IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 July 2024, accepted 15 July 2024, date of publication 22 July 2024, date of current version 30 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3430898

SURVEY

Exploring Predictive Methods for Cardiovascular Disease: A Survey of Methods and Applications

VALERU VISION PAUL[1] **AND JAFAR ALI IBRAHIM SYED MASOOD**[2]
[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
[2]Department of IoT, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

Corresponding author: Jafar Ali Ibrahim Syed Masood (jafarali.s@vit.ac.in)

ABSTRACT Because cardiovascular disease (CVD) is still one of the world's leading causes of death, sophisticated predictive models are required for early detection and prevention. This study examined how to make and compare different CVD prediction models using a large dataset that included biochemical, clinical, and demographic information about each person. During the preprocessing stage, we took great care to ensure the data's accuracy and quality. We have utilized a variety of machine learning algorithms such as random forest, logistic regression, support vector machines, and deep learning neural networks. We assessed the performance of these models using the accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). Our findings show that while more sophisticated algorithms—especially deep learning models—perform better at spotting possible instances of CVD, more conventional models—such as regression—offer significant predictive power. We also investigated the role that feature selection has in improving the interpretability and efficiency of the model. This study highlights the potential of machine learning to transform CVD prediction and emphasizes the importance of using many forms of data to provide a thorough risk evaluation. Our research adds to the continuing efforts in personalized medicine by providing information on creating more precise and effective predictive tools for cardiovascular health management.

INDEX TERMS Predictive modeling in cardiology, cardiovascular risk algorithms, biomarkers in CVD prognostication, computational cardiology, personalized cardiovascular risk models, predictive biometrics in cardiology, wearable technology for cardiovascular monitoring, genetic data analysis in heart disease prediction.

I. INTRODUCTION

To combat the global burden of cardiovascular disease (CVD), the scientific community is continually seeking innovative approaches to predict and prevent this leading cause of global morbidity and mortality. Trauma and stress are regular side effects of the modern lifestyle. In addition to those, additional risk factors for cardiovascular disease include high blood pressure, heart disease, obesity, smoking, type 1 or type 2 diabetes, and a family history of the condition. By recognizing modifiable risk factors and trying to change lifestyle-related risk factors into healthy ones, people can prevent the development of CVD. An early warning sign of a heart attack includes tiredness, sweat, slowness, dizziness, arm pain, and chest discomfort. Early symptom recognition can result in more effective therapies, frequently improving patient outcomes. This is the point at which early prediction becomes crucial. Medical practitioners can advise patients to adopt healthier lifestyles, provide required medicines, or even perform surgical procedures as necessary by forecasting the beginning or progression of heart disease. Our research represents a groundbreaking exploration into predictive modeling for CVD, distinguishing itself by applying novel, nontraditional predictive factors and sophisticated algorithms for machine learning.

This survey paper explores the myriad of predictive methods developed and applied in the battle against CVDs, delving into the heart of this burgeoning field. From traditional

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

statistical models to the latest deep learning architectures, we will navigate a landscape of innovative approaches, each with unique strengths and nuances. Through a comprehensive state-of-the-art analysis, we aim to unravel the intricate tapestry woven by these predictive models, uncovering the factors contributing to their efficacy, interpretability, and real-world applicability.

However, our journey extends beyond mere methodological exploration. We will also shed light on the diverse applications of these predictive models, illuminating their transformative potential across various domains, from risk stratification and early diagnosis to personalized treatment planning and disease management. By bridging the gap between theoretical frameworks and practical implementation, we aspire to empower healthcare professionals and researchers with a comprehensive understanding of the tools available, equipping them to make informed decisions and drive positive change.

In an era where data is key to unlocking unprecedented insights, this survey paper serves as a beacon, guiding us through the intricate landscape of predictive methods for cardiovascular diseases. By synthesizing the collective wisdom of researchers and practitioners, we aim to catalyze a paradigm shift, ushering in a new era of proactive, personalized, and precision-driven cardiovascular care where the power of prediction becomes an indispensable ally in safeguarding the health and well-being of individuals worldwide. These include unrecognized genetic sequences, emerging biomarkers, environmental exposures, and even socioeconomic factors woven together to construct a comprehensive predictive framework for CVD. Using these unique and unknown predictors and advanced modeling methods like neural networks and ensemble learning marks a significant departure from traditional cardiovascular research methods. Our work promises to enhance our understanding of CVD dynamics and usher in a new era of precision cardiology, where prevention strategies are as individualized as the patients are. Through this rare and unparalleled exploration, we aspire to make significant strides in reducing the global impact of cardiovascular diseases, charting a course toward a healthier future.

The primary purpose of this work is to analyze forecasts of heart disorders, with a focus on feature selection and classification enhancement. To address the worldwide issue of cardiovascular diseases (CVD), the leading cause of mortality worldwide, we surveyed the content of this Forum.

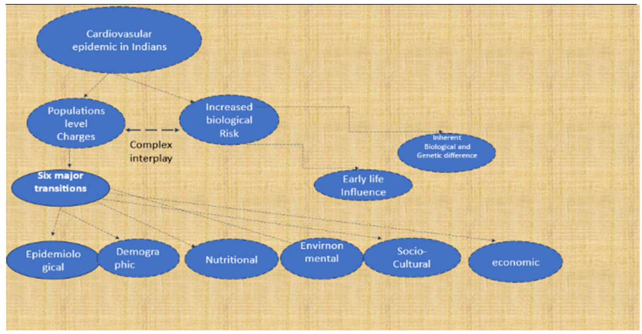

FIGURE 1. Cardiovascular epidemic in indians.

