Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

A Survey of Early Detection of Coronary Artery Disease Prediction using Machine Learning Algorithms

Akhil A Das
Department of Computer Science
College of Engineering Karunagappally
Kerala, India

Aswini M
Department of Computer Science
College of Engineering Karunagappally
Kerala, India

Esha Fathima N
Department of Computer Science
College of Engineering Karunagappally
Kerala, India

Muhammed Ameen
Department of Computer Science
College of Engineering Karunagappally
Kerala, India

Swathi S
Assistant Professor
Department of Computer Science
College of Engineering Karunagappally
Kerala, India

***Abstract*—Coronary Artery Disease (CAD) remains a signifi- cant global health challenge, causing millions of deaths annually. Early detection is crucial for preventing severe complications like heart attacks. However, traditional diagnostic methods often involve invasive procedures, are time-consuming, and may be inaccessible for many individuals. To address this pressing need, we propose a novel machine learning-based CAD prediction model. Leveraging readily available clinical data such as age, cholesterol levels, blood pressure, and blood sugar, our model aims to accurately assess individual CAD risk. Our system utilizes a combination of advanced machine learning algorithms, including Support Vector Machines, Decision Trees, Random Forests, and Multilayer Perceptrons. These algorithms are rig- orously trained and optimized on a comprehensive dataset to achieve high accuracy and reliability. The resulting model is integrated into a user-friendly application, enabling healthcare providers to input patient data and receive rapid, accurate CAD risk assessments. This empowers healthcare professionals to make informed decisions and implement timely interventions, such as lifestyle modifications or medical treatments, to mitigate the risk of heart attacks and improve patient outcomes. By democratizing access to predictive healthcare technology, we envision a future where CAD can be detected and managed more effectively. This innovative approach has the potential to reduce the burden of CAD on healthcare systems, lower associated costs, and ultimately improve public health by preventing severe cardiovascular events and enhancing the quality of life for at-risk populations.**

***Keywords* — Coronary Artery Disease, Support Vector Ma- chines, Decision Trees, Random Forests, Multilayer Percep- tron,Coronary Vascular Disease.**

I. INTRODUCTION

Coronary artery disease (CAD) remains one of the leading causes of mortality worldwide, primarily due to delayed diagnosis and treatment. Conventional diagnostic approaches, such as angiography and stress tests, while effective, are invasive, expensive, and often not accessible to a large segment of the population. This underscores the need for an alternative method that is non-invasive, cost-effective, and efficient in identifying individuals at risk of CAD at an early stage.

This project presents a machine learning-based model designed to predict CAD risk using essential clinical parameters, including cholesterol levels, blood pressure, age, and blood sugar. Several machine learning algorithms, such as Support Vector Machine (SVM), Decision Tree, Random Forest, and Multilayer Perceptron (MLP), are utilized to analyze these features and generate predictions. To enhance the accuracy and reliability of the model, an ensemble learning technique employing soft voting is applied, ensuring a more robust risk assessment system.

The dataset used for this research is compiled from two publicly available sources: the Kaggle CVD dataset, which consists of 70,000 records, and the UCI Heart Disease dataset, which includes 1,025 records. Data preprocessing involves managing missing values, standardizing numerical features, and encoding categorical variables to optimize model performance. The dataset is divided into 80% for training and 20% for testing to ensure a comprehensive evaluation of the model’s effectiveness.

Each machine learning algorithm processes the patient data separately, and the final prediction is determined using a majority voting mechanism. The model's performance is evaluated based on key metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to assess its reliability for clinical use.

Findings suggest that the Random Forest algorithm delivers the highest accuracy due to its ensemble learning capabilities, which help mitigate overfitting and enhance performance. While the Decision Tree model proves effective, it is more prone to overfitting. The MLP model successfully identifies complex patterns but demands significant computational power. Meanwhile, SVM performs well with smaller datasets but faces challenges in handling large-scale feature interactions. The final ensemble model achieves an accuracy exceeding 85%, demonstrating its potential as a valuable tool for early CAD detection and prevention.

