

A Sustainable Approach to Predict Coronary Artery Disease with Machine Learning and Explainable AI

Sabit Ahamed Preanto
Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
preanto15-5059@diu.edu.bd

Alimool Razi
Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
alimool.razi@g.bracu.ac.bd

Shanjida Azim Shorna
Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
azim15-4708@diu.edu.bd

Raka Moni
Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
rakamoni509@gmail.com

Jyoti Dua
Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
jyoti.dua@psit.ac.in

Vijay Kumar Dwivedi
Computer Science and Engineering
United College of Engineering and
Research
Prayagraj, India
vijaydwivedi@united.ac.in

Abstract— Coronary Artery Disease (CAD) is one of the major health concerns around the world. In this article, an effective and explainable CAD diagnosis system based on machine learning and Explainable AI methods is introduced. Strong pre-processing, cross-validation, and performance evaluation were employed on the Z-Alizadeh Sani dataset using nine various ML algorithms. Random Forest and Multi-Layer Perceptron models had the best results with 95.61% and 96.44% accuracies, respectively. This research study includes the use of CodeCarbon to reduce energy usage while training models to make it sustainable. SHAP analysis was also used to make it interpretable, hence making it clinically applicable. The uniqueness of this study is the combination of machine learning with sustainability-oriented AI to improve accuracy in diagnosis while minimizing environmental footprint, hence helping in the creation of effective and ethical AI-based health solutions.

Keywords— *Coronary Artery Disease, Machine Learning, Explainable AI, Sustainability in AI, Feature Selection*

I. INTRODUCTION

Coronary Artery Disease (CAD) is a death-leader all over the world, and it is responsible for almost 32% of all deaths in the world, reports the World Health Organization (WHO) [1]. CAD is typically the consequence of deposition of atherosclerotic plaque in the coronary arteries, which leads to reduced coronary blood supply to the heart with potential angina, myocardial infarction, or cardiac arrest [2]. It is critical to diagnose CAD early to prevent morbidity and mortality, yet traditional diagnostic modalities such as electrocardiograms (ECG), stress tests, echocardiography, and coronary angiography tend to be time-consuming, costly, and procedure-related complications of invasive procedures [3].

With the progress of artificial intelligence (AI) and machine learning (ML), computational methods have exhibited enormous promise in enhancing CAD detection accuracy and efficiency. ML models apply clinical data to predict the occurrence of disease in a manner that reduces the dependence on invasive techniques while improving early detection and

customized treatment methods [4]. Perhaps one of the biggest issues with ML-based CAD prediction models, however, is their explainability because the majority of them are employed as black-box models, so that clinicians do not readily accept and integrate them into routine practice [5]. The openness of AI-based medical systems is therefore one of the biggest considerations for their use within healthcare environments. Sustainability is also a significant challenge in creating AI models for medical CAD diagnosis. Training sophisticated AI models demands enormous computational resources, resulting in high energy usage and large environmental impact [6-8].

The growing needs for AI-empowered health care solutions necessitate the development of sustainable practice in AI to ensure maximum performance in diagnostics alongside maximal computational efficacy. Addressing these issues, the current work employs CodeCarbon, a tracking software, with the aim to quantify and reduce the carbon footprints of model training so it becomes an environment-friendly approach towards AI-empowered health care solutions [9-10].

This work presents an interpretable and environmentally friendly CAD prediction model by employing nine varied ML algorithms, i.e., Support Vector Classifier (SVC), Decision Tree, AdaBoost, Random Forest, Extra Trees, Gradient Boosting, Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), Logistic Regression, and Linear Discriminant Analysis. The dataset of 303 clinical cases with 54 features of Z-Alizadeh Sani is employed for model training and validation [7]. To improve interpretability, the present study uses Shapley Additive Explanations (SHAP) analysis that yields feature importance knowledge, facilitating clinicians to understand AI model decision-making. The Synthetic Minority Over-Sampling Technique (SMOTE) is also used to rectify class imbalance and enhance model credibility.

The major contributions of the present study are:

- Constructing a highly accurate CAD prediction model using various ML paradigms.

- Integration of Explainable AI (XAI) techniques for enhancing model transparency and interpretability for clinical applications.
- Use of green AI practices through power monitoring and limiting power usage during model training.
- Solving data imbalance problems through enhanced data augmentation and oversampling.