II. LITERATURE REVIEW

A. FOCUSING ON THE LAST THREE YEARS

Elsedimy et al. [27]: The paper describes a new method to identify heart disease. It uses a quantum-behaved particle swarm optimization (QPSO) algorithm and a support vector machine (SVM) classification model. This QPSO-SVM model aims to improve the accuracy of cardiovascular disease (CVD) predictions by optimizing feature selection and parameter tuning. They tested it on the Cleveland heart disease dataset and found it performs superior to other models. This paper mostly shows a new way to find heart disease by combining quantum-behaved particle swarm optimization (QPSO) with support vector machines (SVM). The goal is to determine cardiovascular disease (CVD) risk. The suggested QPSO-SVM model prepares the data, uses QPSO to find the best SVM parameters, and uses a self-adaptive threshold method to find the right balance between exploring and exploiting the solution search space. When applied to the Cleveland heart disease dataset, the QPSO-SVM model achieved high prediction accuracy and outperformed other state-of-the-art models in sensitivity, specificity, precision, and F1 score. This paper despite its advancements, the paper suggests the need for further research to address potential limitations, such as the risk of falling into local optima and improving detection rates, indicating a gap in developing more robust and efficient diagnostic models for CVD. The paper emphasizes the importance of early detection and accurate prediction of heart diseases using advanced machine-learning techniques, highlighting the continuous need for improvement in this critical area of medical diagnostics. The use of QPSO for feature selection in CVD prediction is innovative, as it enhances the SVM's ability to handle high-dimensional data.

The model's self-adaptive threshold method for parameter tuning is unique, helping to avoid local minima and ensuring a balance between exploration and exploitation. The main potential gap Despite the model's promising results, the Lack of discussion about its performance on datasets other than Cleveland's could undermine its generalizability. This research paper needs to mention how it would perform in real-world clinical settings where data may be noisier or more imbalanced.

Reshan et al. [25]: In this paper, we discuss the prevalence of heart disease and the importance of accurate prediction systems. Current Systems: Review existing prediction models and their limitations. Hybrid Deep Neural Networks (HDNNs) Emphasize the innovative approach of using HDNNs, which combine CNN, LSTM, and dense layers for heart disease prediction and highlight their performance. Emphasize the superior accuracy of HDNNs compared to traditional methods, as demonstrated on public datasets. While the HDNN system shows improved accuracy, there is an opportunity to explore its scalability and real-world applicability. The paper discusses designing a robust heart disease prediction system using hybrid deep neural networks (HDNNs), which combine multiple neural network architectures to enhance prediction accuracy. It looks at three deep learning models: Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). It also creates a new HDNN model that combines CNN and LSTM with more dense layers. This paper uses two publicly available heart disease datasets, including the Cleveland heart disease dataset, to test the proposed models. We measure the system's performance against conventional systems using various metrics such as sensitivity, Matthews Correlation Coefficient (MCC), F1-measure, accuracy, precision, AUC, and specificity. Assessing the system's performance across diverse demographic groups or integrating it with real-time patient data for dynamic prediction is a potential gap.

Additionally, evaluating the interpretability of the model's predictions could be another unique angle, as explainability is crucial in healthcare decision-making. The paper identifies a gap in applying deep learning approaches for heart disease prediction. It notes that existing studies have yet to apply deep learning techniques like deep ANN, LSTM, CNN, and HDNN, particularly CNN with LSTM models, for heart disease prediction. The proposed research aims to fill this gap by introducing an innovative HDNN approach that utilizes small and large datasets for effective model training and improved heart disease prediction.

Mondal et al.: This paper presents a novel dual-stage stacked machine learning model for predicting the risk of heart diseases, leveraging a dataset of 1190 patients with eleven significant characteristics. Five machine learning classifiers (Random et al. Machine) were used, with hyperparameter tuning via RandomizedSearchCV and GridSearchCV. The best-performing models were refined using a stacking ensemble technique. The stacking model achieved an accuracy of 96%, a recall value of 0.98, and a ROC-AUC score of 0.9623. The model demonstrated high stability and repeatability with a comparable dataset. The proposed model shows potential for accurate and prompt diagnosis of heart diseases, aiding in reducing global fatalities.

The main gap in this research While the paper focuses on achieving high accuracy, it needs more in-depth discussion about model interpretability. Future research could explore techniques to make the model's predictions more transparent and understandable for clinicians and patients. The study primarily relies on internal validation using the same dataset. External validation with independent datasets from different populations or healthcare settings would enhance the model's generalizability. However, the paper mentions eleven significant characteristics, but it does not delve into feature importance analysis. Understanding which features contribute most to the model's predictions is crucial for clinical decision-making. The paper must discuss practical implementation challenges or how the model could be integrated into existing healthcare systems. Addressing these aspects would bridge the gap between research and real-world application. Ethical implications of patient privacy, bias, and fairness should be thoroughly explored. Future work should address these concerns to ensure the responsible deployment of predictive models.

Baghdadi et al.: This paper discusses machine learning techniques for early detection and diagnosis of cardiovascular diseases (CVDs). A proposed Catboost model received an F1-score of 92.3% and an average accuracy of 90.94%, indicating high classification performance. In this paper, We conducted extensive data analysis, including exploratory data analysis and performance evaluation of machine learning models. This study's main contributions include the development of machine learning algorithms for CVD prediction and providing reliable advice to healthcare professionals. The paper discusses the use of machine learning algorithms for early detection and diagnosis of cardiovascular diseases (CVDs), emphasizing the potential to improve clinical practice. The study proposes novel algorithms for automatically selecting key features in early-stage heart disease detection, which could optimize early prediction and intervention. We have yet to extensively explore applying these machine-learning techniques to individual CVD survival prediction in hypertensive patients, which is the main gap. This paper addresses cardiovascular diseases, a leading cause of mortality globally. The research involves comprehensive data analysis, including feature selection and model evaluation. The study proposes a novel approach using machine learning to detect early-stage heart disease, which can lead to better patient outcomes. The paper introduces algorithms for the automatic selection of key features, which is crucial for the accuracy of predictive models. This paper contributes to the field by developing a machine-learning algorithm that could potentially transform the early detection and management of cardiovascular diseases.

