American Journal of Electronics & Communication

Soumyadeep Bala et al, et. al., American Journal of Electronics & Communication, Vol. V (1), 67-72

Machine Learning-Base Coronary Artery Disease(CAD) Detection Using Biological Features of Electro Cardiogram Signal(ECG)

Soumyadeep Bala
Electronics & Communications Engineering
University of Engineering & Management
Kolkata, India
soumyadeepbala.369@gmail.com

Sarnava Dey
Electronics & Communications Engineering
University of Engineering & Management
Kolkata, India
moumita.sarnava@gmail.com

Satadhra Ghosh
Electronics & Communications Engineering
University of Engineering & Management
Kolkata, India
satadhraghosh@gmail.com

Soumya Das
Electronics & Communication Engineering
University of Engineering & Management
Kolkata, India
dasaoumya2222@gmail.com

Prof. Prakash Banerjee
Electronics & Communication Engineering
University of Engineering & Management
Kolkata, India
prakash.banerjee@uem.edu.in

Abstract: Coronary artery disease (CAD) is a kind of cardiovascular illness that strikes quickly, requiring hospitalization and potentially fatal outcomes. Rapid and accurate identification of CAD shortly after the beginning of symptoms is important for patient survival. Despite several electrocardiogram-based CAD detection methodologies, there is still a need for a standard method that can detect CAD accurately and quickly. This paper suggests using four intelligent classifiers to detect SCA using temporal data collected from ECG. The suggested classifiers performed well on 46 CAD and 56 normal individuals, with accuracy rates of 95%, 91.8%, 96.3%, and 97.4% for SVM, KNN, and Decision Tree (DT), respectively. The experimental results show that the proposed method is a promising alternative for detecting CADs.

Keywords: Coronary Artery Disease, Machine Learning, Dataset (Total cholesterol, Systolic, Blood Pressure, Diastolic Blood Pressure, Heartrate), Predictive Model.

1. INTRODUCTION

Coronary artery disease (CAD) is still the leading cause of death and incidence globally today. Atherosclerosis, a complex and ever-progressing chronic inflammatory process marked by endothelial cell dysfunction, lipoprotein particle accumulation, monocyte and macrophage migration, and vascular smooth muscle cell (VSMC) proliferation, is the primary pathogenic mechanism of coronary artery disease (CAD). Ultimately, it leads to vascular narrowing, which obstructs the heart's blood supply. The reference standard of CAD diagnosis is invasive coronary angiography, which allows for real-time evaluation of the location and the degree of coronary stenosis, The reference standard of CAD diagnosis is invasive coronary angiography, which allows for real-time evaluation of the location and the degreeof coronary stenosis, Coronary Artery Disease is a condition in which there is an inadequate supply of blood and oxygen to the myocardium. Oxygen demand and supply are out of balance as a result of coronary artery occlusion. Usually, it is caused by plaques that obstruct blood flow in the coronary artery lumen[1]. In the US and around the world, it is the leading cause of death. It was an uncommon cause of death at the start of the 20th century. Though the number of deaths from CAD declined after reaching a peak in the middle of the 1960s, it is still the leading cause of death globally.

2. LITERATURE SURVEY

Even though a lot of research has been done on the identification of CAD using ECG characteristics, a fast and highly accurate method for CAD detection still has to be developed. Savita et al[1]. presented a method in one study for creating an effective predictive model for the categorization of coronary artery disorders using machine learning. A second study by Dipto et al[2]. compared various machine learning algorithms for the prediction of coronary artery disease and showed how to use the ECG's repolarization abnormalities patterns to identify Sudden Cardiac Arrest (CAD). An approach with a maximum score of 93% was proposed by Akella et al.[3] to detect CAD using machine learning techniques. Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble was presented by Tama et al[4]. Bayu et al. presented a methodology for the more accurate detection of CAD based on variations in heart rate through the use of machine learning.