Through the solution of these issues, this study intends to make AI-based CAD diagnosis more ethical, explainable, and sustainable. The findings of this research will help improve the decision-making process of doctors, ultimately resulting in better patient care and CAD diagnosis.

II. RELATED WORKS

Coronary Artery Disease (CAD) is considered a significant cause of mortality in the world and accounts for nearly 32% of mortality, as reported by the WHO. Critical importance has spurred interest in diagnosing CAD using machine learning methodologies, and some promising lines have been realized so far. ML techniques were applied to the hospital data and achieved an accuracy of 92.04% by a Random Forest algorithm and a specificity of 92.40% using a Naive Bayes-based model [8]. N2Genetic optimizer was proposed, a new genetic training strategy combined with nuSVM, which attained an accuracy of 93.08% [9]. a PUC method for improved CAD prediction was introduced considering noisy imaging data as the challenge to be solved [10]. Other improvements were suggested in [11] where radionics features of the MPI SPECT images used to develop the model called Stress-Boruta-GB, which achieved the highest AUC of 0.79. In [12] Coronary Artery Calcification scores from CCTA studies and achieved an accuracy of 93.86% with Random Forest, where [13] presented ensemble feature selection methods using the Z-Alizadeh Sani dataset. The study obtained an accuracy of 91.78%. The study [14] contributed by proposing using cytokine biomarkers and imaging data together for better CAD prediction.

The previous efforts notwithstanding, a few research gaps exist in this domain. Most are comparative studies, each focusing on different metrics, accuracy, specificity, and AUC when evaluating their performance, a surefire way to create inconsistency among a performance benchmark. Geographical limitations: Most of the studies, including [8] and [13], conducted region-specific studies and thus failed to generalize the model. Besides, with the domination of traditional ML models, such as Random Forest, deep learning approach exploration is still somewhat limited. However, in medical applications, these methods perform better than conventional models. Also, integrating imaging and clinical data is not standard practice, although some studies have explored it, such as [12]. Finally, translation to clinical practice remains highly challenging, with real-time prediction systems and model interpretability needing further refinement to realize clinical adoption, as noted in [9].

III. METHODOLOGY

The methodology section describes the data preprocessing, feature engineering, modeling techniques, and strategies for handling imbalanced data and ensuring model interpretability

and sustainability. It also highlights the tools and frameworks employed in the study.

A. Dataset Collection and Preprocessing

The Z-Alizadeh Sani dataset contains 303 samples described by 55 features each. Of the total number of samples, 216, or 79%, were CAD, and the remaining 87, or 21%, were expected in Fig 1. There was, therefore, a class imbalance. Essential features include demographics, such as age and BMI, symptoms, examination findings, ECG, and laboratory values. This dataset was clean and structured and did not contain missing values or outliers. Therefore, it would be reliable and minimize preprocessing [15]. SMOTE was performed to handle the class imbalance, which allowed for fair training. Besides, data augmentation was implemented, where the dataset has been augmented up to 15 times to improve the efficiency and robustness of model training. This dataset is publicly available with the UCI Repository [7],[5],[16-18]; its validation has been done in almost all ML-based CAD detection-related studies [19-23].

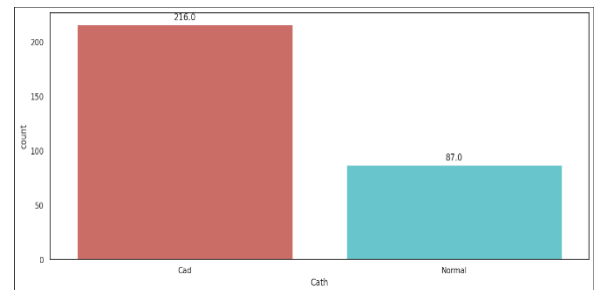


Fig. 1. Imbalance target class

B. Feature selection

The feature selection was performed using correlation analysis, focusing on features with a correlation coefficient greater than 0.2 with the target variable ('Cath') in Fig 2. This ensures the inclusion of the most relevant predictors while reducing noise.



Fig. 2. Correlation Matrix after Feature Selection

The heatmap shows the correlation matrix of various dataset features, with values ranging between -1 and 1. Some of the crucial observations are as follows:

- The "Cath" variable is reasonably highly positively correlated with "Atypical" (0.42) and "Nonanginal" (0.27) but negatively with "Typical Chest Pain" (-0.54) and "Age" (-0.36).