By incorporating multiple machine learning algorithms into an ensemble framework, this study highlights the potential of artificial intelligence in predicting CAD risk. The proposed system provides a data-driven, non-invasive alternative to traditional diagnostic methods, enabling healthcare professionals to detect high-risk individuals early. Future improvements may include expanding the dataset, integrating additional clinical variables, and exploring deep learning methodologies to further refine predictive accuracy and performance

II. LITERATURE REVIEW

A. An Integrated Two-Layered Voting (TLV) Framework for Coronary Artery Disease Prediction Using Machine Learning Classifiers

This paper [1]introduces the TLV (Two-Layer Voting) model, an ensemble method that combines hard and soft voting techniques. In the first layer, feature selection is performed using a combination of soft and hard voting applied to three statistical methods: ANOVA f-test, Chi-squared test, and Mutual Information. The second layer involves comparing the performance of soft and hard voting with a variety of classification algorithms, including Multi-Layer Perceptron, Decision Tree, Support Vector Classifier, and Random Forest. Hyperparameter tuning is employed using the GridSearchCV method to optimize the performance of these algorithms. When applied to the UCI heart disease dataset and the Kaggle CVD dataset, the TLV model with soft voting achieved the highest accuracy of 99.03% and 88.09%, respectively, significantly outperforming existing CAD disease prediction studies.

B. Automatic Identification of Coronary Arteries in Coronary Computed Tomographic Angiography

This paper [2] presents a novel automatic coronary artery identification algorithm designed to improve the accuracy and efficiency of diagnosing cardiovascular disease using CCTA. The algorithm is capable of accurately identifying and segmenting key coronary arteries, including the RCA, PDA, PLB, LCx, LAD, RI, OM1, OM2, and D1, D2. This algorithm adheres to the SCCT's coronary labeling standards and has been successfully implemented in over 100 hospitals for more than a year. Rigorous testing on 892 CCTA datasets has demonstrated its accuracy, with a 95.96% agreement rate with expert-level manual labeling.

C. Topological Transformer Network for Automated Coronary Artery Branch Labeling in Cardiac CT Angiography

This paper [3] proposes a novel Topological Transformer Network (TTN) to address the limitations of existing methods for automated coronary artery branch labeling in cardiac CT angiography. TTN effectively models the overall correlation between branches, capturing subtle differences that traditional methods often miss. To mitigate the class imbalance between main and side branches, a segment-depth loss is introduced. By incorporating a topological encoding that represents the positions of vessel segments within the artery tree, TTN accurately classifies branches. Extensive experiments on a dataset of 325 CCTA demonstrate the superior performance of TTN, particularly in labeling side branches. This innovative approach, distinct from previous methods, has the potential to significantly improve computer-aided diagnosis systems for cardiovascular diseases by assisting clinicians in locating atherosclerotic plaques.

D. Enhancing Coronary Artery Prognosis: A Novel Dual-Class Boosted Decision Trees Strategy for Robust Optimization

This paper [4] proposes an advanced ensemble learning model to predict Chronic Coronary Artery Disease (CAD). By combining multiple machine learning algorithms like Random Forest and SVM, the model achieves higher accuracy and robustness. Precision engineering techniques further optimize the model, enhancing its ability to handle complex relationships within the data and reducing overfitting. However, this approach also introduces complexities, including increased computational requirements and challenges in interpreting the model's decisions. The model's performance heavily relies on the quality of the input data and effective feature selection. The study utilizes a diverse dataset comprising patient records from various sources, which is preprocessed to handle missing values and normalize data. While the model shows promise, addressing challenges like interpretability and computational efficiency is crucial for its practical application in clinical settings. The use of diverse patient data and careful feature selection strengthens the study's credibility and potential impact.

