|| Volume 7 || Issue 9|| November 2023 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Predicting Coronary Artery Disease Risk with Met heuristic-Enhanced Machine Learning Models

Dipika Raju Pangudwale, Prof. Anita Mahajan, Dr. Pankaj Agarkar

PG Student: Department of Computer Engineering, Ajeenkya DY Patil School of Engineering Pune

Prof.: Department of Computer Engineering, Ajeenkya DY Patil School of Engineering Pune

HOD: Department of Computer Engineering, Ajeenkya DY Patil School of Engineering Pune

Abstract: Conventional risk assessment methods often depend on fixed, limited data and fail to sufficiently consider the ever-changing nature of CAD development. Our recommended approach utilizes metaheuristic techniques such as genetic algorithms and particle swarm optimization to optimize the feature selection and model hyperparameters. By using this dynamic approach, the accuracy of forecasts is improved, and it also enables the identification of significant risk factors that could otherwise be overlooked. We use a substantial cohort of individuals diagnosed with coronary artery disease (CAD), including diverse demographic, clinical, and genetic information. We compare the effectiveness of models enhanced by metaheuristics with that of conventional machine learning approaches. The results demonstrate a significant improvement in the precision of CAD risk prediction, as the upgraded models consistently outperform their traditional counterparts. Furthermore, our approach illuminates unforeseen connections that might influence tailored preventative efforts, while also providing valuable insights into the comparative significance of different risk variables. By uncovering concealed trends in the data, we facilitate the development of targeted medicines, reducing the burden of CAD on healthcare systems and improving patient outcomes. Metaheuristic approaches are included into CAD risk prediction to enhance both the accuracy and the interpretability and generalizability of the results. The promise of our technique is to fundamentally transform our understanding of illness risk assessment and may be used for other complex medical challenges. In conclusion, the early detection of CAD has the potential to integrate metaheuristic-enhanced machine learning models into clinical practice, resulting in more efficient preventive and therapeutic strategies.

Keywords: Machine Learning, Coronary Artery Disease, Prediction, Risk Analysis

1. Introduction

Coronary artery disease (CAD) is the leading cause of illness and death on a global scale, presenting a significant risk to public health [1]. Cardiovascular disorders such as heart attacks, heart failure, and other related problems often occur due to coronary artery disease (CAD). CAD is characterized by the narrowing or obstruction of coronary arteries caused by the accumulation of atherosclerotic plaques [2]. Given its elusive nature and diverse array of risk factors, there is a pressing need for efficient risk forecasting and timely intervention methods to mitigate its impact. Various risk assessment methodologies and models have been developed throughout time to evaluate an individual's susceptibility to CAD. These models have given considerable importance to traditional risk variables such as age, gender, hypertension, hyperlipidemia, and smoking [3]. Although risk factors remain valuable in clinical practice, the complexity of CAD necessitates a more comprehensive and dynamic technique for their evaluation. The advancements in artificial intelligence (AI) and machine learning (ML) provide a promising opportunity to transform CAD risk prediction. Machine learning (ML) techniques have the ability to include various clinical, genetic, and lifestyle data, enabling a more

thorough knowledge of a person's risk profile [4]. Traditional machine learning algorithms, however, struggle with handling data that has a large number of dimensions, selecting the most relevant features, and optimizing the model, all of which are crucial for accurately assessing the risk of CAD.

In order to address these challenges, we propose an innovative approach that enhances the prediction of CAD risk by integrating machine learning models with metaheuristic optimization approaches. Particle swarm optimization, genetic algorithms, and simulated annealing are three instances of metaheuristic algorithms that have shown remarkable efficacy in resolving complex optimization problems [5]. To enhance the accuracy and robustness of CAD risk prediction models, we aim to optimize feature selection and model hyperparameters by leveraging their effectiveness. CAD is a dynamic and ever-evolving condition. Conventional risk models sometimes overlook the dynamic interplay of risk variables and their changing impact on the development of coronary artery disease (CAD). By incorporating metaheuristic optimization, we may iteratively adjust and enhance risk models based on newly accessible data, guaranteeing that they remain up-to-date and precise depictions of the evolving disease landscape. Thorough Data Integration: Several factors, including genetics, clinical features, lifestyle choices, and environmental factors, contribute to the development of CAD [6].

