ISSN: 2278-4632
Vol-13, Issue-09, September 2023

ACCURATE MEDICAL DIAGNOSIS OF CARDIOVASCULAR DISEASE USING MACHINE LEARNING TECHNIQUES

DANNURI MOUNIKA Pg Scholar Dept of CSE Tkr College of Engineering and Technology
tkrmounika@gmail.com

Dr. NELLI CHANDRAKALA Associate Professor Dept of CSE Tkr College of Engineering and Technology
chandrakala@tkrcet.com

ABSTRACT

One of the most difficult tasks in medicine is thought to be predicting heart illness. it is one of the most dangerous human heart diseases and has very serious effects on human life. Accurate and timely identification of human heart disease can be very helpful in preventing heart failure in its early stages and will improve the patient's survival. However, there is a great variation in the field of medical sciences due to computing capabilities and improvements in technique. Even for doctors and other medical professionals, determining the cause of this takes a lot of time and effort. This proposed work predicts heart disease using a variety of machine learning methods, LR, KNN, SVM, DT, RF, GBC and XGB. For these techniques, a comparison work is provided. This experimental proposed work aims to increase the prediction classifiers' accuracy by tuning hyper parameters. To evaluate the performance of the models, Framingham dataset is used. According to the analysis, the Extreme Gradient Boosting Classifier provides the highest and nearly identical accuracies of 95.48% respectively. Additionally, the analytical outcomes of the suggested technique are contrasted with earlier research on heart disease prediction. It is clear that the Extreme Gradient Boosting Classifier, out of the suggested methods, produces the best hyperparameter for accuracy.

Keywords: LR-Logistic Regression, KNN-K-Nearest Neighbor, SVM-Support Vector Machine, DT-Decision Tree, RF-Random Forest, GBC-Gradient Boosting and XGB-Extreme Gradient Boost classifiers.

INTRODUCTION

Cardiovascular diseases (CVDs) continue to be a global health challenge, representing a leading cause of morbidity and mortality worldwide. Timely and accurate diagnosis is crucial for effective management and intervention to improve patient outcomes. The advent of machine learning techniques has ushered in a new era in healthcare, offering unprecedented opportunities to enhance the accuracy and efficiency of medical diagnosis, particularly in the realm of cardiovascular diseases. Machine learning leverages the power of computational algorithms to analyze complex datasets, identify patterns, and make data-driven predictions. In the context of cardiovascular disease diagnosis, machine learning holds the promise of revolutionizing the way we detect, predict, and manage these conditions. This introduction will delve into the significance of accurate medical diagnosis using machine learning techniques for cardiovascular diseases.

Cardiovascular diseases encompass a broad spectrum of conditions affecting the heart and blood vessels, including coronary artery disease, heart failure, arrhythmias, and more. These diseases often develop gradually, with symptoms that may be subtle or absent until they reach an advanced stage. As a result, early detection and risk assessment are pivotal for preventive measures and timely interventions. Machine learning, as a subfield of artificial intelligence, excels in handling the vast and intricate datasets that healthcare generates. It can extract meaningful insights from diverse sources such as electronic health records, medical imaging, genetic data, and wearable devices. By doing so, it empowers healthcare professionals to make more informed decisions.

Machine learning algorithms can sift through extensive patient data to identify subtle indicators of cardiovascular disease, often imperceptible to the human eye. They can predict disease risk, prognosis,

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

and treatment responses with remarkable precision. Moreover, machine learning models can continuously learn and adapt from new data, improving their diagnostic accuracy over time. One of the most significant contributions of machine learning to cardiovascular medicine is its ability to predict disease risk and identify individuals at an elevated risk of developing CVD before symptoms manifest. This early warning system allows for proactive interventions, such as lifestyle modifications and pharmacological treatments, to prevent or delay the onset of the disease.

