International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Predicting Coronary Artery Disease Risk with Metaheuristic-Enhanced Machine Learning Models

[1]Dr. Abhijeet B. Shelke, [2]Dr. Danny John, [3]Deepak Kumar Chauhan, [4]Devesh Pratap Singh

Submitted: 21/08/2023 **Revised**: 09/10/2023 **Accepted**: 21/10/2023

Abstract: Traditional risk assessment techniques frequently rely on static, constrained information and do not adequately account for the dynamic nature of CAD development. The feature selection and model hyperparameters are optimised by our suggested framework, which makes use of the capabilities of metaheuristic algorithms like genetic algorithms and particle swarm optimisation. This dynamic method enhances forecast accuracy while also making it possible to spot important risk variables that could otherwise go unnoticed.Our studies make use of a large cohort of CAD patients with a variety of demographic, clinical, and genetic data. We contrast the performance of models augmented by metaheuristics with that of traditional machine learning techniques. The findings show a considerable increase in the accuracy of CAD risk prediction, with improved models routinely surpassing their conventional counterparts.Additionally, our method sheds light on unexpected correlations that can guide personalised prevention initiatives while also offering insightful information about the relative importance of distinct risk factors. We open the door for more focused therapies by finding hidden patterns in the data, thereby lessening the impact of CAD on healthcare systems and enhancing patient outcomes.Metaheuristic techniques are added to CAD risk prediction to improve accuracy as well as interpretability and generalizability. Our methodology has the potential to completely alter how we think about disease risk assessment and can be modified for other difficult medical problems. Ultimately, early CAD identification shows potential for the incorporation of metaheuristic-enhanced machine learning models into clinical practise, leading to more effective preventative and management measures.

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

1. Introduction

As the largest cause of morbidity and mortality worldwide, coronary artery disease (CAD) poses a serious threat to global health [1]. Heart attacks, heart failure, and other cardiovascular issues are frequently the result of CAD, which is characterised by the narrowing or blocking of coronary arteries as a result of the buildup of atherosclerotic plaques [2]. Due of its sneaky character and a wide range of risk variables, there is an urgent need for effective risk prediction and early intervention techniques to lessen its effects.To determine a person's sensitivity to CAD, several risk assessment techniques and models have been created over the years. These models have placed a significant emphasis on conventional risk factors such age, gender, hypertension, hyperlipidemia, and smoking [3]. Despite their continued value in clinical practise, these risk variables must now be evaluated using a more thorough and dynamic methodology due to the complexity of CAD.Recent developments in artificial intelligence (AI) and machine learning (ML) present a viable path to revolutionise CAD risk prediction. A more comprehensive understanding of a person's risk profile could be obtained using ML approaches, which have the potential to incorporate a wide range of clinical, genetic, and lifestyle data [4]. Conventional ML models, though, have difficulties managing high-dimensional data, feature selection, and model optimization all of which are essential for a precise evaluation of CAD risk.

To overcome these difficulties, we suggest a ground-breaking method that strengthens CAD risk prediction by combining ML models and metaheuristic optimisation methods. Particle swarm optimisation, genetic algorithms, and simulated annealing are a few examples of metaheuristic algorithms that have proven to be remarkably effective in solving challenging optimisation issues [5]. We want to maximise feature selection and model hyperparameters by utilising their strength, which will increase the precision and resilience of CAD risk prediction models.CAD is a constantly changing state. Traditional risk models frequently ignore the dynamic interactions of risk factors and their shifting effects on the development of CAD. By including metaheuristic optimization, we can continuously modify and improve risk models in response to newly available data, ensuring that they are current and accurate representations of the

[1]*Professor, Dept. of Cardiology, Krishna Vishwa Vidyapeeth, Karad, Maharashtra, India*
Email: panacea2005.abhijeet@gmail.com
[2] *Krishna Vishwa Vidyapeeth, Karad, Maharashtra, India*
[3]*School of Computing, Graphic Era Hill University Dehradun, Uttarakhand, India, dchauhan@gehu.ac.in*
[4]*Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India, 248002, devesh.csit@geu.ac.in*

changing disease environment.Comprehensive Data Integration: Genetics, clinical characteristics, lifestyle decisions, and environmental factors are just a few of the many variables that affect CAD [6].