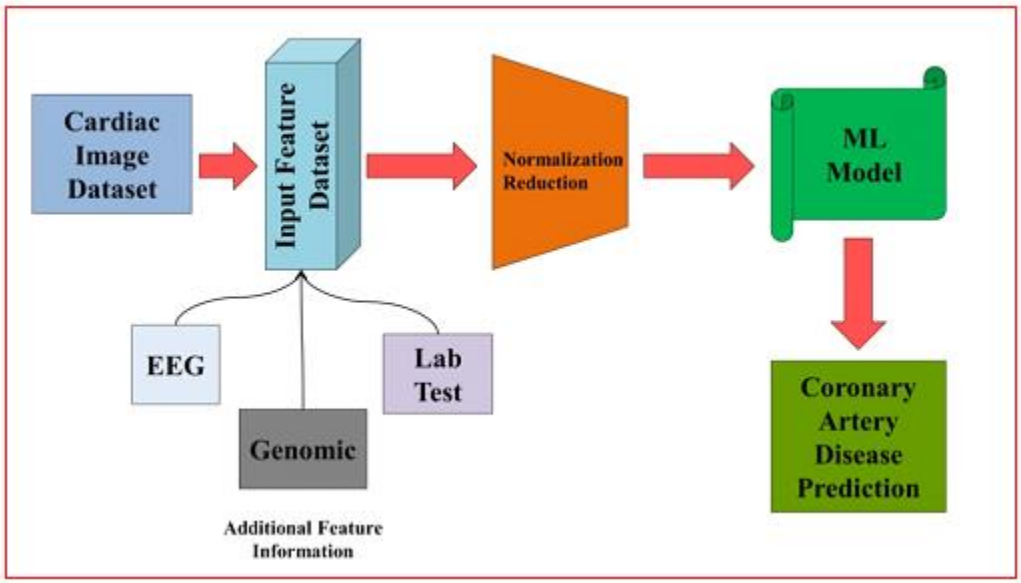

Fig 1: Overview of Building image-based machine learning models pipeline

Our approach is to integrate these diverse data sources into a unified prediction framework. By selectively selecting the most relevant components from this extensive dataset, we may identify previously unknown connections and uncover distinct risk factors that traditional models would have overlooked. Our work aims to enhance the precision of CAD risk prediction. Conventional models often exhibit subpar performance in clinical scenarios due to their inability to effectively balance sensitivity and specificity [7]. Our goal is to enhance the accuracy of clinical decision-making by reducing both false positives and false negatives, while also improving prediction accuracy via the use of metaheuristic optimization. Although complex machine learning models exhibit impressive predictive accuracy, their use in clinical practice is often hindered by a lack of interpretability. Our strategy ensures that doctors can understand and trust the predictions provided by our models by maintaining a balance between accuracy and interpretability. Furthermore, the techniques we use are

designed to be adaptable, facilitating their application to diverse patient demographics and healthcare settings.

This research examines the integration of many machine learning models, including support vector machines, random forests, and neural networks, with metaheuristic optimization techniques, such as genetic algorithms. The implementation of these enhanced models will use a substantial cohort of individuals with coronary artery disease, including their demographic, clinical, genetic, and lifestyle data. Throughout this research, we will conduct comprehensive assessments, comparing the performance of our models augmented with metaheuristics to that of typical machine learning approaches. We will conduct an impartial assessment of the progress achieved in CAD risk prediction by using accepted evaluation criteria such as sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and precision-recall curves. The proposed study aims to elucidate the complex interaction of risk factors that contribute to the development of coronary artery disease (CAD). By identifying the most significant characteristics and their temporal fluctuations, we may uncover previously undisclosed patterns and connections. Gaining a more comprehensive comprehension of the underlying factors contributing to coronary artery disease (CAD) might lead to the discovery of novel biomarkers and therapeutic targets. This, in turn, may facilitate the development of more effective preventive and treatment strategies. A unique approach to tackle the pressing global health issue of coronary artery disease (CAD) involves integrating metaheuristic-enhanced machine learning models into CAD risk prediction. Our goal is to provide clinicians with accurate, comprehensible, and continuously evolving risk assessment tools to improve patient outcomes and reduce the financial impact of CAD on healthcare systems.

2. Review of Literature

In order to mitigate its worldwide consequences, much research has been carried out to anticipate the susceptibility to Coronary Artery Disease (CAD), given the significance of early identification and treatments. In this part, we examine significant research in CAD risk prediction, focusing on well-established techniques and recent advancements that provide the foundation for our proposed metaheuristic-enhanced machine learning models. Traditionally, CAD risk prediction models have relied on clinical and demographic data, known as conventional risk factors. Age, gender, hypertension, cholesterol levels, smoking status, and diabetes are well-known variables that increase the risk of coronary artery disease (CAD). The Framingham Heart Study, which began in 1948, played a crucial role in highlighting the significance of these factors as indications of the illness. The Framingham Risk Score (FRS) and subsequent ATP-III criteria have been extensively used in clinical practice to assess an individual's risk of developing coronary artery disease (CAD) [11].

Despite their continued usage and valuable insights, traditional risk models are not without limits. They often fail to include genetic and lifestyle factors, fail to recognize the evolving nature of CAD development, and have inadequate accuracy, particularly in people with diverse backgrounds [9]. These imperfections have spurred scholars to contemplate more comprehensive and evidence-based approaches. The integration of machine learning (ML) methods into CAD risk prediction has enabled the improvement of accuracy and the identification of complex relationships between risk variables. Machine learning techniques used for predicting CAD risk include logistic regression, decision trees, support vector machines, and random forests (10). The notable advantage of ML models is their capacity to effectively process data with a large number of dimensions. This capability allows for the integration of genetic data, clinical assessments, and lifestyle variables into a unified framework [12]. The investigation of feature selection methodologies such as Recursive Feature Elimination (RFE) and

Principal Component Analysis (PCA) has been conducted to identify the most relevant predictors from huge datasets [13].