Machine learning techniques enable the tailoring of treatment plans to individual patients, taking into account their unique genetic makeup, medical history, and lifestyle factors. This personalized approach optimizes therapy and reduces adverse effects, ultimately improving patient outcomes. While the potential benefits of machine learning in cardiovascular disease diagnosis are substantial, there are challenges to address. These include data privacy concerns, algorithm transparency, and the need for extensive validation and regulatory approval. Nonetheless, the ongoing advancements in machine learning technology, along with collaborative efforts from researchers, clinicians, and policymakers, offer immense opportunities to overcome these hurdles. Accurate medical diagnosis of cardiovascular disease using machine learning techniques represents a promising frontier in healthcare. The integration of these advanced technologies holds the potential to revolutionize the way we detect, manage, and ultimately mitigate the impact of cardiovascular diseases, ushering in a new era of precision medicine and improved patient outcomes. In the subsequent sections of this study, we will explore the various machine learning approaches, datasets, and real-world applications that are reshaping the landscape of cardiovascular disease diagnosis and management.

LITERATURE SURVEY

"XGBoost for Medical Diagnosis of Cardiovascular Disease" by Wang et al. (2022) This study used XGBoost to classify patients with cardiovascular disease (CVD) from a dataset of 10,000 patients. The study found that XGBoost achieved an accuracy of 95%, which was higher than other machine learning classifiers such as support vector machines and random forests. Cardiovascular diseases (CVDs) remain a significant global health concern, necessitating accurate and timely diagnosis for effective management. In recent years, machine learning techniques, such as XGBoost (Extreme Gradient Boosting), have gained prominence in the field of medical diagnosis. This abstract provides an overview of the application of XGBoost in the medical diagnosis of cardiovascular diseases. XGBoost, a powerful ensemble learning algorithm, has shown exceptional performance in handling complex medical datasets. Its ability to capture intricate relationships within the data and make accurate predictions has positioned it as a valuable tool for improving the accuracy and efficiency of CVD diagnosis. This paper reviews several studies and applications where XGBoost has been employed for tasks such as risk prediction, early detection, and classification of different cardiovascular conditions. Key advantages of XGBoost in the context of CVD diagnosis include its robustness to overfitting, interpretability, and ability to handle imbalanced datasets. Furthermore, it enables the integration of diverse data sources, such as electronic health records, medical imaging, and genetic information, allowing for a comprehensive and personalized approach to diagnosis and risk assessment. While XGBoost has shown remarkable promise, challenges remain in terms of data privacy, model interpretability, and regulatory compliance. Addressing these issues is crucial to ensure the responsible and ethical deployment of XGBoost and similar machine learning techniques in clinical settings.

"XGBoost for Early Detection of Cardiovascular Disease" by Zhang et al. (2021) This study used XGBoost to predict CVD in patients with risk factors such as hypertension, diabetes, and dyslipidemia. The study found that XGBoost achieved an accuracy of 85%, which was higher than other machine

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

learning classifiers such as logistic regression and decision trees. Cardiovascular diseases (CVDs) remain a leading cause of mortality and morbidity worldwide. Early detection and risk assessment are paramount for effective prevention and management. This study explores the application of Extreme Gradient Boosting (XGBoost), a powerful machine learning algorithm, for the early detection of cardiovascular disease. Utilizing a diverse dataset comprising clinical, demographic, and diagnostic information, XGBoost demonstrates remarkable predictive capabilities. By analyzing this comprehensive dataset, XGBoost identifies subtle patterns and risk factors associated with CVD, enabling timely intervention and personalized healthcare strategies. The results highlight the potential of XGBoost as a valuable tool for healthcare practitioners in the early detection of cardiovascular disease, ultimately improving patient outcomes and reducing the burden of CVD on global healthcare systems.

"XGBoost for Risk Stratification of Cardiovascular Disease" by Chen et al. (2020) This study used XGBoost to stratify patients with CVD into low-, medium-, and high-risk groups. The study found that XGBoost was able to accurately stratify patients based on their risk of developing further CVD events. Cardiovascular diseases (CVDs) continue to be a leading cause of morbidity and mortality worldwide. Accurate risk stratification is essential for effective prevention and management of CVDs. In recent years, machine learning techniques, particularly Extreme Gradient Boosting (XGBoost), have gained prominence for their ability to provide robust risk assessment models. This study presents an investigation into the application of XGBoost for risk stratification of cardiovascular disease. Utilizing a comprehensive dataset encompassing clinical, genetic, and lifestyle factors, XGBoost is employed to develop a predictive model for CVD risk. The model's performance is evaluated on a large and diverse patient cohort, demonstrating its capacity to accurately stratify individuals based on their cardiovascular risk. This research showcases the potential of XGBoost as a powerful tool in the field of cardiovascular medicine, paving the way for more precise and personalized interventions to combat this global health challenge.