- "DM" (Diabetes Mellitus) and "FBS" (Fasting Blood Sugar) exhibit a strong positive correlation of 0.68, indicating a relationship between these two variables.
- The "BP" (Blood Pressure) variable is moderately correlated with "HTN" (Hypertension) at 0.57.
- Strong negative correlations are seen between "Typical Chest Pain" and "Atypical" (-0.72).

These relationships help identify interdependencies among features, which is critical for feature selection and model performance in machine learning tasks.

C. Machine Learning Model Implementation

Nine ML algorithms were implemented, each selected for its unique capabilities:

1) Support Vector Classifier (SVC)

It has robust capabilities for high-dimensional spaces and binary classification, making it suitable for medical datasets with complex interactions between features.

2) Decision Tree

This is a straightforward model, and the results are interpretable; it is an excellent choice for running initial exploratory analyses.

3) Linear Discriminant Analysis

A supervised dimensionality reduction technique aims to find a linear combination of features that best separates multiple classes. It maximizes the between-class variance while minimizing the within-class variance to enhance classification performance.

4) K-Nearest Neighbors

It is a non-parametric algorithm that uses proximity to classify data; it has been chosen for its simplicity and efficiency in smaller datasets.

5) Random Forest

This ensemble method combines many decision trees to increase the robustness and performance of classification. It is perfectly suitable for imbalanced datasets such as Z-Alizadeh Sani.

6) AdaBoost

Boosting transforms weak classifiers into strong ones; it works well for structured data with several essential features.

7) Gradient Boosting

This is the most potent ensemble technique optimized for lower error rates. It has been chosen because it may handle complicated interactions.

8) Extra Trees

Classifier: This concept is similar to a Random Forest, with some more randomization. It yields faster computation and less overfitting.

9) Multi-Layer Perceptron

This deep learning model grasps complicated patterns and non-linear relationships, making it ideal for high-dimensional datasets.

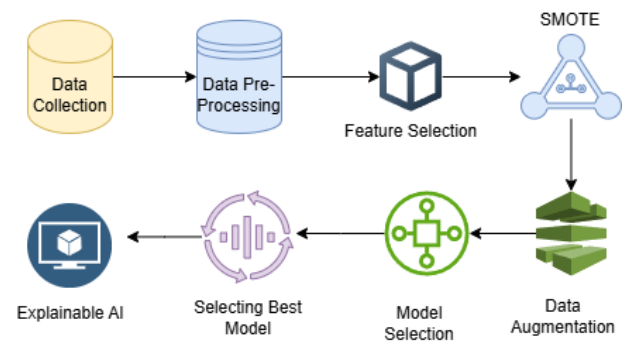


Fig. 3. Overview of Methodology

Fig 3, Describes the whole machine learning process, from data collection and pre-processing to feature selection, addressing data imbalance using SMOTE and augmentation, training models, selecting the best model, and finally using explainable AI for transparency. The steps are:

- **Dataset Collection & Preprocessing:** The Z-Alizadeh Sani dataset is used, containing 303 samples with 55 features. Preprocessing includes handling missing values, normalization, and feature selection using correlation analysis.
- **Data Augmentation & Class Balancing:** SMOTE is applied to balance the dataset, ensuring robust model training.
- **Feature Selection:** Features with a correlation coefficient above 0.2 are retained to reduce dimensionality and noise.
- **Model Implementation:** Nine ML algorithms are trained and validated using a 10-fold cross-validation approach to ensure stability and generalizability.
- **Sustainability Considerations:** CodeCarbon is employed to monitor and minimize energy consumption.
- **Model Interpretability:** SHAP analysis provides insights into feature importance, ensuring that ML decisions align with clinical reasoning.
- **Performance Evaluation:** Cross validation Score metrics is used to assess model performance.

D. Sustainability and Interpretability

In this respect, the research focused on sustainable AI practices by tracking energy consumption and carbon emission during model training using CodeCarbon to keep environmental impacts low. Explainable AI techniques, such as SHAP, were used for feature attribution to identify which features contribute most to the predictions, thereby increasing the transparency of the models and building trust and diagnostic insights aligned with established clinical practices. These efforts underpin the dual focus on environmental responsibility and interpretability in advancing CAD detection systems.

