav, Veer Bahadur Singh, "Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques: Heart Disease Prediction", International Journal of Big Data and Analytics in Healthcare, 2021, pp. 40-56.

[36] J. Nageswara Rao, R. Satya Prasad, "An Enhanced Novel Dynamic Data Processing (ENDDP) Algorithm for Predicting Heart Disease in Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2021, pp. 94-100.

[37] Alaa Khaleel Faieq, Maad M. Mijwil, "Prediction of heart diseases utilising support vector machine and artificial neural network", Indonesian Journal of Electrical Engineering and Computer Science, 2022, pp. 374-380.

[38] Elias Dritsas, Maria Trigka, "Stroke Risk Prediction with Machine Learning Techniques", Sensors 2022, pp. 1-13.

[39] M. Balakrishnan, A. B. Arockia Christopher, "Prediction of Cardiovascular Disease using Machine Learning", Journal of Physics: Conference Series, 2020, pp. 1-9.