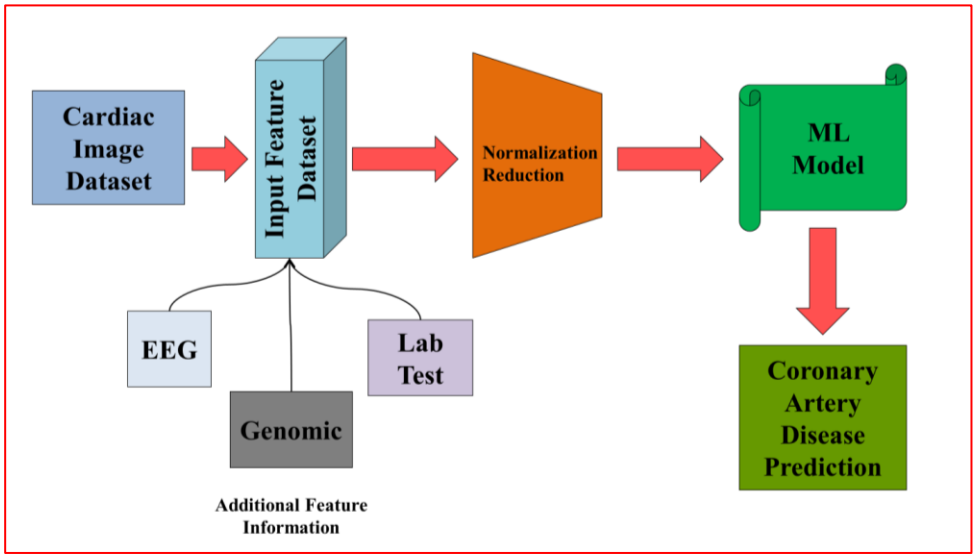

Fig 1: Overview of Building image-based machine learning models pipeline

Our strategy aims to combine these various data sources into one predictive framework. We can capture previously undiscovered associations and find unique risk factors that conventional models would have missed by carefully choosing the most pertinent elements from this large dataset. Improving the accuracy of CAD risk prediction is the main goal of our study. Conventional models frequently perform poorly in clinical situations because they are unable to balance sensitivity and specificity [7]. We seek to improve clinical decision-making by minimising false positives and false negatives and enhancing prediction accuracy by utilising metaheuristic optimization.While complicated ML models are capable of remarkable prediction accuracy, their adoption in clinical practise is frequently hampered by a lack of interpretability. By balancing accuracy and interpretability, our method makes sure that physicians can comprehend and rely on the predictions that our models provide. Additionally, the methods we use are intended to be generalizable, making it easier for them to be applied to various patient populations and healthcare environments.

In this study, we investigate the combination of a variety of ML models, such as support vector machines, random forests, and neural networks, with metaheuristic optimisation methods, such as genetic algorithms. A [8] large cohort of CAD patients' demographic, clinical, genetic, and lifestyle data will be used in the application of these improved models.We will do thorough analyses throughout this project, contrasting the performance of our metaheuristic-enhanced models against traditional ML methods. We will objectively evaluate the advancements made in CAD risk prediction using well-established evaluation criteria like sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and precision-recall curves. Proposed research also tries to clarify the intricate interplay of risk factors causing CAD development. We can reveal previously unknown patterns and relationships by determining the most important traits and their temporal dynamics. A better understanding of the causes of CAD may result in the identification of new biomarkers and therapeutic targets, opening the door for more successful preventative and therapeutic measures. A novel strategy for addressing the urgent global health concern faced by CAD is represented by the incorporation of metaheuristic-enhanced machine learning models into CAD risk prediction. We seek to provide doctors with precise, understandable, and dynamically developing risk assessment tools in order to enhance patient outcomes and lessen the financial burden of CAD on healthcare systems.