Subramani et al. [29]: Focusing on This paper discusses cardiovascular disease (CVD) prediction using machine learning and deep learning techniques, emphasizing the need for AI technologies to predict outcomes in CVD patients. It presents a collection of machine learning models, highlighting their data observation mechanisms and training procedures and comparing them with traditional algorithms. The study incorporates machine learning and deep learning techniques to predict cardiovascular diseases (CVD). It uses a stacking model with a base and meta-learner layers. For feature selection, it uses the SHAP method and classifiers such as Random Forest (RF), Logistic Regression (LR), Multilayer Perceptron (MLP), Extra Trees (ET), CatBoost, and Gradient Boosting Decision Tree (GBDT). The research highlights the potential of AI-based technologies, particularly the Internet of Things (IoT), in enhancing the prediction of CVD outcomes. By accounting for complex data interactions, machine learning models can significantly improve over traditional statistical models. The proposed method achieved nearly 96% accuracy in predicting CVD, outperforming existing methods. According to the study, deep learning could benefit from more medical institution data for developing artificial neural network structures. The paper underscores the importance of advanced computational methods in medical research and their role in improving disease prediction and management. The study could benefit from additional data from a larger number of medical institutions to enhance the development of artificial neural network structures. In this paper, Future work could explore incorporating deep learning algorithms into Internet of Things (IoT) environments, which could increase accuracy in results and utility for hospitals. While the stacking model approach showed promise in addressing the main gap, there is room for exploring more sophisticated machine learning techniques that could offer better performance and interpretability in clinical settings.

Sarra et al. [31] The paper discusses machine learning (ML) advancements for heart disease prediction. It highlights using the Support Vector Machine (SVM) algorithm and a chi-squared (χ^2) statistical method for feature selection to enhance prediction accuracy. It references studies on the global impact of heart disease, emphasizing the need for accurate prediction and early diagnosis to reduce mortality rates. The paper underscores the significance of selecting relevant features in ML models to avoid overfitting and improve performance. We employ the method to identify the most informative attributes for heart disease prediction. The proposed SVM model with χ^2 feature selection is evaluated against traditional models using metrics like accuracy, sensitivity, specificity, and F1-score, demonstrating improved prediction capabilities. The paper shows a heart disease classification model that uses the Support Vector Machine (SVM) algorithm and is improved with a χ^2 statistical optimal feature selection technique to make predictions more accurate. It references studies on the global impact of heart disease, emphasizing the need for accurate prediction and early diagnosis to reduce mortality rates. The paper underscores the significance of selecting relevant features in ML models to avoid overfitting and improve performance. We employ the method to identify the most informative attributes for heart disease prediction. The proposed SVM model with χ^2 feature selection is evaluated against traditional models using metrics like accuracy, sensitivity, specificity, and F1-score, demonstrating improved prediction capabilities.

This review encapsulates the paper's focus on leveraging ML for better heart disease prediction, the critical role of feature selection, and the model's validation through performance metrics. It identifies gaps, such as the need for further research on feature selection methods and the model's application in real-world clinical settings. Despite improvements, the paper needs to extensively discuss the impact of feature selection on different population subsets or the model's scalability to larger, more diverse datasets, which could be crucial for real-world applications. This represents a potential area for further research and development.

Rahim et al. [32]: This paper presents the MaLCaDD framework, which uses highly precise machine learning techniques to predict cardiovascular diseases. It addresses missing values with mean replacement and data imbalances with SMOTE, ensuring data reliability for accurate predictions. The framework employs a feature importance technique for optimal feature selection, reducing computational complexity. Despite the framework's comprehensive approach, there is room for improvement in real-time data analysis and integration with wearable health monitoring devices for continuous cardiovascular risk assessment. The main gap in this paper, which extensively covers data preprocessing, feature selection, and prediction accuracy, is that it needs an exploration of real-time data integration with wearable health monitoring devices. Continuous cardiovascular risk assessment using wearable devices could enhance the framework's practical applicability.

Al-Alshaikh et al.: The study introduces a machine learning-based heart disease prediction model (ML-HDPM) that combines genetic algorithms, recursive feature elimination, and deep learning techniques to enhance prediction accuracy; the study introduces a prediction model using genetic algorithms and recursive feature elimination to enhance robustness and address data imbalance. The model achieves high performance with an accuracy rate of 95.5%, precision rate of 94.8%, sensitivity (recall) of 96.2%, and specificity of 89.7%. The model employs a multilayer deep convolutional neural network (MLDCNN) and adaptive elephant herd optimization method (AEHOM) for training. Techniques like under-sampling, clustering, and oversampling (USCOM) are used to address data imbalance, improving the model's predictive capabilities; the findings highlight the potential of ML-HDPM to transform heart disease prediction, aiding healthcare practitioners in providing precise diagnoses and improving patient care outcomes. The paper includes numerous references to previous research and methodologies in heart disease prediction. This paper identifies the main gap in researchers focused on predicting heart disease; the performance metric—specifically prediction accuracy—remains suboptimal. Improving the accuracy of heart disease prediction models is crucial to prevent life-threatening situations and ensure effective treatment. Some existing models lack external validation. It is essential to validate these models using independent datasets to assess their generalizability and robustness. Addressing imbalanced datasets is critical. Biased predictions can occur when the dataset has an unequal distribution of positive and negative cases. Researchers should explore techniques to handle data imbalance effectively. Most models focus on short-term prediction. Integrating clinical data (such as patient history, lifestyle, and comorbidities) into predictive models could enhance their accuracy and practical utility.