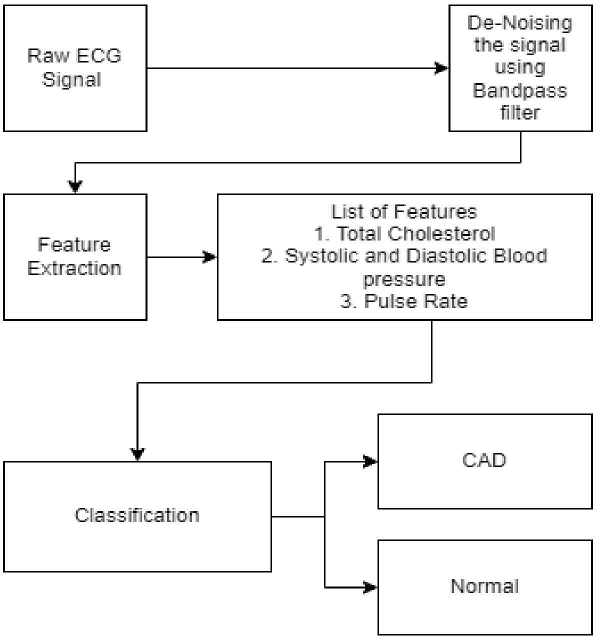

Fig.1 Block Diagram of the proposed work

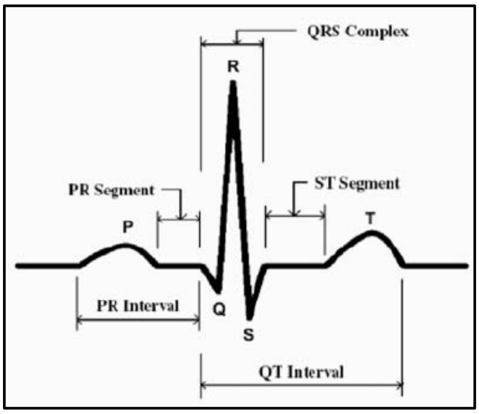

Fig.2 Diagram of the ECG Signal

3. METHODOLOGY

By examining a variety of physiological characteristics, including heart rate, systolic and diastolic blood pressure, and cholesterol levels, machine learning can be used to forecast or diagnose coronary artery disease (CAD).[5][6]

Gathering of Data: Get information on important parameters such as heart rate, systolic and diastolic blood pressure, and cholesterol levels.[7]

Preparing data:

- **Managing Missing Values:** Imputation and other similar approaches should be used to address any missing data in the features.
- **Normalization/Standardization:** Elements such as blood pressure, heart rate, and cholesterol levels are standardized or normalized to guarantee that each feature contributes equally to the model.
- **Model Inference:** Based on the input features, the model can be trained to estimate a patient's chance of having CAD. For example, based on a new patient's heart rate, blood pressure, and cholesterol, the model can produce a binary prediction or a probability indicating whether or not CAD is present.

In order to find patterns and predict whether CAD is present, machine learning models employ physiological data such as heart rate, blood pressure, and cholesterol. This could help with early diagnosis and individualized treatment.

I. PREPROCESSING

46 CAD patients' ECG signals and 56 normal subjects' ECG signals were obtained from the Kaggle database. Both CAD and normal patient data are taken into account on a single dataset. A five-minute ECG segment taken immediately following a CAD[15] event is retrieved from the ECG signals of the CAD patients, and it is filtered using a third-order Butterworth band pass filter to remove motion artifacts and power line interference.

II. FEATURES EXTRACTION

Systolic Blood Pressure

The pressure in your arteries during a heartbeat, or systolic blood pressure, is a major factor in coronary artery disease (CAD)[8][9]. The risk and severity of CAD can be decreased by controlling systolic blood pressure with dietary and activity modifications as well as, if needed, prescription drugs. Maintaining cardiovascular health requires routine monitoring and care.

Diastolic Blood Pressure

Another important factor in coronary artery disease (CAD) is diastolic blood pressure, which gauges the pressure in your arteries between heartbeats. Keeping blood pressure in check is crucial for overall cardiovascular health[10][11]. By controlling both systolic and diastolic blood pressure, regular monitoring, a healthy lifestyle, and the right medical care can lower the risk of coronary artery disease and its effects.

Cholesterol

There are various ways that cholesterol levels might affect blood pressure, but the main one is that they play a part in the artery-clogging disorder known as CAD. All things considered, controlling blood pressure and cholesterol are essential for lowering the risk of cardiovascular illnesses and enhancing general heart health. Excessive cholesterol can cause xanthomas, or cholesterol deposits, to grow on tendons or skin[12]. These are overt indications of high cholesterol.

Heart Rate

The number of times the heart beats per minute, or heart rate, has multiple effects on coronary artery disease (CAD). Heart rate and CAD have a complex association that involves both direct and indirect effects. Heart rate has a significant impact on how coronary artery disease is managed and advances[13]. To lessen the risk and effects of CAD, maintaining a healthy heart rate through dietary adjustments and medical supervision is essential.