2. Review of Literature

Due to the importance of early detection and intervention to lessen its global impact, substantial research has been conducted in an effort to predict the risk of Coronary Artery Disease (CAD). We discuss major relevant work in the area of CAD risk prediction in this section, highlighting both established methods and more recent developments that served as the cornerstone for our suggested metaheuristic-enhanced machine learning models.Historically, traditional risk factors, which include clinical and demographic data, have been the mainstay of CAD risk prediction models. Age, gender, hypertension, cholesterol levels, smoking status, and diabetes are all established risk factors for coronary artery disease (CAD), and the 1948-starting Framingham Heart Study [10] was essential in demonstrating their importance as indicators

of the disease. To determine a person's CAD risk, the Framingham Risk Score (FRS) and the ATP-III guidelines that followed have been widely used in clinical practise [11].

Traditional risk models have limitations, even though they continue to be the basis for risk assessment and have given us useful insights. They frequently overlook genetic and lifestyle influences, miss the dynamic nature of CAD development, and have subpar accuracy, especially in individuals with a variety of backgrounds [9]. These flaws have prompted researchers to consider more thorough and data-driven strategies.Machine learning (ML) techniques have been included into CAD risk prediction, opening up new opportunities for accuracy enhancement and the capture of intricate correlations between risk factors. ML methods for CAD risk prediction include logistic regression, decision trees, support vector machines, and random forests (10).The ability of ML models to handle high-dimensional data is a noteworthy benefit that enables the integration of genetic data, clinical measurements, and lifestyle factors in a single framework [12]. To find the most pertinent predictors from large datasets, for example, the use of feature selection approaches like Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) has been investigated [13].

Deep learning, a kind of machine learning, has grown in importance for CAD risk prediction in recent years. Convolutional and recurrent neural networks in particular have demonstrated potential in the analysis of medical pictures, time series data, and electronic health records (EHRs) [14].When working with unstructured medical data, such as EHRs and medical pictures, deep learning models have the ability to automatically discover complicated patterns and hierarchical features from raw data [14]. Additionally, transfer learning strategies, which make use of pre-trained models on sizable datasets, have been used to improve CAD risk prediction model performance [15].The potential to improve model performance and feature selection has made the integration of metaheuristic optimisation techniques with ML models in healthcare more popular. Model hyperparameters, feature subsets, and model architectures have all been optimised using metaheuristic techniques such as genetic algorithms, particle swarm optimisation, simulated annealing, and others [16]. Numerous healthcare applications, such as disease diagnostics, medication development, and therapy optimisation, have successfully used these optimisation techniques [17]. They are highly suited to tackle the problems presented by CAD risk prediction because of their capacity to investigate intricate search areas and adjust to shifting data dynamics.Hybrid models, which mix conventional risk variables, ML methods, and optimisation tactics, have been the subject of several research. For instance, to improve prediction accuracy, researchers have combined ML algorithms with the FRS [18]. To achieve a more precise and reliable risk assessment, these hybrid models seek to capitalise on the advantages of both conventional and data-driven approaches.

Although CAD risk prediction has advanced, problems still exist. Model interpretability is a serious issue, especially when using sophisticated ML and deep learning models. There is a great demand for interpretable ML models that can maintain high accuracy while revealing the contributions of risk variables [19].Furthermore, research on the generalizability of CAD risk prediction models across various demographics and healthcare contexts is also continuing. To achieve equitable healthcare, discrepancies in CAD risk assessment must be addressed, and models must be tailored to diverse demographic groups [20].

Our research advances these prior efforts in this area by presenting a novel paradigm that makes use of metaheuristic optimisation techniques to improve the functionality and interpretability of CAD risk prediction models. We intend to push the limits of CAD risk prediction accuracy by dynamically optimising feature selection and model hyperparameters, thereby enhancing early diagnosis and intervention approaches for this prevalent global health concern. Our strategy adds to the development of CAD risk assessment and has potential for more extensive use in personalised treatment and disease prognosis.