E. A Novel Early Detection and Preventation of Coronary Heart Disease Framework Using Hybrid Deep Learning Model and Neural Fuzzy Inference System

This study [5] presents O-SBGC-LSTM, an advanced deep learning model aimed at enhancing early diabetes detection and prevention. By combining Graph Convolutional Neural

TABLE I
Summary of Engagement Detection Studies

Title	Dataset Used	Pros	Cons	Performance Metrics
TLV + ML [1]	UCI's heart disease dataset & Kaggle's CVD dataset	Better Features and Improved Results.	Computational complexity and Data sensitivity.	Accuracy:99.03%
Automatic CAD CT Angiography [2]	CCTA datasets based on SCCT standards.	Fast, 1- minute coronary artery identification	The algorithm doesn't identify certain arteries due to low clinical significance.	Accuracy:95.96%
TTN + Cardiac CT Angiography [3]	CCTA dataset with 325 subjects	Captures artery branch correlations, enhancing side branch analysis.	Main branches accuracy dips slightly, prioritizing side branches.	Recall: 89.4%, precision: 86.9% and F1 score of 88.0%
Boosting CAD Prognosis with Dual-Class DT [4]	Heart disease dataset from IEEE DataPort with 1,190 records & 14 feautres	Highly accurate for real-time monitoring.	Less efficient with large datasets.	AUC : 0.991
Hybrid DL Model and Neural Fuzzy Inference System [5]	CHD and diabetes data from NID.	Efficient in handling both spatial and temporal data.	Computational complexity.	Accuracy: 98%
Lesion Degree Ranges based on DL [6]	Dataset of ICA images from 42 patients.	High accuracy with AUC up to 98.1%, F-measure of 92.7%	Decreased accuracy with lesion ranges below 99%.	Accuracy:98.1%
Novel Hybrid Harris Hawks Approach [7]	Heart disease dataset with 270 instances & 14 features.	Improved exploration and exploitation	Susceptibility to local optima.	Accuracy:94.74%
Predicting CHD using an improved LightGBM model [8]	Data from the Framingham Heart Institute	Reduces overfitting using OPTUNA's efficient pruning and sampling	Performance may vary by dataset	Accuracy: 94%
Leveraging Regression analysis to predict overlapping symptoms of CVD [9]	2,621 UAE medical records for CVD patterns by age, symptoms.	Handles complex, overlapping symptoms to improve CVD prediction	Predictive performance depends the quality of patient records.	Accuracy:91%
CresFormer-based Heart Disease Classification [10]	Dual-lead ECG signals data	Dual feature extraction enhances diagnosis accuracy with efficient ECG processing.	Computational complexity due to multiple deep layer	Accuracy: 97%
QT Interval Time Series and ST-T Waveform [11]	ECG dataset:107 healthy, 93 CAD patients.	Efficient use of single-lead ECG data for non-invasive diagnosis	Results may vary for other datasets.	Accuracy: 96.16%

Networks (GCNNs) with Long Short-Term Memory (LSTM) networks, the model captures intricate spatial and temporal data patterns. To further improve its performance, the Eurygaster Optimization Algorithm (EOA) is employed for hyperparameter tuning. The model's hierarchical temporal design facilitates efficient learning of high-level semantic features, while a fuzzy-based inference system offers personalized prevention strategies, enabling individuals to adopt proactive health measures. Achieving an accuracy rate exceeding 98% in various evaluations, this model significantly outperforms conventional machine learning approaches, making it a highly promising solution for managing diabetes at an early stage.

F. Coronary Artery Disease Classification with Different Lesion Degree Ranges based on Deep Learning

This paper [6] explores the performance of deep learning techniques for binary classification of Invasive Coronary Angiography (ICA) images, focusing on varying lesion degrees. An annotated ICA image dataset, containing ground truth, lesion locations, and seven severity levels, was employed. The images were divided into "lesion" and "non-lesion" patches to analyze how binary classification accuracy is affected by different lesion degree ranges within the positive class. Five Convolutional Neural Network architectures (DenseNet-201, MobileNet-V2, NasNet-Mobile, ResNet-18, and ResNet-50) were trained on various input images incorporating different lesion degree ranges. Four experimental setups were designed, with and without data augmentation, to assess F-measure and Area Under the Curve (AUC) metrics. The results yielded an F-measure of 92.7% and an AUC of 98.1%. However, the study revealed a significant decrease in accuracy, around 15%, when classifying lesions with less than 99% severity.