Guarneros-Nolasco et al.: The study aims to identify key risk factors for cardiovascular diseases (CVDs) using machine learning algorithms (MLAs) to improve diagnosis and prediction. The researchers analyzed the performance of ten MLAs on four datasets, focusing on top-two and top-four attributes for CVD prediction and diagnosis.

The study identified three main attributes (arrhythmia, tachycardia, and others) as significant for CVD diagnosis and prediction. The ML classifiers performed appropriately, suggesting their potential for enhancing CVD diagnosis, especially in regions with limited medical staff. After studying this paper, we Identify the gap, which paper focuses on identifying risk factors for cardiovascular diseases (CVDs) using machine learning algorithms (MLAs). However, it does not explore the impact of socioeconomic factors (such as income, education, and lifestyle) on CVD prediction. Investigating how socioeconomic variables interact with clinical attributes could enhance the accuracy of CVD prediction models. The paper lacks a discussion on ethical implications related to CVD prediction. Addressing issues like bias, fairness, and interpretability is crucial. The study evaluates MLAs on existing datasets. Including an external validation using real-world patient data would strengthen the findings. While the study identifies significant attributes, it does not consider feature importance. Understanding which features contribute most to prediction accuracy is essential.

DeGroat et al. [1]: Looking back at the last three years, new studies have used the power of statistical analysis along with advanced AI and machine learning methods to find transcriptomic biomarkers linked to cardiovascular diseases (CVDs) and create a reliable prediction model. To solve the problem of combining different kinds of data, the authors created the Clinically Integrated Genomics and Transcriptomics (CIGT) format. This makes it possible to combine transcriptomics, clinical, and demographic data without problems. They use algorithms to find important biomarkers, such as recursive feature elimination, Pearson correlation, chi-square, and ANOVA tests. The identified biomarkers undergo rigorous validation through a literature review and cross-referencing with patient clinical records. The study then employs a group of machine learning models, such as Random Forest, Support Vector Machine, XG Boost, and k-Nearest Neighbors, combined with a soft voting classifier to accurately predict the presence of CVD based on the identified biomarkers. This innovative methodology uncovers novel biomarkers and demonstrates the potential of AI and ML techniques in facilitating early disease detection and paving the way for personalized interventions in precision medicine. The intricate interactions between these variables can impact the development of various health issues. Standard statistical methods cannot fully account for the complex causal relationships between risk factors. This survey pinpoints a specific gap: their existing AI/ML model and approach can only predict cardiovascular disease in binary terms, contrasting it with a state of health. It cannot perform multiclass prediction and classification of different disease types and subtypes. The authors say we need fresh, unsupervised clustering methods that mix various kinds of genomic data, like transcriptomics, genomic variants, epigenomics, and patient demographics, to map and group patients with multiple diseases based on the key features of these mixed data sets.

Future work must address this gap, which limits us to binary prediction of a single disease. We must overcome this limitation to enable multiclass disease prediction for precision medicine applications.

The recursive feature elimination classifier prioritized transcriptome features based on their relationship to the case (''principal investigator'') and ''principle.''

Shobana and Bushra [2] proposed to explore the Ground tests for the NASA X-59 aircraft have demonstrated significant progress in the measurement of subdued sonic booms, thereby establishing a foundation for less noisy supersonic transportation. Scientists are making notable advancements in comprehending the correlation between human emotions and artificial intelligence, a field known as affective computing. Researchers are making strides in understanding how AI can recognize and respond to human emotions, leading to more empathetic and context-aware systems. This paper employed environmental DNA (eDNA). This study utilized eDNA techniques, efficiently detecting NISs even when their abundance is minimal. DNA is a genetic material that organisms release into the environment (such as water) and is useful for identifying different species. Golden Mussel Detection: This study focused on highly invasive golden mussels (Limonene fortunes). This paper proposes an eDNA tool for early detection of this mussel based on the mitochondrial cytochrome c oxidase subunit I gene (COI).

Siva Kumar and Pramod [3] This paper proposes sustainable aviation fuel (SAF): The industry contributes 2–3% of global CO_2 emissions. SAF, made from biological or nonbiological sources, aims to reduce aviation's carbon footprint. By 2040, SAFs need to meet 13–15% of aviation fuel demand to achieve net-zero emissions by 2050. These sensors allow real-time monitoring of plant health, soil conditions, and environmental factors. Data on crop growth and stress can enhance sustainable agriculture practices. AI models that can produce original text, audio, or image content are known as ''generative AI models.'' These models have the power to transform problem-solving and creativity completely. Innovations in gene editing, synthetic biology, and personalized medicine are revolutionizing molecular biology. These advancements could transform healthcare and address pressing global health challenges. Transforming real-time information processing: The emergence of cutting-edge AI The development of an edge computing device has addressed the constraints of traditional cloud computing. It allows for faster data processing at lower computational costs. The book ''Beyond Moore's Law: A Novel Approach to Creating Extremely Adaptable Electronics'' explores this concept.

These breakthroughs span various fields, from aviation and computing to materials science and energy. Exciting times lie ahead as technology continues to evolve. These studies reflect the dynamic and multifaceted nature of recent research in predictive modeling for CVD. These papers are at the cutting edge of efforts to make more accurate, personalized, and complete models for predicting CVD. They use various data types and advanced analytical methods, such as genomic and environmental factors, SDOH, and unstructured healthcare data. This progression enhances our understanding of CVD and paves the way for innovative prevention and intervention strategies tailored to individual risk profiles.