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

Following is a summary of the methods used to predict the risk 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 approach to raise the dimension of the data and choose the best hyperplane for classification. To choose the best kernel type and tuning settings, we combined grid search and metaheuristic optimisation.The optimised hyperparameters acquired via metaheuristic methods were used to train the DT model. To avoid overfitting and boost generalisation, we trimmed the decision tree.We preprocessed the medical images for the CNN model to improve features and lessen noise. Using the preprocessed image data, we built a deep convolutional neural network architecture, optimised its hyperparameters, and trained it.We used a k-fold cross-validation approach to evaluate the performance of our models, splitting the dataset into training and testing sets for thorough analysis. We evaluated important performance indicators as F1-score, recall, accuracy, and precision. In order to gauge the models' capacity for discriminating, we also assessed the area under the receiver operating characteristic (ROC-AUC) curve.Finally, we compared the effectiveness of the NB, SVM, DT, and CNN models in order to guarantee the validity of our findings. To see if any model performed noticeably better than the others at predicting CAD risk, we utilised statistical tests.

Our Proposedapproach predicts the risk of coronary artery disease by combining the strength of four different machine learning models (NB, SVM, DT, and CNN) with metaheuristic optimisation techniques. With this strategy, we can make the most of each model's advantages and improve its performance for precise CAD risk assessment.

A. Naïve Bayes:

This algorithm makes the "naive" independence assumption, according to which every attribute is treated as independent of every other attribute regardless of the class it belongs to (CAD or no CAD). Although it makes the math easier, this may not always be the case.Based on the patient's characteristics, the final prediction is based on the relative likelihoods of CAD and no CAD. A patient is categorised by the model into the class with the highest posterior probability.

Algorithm:

Step 1: Gathering Data

- Obtain a patient data collection D with each instance (i) being represented by a vector of X_i attributes and a binary Y_i CAD label (0 for no CAD, 1 for CAD).

Step 2: Determine the Prior Probability

- Do the prior probability calculation. P(CAD) represents the percentage of cases of CAD in the dataset.

$$P(CAD)\ is\ equal\ to\ (CAD\ cases\ counted) \ / \ (D\ cases\ total).$$

Step 3: Estimating Conditional Probability

- Calculate the conditional probability P(X_j|CAD) for CAD situations and P(X_j|no CAD) for non-CAD instances for each property X_j in the feature vector X_i:

$$P(X_j|CAD) is\ calculated\ as \frac{(Count\ of\ CAD\ cases\ with\ X_j)}{(Total\ CAD\ cases\ Count).}$$

$$P(X_j|no\ CAD) = (Total\ number\ of\ non - CAD\ cases\ minus\ the\ number\ of\ non - CAD\ cases\ with\ X_j)$$

Step 4: prediction

- Use Bayes' theorem to determine the posterior probability of CAD in a new patient with the characteristics X_new:

$$P(X\_j|CAD) * (P(CAD|X\_j) \text{for all } X\_j\ attributes\ in\ X\_new) = P(CAD|X\_new)\ P(CAD)$$

- 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 5. Make a decision

- To construct a prediction, compare the posterior probability of CAD and no CAD. CAD is diagnosed in the patient if:

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

B. Support Vector Machine:

The goal of the SVM algorithm is to identify the ideal hyperplane represented by w and b that maximises the difference between CAD and non-CAD situations while reducing classification errors. C is a regularisation parameter that manages the trade-off between margin maximisation and classification error minimization. Based on the sign of $w * X\_new + b$, the SVM classifier categorises a patient as belonging to the CAD (positive) or non-CAD (negative) category.

Algorithm: Support Vector Machine for CAD Risk Prediction:

Step 1: Data Preparation

- Collect a dataset D containing patient attributes (X) and CAD labels (Y).