G. Coronary Artery Disease Prediction: A Novel Hybrid Harris Hawks Approach

This paper [7]introduces a hybrid machine learning framework that leverages the Harris Hawks Optimization (HHO) technique to enhance CAD prediction accuracy and adaptability. HHO's efficient exploration and exploitation of the feature space, combined with machine learning algorithms, improves model performance and enables feature selection. The framework is adaptable to diverse datasets and can automate parameter tuning, making it scalable and flexible. However, its complexity, computational requirements, and dependence on data quality are limitations. While the model shows promise, challenges like interpretability and generalization need to be addressed for wider clinical adoption. The paper reports an accuracy of [Insert accuracy percentage from the paper here] for the proposed hybrid model, which is a significant improvement over existing methods.

H. Predicting Coronary Heart Disease Using an Improved LightBGM Model

In this paper [8]novel prediction model, HY_OptGBM, to address the critical challenge of early Coronary Heart Disease (CHD) detection. Leveraging an optimized LightGBM classifier, our model aims to forecast the onset of CHD. To enhance its predictive accuracy, we meticulously fine-tuned hyperparameters using OPTUNA and incorporated Focal Loss (FL) for a more robust loss function. We rigorously evaluated the model's performance on the Framingham Heart Study dataset, employing a comprehensive range of metrics including precision, recall, F-score, accuracy, MCC, sensitivity, specificity, and AUC. Our results are promising, with the model achieving an impressive AUC of 97.8%, outperforming existing models. This breakthrough demonstrates the potential of our approach to significantly improve early CHD detection, ultimately leading to reduced healthcare costs and improved patient outcomes.

I. Leveraging Regression Analysis to Predict Overlapping Symptoms of Coronary Vascular Disease

This paper [9]investigating the potential of deep learning-based regression analysis for early prediction of cardiovascular diseases (CVDs). Our approach involved training a long short-term memory (LSTM) network on a dataset of 2,621 medical records from UAE hospitals, encompassing information on age, symptoms, and CVD history. We found that pairing diseases with overlapping symptoms significantly improved prediction accuracy. For instance, coronary heart disease prediction accuracy increased from 71.5% to 84.4% when combined with dyspnea. By progressively incorporating additional symptoms like chest pain, cyanosis, weakness, fatigue, hemoptysis, and chest discomfort, we achieved a peak accuracy of 91% with the inclusion of fever. These results underscore the effectiveness of our proposed method in early CVD prediction, as validated across various evaluation benchmarks.

J. Intra-Patient and Inter-Patient Multi-Classification of Severe Cardiovascular Disease based on CResFormer

CResFormer is a novel deep learning model designed to significantly improve the diagnosis [10] of severe cardiovascular diseases like coronary artery disease, myocardial infarction, and congestive heart failure. It efficiently processes dual-lead electrocardiogram (ECG) signals without extensive preprocessing. The model combines the strengths of convolutional neural networks (CNNs) for dimensionality reduction, residual networks (ResNets) for feature preservation, and transformer encoders with multi-headed attention for feature interdependence analysis. CResFormer outperforms existing models on public datasets like MIT-BIH, PTBDB, and INCART, achieving high accuracy rates of 99.84% for intra-patient and 97.48% for inter-patient multi-class classification. Its robustness to noise and potential for automated disease detection in clinical and resource-constrained settings are promising. However, challenges remain in computational complexity and diagnosis time. Future research aims to simplify the model, improve its ability to process multi-lead ECG signals, and validate its performance using real-world clinical data. CResFormer represents a significant step forward in leveraging deep learning for accurate and efficient cardiovascular disease diagnosis.