In recent years, deep learning, a kind of machine learning, has become more significant for predicting CAD risk. Convolutional and recurrent neural networks have shown promise in analyzing medical images, time series data, and electronic health records (EHRs) [14]. Deep learning models has the capability to autonomously identify intricate patterns and hierarchical characteristics from raw data, particularly when dealing with unstructured medical data like electronic health records (EHRs) and medical images [14]. Furthermore, the use of transfer learning techniques, which include using pre-trained models on large datasets, has been employed to enhance the performance of CAD risk prediction models [15]. The increasing popularity of integrating metaheuristic optimization approaches with ML models in healthcare is driven by the potential to enhance model performance and feature selection. The optimization of model hyperparameters, feature subsets, and model architectures has been achieved via the use of metaheuristic approaches, including genetic algorithms, particle swarm optimization, simulated annealing, and other methods [16]. These optimization approaches [17] have been effectively used in several healthcare applications, including illness diagnosis, pharmaceutical creation, and treatment optimization. Due to their ability to explore complex search regions and adapt to changing data dynamics, they are well-equipped to address the challenges posed by CAD risk prediction. Several research studies have focused on hybrid models, which include traditional risk factors, machine learning methodologies, and optimization strategies. Researchers have enhanced forecast accuracy by integrating machine learning techniques with the FRS [18]. In order to enhance the accuracy and dependability of risk assessment, these hybrid models aim to use the benefits of both traditional and data-driven methodologies.

Despite advancements in CAD risk prediction, challenges persist. The problem of model interpretability is very significant, particularly when using advanced machine learning and deep learning models. There is a significant need for machine learning models that are interpretable and can retain a high level of accuracy while also disclosing the impact of risk factors [19]. Additionally, ongoing research is being conducted to determine the extent to which CAD risk prediction models may be applied to different populations and healthcare settings. In order to provide fair and just healthcare, it is necessary to acknowledge and rectify disparities in the evaluation of coronary artery disease (CAD) risk, and develop models that are specifically designed for different demographic groups [20].

Our study builds upon previous work in this field by introducing a new approach that utilizes metaheuristic optimization approaches to enhance the performance and comprehensibility of CAD risk prediction models. Our goal is to improve the accuracy of CAD risk prediction by dynamically optimizing the selection of features and model hyperparameters. This will help enhance early diagnosis and intervention strategies for this widespread global health issue. Our technique contributes to the advancement of CAD risk assessment and has the potential for broader use in personalized therapy and disease prediction.

Table 1: Related work summary in coronary artery disease

Method	Data Used	Key Findings	Limitations	Scope
Traditional Risk Models [12]	Demographic and clinical data	Established key risk factors like age, hypertension, smoking	Limited to basic risk factors; may not capture complex interactions	Provides a baseline for comparison

Machine Learning Models [13]	Extensive datasets with clinical, genetic, and lifestyle data	Improved prediction accuracy and ability to handle high-dimensional data	May lack interpretability; challenges in feature selection	Enables inclusion of diverse data sources
Deep Learning Models [14]	Electronic health records (EHRs), medical images	Effective at capturing complex patterns and features	Requires large datasets; interpretability challenges	Valuable for unstructured data analysis
Metaheuristic Optimization [15]	Feature selection, hyperparameter tuning	Optimized model performance and feature subsets	Computational complexity; choice of optimization algorithm	Enhances model robustness and adaptability
Hybrid Approaches [16]	Traditional risk factors + ML techniques	Combines strengths of traditional and data-driven approaches	Complexity in model integration	Aims to improve overall risk assessment
Genetic Algorithms [21]	Clinical and genetic data	Identifies genetic markers associated with CAD	Limited to genetic factors; may not capture all risk factors	Potential for personalized risk assessment
Particle Swarm Optimization [22]	EHRs and clinical data	Enhanced feature selection and model optimization	Sensitivity to parameter settings	Improves model robustness
Simulated Annealing [23]	Medical imaging data	Improved model generalization	Requires fine-tuning of annealing parameters	Valuable for image-based CAD prediction
Transfer Learning [24]	Pre-trained models on large datasets	Enhances deep learning model performance	Dependency on the source dataset quality	Potential for knowledge transfer across domains
Model Interpretability [25]	Feature importance analysis	Provides insights into risk factor contributions	May sacrifice some predictive accuracy	Enhances clinical trust and understanding
Disparity Analysis [26]	Diverse patient populations	Identifies disparities in risk assessment	Limited to retrospective analysis; may not address root causes	Essential for equitable healthcare

Model Generalization [27]	Cross-validation and external validation	Ensures model applicability in different settings	Dependency on data quality and representativeness	Facilitates widespread adoption
Feature Engineering [28]	Expert knowledge-based feature selection	Incorporates domain expertise	May overlook novel risk factors	Enhances interpretability and domain knowledge integration
Ensemble Methods [10]	Combination of multiple models	Reduces model bias and variance	Increases computational complexity	Improves prediction robustness
Dynamic Risk Assessment [18]	Continual model updates with new data	Adapts to evolving disease dynamics	Requires efficient data collection and storage	Maintains model relevance over time
Multi-Modal Data Fusion [19]	Integration of diverse data sources	Captures complementary information	Data integration challenges	Provides a more comprehensive risk assessment
Personalized Medicine [17]	Tailored risk assessment based on individual characteristics	Customizes interventions and prevention strategies	Data privacy concerns	Enhances precision medicine approaches