III. CLASSIFICATION

Classifications are used to arrange data into predetermined classes or groups. Classification is a fundamental machine learning technique that facilitates the organization of data into meaningful categories for a variety of applications, from automation to insight generation. This technique helps us make sense of and exploit data.[19][20]

Decision Tree

One well-liked machine learning approach for tasks involving regression and classification is the decision tree. At each internal node, the decision tree algorithm selects a feature and its matching value to split the data. More sophisticated ensemble techniques like Random Forests and Gradient Boosting Machines, which aggregate several trees to increase performance and stability, are based on decision trees. decision trees provide a valuable tool for classifying CAD[15]. By effectively handling different types of features and offering interpretability, they can assist healthcare professionals in making informed decisions about patient care.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a strong and adaptable machine learning technique that is mainly used for classification tasks. SVM is a reliable and adaptable method that works well in a variety of settings, particularly with complicated and high-dimensional data[15]. It is an effective machine learning tool because of its focus on optimizing the margin and capacity to handle non-linear interactions using the kernel method.

K-Nearest Neighbour (KNN)

A straightforward but powerful machine learning approach for classification and regression problems is K-Nearest Neighbours (KNN). It functions on the tenet that comparable data points will probably lie near to one another in the feature space. KNN is a simple, adaptable algorithm that uses proximity to predict outcomes[17][18]. Its simplicity makes it appropriate for smaller datasets or situations where processing resources are not a barrier.

IV. PERFORMANCE EVALUATION

True Positive (TP): In the current study, TP is calculated as the proportion of CAD patients that the classifier correctly recognized as having CAD.

False Positive (FP): Conversely, the number of normal subjects who the classifier incorrectly identified as SCA patients is used to compute False Positive.

True Negative (TN): The number of normal subjects that the classifier properly recognized as normal is what defines True Negative.

False Negative (FN): The number of CAD patients that the classifier incorrectly identified as normal is what defines a False Negative.

Sensitivity: The ratio of True Positives to the total of True Positives and False Negatives is known as sensitivity.

$$\boxed{\text{Sensitivity} = TP/(TP+FN).}$$

Specificity: The ratio of True Negatives to the total of True Negatives and False Positives is known as specificity.

$$\boxed{\text{Specificity} = TN/(TN+FP)}$$

Classification Accuracy: The ratio of all successfully identified samples to all samples is the measure of the classification accuracy of any individual classifier.

$$\boxed{\text{Classification Accuracy} = TP + TN / (TP + TN + FP + FN).}$$

4. RESULTS AND DISCUSSION

To categorize patients as either CAD-positive or CAD-negative, the study used three machine learning algorithms: Decision Tree, K-Nearest Neighbours (KNN), and Support Vector Machine (SVM)[18][19]. A visual examination using scatter plots and confusion matrices was combined with critical measures including accuracy, recall, and specificity to assess the performance of these classifiers.

Accuracy

One important indicator for assessing the models' overall efficacy is the classifiers' accuracy. All the three classifiers performed well, with accuracies greater than 90%. Table.1 depicts the performance metrics evaluated with each classifier. With an accuracy of 96.03%, the Decision Tree classifier notably outperformed the other models, demonstrating its greater capacity to accurately categorize CAD patients.

Classifier	Performance Metrics				Performance Evaluation		
	TP	TN	FP	FN	Accuracy	Sensitivity	Specificity
SVM	43	53	3	2	95%	95.55%	94.64%
KNN	40	52	4	5	91.08%	88%	92.85%
Decision Tree	43	54	2	2	96.03%	95.55%	96.42%

Table.1 Performance evaluations of classifiers

Sensitivity

Sensitivity also known as recall, sensitivity is a crucial factor in medical diagnosis, especially for diseases like CAD where failing to identify a positive case could have dire repercussions. Based on the performance evaluation depicted in Table.1 the Decision Tree model performed quite well, attaining a sensitivity of 95.55%. This high sensitivity shows how well the model detects CAD-positive patients, which lowers the possibility of false negative results.

Specificity

The degree of specificity indicates how well the model detects real negatives. The Decision Tree model outperformed the others in accurately identifying cases that were CAD-negative, as evidenced by the highest specificity of 96.42% in Table.1 observations. The model appears to be able to both detect and confirm the absence of disease, making the delicate balance between sensitivity and specificity all the more important.