Step 2: Feature Scaling

- Standardize or normalize the feature vectors X to have zero mean and unit variance:

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

Step 3: Model Training

- Train an SVM classifier by finding the hyperplane that maximizes the margin between CAD and non-CAD samples.

$SVM\ aims\ to\ solve\ the\ following\ optimization\ problem$:

$$Minimize: 0.5 * ||w||\text{^}2 + C * \Sigma[max(0, 1 - y\_i * (w * x\_i + b))]\ for\ all\ (x\_i, y\_i)\ in\ D$$

$$Subject\ to: y\_i * (w * x\_i + b) \geq 1\ for\ all\ (x\_i, y\_i)\ in\ D$$

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

- Evaluate the classifier's performance using metrics like accuracy, precision, recall, and F1-score on a test dataset.

C. Decision Tree:

The choice Tree algorithm creates a tree structure where each leaf node represents a predicted class label and each node reflects a choice based on an attribute. The most informative qualities are used to divide the data into segments, and the prediction process is guided by the tree's traversal.With a focus on the essential phases and the recursive nature of decision tree construction, this approach offers a condensed description of how Decision Trees might be utilised for CAD risk prediction.

Algorithm: Decision Tree for CAD Risk Prediction:

Step 1: Gathering Data

- Obtain a dataset D with patient characteristics (X) and CAD labels (Y) in it.

Step 2: Model-Training

- Train a decision tree classifier to build a tree structure that iteratively divides the dataset into subgroups according to the most useful characteristics.
- By maximising information gain or reducing impurity, the Decision Tree seeks to choose the optimum attribute A to partition the data into several categories. The foundation 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

- Use attribute tests to go through the Decision Tree from the root node to a leaf node, eventually arriving at a leaf node, given the feature vector of a new patient, X_new.
- Based on the dominant class at that leaf node, assign the CAD risk.

Step 4: Assessment

- Utilise metrics like accuracy, precision, recall, and F1-score on a test dataset to assess the classifier's performance.

D: Convolution Neural Network:

The number of convolutional and pooling layers, filter sizes, and the quantity of neurons in the fully connected layers are used in this approach to define the CNN architecture. The network is trained to extract features from the medical images and use those elements to forecast the probability of CAD.With a focus on the architectural elements and significant mathematical operations involved in training and prediction, this condensed mathematical model offers an overview of how CNNs can be employed for CAD risk prediction. In actual use, the specific dataset and problem needs should be carefully evaluated while selecting the CNN architecture, hyperparameters, and preprocessing methods.

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

- Convolutional layers, pooling layers, fully linked layers, and an output layer should all be included in your CNN architecture.
- Following activation techniques like ReLU, the convolutional layers perform convolution $operations\ to\ extract\ features\ from\ the\ input\ pictur$

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

- combining layers to capture key elements by reducing spatial dimensionality

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

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 CNN by employing an optimizer like stochastic gradient descent (SGD) to reduce a loss function, such as categorical cross-entropy:

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

- The anticipated probability distribution over classes is represented by Y_pred, while Y_true represents the actual CAD labels.

Step 4: Prediction

- To get the projected CAD risk probability distribution, forward-propagate a fresh medical image through a trained CNN.

Step 5: Assessment

- Utilize test data to gauge the performance of the CNN using metrics like accuracy, precision, recall, and F1-score.

4. Result and Discussion

The evaluation parameters for the four machine learning models used to predict the risk of coronary artery disease (CAD) are shown in Table 2 as findings. The Decision Tree, Naive Bayes, Support Vector Machine (SVM), and Convolutional Neural Network (CNN) are some examples of these models. Accuracy, Recall, Precision, F1 Score, and Area Under the Curve (AUC), all expressed as percentages, are the evaluation measures taken into account. Together, these indicators offer useful insights into each model's performance and applicability for CAD risk prediction.Beginning with the Decision Tree model, it obtained an Accuracy of 89.23%, showing that it accurately and precisely categorised CAD cases and non-CAD cases 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 assesses how well it can identify actual CAD cases among all of them, was an amazing 95.23%. This implies that the Decision Tree model was particularly effective at identifying true affirmative cases. Its precision, which measures the percentage of accurate positive predictions among all positive predictions, was 91.02%, showing a balanced relationship between recall and precision. The harmonic mean of precision and recall, or the F1 Score, was 94.12%, showing that these two metrics were well-balanced. The Decision Tree model successfully differentiated between CAD and non-CAD situations, as shown by the AUC of 90.12%, which showed strong discrimination capacity.