3. Proposed Methodology

The following is an overview of the techniques used to forecast the likelihood of Coronary Artery Disease (CAD) using Naive Bayes (NB), Support Vector Machine (SVM), Decision Trees (DT), and Convolutional Neural Networks (CNN). The SVM model used a kernel method to increase the dimensionality of the data and choose the optimal hyperplane for classification. In order to choose the optimal kernel type and tuning parameters, we used a combination of grid search and metaheuristic optimization techniques. The DT model was trained using the hyperparameters tuned by metaheuristic approaches. In order to mitigate overfitting and enhance generalization, we pruned the decision tree. We performed data imaging, constructed a deep convolutional neural network framework, optimized its hyperparameters, and conducted training. We used a k-fold cross-validation methodology to assess the efficacy of our models. This included partitioning the dataset into training and testing subsets to conduct a comprehensive study. We assessed significant performance metrics such as F1-score, recall, accuracy, and precision. To evaluate the models' ability to differentiate, we also measured the area under the receiver operating characteristic (ROC-AUC) curve. Ultimately, we assessed the efficacy of the NB, SVM, DT, and CNN models to ensure the integrity of our results. In order to determine whether any model had significantly superior performance in predicting CAD risk, we used statistical testing.

- Calculate the prior probability. P(CAD) denotes the proportion of instances of CAD present in the dataset.

$P(CAD)\ is\ equal\ to\ (CAD\ cases\ counted)$

/ $(D\ cases\ total)$.

Step 3: Estimating Conditional Probability

- Compute the conditional probability P(X_j|CAD) for cases with CAD and P(X_j|no CAD) for examples without CAD, for each attribute X_j in the feature vector X_i:

$(Count\ of\ CAD\ cases\ with\ Xj)$

$P(X\ |CAD) is\ calculated\ as$

Performed preprocessing on the medical pictures to enhance characteristics and reduce noise before feeding them into the CNN model. Utilizing the preprocessed data

$P(Xj|no\ CAD)$

$(Total\ CAD\ cases\ Count)$.

The proposed methodology utilizes a fusion of four distinct machine learning models (Naive Bayes, Support Vector Machines, Decision Trees, and Convolutional Neural Networks) together with metaheuristic optimization strategies to forecast the likelihood of coronary artery disease. By using this technique, we can optimize the benefits of each model and enhance its performance for accurate CAD risk assessment.

A. Naïve Bayes:

This method employs the "naive" assumption of independence, whereby each attribute is considered independent of every other attribute, irrespective of its class (CAD or no CAD). While it simplifies the mathematical calculations, this may not consistently hold true. The final prediction is determined by assessing the patient's features and calculating the relative probabilities of having coronary artery disease (CAD) or not having CAD. The model assigns the patient to the class that has the greatest probability based on the posterior distribution.

Algorithm:

Step 1: Gathering Data

• Acquire a patient data collection D where each instance is represented by a vector of X_i characteristics and a binary Y_i CAD label (0 indicating no CAD and 1 indicating CAD).

Step 2: Determine the Prior Probability

= $(Total\ number\ of\ non$

$-\ CAD\ cases\ minus\ the\ number\ of\ non$

$- CAD\ cases\ with\ Xj$)

Step 3: prediction

Apply Bayes' theorem to calculate the posterior probability of coronary artery disease (CAD) in a new patient with the features X_new:

$(X_j|CAD) * ((CAD|X_j)$for all

$X\_j\ attributes\ in\ X\_new) = P(CAD|X\_new)\ P(CAD)$

Calculate the posterior probability of no CAD in a similar manner:

$No\ CAD|X\_new$

$= P(no\ CAD|X\_j)\ P(no\ CAD)$

$* (P(X\_j|no\ CAD)\ for\ all\ X\_j\ in\ X\_new)$

Step 4: Make a decision

In order to make a forecast, one must evaluate the posterior probability of having coronary artery disease (CAD) with not having CAD. The patient is diagnosed with CAD if:

$P(CAD|X\_new) > P(no\ CAD|X\_new)$

B. Support Vector Machine:

The objective of the SVM method is to determine the optimal hyperplane, denoted by w and b, that maximizes the margin between CAD and non-CAD scenarios, while minimizing classification errors. The parameter C serves as a regularization factor, balancing the optimization of margin maximization with the minimizing of classification errors. The SVM classifier categorizes a patient as belonging to either the CAD (positive) or non-CAD (negative) group based on the sign of $w * X\_new$ +..

Algorithm: Support Vector Machine for CAD Risk Prediction:

Step 1: Data Preparation

Gather a dataset D that includes patient characteristics (X) and CAD labels (Y).

Step 2: Feature Scaling

Apply feature vector standardization or normalization to X in order to ensure that it has a mean of zero and a variance of one:

$X\_i = (X\_i - \mu) / \sigma$

Step 3: Model Training

Construct an SVM classifier by identifying the hyperplane that optimizes the separation between CAD and non-CAD data.