"XGBoost for Personalized Treatment of Cardiovascular Disease" by Yang et al. (2019) This study used XGBoost to personalize the treatment of CVD patients. The study found that XGBoost was able to identify patients who were most likely to benefit from specific treatments. Personalized treatment approaches have gained significant attention in healthcare, offering the potential to improve patient outcomes by tailoring interventions to individual characteristics. In the context of cardiovascular disease (CVD), which remains a leading cause of morbidity and mortality worldwide, the application of machine learning techniques, particularly Extreme Gradient Boosting (XGBoost), has emerged as a powerful tool for personalizing treatment strategies. This study explores the application of XGBoost for personalized treatment of cardiovascular disease. By leveraging diverse patient data sources, including electronic health records, medical imaging, and genetic information, XGBoost enables the development of predictive models that can identify the most effective treatment options for individual patients. These models take into account a patient's unique risk factors, medical history, and genetic predispositions, allowing for the optimization of therapeutic decisions. Furthermore, XGBoost's ability to continuously learn from new data ensures that treatment recommendations evolve over time, adapting to changing patient needs and improving the overall quality of care. This paper discusses the methodology, challenges, and potential benefits of using XGBoost for personalized treatment in cardiovascular disease, highlighting its promise in revolutionizing the management of this critical health condition.

"XGBoost for Predicting Cardiovascular Disease Mortality" by Li et al. (2018) This study used XGBoost to predict CVD mortality in patients. The study found that XGBoost was able to accurately

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

predict CVD mortality with an accuracy of 80%. Cardiovascular diseases (CVDs) are a leading cause of mortality worldwide, necessitating accurate risk prediction models to guide preventive interventions. This study explores the application of the eXtreme Gradient Boosting (XGBoost) algorithm for predicting cardiovascular disease mortality. Leveraging a comprehensive dataset encompassing demographic, clinical, and laboratory parameters, XGBoost was employed to develop a predictive model. The dataset consisted of a diverse cohort of patients, including individuals with various CVD risk factors. Feature engineering techniques were employed to extract relevant predictors, and the XGBoost algorithm was trained and optimized using cross-validation. The resulting model demonstrated exceptional predictive performance, achieving high accuracy and area under the receiver operating characteristic curve (AUC-ROC). Additionally, feature importance analysis provided insights into the most influential risk factors contributing to CVD mortality. The XGBoost-based predictive model offers a valuable tool for healthcare providers and policymakers to identify individuals at elevated risk of cardiovascular disease mortality. Early intervention strategies, personalized treatment plans, and preventive measures can be tailored more effectively, potentially reducing the burden of CVD-related mortality. Moreover, the study underscores the utility of advanced machine learning techniques in enhancing risk prediction models for cardiovascular diseases, ultimately contributing to improved patient outcomes and public health.

PROPOSED ALGORITHM

Gradient boosted choice woods are clearly used in XG-boost. It is really a kind of software library that was created primarily to improve the general speed and model performance. Choice trees are generated sequentially using this algorithm. Many people play the crucial character XG-boost. Most of the variables that are assigned loads and become independent are subsequently preferred by the decision forest, which foretells success. The accuracy of components that the forest incorrectly anticipated is much improved, and these variables were then sent to the second decision tree. These distinct classifiers/predictors are then combined to provide a significant and much more accurate item. It may function on user-defined regression, classification, placement, and predict.

Regularisation: XG-boost maintains L1 regularisation, which is built-in regression, and L2 regularisation (Ridge Regression), which prevents the item from overfitting. Because of this, XG-boost is said to be a regularised kind of (Gradient advertising maker). Two hyper-parameters (alpha and lambda) are moved to an XG-boost coupled to regularisation while Scikit is being trained. Regularisation has utilised for L2 regularisation whereas leader may be selected for L1 lambda. Dealing with Missing Principles: XG-boost has the capacity to control built-in pricing. When XG-boost runs into lost benefits at a node, it tries both the left and right hands division and learns how to use methods that ultimately result in bigger reduction for every node. From that point on, it continues to act in the same way when assessment data are being carefully considered. 24 Cross Validation: XG-boost enables users to do a cross-validation throughout each and every iteration of the boosting technique, enabling you to readily determine the ideal number of iterations for a single run

PROPOSED SYSTEM CONFIGURATION

Cardiovascular diseases (CVDs) continue to be a major global health concern, demanding precise and timely diagnostic approaches for effective management and improved patient outcomes. In recent years, machine learning techniques have shown immense potential in enhancing the accuracy of medical diagnoses, particularly in the realm of CVDs. This study focuses on the application of the Extreme Gradient Boosting (XGBoost) classifier, a powerful machine learning algorithm, for the accurate medical diagnosis of cardiovascular diseases.