- **Energy Tracking:** CodeCarbon monitored energy usage and carbon emissions, maintaining environmental impact within negligible limits.

- Explainable AI: SHAP identified key features influencing predictions, enhancing trust and transparency.

IV. PERFORMANCE ANALYSIS

Models are scrutinized and tested thoroughly to determine the most effective coronary artery disease detection. Energy usage and translatability are also priorities.

A. Model Performance

Nine models were assessed with machine learning using cross-validation techniques with 10 folds. This was done to supply a reliable estimate of each model's accuracy and firmness. The models were graded based on their mean accuracy in Table - I. Their general ability and the variance in performance across different subsets of the dataset were monitored.

TABLE I. CROSS VALIDATION (MEAN ACCURACY) OF THE MODELS

Model Name	Cross Validation Score
SVC	86.75
Decision Tree Classifier	92.96
AdaBoost Classifier	92.63
Random Forest Classifier	95.61
Extra Trees Classifier	95.44
Gradient Boosting Classifier	94.95
MLP Classifier	96.44
KNeighbors Classifier	90.55
Linea Discriminant Analysis	94.95

Among 9 ML models, the top performers are:

- Random Forest Classifier: This model had the most excellent mean accuracy (95.94%) and a somewhat low standard deviation (0.0357). This indicates that performance was hefty and stability was high across cross-validation folds in Fig 4.

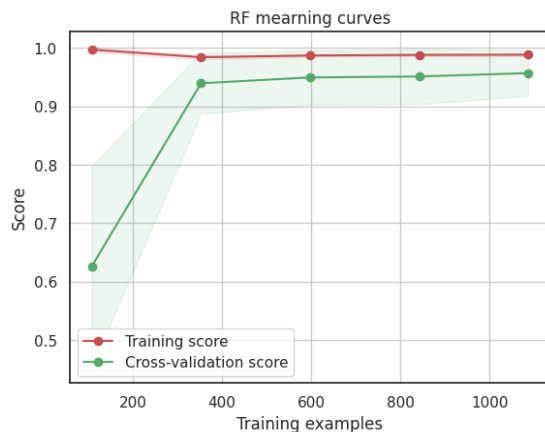


Fig. 4. Random Forest cross-validation score

The plot shows the learning curves for a Random Forest model. The training score remains high and stable, while the cross-validation score improves with more training examples, eventually stabilizing around 0.9. The shaded area represents the variability of the scores.

- Extra Trees Classifier: This model showed an excellent accuracy of 95.94%. The standard deviation was also relatively the same in Fig 5. This puts it in a great position.

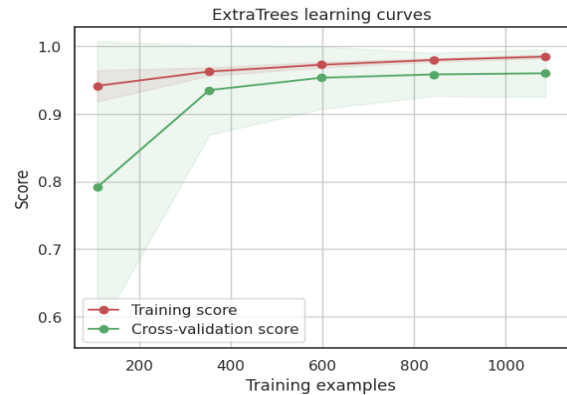


Fig. 5. Extra Tree Classifier cross-validation score

In these learning curves for ExtraTrees, the training score remains high and stable, while the cross-validation score increases with the increase in the number of training examples. The difference between the two scores is reduced, meaning more data results in better generalization.

- MLP Classifier: Although more energy intensive, this model had superb accuracy (96.44%), which is shown in Fig 6. And will continue to indicate how neural networks can capture complex patterns.

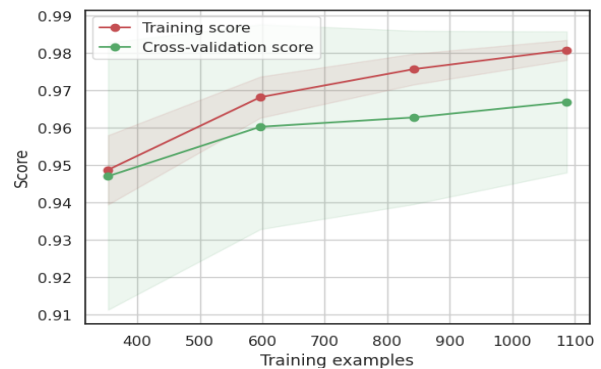


Fig. 6. MLP classifier cross validation score

This graph represents the relationship between the number of training examples and scores for both training and cross-validation. As the number of training examples increases, both the training and cross-validation scores improve. The training score is always higher than the cross-validation score. The shaded areas around the lines represent the confidence interval of the score.