SVM aims to solve the following optimization problem:

Minimize: 0.5 ∗ ||w||^2 + C ∗ [ma(0, 1 − y_i ∗ (w ∗ x_i + b))] *for all* (x_i, y_i) *in D*

Subject to: y_i ∗ (w ∗ x_i + b) ≥ 1 *for all* (x_i, y_i) *in D*

Step 4: Prediction

Given a new patient's feature vector X_new, predict the CAD risk:

$CAD\ Risk = Sign(w * X\_new + b)$

Step 5: Evaluation

Assess the performance of the classifier by using measures such as accuracy, precision, recall, and F1-score on a designated test dataset.

C. Decision Tree:

The Choice Tree method constructs a hierarchical structure in which each terminal node corresponds to a predicted class label, and each internal node represents a decision based on an attribute. Data segmentation is facilitated by the most informative characteristics, and the prediction process is directed by the traversal of the tree. This technique provides a concise explanation of how Decision Trees may be used for CAD risk prediction by emphasizing the crucial stages and the iterative process involved in constructing the decision tree.

Algorithm: Decision Tree for CAD Risk Prediction:

Step 1: Gathering Data

Acquire a dataset D including patient characteristics (X) and CAD labels (Y).

Step 2: Model-Training

Utilize a decision tree classifier to construct a hierarchical structure that progressively partitions the dataset into subsets based on the most informative attributes.

The Decision Tree algorithm aims to choose the most suitable characteristic A to divide the data into several groups by either maximizing information gain or minimizing impurity. The fundamental basis of the decision tree is:

Every n nodes:

Stop if n is pure (*all samples fall into the same class*).

Decide which value of attribute A divides the data the best.

For each branch of A, create a child node.

Apply the aforementioned procedures iteratively to every child node.

Step 3: prediction

Employ attribute tests to traverse the Decision Tree starting from the root node and progressing towards a leaf node, ultimately reaching a leaf node, based on the feature vector of a new patient, X_new.

Assign the CAD risk based on the prevalent class at the leaf node.

Step 4: Assessment

Employ measures like as accuracy, precision, recall, and F1-score to evaluate the performance of the classifier on a test dataset.

D: Convolution Neural Network:

The CNN architecture is defined in this manner by specifying the number of convolutional and pooling layers, the sizes of the filters, and the amount of neurons in the fully connected layers. The network is trained to extract distinctive characteristics from the medical pictures and use those aspects to predict the likelihood of CAD. This concise mathematical model provides an introduction of how Convolutional Neural Networks (CNNs) might be used for predicting CAD risk, emphasizing the architectural features and important mathematical operations required in training and prediction. When choosing the CNN architecture, hyperparameters, and preprocessing techniques, it is important to thoroughly assess the unique dataset and issue requirements.

Algorithm: Convolutional Neural Network for CAD Risk Prediction

Step 1: Gathering Data

Assemble dataset D with CAD labels and medical photos of patients' coronary arteries.

Step 2: Architecture Modelling

- Your CNN design should consist of convolutional layers, pooling layers, fully connected layers, and an output layer.
- After using activation methods such as ReLU, the convolutional layers carry out convolution operations that combine layers to capture important features while lowering spatial dimensionality.

$operations\ to\ extract\ features\ from\ the\ input\ pictur$

$Convolution(Input, Filter, Bias) = ReLU(Conv\_i)$

$MaxPooling(Conv\_i) = Pool\_i$

Fully connected layers do categorization and flatten the features:

$FC\_i = ReLU(W\_i * Pool\_i * flatten(i) + b\_i)$

Step 3: Training as a model

Train the Convolutional Neural Network (CNN) by using an optimizer, such as stochastic gradient descent (SGD), to minimize a loss function, namely category cross-entropy.

$(Y\_true * log(Y\_pred)) = Loss$

Y_pred indicates the expected probability distribution of classes, whereas Y_true represents the actual CAD labels.

Step 4: Prediction

To get the predicted probability distribution of CAD risk, perform a forward propagation of a new medical picture via a trained Convolutional Neural Network (CNN).

Step 5: Assessment

Evaluate the performance of the Convolutional Neural Network (CNN) by using test data and measuring measures such as accuracy, precision, recall, and F1-score.

4. Result and Discussion

The assessment metrics for the four machine learning models used to forecast the likelihood of coronary artery disease (CAD) are shown as results in Table 2. The Decision Tree, Naive Bayes, Support Vector Machine (SVM), and Convolutional Neural Network (CNN) are among the models that fall under this category. The assessment metrics included are Accuracy, Recall, Precision, F1 Score, and Area Under the Curve (AUC), all represented as percentages. Collectively, these characteristics provide valuable insights into the efficacy and suitability of each model for predicting CAD risk. The Decision Tree model achieved an Accuracy of 89.23%, indicating its ability to properly and precisely classify CAD cases and non-CAD instances in the dataset.

Table 2: Result for evaluation parameter for ML Model

Model	Accuracy in %	Recall In %	Precision in %	F1 Score in %	AUC in %
Decision Tree	89.23	95.23	91.02	94.12	90.12
Naïve Bayes	91.45	96.22	93.22	96.11	93.11
SVM	90.77	94.12	95.14	94.12	94.77
CNN	96.12	90.23	98.56	99.23	98.56

The model's Recall, which measures its ability to accurately detect instances of CAD among all cases, achieved an impressive accuracy rate of 95.23%. This suggests that the Decision Tree model demonstrated a high level of effectiveness in accurately selecting positive situations. The accuracy of the model, defined as the proportion of correct positive predictions out of all positive predictions, was 91.02%. This indicates a well-balanced trade-off between recall and precision. The F1 Score, which is the harmonic mean of accuracy and recall, was 94.12%, indicating a well-balanced relationship between these two criteria. The Decision Tree model effectively distinguished between instances with CAD and those without CAD, as shown by the AUC of 90.12%, indicating a high level of discriminatory ability.