Asif et al. [4] In this paper, they used the UCI dataset for both model training and testing. This study offers a thorough examination of numerous machine-learning techniques for predicting CVD. We compare twelve distinct algorithms and assess their performance using the F1 score, ROC-AUC23, accuracy, precision, sensitivity, and specificity. This study highlights the need for a more precise prediction model for cardiovascular illnesses. While the paper achieves a high accuracy of 92% with ensemble voting classifiers, further research could focus on improving precision and specificity, particularly in practical healthcare applications where reducing false positives is crucial. The article also suggested comparing grid and random search techniques for hyperparameter optimization could improve computational efficiency. These findings provide opportunities for further research into more effective algorithms that can detect cardiovascular disease more quickly and accurately.

Zhang et al. [5] discusses an intelligent diagnosis system for cardiovascular disease using case-based reasoning (CBR). The gap identified is the need for improved diagnostic accuracy and reliability in cardiovascular disease diagnosis. This paper proposes a CBR-based system that utilizes pulse wave data and reflective learning to enhance the diagnostic process. The system aims to provide auxiliary decision-making support for doctors and reduce diagnostic errors. This gap is addressed by comparing different weight allocation methods for case diagnosis and demonstrating that introspective learning can significantly improve diagnostic performance. This paper suggests that further research is needed to optimize parameters such as the number of nearest neighbors in case retrieval and the threshold of matching attributes to enhance the method's effectiveness.

Sembina et al. [6]: This article discusses an intelligent diagnosis system for cardiovascular disease using case-based reasoning (CBR). The identified gap is a need for improved diagnostic accuracy and reliability in cardiovascular disease diagnosis. This paper proposes a CBR-based system that utilizes pulse wave data and reflective learning to enhance the diagnostic process. The system aims to provide auxiliary decision-making support for doctors and reduce diagnostic errors. We address this gap by comparing different weight allocation methods for case diagnosis and demonstrating that introspective learning can significantly improve diagnostic performance. To improve the method's effectiveness, this paper suggests further research to optimize parameters such as the number of nearest neighbors in case retrieval and the threshold of matching attributes.

Ghorashi et al. [7]: This research paper examines the practical application of these models in medical software to support early diagnosis and preventative healthcare. The report identified a research gap in the need for an improved case-based reasoning (CBR) system for diagnosing cardiovascular illnesses. The proposed system uses genetic algorithms and reflective learning to change the weights of attributes in the CBR process on the fly. This makes the diagnosis more accurate. This strategy seeks to prevent diagnostic errors and provide physicians with greater help when making decisions. In this paper, we have addressed this gap. This research introduces a more accurate and adaptive diagnostic technique to close the gap caused by the current limits of CBR systems.

Incidence. The goal is to improve the accuracy of disease diagnosis by identifying comparable symptoms.

Pasha et al. [8]: his work investigates using different machine learning algorithms to predict heart disease using a Kaggle dataset that includes characteristics such as age, sex, blood pressure, and cholesterol. The authors found that traditional methods like support vector machines (SVMs), K-nearest neighbor (KNNs), decision trees (DTs), and artificial neural networks (ANNs) do not work as well with large datasets. They did this by comparing TensorFlow Kera's performance with these algorithms. They suggested utilizing Tensor Flow Keras and an ANN to increase the prediction accuracy. The study's results demonstrate that the ANN model, particularly the binary model, achieved an accuracy of 85.24%, outperforming other algorithms. This paper details the steps involved in creating and training the model, including data splitting, model creation with specific activation functions, and training for a set number of epochs. The authors conclude that deep learning techniques, especially ANN with TensorFlow Kera's, provide a more accurate prediction of cardiovascular diseases than other machine learning techniques.

Additionally, the study examines the accuracy levels of the various algorithms and discusses the significance of F1 scores, precision, and recall in predicting disease. In this work, we use big datasets to predict cardiovascular illness and find gaps in the accuracy and performance of different machine learning systems. It states that methods like Decision Trees (DT), K-nearest neighbor (KNN), and Support Vector Machines (SVM) do not function well with large datasets. To increase prediction accuracy, the study investigates the use of tensor flow keras and artificial neural networks (ANN). This research analyzes the Kaggle dataset, which includes characteristics linked to heart disease, such as age, gender, blood pressure, cholesterol, etc. The goal is to improve the accuracy and consistency of the prognosis for cardiovascular disease.

TABLE 1. Summary of previous year literature.