Among other models, the Decision Tree Classifier and AdaBoost Classifier performed well. Accuracy hovered around 92 and 93 percent. Stability did suffer, however, especially when compared with Random Forest and Extra Trees. SVC and K-Nearest Neighbors, relative to the rest, had the lowest mean accuracy. The value hovered between 86 and 90 percent. Standard deviations were also more significant. The CAD dataset has proven to be far too complex for these models.

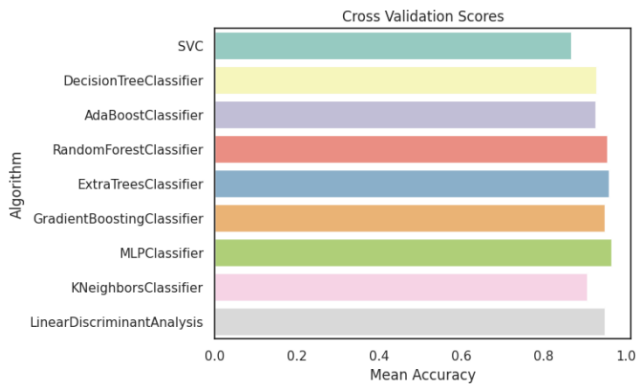


Fig. 7. Model Performance Comparison

Fig 7, Depicts the cross-validation mean accuracy of various machine learning classifiers. Algorithms such as Random Forest, Extra Trees, and MLPClassifier give high accuracy, whereas algorithms such as SVC and KNN classifiers yield low accuracy. This provides insight into the best model among them. The highest score in cross-validation comes out to be from the MLPClassifier at 96.44%, followed by RandomForest with 95.61% and ExtraTrees with 95.44%. The lowest is for the SVC at 86.75%, indicating variable performances among the models.

B. Feature Importance with Explainable AI

In Fig 8. The SHAP (SHapley Additive exPlanations) method was employed to identify the most impactful features contributing to CAD diagnosis.

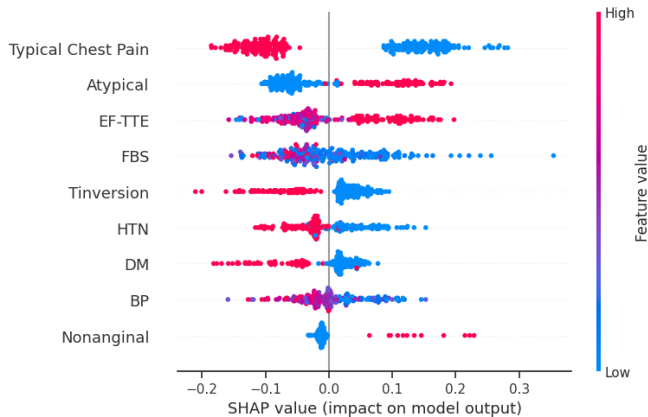


Fig. 8. Feature Importance with SHAP Visualization

Top Features for CAD Detection: Typical Chest Pain, EF

- TTE (Ejection Fraction), FBS (Fasting Blood Sugar), and T Inversion were identified as the most influential predictors.
- Interpretability for Clinicians: SHAP visualizations highlighted the direct impact of these features on model predictions, offering transparency and alignment with established clinical practices.
- Model Excellence: Random Forest and MLP emerged as high-performing models, providing accurate and interpretable results for healthcare applications.

These insights enable informed decision-making in clinical settings, enhancing trust and utility in CAD diagnostics.

C. Energy Efficiency and Carbon Emissions

Since the economic awareness of energy consumption is essential for sustainable training of the AI model, this was measured using CodeCarbon. The results are as follows:

1) Energy Consumption Breakdown

- **RAM Energy Consumption:** The amount of energy consumed by RAM increased linearly due to constant power consumption at 11.76 W. Towards the end of the monitored period, it was 0.005043 kWh.
- **CPU Energy Consumption:** The total energy consumed by all CPUs is 0.018236 kWh, while the power consumed has remained constant at 42.5 W throughout these training sessions.