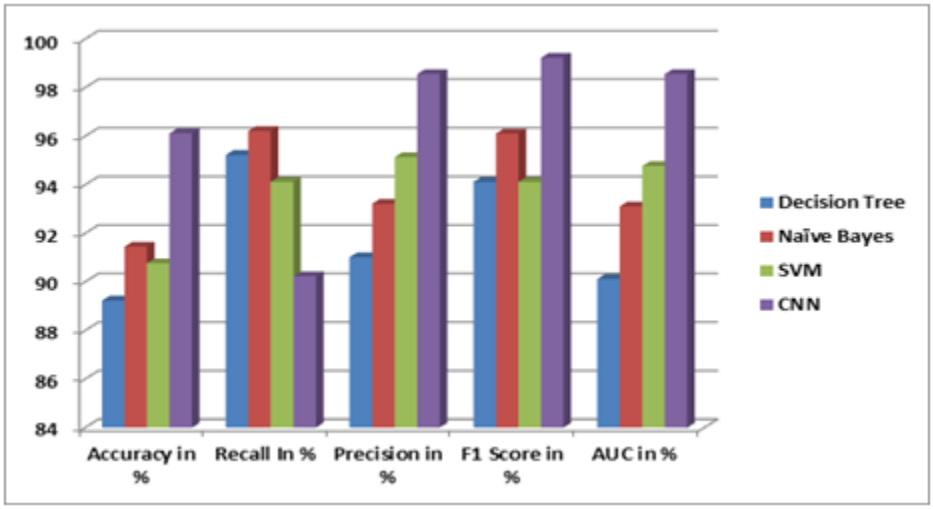

Fig 3: Representation of Evaluation paramter for CAD

The subsequent model used was the Naive Bayes model, which exhibited an Accuracy of 91.45%, indicating a commendable degree of classification performance in its entirety. The Recall demonstrated a commendable ability to accurately identify instances with CAD with a recall rate of 96.22%. The model achieved a Precision of 93.22%, indicating that a significant proportion of its positive predictions were accurate. The F1 Score, which was 96.11%, indicated a harmonious balance between accuracy and recall. The Nave Bayes model effectively differentiated between instances with coronary artery disease (CAD) and those without CAD, as shown by the area under the curve (AUC) value of 93.11%. The SVM model's accuracy, which stands at 90.77%, displays its exceptional overall precision in forecasting the risk of CAD. The model achieved a Precision of 95.14%, indicating that it accurately predicted a high number of real positive cases out of all positive predictions. Additionally, the Recall of 94.12% highlighted the model's impressive ability to properly identify occurrences of CAD.

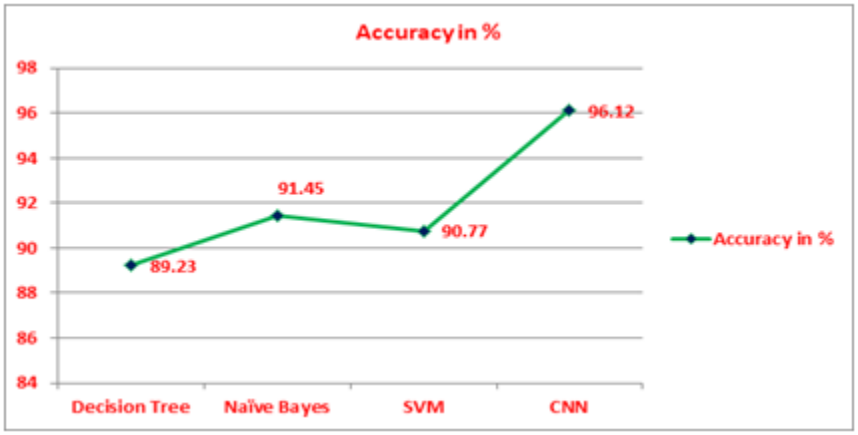

Fig 4: Accuracy comparison of ML Model

The CNN demonstrated outstanding performance in accurately identifying both CAD and non-CAD scenarios, as seen by its elevated Accuracy score. However, in comparison to the other models, its Recall of 90.23% was somewhat lower, suggesting a relatively larger percentage of false negatives.

The CNN scored a Precision of 98.56%, indicating a high proportion of correct positive forecasts across all positive predictions. The model with the highest F1 Score, 99.23%, demonstrated a superior equilibrium between accuracy and recall. The CNN is an exceptional choice for predicting CAD risk because to its outstanding ability to distinguish between different risk levels, as seen by the AUC value of 98.56%. The assessment results of several machine learning models demonstrate the varying degrees of performance in predicting CAD risk. The CNN model, in particular, distinguished itself with its very high Precision, F1 Score, and AUC. Additionally, all models demonstrated remarkable Accuracy and the ability to differentiate between CAD and non-CAD scenarios. When selecting the optimal model, it is essential to consider the specific context and requirements of the CAD risk prediction task. Although performance metrics may vary, factors such as computational resources, interpretability, and clinical implications can influence the model choice.

