IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Impact Factor: 5.731
Volume 5, Issue 12, December 2025

A Survey on Predicting Heart Disease with Machine Learning Techniques

Prof. Sumarani H and Dr. Sreedhar Kumar S
Assistant Professor, Department of CSE
Professor, Department of CSE
C Byre Gowda Institute of Technology, Kolar, Karnataka, India
SIR MVIT, Bengaluru, Karnataka, India
sumaemscbit@gmail.com sree.dr.2018@gmail.com

Abstract: *In datasets, but most of these algorithms easily tend to overfit. The models created often overfit the test is simple to do and can save lives. Recent progress has improved the sensitivity of heart failure detection on test datasets while maintaining this accuracy on test the arteries that feed the heart muscle become blocked, this causes coronary heart disease as blood low in oxygen is prevented from being efficiently delivered to the heart. Early detection of heart disease is possible and it can save on long-term health care costs ó press on the graphic to enlarge it mortality. When health worldwide. Early identification and precise diagnosis of cardiac diseases are important for improving patient outcome and decreasing Abstract Hundreds of thousands of people die each year of heart disease, making it a major concern in public machine learning methods, including DT, RF, SVM, and PCA, are utilized. acquire rapid and accurate diagnoses. In this paper, supervised study, we introduce a novel diagnostic system to solve this problem, which is high-intelligent along with high-performance for both the training and testing datasets. Machine Learning (ML) techniques have already demonstrated high potential in assisting health care providers this.*

Keywords*:* Coronary heart disease, Blood pressure, Healthcare, Machine Learning, Imbalanced data, Decision tree, Random Forest, SVM, K-means Clustering, KNN

I. INTRODUCTION

The early identification and precise diagnosis of cardiac disease improve patient outcomes and lower mortality rates by enabling timely intervention and care. Quick recognition allows for prompt action, often preventing the progression of the disease and its related complications. This practice enables healthcare providers to tailor treatment plans, manage risk factors, and enhance long-term disease management. Additionally, it results in cost savings and has a positive impact on public health by alleviating the overall burden of cardiovascular disease and mortality. Historically, diagnosing heart disease has relied heavily on clinical assessments.

Recent advancements in machine learning (ML) and the availability of extensive healthcare datasets have empowered healthcare professionals to identify and diagnose cardiac conditions with greater accuracy. ML algorithms can autonomously extract key information and generate predictions based on data patterns and correlations. These algorithms have the potential to reveal trends associated with heart disease within substantial volumes of medical, imaging, and genetic data [4]. This paper aims to provide a comprehensive review of the existing literature on the detection of heart disease through machine learning methodologies. The study intends to enhance understanding and promote progress in this vital research area by assessing and summarizing the methodologies, datasets, performance metrics, and challenges faced in developing reliable prediction models [5, 6]. In this paper, we will begin with a review of relevant literature on heart disease detection using ML techniques.

The dataset we are examining is the Framingham dataset. We will review various machine learning methods previously applied, including both supervised techniques [7]. Additionally, we will evaluate the dataset used for training and testing, focusing on its diversity and representativeness [8]. We will also explore the data collection and preprocessing methods utilized in studies related to heart disease detection. Key topics such as handling missing data, addressing outliers, feature selection, and considerations for data augmentation and balancing class distribution will be analyzed, as these elements greatly impact the performance and generalizability of machine learning models [9, 10]. Evaluating performance is crucial for developing effective heart disease detection models. In our study, we will examine several criteria Among these metrics are the area under the receiver operating characteristic curve, accuracy, confusion matrix, sensitivity, specificity, precision, recall, F1-score, and F-score. Furthermore, we will explore cross- validation techniques and model selection strategies to ensure accurate and dependable predictions

II. LITERATURE SURVEY

Forecasting Cardiovascular and Liver Diseases Using a Hybrid Machine Learning Algorithm that Integrates Particle Swarm Methods like particle swarm optimization (PSO) and feature selection techniques have been explored to enhance classification accuracy. The proposed hybrid model aims to improve classification performance compared to existing approaches. Various metrics, including classification accuracy, error rate, correctness, recall, and F1 score, are used to evaluate the efficacy of the hybrid model in this study, which combines datasets from the UCI Machine Learning Repository. The integration of machine learning classifiers and feature engineering based on entropy is employed to predict cardiovascular diseases, as noted by Rajkamal Rajendran et al. [12]. The experimental results demonstrate that the suggested ensemble model (logistic regression + naive Bayes) achieves superior outcomes when implementing the new pipeline. Metrics such as AUC, accuracy, specificity, precision, and F1 score reflect the ensemble model's effectiveness in accurately diagnosing cardiac conditions.