Author name	journal-publish years	Research Insights	Problem Addressed	Technique used	D-merits/Limitations
E. I. Elsedimy	Springer-2023	New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization	improve the prediction of cardiovascular disease (CVD) by proposing a novel heart disease detection model	Data preprocessing, Qpso Algorithm, Svm classification, Self-adaptive Threshold	Model Generalization, Optimization Limitations, Dataset Specificity, Parameter Turning, comparative analysis
MANA SALEH AL RESHAN	IEEE Accesses -2023	A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks	This paper addressed the HD prediction Challenge, Complexity in Diagnosis, the need for Technology Aid, and Advancement in the DL System.	Deep Learning Models, Hybrid Architecture, Feature Learning, Data Preprocessing, Performance Evaluation.	Limited feature selection compatibility, Vulnerability to datasets with missing data, Lack of comprehensive testing on real-world datasets
SUBHASH MONDAL	IEEE Accesses-2023	An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches	Global Mortality, Effective Risk, Machine Learning Approaches, Dataset Utilization	Machine Learning Classifiers, Hyperparameter Tuning, Stacking Ensemble Technique	Dataset Limitations, Model Complexity, Validation Constraints, Future Work
Sivakannan Subramani	Frontiers in medicine-2023	Cardiovascular disease prediction by machine learning incorporation with deep learning	CVD Prediction Challenge, Machine Learning Integration, Deep Learning Potential, IoT and Healthcare	Machine Learning Models, Stacking Models, Feature Selection, performance Metrics Analysis	Data Limitations, Generalizability, Technological barriers,
Raniya R. Sarra	MDPI-2022	Enhanced Heart Disease Prediction Based on Machine Learning and χ 2 Statistical Optimal Feature Selection Model	Automatic Prediction, Feature Selection, Model Performance	K-Models clustering with Huang Starting, Machine learning models-Random forest, Decision Tree classifier(DT), performance Evaluation	Improved Accuracy, Optimal Feature Selection, Reduced complexity, Benchmarking
AQSA RAHIM	IEEE Accesses-2021	An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases	Data Imbalance, Missing Values, Feature Selection, Integration of Techniques, Computational Complexity, Accuracy of Prediction:	Neural Networks, Hybrid Techniques, Feature Selection, Data Mining, Performance Evaluation, MaLCaDD	Data Limitations, Model Complexity, Generalizability, Computational Resources, Overfitting Risk, Validation Methods, Adaptability, Clinical Integration
HalahA.Al-Alshaikh	Scientific Reports - Nature-2022	Comprehensive evaluation and performance analysis of machine learning in the heart disease prediction	Challenges in Predicting Heart Disease, Data Imbalance, Feature Selection, Integration of Techniques, Use of Deep Learning	Genetic Algorithm (GA), Recursive Feature Elimination Method (RFEM),	Complexity, Data Imbalance, Generalizability, Interpretability, Validation, Clinical Integration, Cost and Accessibility
Luis Rolando Guarneros-Nolasco	MDPI-2021	Identifying the Main Risk Factors for Cardiovascular Diseases Prediction Using Machine Learning Algorithms	Global Impact of Cardiovascular Diseases (CVDs), Challenges in CVD Diagnosis, Utilization of Machine Learning (ML) in CVD Prediction,	Machine Learning Algorithms, Dataset Analysis, Attribute Selection, Evaluation Metrics, Validation Techniques, Performance Analysis	Data Limitations, Algorithmic Limitations, Feature Selection, Validation Methods

Kresoja et al. [9]: The study offers cardiologists a thorough manual on using machine learning (ML) to predict the prognosis of cardiovascular illnesses. Machine learning (ML) algorithms can help interpret complicated data for clinical decision-making, and they also address the abundance of available clinical and scientific data, which exceeds human analytical capabilities. By outlining fundamental ideas, possible dangers, and contemporary cardiology applications—such as omics, imaging, and basic science—the study seeks to equip medical professionals and researchers with the difficulties presented by machine learning. This article also delves into the basic concepts of ML, distinguishing it from artificial intelligence (AI) and conventional statistical methods and emphasizing its focus on its predictive power over interpretability. It covers various ML approaches, such as supervised and unsupervised learning, and discusses their applications in cardiology. Additionally, the paper addresses the limitations of ML, including the quality of data, outcome prediction challenges, and the need for meticulous validation in clinical trials. We identified the key gaps in the paper "A Cardiologist's Guide to Machine Learning in Cardiovascular Disease Prognosis and Prediction" by highlighting the importance of understanding ML for safe and effective use in medical practice.

1) DATA QUALITY
The paper highlights the importance of high-quality data for training ML models. Variance in clinical practice data and differences in imaging standards can affect model performance.

2) INTERPRETABILITY
More interpretations of ML models are needed, particularly for healthcare professionals. According to the article, knowledge of ML decisions is essential for their safe and thoughtful daily application.

3) CLINICAL TRIALS
This study highlights the Lack of randomized clinical trials demonstrating the superiority of machine learning-guided decision-making over traditional methods. Clinical trials require rigorous validation.

4) INTEGRATION OF ML
This paper discusses the challenge of integrating ML into everyday clinical practice and the need for physicians to understand it for its effective use.

5) RESULTS AND DISCUSSIONS
The paper discusses the importance of developing accurate predictive models for early detection and prevention of cardiovascular disease (CVD) using machine learning algorithms and various data sources.

It mentions that more advanced algorithms, especially deep learning neural networks, perform better at detecting potential CVD cases than traditional models like regression. This paper highlights the potential of machine learning to transform CVD prediction by utilizing diverse data sources, including demographic, clinical, biochemical, genetic, environmental, and socioeconomic factors.

The authors claim their approach is novel and different from traditional cardiovascular research. They combine unique data sources with advanced modeling techniques, such as neural networks and ensemble learning.

The paper suggests that future research should incorporate novel data types and improve the accuracy and personalization of CVD prediction models.

III. CONCLUSION

This study's exploration of predictive cardiovascular disease (CVD) methods underscores the remarkable potential of machine learning algorithms. This research demonstrated that advanced models, particularly deep learning neural networks, outperform traditional predictive approaches by integrating diverse data, including demographic, clinical, and biochemical parameters. This advancement enhances the accuracy of CVD predictions and contributes significantly to personalized medicine, paving the way for more efficient and precise cardiovascular health management tools.

FUTURE DIRECTIONS

This study emphasizes feature selection's crucial role in improving models' interpretability and efficiency. Incorporating novel data types, such as genetic sequences and environmental factors, offers a comprehensive framework for CVD risk assessment. As the quest for innovative predictive models continues, this research provides valuable insights and a solid foundation for future endeavors to combat the global burden of cardiovascular diseases through early detection and intervention strategies.