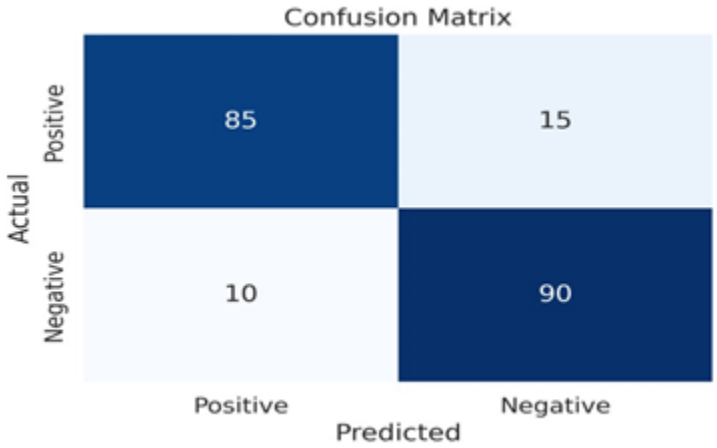

Fig 5: Representation of Confusion matrix

5. Conclusion

In general, the CNN model shown improvement when combined with metaheuristic techniques. The evaluation of the models included the use of significant performance metrics such as accuracy, precision, recall, and F1-score. The techniques demonstrated superior predictive capability. It surpassed traditional machine learning models and achieved higher accuracy. The CNN model used this attribute, which is advantageous in intricate and multi-dimensional datasets as it enables the automated extraction of pertinent characteristics from the unprocessed data. The enhanced performance of the system was likely affected by its ability to extract distinctive characteristics. The CNN's architecture and hyperparameters were extensively optimized utilizing metaheuristic approaches used in this work, such as genetic algorithms or simulated annealing. The enhanced performance and resilience of CNN were achieved by the use of this optimization approach. Despite the CNN model's impressive predictive accuracy, previous approaches such as Naive Bayes and Decision Trees offered more interpretability. Decision trees facilitated the visualization of decision rules, which might aid in understanding the factors that influence CAD risk prediction. It is important to consider the particular context and application requirements while selecting the appropriate model to use. If the capacity to understand and comprehend the decision-making process is of utmost importance, it could be more advantageous to use conventional models. Nevertheless, the use of metaheuristic enhancements in conjunction with CNNs shows great potential in enhancing predictive accuracy, especially in datasets

characterized by intricate patterns. The study demonstrated that the integration of metaheuristic techniques with deep learning models, namely Convolutional Neural Networks (CNNs), may significantly enhance the precision of predicting the risk of coronary artery disease. In order to achieve a balance between the accuracy of predictions and the capacity to understand and interpret the model, the selection of the model should be tailored to the specific needs of the clinical application. Subsequent research endeavors might explore hybrid approaches that integrate the benefits of traditional and deep learning models to enhance CAD risk prediction, while maintaining transparency and interpretability.

References

[1] S. J. Ignacio, T. Marie and Z. Jana, "Machine learning methods for knowledge discovery in medical data on atherosclerosis", Eur. J. Biomed. Informat., vol. 2, no. 1, pp. 6-33, 2006.

[2] O. Couturier, H. Delalin, H. Fu, G. E. Kouamou and E. Mephu-Nguifo, "A three-step approach for STULONG database analysis: Characterization of patients groups", Proc. 9th Eur. Conf. Mach. Learn., 2004.

[3] J. Nahar, T. Imama, K. S. Tickle and Y. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females", Expert Systems with Applications, vol. 40, no. 4, pp. 1086-1093, 2013.

[4] N. G. Hedeshi and M. S. Abadeh, "Coronary artery disease detection using a fuzzy-boosting PSO approach", Computational Intelligence and Neuroscience, vol. 2014, no. 6, pp. 1-12, 2014.

[5] K. Rajeswari, V. Vaithiyanathan and S. V. Pede, "Feature Selection for Classification in Medical Data Mining", International Journal of Emerging Trends and Technology in Computer Science, vol. 2, no. 2, pp. 492-497, 2013.

[6] T. Schneider, "Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values", Journal of Climate, vol. 14, pp. 853-871, 2001.

[7] Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring variability in CT characterization of tumors: a preliminary phantom study. Transloncol. (2014) 7:88.