K. Enhanced Automated Diagnosis of Coronary Artery Disease using Features Extracted from QT Interval Time Series and ST-T Waveform

This paper [11] explores the use of machine learning and deep learning techniques to enhance the detection of Coro- nary Artery Disease (CAD). A dataset comprising 5-minute single-lead electrocardiograms (ECGs) and clinical data from 107 healthy individuals and 93 CAD patients was analyzed. Features extracted from QT intervals, RR intervals, and ST-T waveforms were evaluated for their ability to classify CAD. Various machine learning models, including Gaussian Naive Bayes, Support Vector Machine, Extreme Gradient Boosting, and a Residual Neural Network (ResNet-18) for feature extraction, were employed. The most promising results were achieved by combining features from all three data sources, along with clinical information, resulting in an accuracy of 96.16%, sensitivity of 95.75%, and specificity of 96.40%. These results highlight the importance of QT interval and ST-T waveform features in improving automated CAD diagnosis.

III. CONCLUSION

This research explored the application of machine learning algorithms to predict the onset of coronary artery disease (CAD). We utilized Support Vector Machines, Decision Trees, Random Forest, and Multilayer Perceptrons to develop a model capable of accurately predicting individual risk based on factors like cholesterol levels, blood pressure, and age. Our findings suggest that machine learning holds significant potential for early detection and proactive intervention in CAD. The developed model, projected to achieve an accuracy rate surpassing 85%, has the potential to revolutionize clinical practice. By precisely identifying individuals at risk, healthcare providers can implement preventive strategies and personalized treatment plans, ultimately enhancing patient outcomes and optimizing the healthcare system. Further research is necessary to refine the model's accuracy and explore additional factors that may influence CAD risk. By continuously advancing machine learning in this field, we can aspire to a future where CAD can be effectively managed and prevented.

REFERENCES

[1] D. Y. Omkari and K. Shaik, “An integrated two-layered voting (tlv) framework for coronary artery disease prediction using machine learning classifiers,” IEEE Access, vol. 12, pp. 56275–56290, 2024.

[2] N. Kaushik, S. G, M. G, A. Ramesh, and V. Gopal BT, “Coronary artery disease detection in early stages using machine learning,” in 2024 5th International Conference for Emerging Technology (INCET), 2024.

[3] A. R. Vijayaraj and S. Pasupathi, “Nature inspired optimization in context-aware-based coronary artery disease prediction: A novel hybrid harris hawks approach,” IEEE Access, vol. 12, pp. 92635–92650, 2024.

[4] L. S. Nagra, S. Thapa, P. Rahi, T. Sharma, H. Ashok, and V. Pathania, “An accurate ensemble machine learning model with precision engineering for chronic coronary artery disease prognosis,” in 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), 2024.

[5] A. Phoemsuk and V. Abolghasemi, “Deep learning for predicting coronary artery disease using convolutional neural network,” in IEEE Conference on Artificial Intelligence (CAI), 2024.

[6] G. P. A. S. H. S. S, V. Krishna Jyothis and A. J, “Ml in cardiovascular disease risk assessment :a comparative study,” in 2024 Control Instrumentation System Conference (CISCON), 2024.

[7] P. V. Adhishayaa, V. Gomathi, and K. Mahendran, “Hybrid deep learning model for coronary artery disease pre- diction,” in 2023 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–6, 2023.

[8] G. S. Sajja, M. Mustafa, K. Phasinam, K. Kaliyaperumal, R. J. M. Ventayen, and T. Kassanuk, “Random forest and soft voting ensemble for coronary artery disease classification,” in 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021.

[9] M. Pouladian, M. R. H. Golpayegani, A. A. Tehrani-Fard, and M. Bubvay-Nejad, “Application of multi-layer perceptron neural networks in cad prediction,” IEEE Transactions on Biomedical Engineering, vol. 52, no. 4, pp. 743–747, 2005.

[10] M. Pouladian, M. R. H. Golpayegani, A. A. Tehrani-Fard, and M. Bubvay-Nejad, “Application of multi-layer perceptron neural networks in cad prediction,” IEEE Transactions on Biomedical Engineering, vol. 52, no. 4, pp. 743–747, 2005.

[11] M. Abdar, E. Nasarian, X. Zhou, G. Bargshady, V. N. Wijayaningrum, and S. Hussain, “Performance improvement of decision trees for diagnosis of coronary artery disease using multi filtering approach,” in 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019.