REFERENCES

[1] W. DeGroat, H. Abdelhalim, K. Patel, D. Mendhe, S. Zeeshan, and Z. Ahmed, ''Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel Nexus of machine learning techniques for precision medicine,'' *Sci. Rep.*, vol. 14, no. 1, Jan. 2024.

[2] G. Shobana and S. N. Bushra, ''Prediction of cardiovascular disease using multiple machine learning platforms,'' in *Proc. Int. Conf. Innov. Comput., Intell. Commun. Smart Electr. Syst. (ICSES)*, Sep. 2021, pp. 1–7.

[3] S. Sivakumar and T. C. Pramod, ''Comprehensive analysis of heart disease prediction: Machine learning approach,'' in *Proc. IEEE 3rd Global Conf. Advancement Technol. (GCAT)*, Oct. 2022, pp. 1–7.

[4] M. A.-A.-R. Asif, M. M. Nishat, F. Faisal, R. R. Dip, M. H. Udoy, M. F. Shikder, and R. Ahsan, ''Performance evaluation and comparative analysis of different machine learning algorithms in predicting cardiovascular disease,'' *Eng. Lett.*, vol. 29, no. 2, 2021.

[5] C. Zhang, Q. Zhu, and H. Li, ''Intelligent diagnosis system of cardiovascular disease based on case-based reasoning,'' in *Proc. Int. Conf. Intell. Innov. Technol. Comput., Electr. Electron. (IITCEE)*, Jan. 2023, pp. 418–422.

[6] G. Sembina, A. Aitim, and M. Shaizat, ''Machine learning algorithms for predicting and preventive diagnosis of cardiovascular disease,'' in *Proc. Int. Conf. Smart Inf. Syst. Technol. (SIST)*, Apr. 2022, pp. 1–5.

[7] S. Ghorashi, K. Rehman, A. Riaz, H. K. Alkahtani, A. H. Samak, I. Cherrez-Ojeda, and A. Parveen, ''Leveraging regression analysis to predict overlapping symptoms of cardiovascular diseases,'' *IEEE Access*, vol. 11, pp. 60254–60266, 2023.

[8] S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, ''Cardiovascular disease prediction using deep learning techniques,'' *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 981, no. 2, Dec. 2020, Art. no. 022006.

[9] K.-P. Kresoja, M. Unterhuber, R. Wachter, H. Thiele, and P. Lurz, ''A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction,'' *Basic Res. Cardiol.*, vol. 118, no. 1, p. 10, Mar. 2023.

[10] C. Krittanawong, H. U. H. Virk, S. Bangalore, Z. Wang, K. W. Johnson, R. Pinotti, H. Zhang, S. Kaplin, B. Narasimhan, T. Kitai, U. Baber, J. L. Halperin, and W. H. W. Tang, ''Machine learning prediction in cardiovascular diseases: A meta-analysis,'' *Sci. Rep.*, vol. 10, no. 1, p. 16057, Sep. 2020.

[11] M. Milosevic, Q. Jin, A. Singh, and S. Amal, ''Applications of AI in multi-modal imaging for cardiovascular disease,'' *Frontiers Radiol.*, vol. 3, Jan. 2024, Art. no. 1294068.

[12] C. Onnis, M. van Assen, E. Muscogiuri, G. Muscogiuri, G. Gershon, L. Saba, and C. N. De Cecco, ''The role of artificial intelligence in cardiac imaging,'' *Radiolog. Clinics*, vol. 62, no. 3, pp. 473–488, 2024.

[13] V. Ho, C. B. Johnson, I. Ghanzouri, S. Amal, S. Asch, and E. Ross, ''Physician- and patient-elicited barriers and facilitators to implementation of a machine learning–based screening tool for peripheral arterial disease: Preimplementation study with physician and patient stakeholders,'' *JMIR Cardio*, vol. 7, Nov. 2023, Art. no. e44732.

[14] J. Qiu, J. Zhu, S. Liu, W. Han, J. Zhang, C. Duan, M. A. Rosenberg, E. Liu, D. Weber, and D. Zhao, ''Automated cardiovascular record retrieval by multimodal learning between electrocardiogram and clinical report,'' in *Machine Learning for Health (ML4H)*, 2023, pp. 480–497.

[15] S. Amal, R. Kramer, D. Sawyer, J. B. Rabb, A. S. Maurais, C. S. Ross, A. Iribarne, and R. L. Winslow, ''Predicting adverse events in the cardiothoracic surgery intensive care unit using machine learning: Results and challenges,'' *medRxiv*, Dec. 2022.

[16] T. Shahzad, S. Saleem, S. Usman, J. Mirza, Q.-U. Islam, K. Ouahada, and T. Marwala, ''System dynamics of active and passive postural changes: Insights from principal dynamic modes analysis of baroreflex loop,'' *Comput. Biol. Med.*, vol. 100, pp. 27–35, Sep. 2018.

[17] T. Shahzad, S. Saleem, T. Marwala, and K. Ouahada, ''Variations in information flow patterns following intracranial hypertensive events in traumatic brain injured patients,'' *IEEE Access*, vol. 9, pp. 121137–121145, 2021.

[18] I. Haq, T. Mazhar, R. N. Asif, Y. Y. Ghadi, N. Ullah, M. A. Khan, and A. Al-Rasheed, ''YOLO and residual network for colorectal cancer cell detection and counting,'' *Heliyon*, vol. 10, no. 2, Jan. 2024, Art. no. e24403.

[19] I. Haq, T. Mazhar, R. Naz Asif, Y. Yasin Ghadi, R. Saleem, F. Mallek, and H. Hamam, ''A deep learning approach for the detection and counting of colon cancer cells (HT-29 cells) bunches and impurities,'' *PeerJ Comput. Sci.*, vol. 9, p. e1651, Dec. 2023.