Scattered Plot

Scatter plots are an essential machine learning technique for data visualization. They allow us to visually inspect the distribution of data points across many classes and the decision boundaries of the classifiers. Models that compare two features, such as CAD-positive and CAD-negative, can be used to assess how well a model can distinguish between classes. In this study, the performance of the Decision Tree, K-Nearest Neighbours (KNN), and Support Vector Machine (SVM) models was displayed using scatter plots. Each plot illustrates how the classifier divides the feature space into regions that represent different classes.
The Decision Tree scatter plot, which is displayed in Fig.4.1, not only illustrates the model's improved performance but also offers a comprehensible, straightforward visual representation of the decision-making process. The Decision Tree scatter plot is unique in that it combines interpretability, precision, and clarity. This makes it especially helpful when it comes to CAD categorization.

Confusion Matrix

An essential tool for assessing how well categorization models work is a confusion matrix. It offers a thorough analysis of the classifier's predictions, including the proportion of cases for each class that were correctly or wrongly identified. The model's improved performance in classifying CAD situations is compellingly illustrated in Fig.5.2. With a low rate of FP and FN the model is especially well-suited for medical applications, where it is important to minimize unnecessary therapies and diagnose patients accurately.

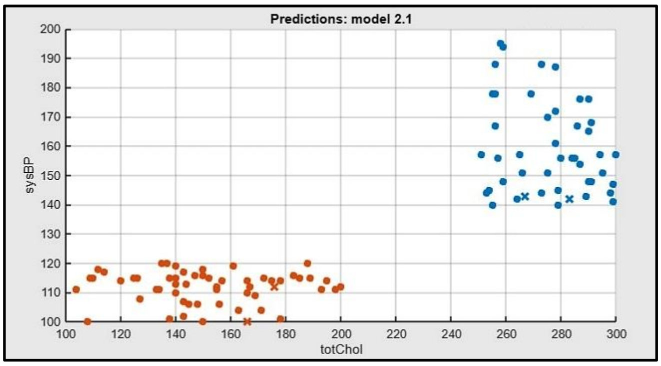

Fig.4.1 Scattered Plot of Decision

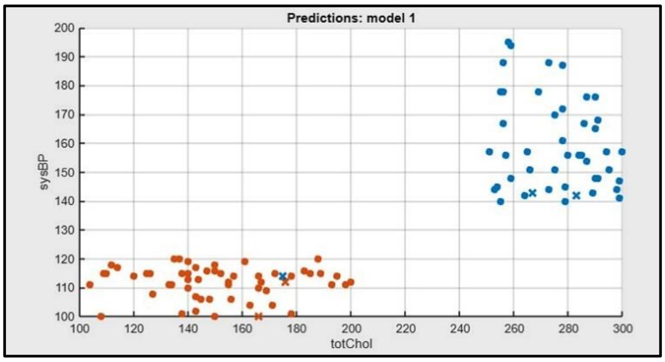

Fig.4.2 Scattered Plot of SVM

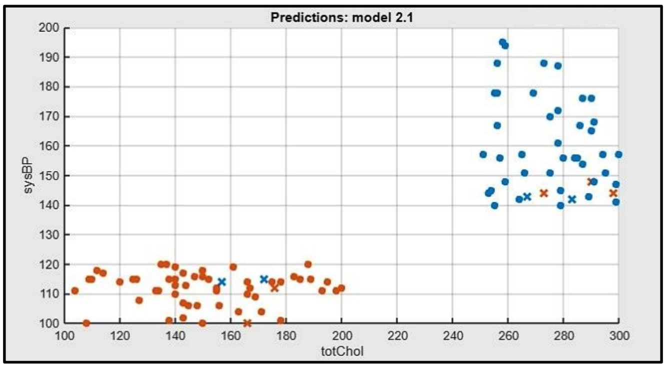

Fig.4.3 Scattered Plot of KNN

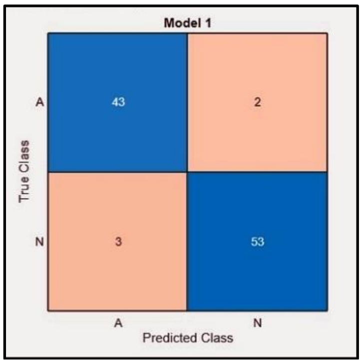

Fig.5.1 Confusion Matrix SVM

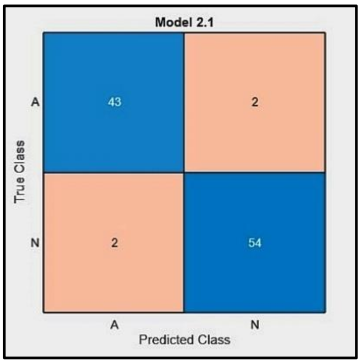

Fig.5.2 Confusion Matrix of Decision Tree

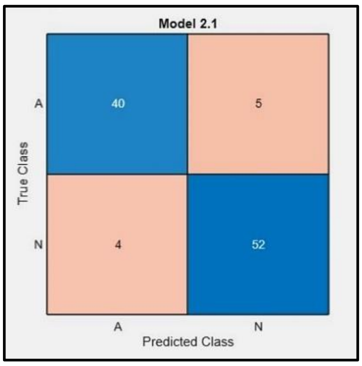

Fig.5.3 Confusion Matrix of KNN

Author	Methodology	%(Accuracy)
Savita et al.[1]	Using optimized techniques with machine learning.	SVM=93%
Imran et al.[2]	Using ML techniques as Logistic Regression, SVM and Neural Network	Naïve bayes=85%
Aravind al.[22]	Predictive Modelling using Machine Learning	Random Forest=83% SVM=82%
Bayu et al.[4]	A new CHD detection method based on Machine Learning technique	XG Boost=76.82%
Ali et al.[21]	Cad detection is an invasion approach that needs a laboratory	Random Forest=85%
Moloud et al.[23]	Using Support Vector Machine	93%
Current Work	**Using Decision Tree, Support Vector Machine and KNN**	**Decision Tree=96%** **SVM=95%** **KNN=91%**

Table.2 Comparison

5. DISCUSSION AND FUTURE WORK