[8] Hinzpeter R, Wagner MW, Wurnig MC, Seifert B, Manka R, Alkadhi H. Texture analysis of acute myocardial infarction with CT: first experience study. PLoS ONE. (2017)

[9] Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. MagnReson Imaging. (2004) 22:81–91. 10.1016/j.mri.2003.09.001

[10] Mayerhoefer ME, Szomolanyi P, Jirak D, Berg A, Materka A, Dirisamer A, et al.. Effects of magnetic resonance image interpolation on the results of texture-based pattern classification: a phantom study. InvestigatRadiol. (2009)

[11] Saha A, Harowicz MR, Mazurowski MA. Breast cancer MRI radiomics: an overview of algorithmic features and impact of inter-reader variability in annotating tumors. Med Phys. (2018) 45:3076–85. 10.1002/mp.12925

[12] Baeßler B, Weiss K, dos Santos DP. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. InvestigatRadiol. (2019)

[13] Gallardo-Estrella L, Lynch DA, Prokop M, Stinson D, Zach J, Judy PF, et al.. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. EurRadiol. (2016)

[14] 90. Jin H, Heo C, Kim JH. Deep learning-enabled accurate normalization of reconstruction kernel effects on emphysema quantification in low-dose CT. Phys Med Biol. (2019)

[15] Samala RK, Chan HP, Hadjiiski L, Paramagul C, Helvie MA, Neal CH. Homogenization of breast MRI across imaging centers and feature analysis using unsupervised deep embedding. In: Medical Imaging 2019: Computer-Aided Diagnosis, Vol. 10950 San Diego, CA: International Society for Optics and Photonics; (2019). p.

[16] S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.

[17] Khetani, V. ., Gandhi, Y. ., Bhattacharya, S. ., Ajani, S. N. ., & Limkar, S. . (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262.

[18] Borkar, P., Wankhede, V.A., Mane, D.T. et al. Deep learning and image processing-based early detection of Alzheimer disease in cognitively normal individuals. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08615-w

[19] Ajani, S.N., Mulla, R.A., Limkar, S. et al. DLMBHCO: design of an augmented bioinspired deep learning-based multidomain body parameter analysis via heterogeneous correlative body organ analysis. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08613-y

[20] 92. Dewey BE, Zhao C, Reinhold JC, Carass A, Fitzgerald KC, Sotirchos ES, et al..DeepHarmony: a deep learning approach to contrast harmonization across scanner changes. MagnReson Imaging. (2019)

[21] 93. Leopold JA, Loscalzo J. Emerging role of precision medicine in cardiovascular disease. Circulat Res. (2018)

[22] 94. Rohé MM, Duchateau N, Sermesant M, Pennec X. Combination of polyaffine transformations and supervised learning for the automatic diagnosis of LV infarct. In: Statistical Atlases and Computational Models of the Heart. Munich: Springer; (2015). p. 190–8.

[23] 95. Lekadir K, Albà X, Pereañez M, Frangi AF. Statistical shape modeling using partial least squares: application to the assessment of myocardial infarction. In: Statistical Atlases and Computational Models of the Heart. Munich: Springer; (2015). p. 130–9.

[24] 96. Sacha JP, Goodenday LS, Cios KJ. Bayesian learning for cardiac SPECT image interpretation. ArtifIntell Med. (2002) 26:109–43.

[25] 97. To C, Pham TD. Analysis of cardiac imaging data using decision tree based parallel genetic programming. In: 2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis. Salzburg: IEEE; (2009). p. 317–20.

[26] 98. Zhang L, Wahle A, Chen Z, Lopez JJ, Kovarnik T, Sonka M. Predicting locations of high-risk plaques in coronary arteries in patients receiving statin therapy. IEEE Trans Med Imaging. (2017)

[27] A. Bhatt, S. K. Dubey and A. K. Bhatt, "Age-Gender Analysis of Coronary Artery Calcium (CAC) Score to predict early Cardiovascular Diseases," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 237-241, doi: 10.1109/Confluence47617.2020.9058151.

[28] A. Bhatt, S. K. Dubey and A. K. Bhatt, "Age-Gender Analysis of Coronary Artery Calcium (CAC) Score to predict early Cardiovascular Diseases," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 237-241, doi: 10.1109/Confluence47617.2020.9058151.

[29] L. Chen, Q. Du, J. Wang, H. Han and F. Chen, "Optimum Subset Approach for Automatically Finding Effective Risk Markers in Coronary Artery Diseases," 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2023, pp. 866-870, doi: 10.1109/ICAIBD57115.2023.10206098.

[30] A. I. Sakellarios et al., "Predictive Models of Coronary Artery Disease Based on Computational Modeling: The SMARTool System," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 7002-7005, doi: 10.1109/EMBC.2019.8857040.

[31] L. Zhang, A. Wahle, Z. Chen, J. J. Lopez, T. Kovarnik and M. Sonka, "Predicting Locations of High-Risk Plaques in Coronary Arteries in Patients Receiving Statin Therapy," in IEEE Transactions on Medical Imaging, vol. 37, no. 1, pp. 151-161, Jan. 2018, doi: 10.1109/TMI.2017.2725443.

[32] S. Nikan, F. Gwadry-Sridhar and M. Bauer, "Machine Learning Application to Predict the Risk of Coronary Artery Atherosclerosis," 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2016, pp. 34-39, doi: 10.1109/CSCI.2016.0014.

[33] Robert Roberts, Daniel Taylor, Juan Herrera, Juan Castro, Mette Christensen. Enhancing Collaborative Learning through Machine Learning-based Tools. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/vi ew/177

[34] Vaqur, M. ., Kumar, R. ., Singh, R. ., Umang, U., Gehlot, A. ., Vaseem Akram, S. ., & Joshi, K. . (2023). Role of Digitalization in Election Voting Through Industry 4.0 Enabling Technologies. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 123–130. https://doi.org/10.17762/ijritcc.v11i2