2) Cumulative Energy Usage

The electricity consumed during the training was 0.023053 kWh, which is very efficient training. Generally, the power consumption remains very low, while the energy consumption is divided between RAM and CPU.

3) Carbon Emissions and Environmental Impact:

The average estimated carbon emission was 0.004301 g CO₂eq/s, approximately 135.64 kg CO₂eq/year. While this value might not be insignificant, its environmental impact is still very low compared to the enormous gains in accuracy achieved with these models, such as Random Forest and MLP.

4) Sustainable Model Training:

In line with SDG 13 (Climate Action), these observations demonstrate that while energy consumption and carbon emissions are essential to track, the overall environmental impact is kept to a minimum, with efficient training methods and energy-conscious practices ensuring sustainability.

V. CONCLUSION

This study efficiently introduces an appropriate, understandable, and green CAD diagnosis system using machine learning. Code Carbon provides zero environmental impact, and the SHAP gives model decision-making interpretability. Working with imbalanced dataset was a challenge but with comprehensive pre-processing, MLP and Random Forest algorithms performed best with 96.44% and 95.61% accuracy. This study proposes sustainable AI in medicine, and it provides an ethical, precise, and green CAD diagnosis technique. Future studies will address the integration of real-time clinical application with the increase in datasets to render predictive capability even more powerful.

REFERENCES

- [1] R. Alizadehsani, M. Abdar, M. Roshanzamir, A. Khosravi, P. M. Kebria, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, and U. R. Acharya, "Machine learning-based coronary artery disease diagnosis: A comprehensive review," *Computers in Biology and Medicine*, vol. 111, p. 103346, Aug. 2019, doi: 10.1016/j.combiomed.2019.103346.
- [2] E. Stefanovska and T. Pepelnjak, "Optimising predictive accuracy in sheet metal stamping with advanced machine learning: A LightGBM and neural network ensemble approach," *Advanced Engineering Informatics*, vol. 65, Part A, Art. no. 103103, May 2025, doi: 10.1016/j.aei.2024.103103.