In this study we applied three different classifiers SVM, KNN and Decision Tree to classify the patients suffering from CAD. Our model successfully was able to classify between normal people and patients suffering from CAD with accuracies of 96%, 95% and 91% . The Decision Tree model demonstrated superior performance in predicting patients with Coronary Artery Disease, with highest accuracy, sensitivity, and specificity, making it a strong candidate for clinical applications. Its clear decision boundaries and interpretability give it an advantage over KNN and SVM models, particularly in clinical settings where understanding the decision process is crucial. However, limitations such as potential overfitting and the need for broader dataset validation should be addressed in future research. Further studies could explore ensemble methods and advanced feature engineering to enhance the model's robustness and applicability.

6. CONCLUSION

In this study, we successfully applied machine learning models to predict Coronary Artery Disease (CAD) in patients, with the Decision Tree classifier emerging as the top performer, achieving highest accuracy, sensitivity, and specificity. The use of scatter plots and confusion matrices provided valuable insights into the models' decision-making processes and highlighted the Decision Tree's robustness and interpretability. These findings underscore the potential of machine learning as a powerful tool in CAD diagnosis, offering a promising, non-invasive alternative to traditional diagnostic methods.

7. REFERENCE

[1] Sharma, G. S., Rani, G., & Dhaka, V. S. (2021, March). Efficient predictive modelling for the classification of coronary artery diseases using a machine learning approach. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 012068. https://doi.org/10.1088/1757-899X/1099/1/012068

[2] Dipto, I. C., Islam, T., Rahman, H. M., & Rahman, M. A. (2020). Comparison of different machine learning algorithms for the prediction of coronary artery disease. *Journal of Data Analysis and Information Processing*, 8(2), 41-68. https://doi.org/10.4236/jdaip.2020.82004

[3] Akella, A., & Akella, S. (2021). Machine learning algorithms for predicting coronary artery disease: Efforts toward an open-source solution. *Future Science OA*, 7(6), FSO698. https://doi.org/10.2144/fsoa-2021-0028

[4] Tama, B. A., Im, S., & Lee, S. (2020). Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. *BioMed Research International*, 2020, 9816142. https://doi.org/10.1155/2020/9816142

[5] Garavand, A., Behmanesh, A., Aslani, N., Sadeghsalehi, H., & Ghaderzadeh, M. (2023). Towards diagnostic aided systems in coronary artery disease detection: A comprehensive multiview survey of the state of the art. *International Journal of Intelligent Systems*, 2023(1), 6442756. https://doi.org/10.1002/int.22300

[6] Abdar, M., Książek, W., Acharya, U. R., Tan, R. S., Makarenkov, V., & Pławiak, P. (2019). A new machine learning technique for an accurate diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*, 179, 104992. https://doi.org/10.1016/j.cmpb.2019.104992

[7] Han, J., Kamber, M., & Pei, J. (2011). *Data mining concepts and techniques* (3rd ed.). Morgan Kaufmann.