In their paper 'A systematic literature review of machine learning applications for diagnosing cardiovascular diseases', Md. Manjurul It and colleagues [13] examine how well machine learning algorithms work when the dataset is unbalanced, with one class being significantly more prevalent. It emphasizes how important it is to identify patterns, methods, opportunities, and gaps in the use of machine learning for the identification of heart disease. Furthermore, a predictive study of cardiac diseases utilizing machine learning approaches is presented by Ramesh TR et al. [14]. They address the difficulties posed by dimensionality and high- dimensional data, and they offer comprehensive insights into the training and testing procedures for machine learning models.

To identify people who are at high risk, a variety of machine learning algorithms are used in the categorization and prediction of cardiovascular illnesses. Heart disease may be caused by a number of factors, including blood pressure, diabetes, cholesterol, obesity, tobacco use, and family history. Alireza Ghasemieh et al. created a novel machine learning model based on Stacking Ensemble Learner to predict the requirement for emergency hospital readmission among cardiac patients [15]. An XgBoost-based stacking ensemble learner (SEL) that integrates ensemble learning techniques was used in this investigation. Achieving the maximum detection efficiency is the aim.

This study's contributions include the creation of the SEL model, the use of a sizable, private dataset from the MIT Laboratory for Computational Physiology, and a unique behavioral trait-based classification method for emergency readmission. Tiwari et al. have designed and trained a ConvNet model utilizing Hybrid Constant-Q Transform (HCQT) to categorize cardiac sounds [19]. Each phonocardiogram signal's acoustic characteristics, including its major MFCC features, are fed into five-layer regularized ConvNets, which extract the CQT, HCQT, and Variable-Q Transform (VQT). This is the first use of HCQT for PCG signals, according to a survey of the pertinent literature.

Because of its ability to withstand overfitting, the method applied here has enhanced the ability to handle non-linear connections in complex datasets. The study in [23] focused on K Nearest Neighbors and Logistic Regression, predicting cardiac disease using machine learning approaches with an accuracy of 88.5%. Despite its ease of use and interpretability, logistic regression may have trouble interpreting complicated feature interactions and non-linear decision boundaries. In the study conducted in [24], the CQT, Variable-Q Transform (VQT), and HCQT are retrieved from the Support Vector Machine (SVM) features that are fed into five-layer regularized ConvNets.

This is the first use of HCQT for PCG signals, according to a survey of the pertinent literature.approach, which in their study focused on the prediction of heart disease had an accuracy of 64.4%. Effective modeling of complex interactions is ensured by the SVM method's flexibility in kernel selection and added advantage in handling high- dimensional data. With an accuracy of 87%, K Nearest Neighbors (KNN) was used in [25] to predict heart disease. Among KNN's advantages are its simplicity and lack of a training step, which make it both beginner-friendly and computationally efficient for producing predictions. In [26], a web-based method for predicting cardiac illness was suggested using the Decision Tree algorithm, which achieved the maximum accuracy rate of 99%.

Its advantages include its capacity to handle non-linear interactions within complicated datasets and its interpretability, which facilitates communication with stakeholders and medical professionals. Artificial Neural Networks (ANN) were used in an article in [27] to predict cardiac disease, and the accuracy was 92.30%. ANN's benefits include its capacity to recognize intricate non-linear relationships and extract key characteristics from unprocessed data. A hybrid framework called HRFLM was used in [28] to improve the accuracy of heart disease prediction to 88.4% by combining Random Forest and linear models. This method improves efficiency and resilience against overfitting by utilizing the advantages of both models. [29] highlighted the significance of using the K Nearest Neighbors (KNN) method.

First, Random Forest Regression and K Nearest Neighbors are two models that are commonly used. Furthermore, overfitting, complexity, and computer resource requirements are the main problems. Moreover, a hybrid approach that combines random forest and linear models is available to improve the prediction accuracy of cardiac disease. In the end, interpretability, adaptability, simplicity, and striking a precise balance between accuracy and model complexity are the most important criteria to take into account when choosing a suitable model for predicting cardiac disorders.