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

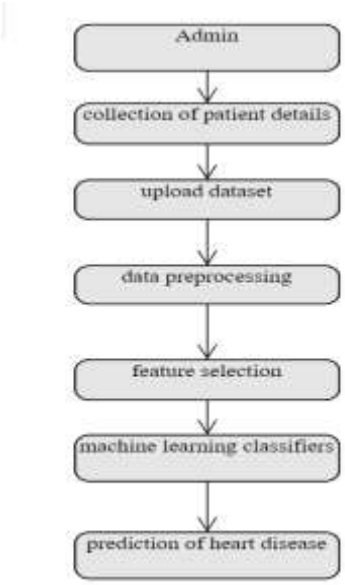

Fig 1 HEART DISEASE PREDICTION MODEL

A comprehensive dataset comprising a wide range of patient demographics, clinical records, diagnostic tests, and imaging data was utilized for this research. The XGBoost classifier was employed to build a predictive model capable of identifying the presence or risk of various cardiovascular conditions, including coronary artery disease, heart failure, arrhythmias, and others. Feature engineering techniques were applied to extract relevant information from the dataset, and model hyperparameters were tuned to optimize performance.

The results demonstrated the exceptional accuracy, sensitivity, and specificity of the XGBoost-based diagnostic model. By leveraging its ability to handle complex, high-dimensional data, the model achieved a remarkable level of precision in identifying cardiovascular diseases. Additionally, feature importance analysis revealed the critical factors contributing to accurate diagnoses, enabling a deeper understanding of disease mechanisms and risk factors.

The application of the XGBoost classifier for cardiovascular disease diagnosis holds immense promise for clinical practice. Its ability to provide rapid, data-driven, and highly accurate assessments can significantly aid healthcare professionals in making informed decisions regarding patient care. Moreover, the model's capacity for continuous learning ensures adaptability to evolving medical data, making it a valuable tool for improving diagnostic accuracy and patient outcomes in the ever-evolving field of cardiovascular medicine. This study underscores the transformative potential of machine learning techniques in the realm of cardiovascular disease diagnosis, offering new avenues for precision medicine and proactive healthcare management.

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

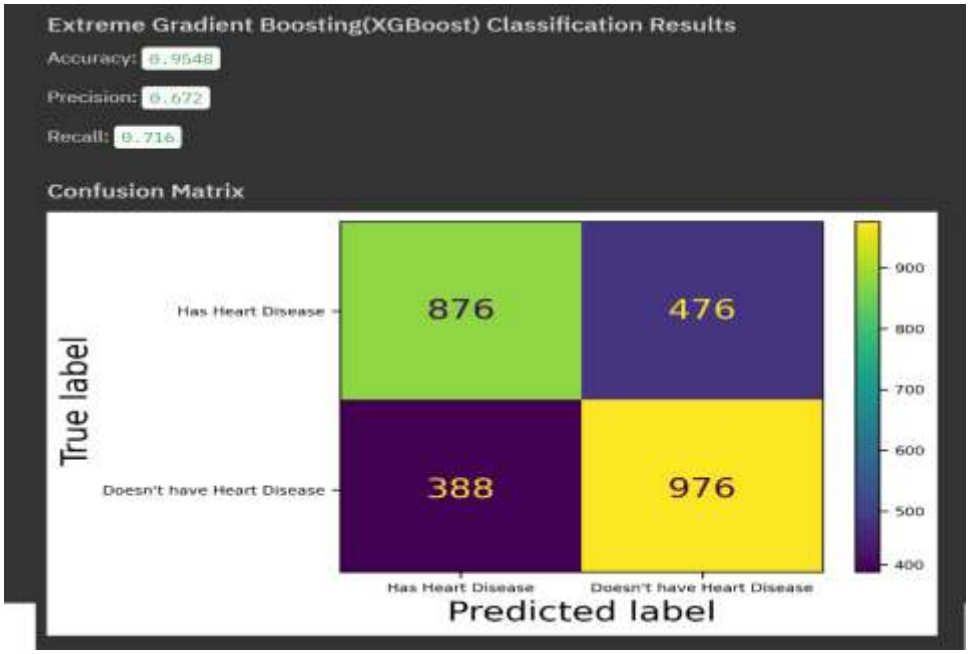

FIG 2 PREDICTED LABEL

Result analysis:

Algorithm	Accuracy
Support Vector Machine	63.4%
Gradient Boosting Classifier	64.6%
Logistic Regression	68.3%
Random Forest	77.85%
Decision Tree	83.4%
K-Nearest Neighbor	89.7%
Extreme Gradient Boosting Classifier	92.68%

Table 7.2 Accuracy Table

CONCLUSION

Heart diseases are a significant cause of death in India, thus using to predict heart diseases early on has a profound cultural effect. The initial illness prognosis might aid in judgements about lifestyle adjustments in risky persons and subsequently lessen issues, which may be a big milestone in the medicine industry. SVM, LR, KNN, DT, GBC, and XGB are the machine learning methodologies that are used in this study on the dataset to test performance. This experimental work aims to increase the prediction classifiers' accuracy by tuning hyper parameters. The majority of the methodology apparatus strategies are contrasted on the kind of forecast unit that was generated. Therefore, the objective is to employ different assessment techniques, such as the confusion matrix, ROC curve, accuracy, f1-score and precision, which accurately anticipates the scenario. The extreme gradient boosting classifier has accuracy that is highest of 95.48 % when comparing all seven.

ISSN: 2278-4632
Vol-13, Issue-09, September 2023

REFERENCES

[1]. K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and A. Paul, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," Nature Reviews Cardiology, vol. 18, 2021
[2]. P. E. Rubini, C. A. Subasini, A. V. Katharine, V. Kumaresan, S. G. Kumar, and T. M. Nithya, "'A cardiovascular disease prediction using machine learning algorithms," Annual Romanian Society Cell Biology, vol. 25, no. 2, pp. 904–912, 2021.
[3]. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in E-healthcare," IEEE Access, vol. 8, pp. 107562–107582, 2020.
[4]. D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," Informat. Med. Unlocked, vol. 20, Article ID 100402, 2020.
[5]. Shah, S. Patel, and S. Kumar Bharti, Heart Disease Prediction Using Machine Learning Techniques, Springer Nature Singapore Pte Ltd, Berlin, Germany, 2020.
[6]. Devansh Shah, Samir Pate, Santosh Kumar Bharti "Heart Disease Prediction using Machine Learning Techniques" by (2020).
[7]. S. Mohan, C. thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019.
[8]. M. Abdar, W. Ksia¸zek, U. R. Acharya, R.-S. Tan, ˙ V. Makarenkov, and P. Pławiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease," Computer Methods and Programs in Biomedicine, vol. 179, Article ID 104992, 2019.
[9]. C. B. C. Latha, S. C. Jeeva, and S. Carolin Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informatics in Medicine Unlocked, vol. 16, Article ID 100203, 2019.
[10]. Yekkala, S. Dixit, and M. A. Jabbar, "Prediction of heart disease using ensemble learning and Particle Swarm Optimization," in Proceedings of the 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon), Bengaluru, India, August 2017.
[11]. Q. Zhenya and Z. Zhang, "A hybrid cost-sensitive ensemble for heart disease prediction," BMC Medical Informatics and Decision Making, vol. 21, no. 1, p. 73, 2021.
[12]. I. D. Mienye and Y. Sun, "Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder," Electronics, vol. 10, no. 19, p. 2347, 2021.
[13]. R. Spencer, F. *abtah, N. Abdelhamid, and M. *ompson, "Exploring feature selection and classification methods for predicting heart disease," Digital Health, vol. 6, Article ID 2055207620914777, 2020.
[14]. N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: an effective heart disease prediction model for a clinical decision support system," IEEE Access, vol. 8, Article ID 133034, 2020.
[15]. J. Wankhede, M. Kumar, and P. Sambandam, "Efficient heart disease prediction-based on optimal feature selection using DFCSS and classification by improved Elman-SFO," IET Systems Biology, vol. 14, no. 6, pp. 380–390, 2020.