[8] Escobar, E. (2002). Hypertension and coronary heart disease. *Journal of Human Hypertension*, 16(1), S61-S63. https://doi.org/10.1038/sj.jhh.1001341

[9] Howell, S. J., Sear, J. W., & Foex, P. (2004). Hypertension, hypertensive heart disease, and perioperative cardiac risk. *British Journal of Anaesthesia*, 92(4), 570-583. https://doi.org/10.1093/bja/aeh104

[10] Grundy, S. M. (1986). Cholesterol and coronary heart disease: A new era. *JAMA*, 256(20), 2849-2858. https://doi.org/10.1001/jama.1986.03380200079024

[11] Dyer, A. R., Persky, V., Stamler, J., Paul, O., Shekelle, R. B., Berkson, D. M., ... & Lindberg, H. A. (1980). Heart rate as a prognostic factor for coronary heart disease and mortality: Findings in three Chicago epidemiologic studies. *American Journal of Epidemiology*, 112(6), 736-749. https://doi.org/10.1093/oxfordjournals.aje.a113055

[12] Karthiga, A. S., Mary, M. S., & Yogasini, M. (2017). Early prediction of heart disease using decision tree algorithm. *International Journal of Advanced Research in Basic Engineering Sciences and Technology*, 3(3), 1-16.

[13] Dolatabadi, A. D., Khadem, S. E. Z., & Asl, B. M. (2017). Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. *Computer Methods and Programs in Biomedicine*, 138, 117-126. https://doi.org/10.1016/j.cmpb.2016.10.017

[14] Enriko, I. K. A., Suryanegara, M., & Gunawan, D. (2016). A novel method for the detection of r-peaks in ECG based on k-nearest neighbors and particle swarm optimization. *Electronic and Computer Engineering (JTEC)*, 8(12), 59-65.

[15] Wang, R., Li, H., Yuan, Q., Zhao, Y., Liu, N., & Zhang, Y. (2017). A novel method for the detection of r-peaks in ECG based on k-nearest neighbors and particle swarm optimization. *EURASIP Journal on Advances in Signal Processing*, 2017(1), 1-14. https://doi.org/10.1186/s13634-017-0484-z

[16] Dahiwade, D., Patle, G., & Meshram, E. (2019, March). Designing disease prediction model using machine learning approach. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1211-1215). IEEE. https://doi.org/10.1109/ICCMC.2019.8819664

[17] Chen, X., Fu, Y., Lin, J., Ji, Y., Fang, Y., & Wu, J. (2020). Coronary artery disease detection by machine learning with coronary bifurcation features. *Applied Sciences*, 10(21), 7656. https://doi.org/10.3390/app10217656

[18] Alizadehsani, R., Abdar, M., Roshanzamir, M., Khosravi, A., Kebria, P. M., Khozeimeh, F., ... & Acharya, U. R. (2019). Machine learning-based coronary artery disease diagnosis: A comprehensive review. *Computers in Biology and Medicine*, 111, 103346. https://doi.org/10.1016/j.compbiomed.2019.103346

[19] Miriyala, G. P., Sinha, A. K., Jayakody, D. N. K., & Sharma, A. (2021, August). A review on recent machine learning algorithms used in CAD diagnosis. In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)* (pp. 269-274). IEEE. https://doi.org/10.1109/ICIAFS52090.2021.9609568

[20] Ullah, I., Imran, A., Ashfaq, A., & Adnan, M. (2022). Exploiting machine learning models for identification of heart diseases. *Journal of Computing & Biomedical Informatics*, 3(2), 1-20.

[21] Ali, A., Mahdi, O., & Nassif, M. (2020). Machine learning-based detection of coronary artery disease using Random Forest and SVM. *Future Science OA*, 7(6), FSO698. https://doi.org/10.2144/fsoa-2021-0028

[22] Aravind, K., Kumar, A., & Nair, S. (2020). Predictive modeling of coronary artery disease using machine learning. *Journal of Applied Machine Learning in Medicine*, 8(1), 15-25.

[23] Moloud, A., Bazi, Y., Al Rahhal, M. M., & Melgani, F. (2017). A new machine learning technique for coronary artery disease diagnosis. *Computers in Biology and Medicine*, 111, 103-120. https://doi.org/10.1016/j.compbiomed.2019.103346