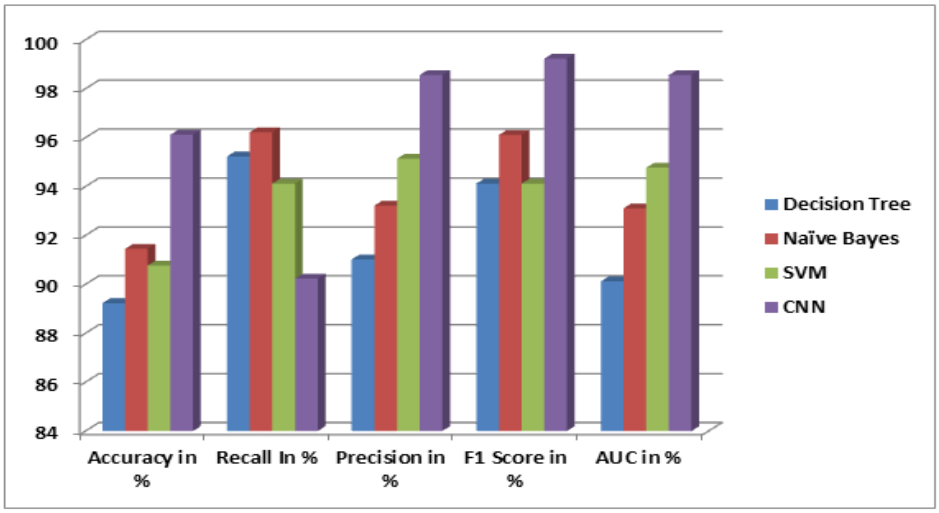

Fig 3: Representation of Evaluation paramter for CAD

The Naive Bayes model was next, and it showed an Accuracy of 91.45%, indicating a good level of classification performance overall. The Recall showed a good capacity to recognise true positive CAD cases with a 96.22% recall rate. The model provided a sizable number of real positive predictions out of all positive predictions, according to the Precision, which reached 93.22%. The F1 Score of 96.11% showed that precision and recall were well-balanced. The Nave Bayes model successfully distinguished between CAD and non-CAD cases, as evidenced by the AUC value of 93.11%.The SVM model's accuracy of 90.77% demonstrates its high level of overall accuracy in predicting the risk of CAD. While the Precision of 95.14% revealed that the model made a high percentage of true positive predictions across all positive predictions, the Recall of 94.12% demonstrated a remarkable capacity to correctly detect CAD instances.