Researchers are examining many models to find the most reliable and accurate ones since the use of machine learning in heart disease prediction has greatly progressed. The literature now in publication shows that ensemble models, like Random Forest, are incredibly powerful. For complex datasets that might contain missing values, the combination of several decision trees makes a great model. Random Forest's strength is its capacity to reduce overfitting, a problem in which models perform well on training data but badly on unknown data. As a result, Random Forest has become a strong contender for the prediction of heart disease [38]. Another model that is often used in research investigations is K-Nearest Neighbors (KNN). It doesn't need a training phase and is quite simple. But KNN has its other information for forecasting results. Logistic regression is another popular model that is well- known for being simple to understand. It has trouble, nevertheless, with non-linear feature connections. Although logistic regression is a useful tool for simpler situations, it may not work well for more complicated patterns, which are frequently found in data related to heart disease. Support Vector Machines (SVM) typically produce better outcomes in these situations [30] [31]. In particular, SVM works especially well in these circumstances.

To take advantage of their unique characteristics, other researchers have begun combining several machine learning models. For instance, accuracy has increased when Random Forest and Linear methods are combined. Random Forest provides simplicity, interpretability, and robustness for heart disease prediction when combined with linear models. Decision trees, on the other hand, are simple and easy to comprehend. This quality is especially helpful in therapeutic settings where openness is essential since it makes decision-making easier. However, decision trees might overfit, especially when dealing with tiny datasets. Techniques like pruning are commonly used to lessen this difficulty [31]. All of the data points to the lack of a single method for predicting heart disease. The dataset determines which model is used.

III. CONCLUSION

Through extensive trials, machine learning models are evaluated to predict heart illnesses, confirming the efficacy of machine learning approaches in offering significant insights into the potential of predictive analytics in clinical contexts. With an accuracy of up to 91.51%, the Support Vector Machine (SVM) model is remarkably accurate [17]. Each model's accuracy scores demonstrate the potential of advanced computational methods in healthcare, guaranteeing well-informed choices and enhanced patient outcomes via predictive analytics. In order to promote early diagnosis and self- management of cardiovascular diseases, our research, which demonstrates positive and faultless outcomes from our predictive analysis, calls for the integration of these predictive models into routine clinical procedures. healthcare practitioners, with the use of machine learning.

REFERENCES

[1]. Coffey, S., Roberts-Thomson, R., Brown, A., Carapetis, J., Chen, M., Enriquez-Sarano, M., ... & Prendergast, B. D. (2021). Global epidemiology of valvular heart disease. Nature Reviews Cardiology, 18(12), 853-864.