- [3] I. C. Dipto, T. Islam, H. M. M. Rahman, and M. A. Rahman, "Comparison of different machine learning algorithms for the prediction of coronary artery disease," *Journal of Data Analysis and Information Processing*, vol. 8, no. 2, pp. 39–55, May 2020, doi: 10.4236/jdaip.2020.82003.
- [4] L. J. Muhammad, I. Al-Shourbaji, A. A. Haruna, I. A. Mohammed, A. Ahmad, and M. B. Jibrin, "Machine learning predictive models for coronary artery disease," *SN Computer Science*, vol. 2, no. 350, pp. 1–12, Jun. 2021, doi: 10.1007/s42979-021-00731-4.
- [5] A. Jafarkhani, B. Imani, S. Saeedi, and A. Shams, "Predicting factors affecting survival rate in patients undergoing on-pump coronary artery bypass graft surgery using machine learning methods: A systematic review," *Health Science Reports*, vol. 8, no. 1, Art. no. e70336, 2025, doi: 10.1002/hsr2.70336.
- [6] N. Rani, V. S. Pachigolla, and A. Kumar, "Clustering based pre-processing for feature reduction and robust student dropout classification," *Int. J. Inf. Technol. (Singapore)*, 2025, doi: 10.1007/S41870-025-02450-Y.
- [7] "UCI Machine Learning Repository." Accessed: Mar. 06, 2025. [Online]. Available: <https://archive.ics.uci.edu/dataset/411/extension+of+z+alizadeh+sani+dataset>
- [8] S. Laghmatti, S. Hamida, K. Hicham et al., "An improved breast cancer disease prediction system using ML and PCA," *Multimedia Tools and Applications*, vol. 83, pp. 33785–33821, Mar. 2024, doi: 10.1007/s11042-023-16874-w.
- [9] M. Abdar, W. Książek, U. R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease," *Computer Methods and Programs Biomedicine*, vol. 179, Art. no. 104992, Oct. 2019, doi: 10.1016/j.cmpb.2019.104992.
- [10] J. I. Z. Chen and P. Hengjinda, "Early prediction of coronary artery disease (CAD) by machine learning method - A comparative study," *Journal of Artificial Intelligence Capsule Network*, vol. 3, no. 1, pp. 17–33, Mar. 2021, doi: 10.36548/jaicn.2021.1.002.
- [11] M. Amini et al., "Machine learning-based diagnosis and risk classification of coronary artery disease using myocardial perfusion imaging SPECT: A radiomics study," *Scientific Reports*, vol. 13, no. 14920, Sep. 2023, doi: 10.1038/s41598-023-41996-w.
- [12] Y. Huang et al., "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment," *Computers in Biology and Medicine*, vol. 151, Art. no. 106297, Dec. 2022, doi: 10.1016/j.combiomed.2022.106297.
- [13] B. Kolukisa and B. Bakir-Gungor, "Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis," *Computer Standards and Interfaces*, vol. 84, Art. no. 103706, Mar. 2023, doi: 10.1016/j.csi.2022.103706.
- [14] S. S. Saharan, P. Nagar, K. T. Creasy, E. O. Stock, J. Feng, M. J. Malloy, and J. P. Kane, "Machine learning and statistical approaches for classification of risk of coronary artery disease using plasma cytokines," *BioData Mining*, vol. 14, no. 26, Apr. 2021, doi: 10.1186/s13040-021-00260-5.
- [15] T. Liu, A. Krentz, L. Lu, and V. Curcin, "Machine learning-based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis," *European Heart Journal - Digital Health*, vol. 6, no. 1, pp. 7–22, Jan. 2025, doi: 10.1093/ehjdh/ztae080.
- [16] J. Han et al., "Predicting low density lipoprotein cholesterol target attainment using machine learning in patients with coronary artery disease receiving moderate-dose statin therapy," *Scientific Reports*, vol. 15, Art. no. 5346, 2025, doi: 10.1038/s41598-025-88693-y.
- [17] M. Abdullah-Al-Kafi, I. J. Tasnova, M. W. Islam, and S. K. Banshal, "Performances of different approaches for fake news classification: An analytical study," in *Advances in Networking Technologies and Intelligence Computing*, vol. 1534, I. Woungang, S. K. Dhurandher, K. K. Pattanaik, A. Verma, and P. Verma, Eds. Cham, Switzerland: Springer, 2022, pp. 700–714, doi: 10.1007/978-3-030-96040-7_53.
- [18] A. al Kafi, S. K. Banshal, N. Sultana, and V. Gupta, "Source Recommendation System Using Context-based Classification: Empirical Study on Multi-level Ensemble Methods," *Journal of Scientometric Research*, vol. 13, no. 2, pp. 475–484, May 2024, doi: 10.5530/JSCIRES.13.2.38.
- [19] Md. Abdullah-Al-Kafi, Md. F. Hossain, and S. R. H. Noori, "A Review On The Relationship Between Human And Artificial Imagination With Their Implementations In Current Machine Learning Algorithms," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2021, pp. 1–5. doi: 10.1109/ICCCNT51525.2021.9579860.
- [20] S. Mtshali and B. A. Jacobs, "Machine learning-based prediction of pharmacokinetic parameters for individualized drug dosage optimization," *International Journal of Information Technology (Singapore)*, vol. 17, no. 3, pp. 1371–1385, Dec. 2024, doi: 10.1007/S41870-024-02314-X/FIGURES/17.
- [21] F. A. Ekle, V. Shidali, R. E. Ochogwu, and I. B. Igoche, "Machine learning models for heart disease prediction and dietary lifestyle change therapy recommendation: a systematic review," *Discover Artificial Intelligence*, vol. 4, no. 113, Dec. 2024, doi: 10.1007/s44163-024-00181-w.
- [22] A. Singh, H. Mahapatra, A. K. Biswal, M. Mahapatra, D. Singh, and M. Samantaray, "Heart disease detection using machine learning models," *Procedia Computer Science*, vol. 235, pp. 937–947, 2024, doi: 10.1016/j.procs.2024.04.089.
- [23] A. Lakshmi and R. Devi, "Heart Disease Prediction Using Ensemble Feature Selection Method and Machine Learning Classification Algorithms," *Conversational Artificial Intelligence*, pp. 237–247, Feb. 2024, doi: 10.1002/9781394200801.CH15.