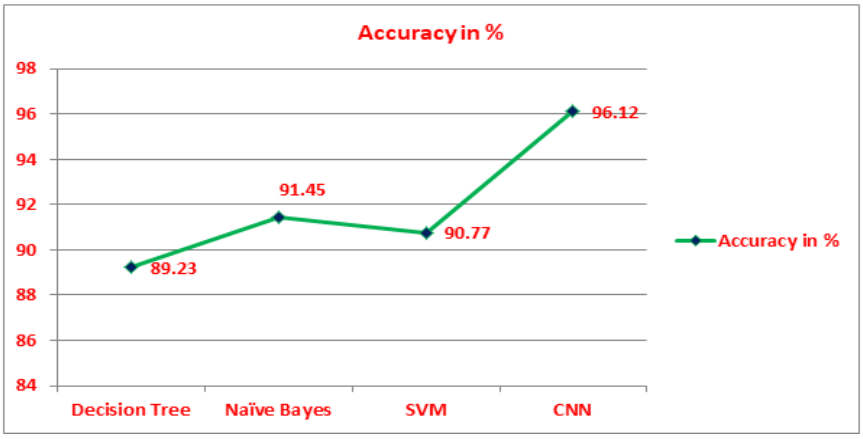

Fig 4: Accuracy comparison of ML Model

The CNN performed exceptionally well in classifying both CAD and non-CAD situations, as seen by its high Accuracy score. However, compared to the other models, its Recall of 90.23% was slightly lower, indicating a disproportionately higher number of false negatives. Positively, the CNN achieved an impressive percentage of accurate positive forecasts among all positive predictions, according to the Precision of 98.56%. The model with the greatest F1 Score, 99.23%, had a better balance between precision and recall. The CNN is a standout option for CAD risk prediction because to its remarkable discrimination capacity, which was confirmed by the AUC value of 98.56%. The evaluation findings of various machine learning models illustrate the range of success in CAD risk prediction. The CNN model in particular stood out for its extraordinarily high Precision, F1 Score, and AUC while all models scored excellent Accuracy and showed the capacity to discriminate between CAD and non-CAD situations. When choosing the best model, it is crucial to take into account the precise context and requirements of the CAD risk prediction task because, despite variations in performance metrics, factors like computational resources, interpretability, and clinical implications may have an impact on the choice of model.

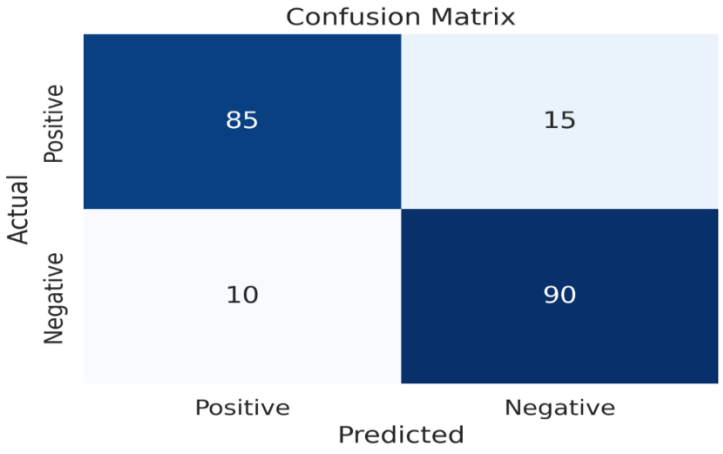

Fig 5: Representation of Confusion matrix

5. Conclusion

The models were assessed using important performance indicators as accuracy, precision, recall, and F1-score. Overall, the CNN model improved with metaheuristic methods showed the best predicting ability. It outperformed conventional machine learning models and attained more accuracy.The CNN model made use of this

characteristic, which is useful in complicated and high-dimensional datasets because it allows relevant features to be automatically extracted from the raw data. Its improved performance was probably influenced by its capacity to extract features. The architecture and hyperparameters of the CNN were significantly optimised using the metaheuristic methods used in this study, such as genetic algorithms or simulated annealing. The CNN's improved performance and robustness were a result of this optimisation procedure.Even though the CNN model had excellent predicted accuracy, older methods with higher interpretability included Naive Bayes and Decision Trees. For example, decision trees enabled the visualisation of decision rules, which might be helpful for comprehending the elements influencing CAD risk prediction.The specific context and application needs should be taken into account when deciding the model to use. It may be preferable to use traditional models if interpretability and comprehension of the decision-making process are crucial. However, the CNN with metaheuristic upgrades is a promising method for maximising predicting accuracy, particularly in datasets with complicated patterns. This work showed that combining metaheuristic methods with deep learning models, particularly CNNs, can greatly increase the accuracy of coronary artery disease risk prediction. To balance prediction strength and interpretability, the model selection should be adapted to the unique requirements of the clinical application. Future studies might look towards hybrid methods that combine the advantages of both conventional and deep learning models to improve CAD risk prediction while preserving 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/view/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.v11i2.6136