[2]. De Hert, M., Detraux, J., & Vancampfort, D. (2022). The intriguing relationship between coronary heart disease and mental disorders. Dialogues in clinical neuroscience.
[3]. Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2022). Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation, 145(8), e153-e639.
[4]. Katarya, R., & Srinivas, P. (2020, July). Predicting heart disease at early stages using machine learning: A survey. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 302-305). IEEE.
[5]. Liu, A., Diller, G. P., Moons, P., Daniels, C. J., Jenkins, K. J., & Marelli, A. (2023). Changing epidemiology of congenital heart disease: effect on outcomes and quality of care in adults. Nature Reviews Cardiology, 20(2), 126-137.
[6]. Bhatt, C. M., Patel, P., Ghetia, T.
[7]. Aluru, J. S., Barsouk, A., Saginala, K., Rawla, P., & Barsouk, A. (2022). Valvular heart disease epidemiology. Medical Sciences, 10(2), 32.
[8]. Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. Artificial Intelligence in Medicine, 128, 102289.
[9]. Sekar, J., Aruchamy, P., Sulaima Lebbe Abdul, H., Mohammed, A. S., & Khamuruddeen, S. (2022). An efficient clinical support system for heart disease prediction using TANFIS classifier. Computational Intelligence, 38(2), 610-640.
[10]. Zhou, Y., Zhu, X., Cui, H., Shi, J., Yuan, G., Shi, S., & Hu, Y. (2021). The role of the VEGF family in coronary heart disease. Frontiers in cardiovascular medicine, 8, 738325.
[11]. Behera, M. P., Sarangi, A., Mishra, D., & Sarangi, S. K. (2023). A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine. Procedia Computer Science, 218, 818-827.
[12]. Rajendran, R., & Karthi, A. (2022). Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers. Expert Systems with Applications, 207, 117882.
[13]. Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. Artificial Intelligence in Medicine, 128, 102289.
[14]. Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, 132-148.
[15]. Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., & Kashef, R. (2023). A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. Decision Analytics Journal, 7, 100242.
[16]. Pathan, M. S., Nag, A., Pathan, M. M., & Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthcare Analytics, 2, 100060.
[17]. Wang, J., Rao, C., Goh, M., & Xiao, X. (2023). Risk assessment of coronary heart disease based on cloud-random forest. Artificial Intelligence Review, 56(1), 203-232.
[18]. Emmons-Bell, S., Johnson, C., & Roth, G. (2022). Prevalence, incidence and survival of heart failure: a systematic review. Heart, 108(17), 1351- 1360.
[19]. Tiwari, S., Jain, A., Sharma, A. K., & Almustafa, K. M. (2021). Phonocardiogram signal based multi-class cardiac diagnostic decision support system. IEEE Access, 9, 110710-110722.
[20] Absar, N., Das, E. K., Shoma, S. N., Khandaker, M. U., Miraz, M. H., Faruque, M. R. I., Pathan, R. K. "The Efficacy of Machine Learning- Supported Smart System for Heart Disease Prediction." Healthcare, vol. 10, no. 6, 2022, p. 1137.
[21] Chandrika, L., and Madhavi, K. "A Hybrid Framework for Heart Disease Prediction Using Machine Learning Algorithms." E3S Web of Conferences, vol. 309, EDP Sciences, 2021, p. 01043.
[22] Indra Kumari, R., Poongodi, T., and Jena, S. R. "Heart Disease Prediction Using Exploratory Data Analysis." Procedia Computer Science, vol. 173, 2020, pp. 130-139.
[23] Jagtap, A., Malewadkar, P., Baswat, O., and Rambade, H. "Heart Disease Prediction Using Machine Learning." International Journal of Research in Engineering, Science and Management, vol. 2, no. 2, 2019, pp. 352-355.

[24] Jindal, H., Agrawal, S., Khera, R., Jain, R., and Nagrath, P. "Heart Disease Prediction Using Machine Learning Algorithms." IOP Conference Series: Materials Science and Engineering, vol. 1022, no. 1, IOP Publishing, 2021, p. 012072.

[25] Rahman, M. M. "A Web-Based Heart Disease Prediction System Using Machine Learning Algorithms." Network Biology, vol. 12, no. 2, 2022, p.64.

[26] Sharma, V., Yadav, S., and Gupta, M. "Heart Disease Prediction Using Machine Learning Techniques." 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), IEEE, 2020, pp. 177- 181.

[27] Srivastava, K., and Choubey, D. K. "Heart Disease Prediction Using Machine Learning and Data Mining." International Journal of Recent Technology and Engineering, vol. 9, no. 1, 2020, pp. 212-219.

[28] Williams, R., Shongwe, T., Hasan, A. N., and Rameshar, V. "Heart Disease Prediction Using Machine Learning Techniques." 2021 International Conference on Data Analytics for Business and Industry (ICDABI), IEEE, 2021, pp. 118-123.

[29] Yaswanth, R., and Riyazuddin, Y. M. "Heart Disease Prediction Using Machine Learning T.

[30] Ingole, B. S., Patel, P., Mullankandy, S., & Talegaonkar, R. (2024). AI-Driven Innovation in Medicare: Revolutionizing Senior Care and Chronic Disease Management with Data-Driven Insights. IJRAR - International Journal of Research and Analytical Reviews (IJRAR), 11(3), 565-571.

[31] Patel, N. A., Suhas, S., Ingle, P. S., Patsamatla, S. K., Omotunde, H., & Ingole, B. S. (2024). Integration of Blockchain and AI for Enhancing Data Security in Healthcare: A Systematic Review. Library Progress International, 44(3), 2020-2029.

[32] Ingole, B. S., Ramineni, V., Krishnappa, M. S., & Jayaram, V. (2024). AI-Driven Innovation in Medicaid: Enhancing Access, Cost Efficiency, and Population Health Management. International Journal of Healthcare Information Systems and Informatics (IJHISI), 1(1), 9-17.

[33] Patel, P. (2024). Impact of AI on Manufacturing and Quality Assurance in Medical Device and Pharmaceuticals Industry. In International Journal of Innovative Technology and Exploring Engineering (Vol. 13, Issue 9, pp. 9–21).