[20] Z. Hasan, A. Fatima, T. Shahzad, S. Abbas, T. M. Ghazal, M. M. Al-Sakhnini, M. A. Khan, and A. Ahmed, ''Nanomedicine: Treatment of chronic disease using gold nano thermo robot (GNTR) empowered with nanotechnology approaches,'' *IEEE Access*, vol. 12, pp. 8552–8584, 2024.

[21] N. Nghiem, J. Atkinson, B. P. Nguyen, A. Tran-Duy, and N. Wilson, ''Predicting high health-cost users among people with cardiovascular disease using machine learning and nationwide linked social administrative datasets,'' *Health Econ. Rev.*, vol. 13, no. 1, p. 9, Feb. 2023.

[22] H. A. Al-Alshaikh, R. C. Poonia, A. K. J. Saudagar, M. Yadav, H. S. AlSagri, and A. A. AlSanad, ''Comprehensive evaluation and performance analysis of machine learning in heart disease prediction,'' *Sci. Rep.*, vol. 14, no. 1, p. 7819, Apr. 2024.

[23] L. R. Guarneros-Nolasco, N. A. Cruz-Ramos, G. Alor-Hernández, L. Rodríguez-Mazahua, and J. L. Sánchez-Cervantes, ''Identifying the main risk factors for cardiovascular diseases prediction using machine learning algorithms,'' *Mathematics*, vol. 9, no. 20, p. 2537, Oct. 2021.

[24] M. Swathy and K. Saruladha, ''A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques,'' *ICT Exp.*, vol. 8, no. 1, pp. 109–116, Mar. 2022.

[25] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, and A. Shaikh, ''A robust heart disease prediction system using hybrid deep neural networks,'' *IEEE Access*, vol. 11, pp. 121574–121591, 2023.

[26] L. Yahaya, N. David Oye, and E. Joshua Garba, ''A comprehensive review on heart disease prediction using data mining and machine learning techniques,'' *Amer. J. Artif. Intell.*, vol. 4, no. 1, p. 20, 2020.

[27] E. I. Elsedimy, S. M. M. AboHashish, and F. Algarni, ''New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization,'' *Multimedia Tools Appl.*, vol. 83, no. 8, pp. 23901–23928, Aug. 2023.

[28] S. Mondal, R. Maity, Y. Omo, S. Ghosh, and A. Nag, ''An efficient computational risk prediction model of heart diseases based on dual-stage stacked machine learning approaches,'' *IEEE Access*, vol. 12, pp. 7255–7270, 2024.

[29] S. Subramani, N. Varshney, M. V. Anand, M. E. M. Soudagar, L. A. Al-Keridis, T. K. Upadhyay, N. Alshammari, M. Saeed, K. Subramanian, K. Anbarasu, and K. Rohini, ''Cardiovascular diseases prediction by machine learning incorporation with deep learning,'' *Frontiers Med.*, vol. 10, Apr. 2023, Art. no. 1150933.
[30] V. Paul, J. A. Ibrahim, and A. Sharma, ''Utilizing high computational ML model for lung cancer prediction with accurate data analysis,'' in *Proc. AIP Conf.*, 2023, vol. 2855, no. 1.
[31] R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, ''Enhanced heart disease prediction based on machine learning and χ^2 statistical optimal feature selection model,'' *Designs*, vol. 6, no. 5, p. 87, Sep. 2022.
[32] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, and A. W. Muzaffar, ''An integrated machine learning framework for effective prediction of cardiovascular diseases,'' *IEEE Access*, vol. 9, pp. 106575–106588, 2021.

VALERU VISION PAUL received the Bachelor of Technology degree in information technology from the SSN College of Engineering and Technology, JNTU Kakinada University, Ongole, Andhra Pradesh, India, and the Master of Technology degree in computer science and engineering from the Qis College of Institute of Technology, JNTUK.

JAFAR ALI IBRAHIM SYED MASOOD received the Bachelor of Technology degree in information technology from the Syed Ammal Engineering College, Ramanathapuram, Anna University, Chennai, Tamil Nadu, India, the Master of Technology degree in computer systems and networks from the Dr. MGR Educational and Research Institute, Chennai, Tamil Nadu, India, and the Doctor of Philosophy degree in information and communication engineering from Anna University, Chennai.

He was a full-time Research Fellow with Anna University, Chennai. He is currently a Faculty Member of the Internet of Things Department, School of Computer Science and Engineering, Vellore Institute of Technology (Deemed to be University), Vellore, Tamil Nadu, India. He possesses nearly seven years of industrial experience, both administrative and technical perception in IT security, bio-metric deployment, IT administration, network security, web services and security, and service oriented architecture environments. Also, he possesses six years of research and teaching experience in health applications, genetic data analysis, computational and machine intelligence, and IoT-enabled recommender systems. He has presented his paper at many international conferences and published 86 journal articles. He has published over 100 national and international patents and holds 38 patent grants. His research interests include cloud computing, clinical research, medical informatics, the Internet of Things, intelligent systems development, machine learning, ontology development, big data, and data mining.

Dr. Masood serves as a reviewer, an editorial board member, a technical committee member, and an advisor in many international conferences and journals. He is a consultant in some IT industries from Malaysia and the U.K. He is an active member of the Academic Council, Board of Studies, Board of Research, and Governing Council for various reputed institutions in India and Abroad. Also, he is a reviewer of some reputed international journals and ACM conferences. He is a lifetime member of various reputable and autonomous technical bodies from different countries. He visited Japan, Hong Kong, Malaysia, Singapore, Thailand, Indonesia, South Korea, the Philippines, Sri Lanka, and Gulf countries for research.