International Journal of Engineering & Applied Sciences (IJEAS)
Vol.16, Issue 1 (2024) 16-32
http://dx.doi.org/10.24107/ijeas.1380819
Int J Eng Appl Sci 16(1) (2024) 16-32

Early Diagnoses of Acute Coroner Syndrome Based on Machine Learning Model

Umut Utku Tiryaki [a], Gul Karaduman [b], Sare Nur Cuhadar [c], Ahmet Uyanik [d], Habibe Durmaz [e]*

[a,c,e] Karamanoğlu Mehmetbey University, Department of Electrical and Electronics Engineering, Karaman, Türkiye
[b]Karamanoğlu Mehmetbey University, Vocational School of Health Services, Karaman, Türkiye*
[d]Konya Meram State Hospital, Konya, Türkiye

✉: : gulk@bu.edu [b*], durmazgul@gmail.com [e]. *: 0009-0000-5028-0782 [a], 0000-0002-2776-759X [b*], 0000-0003-4461-877X [c], 0000-0001-5037-1019 [d], 0000-0002-5929-861X [e]*

Received:27.10.2023, Revised: 05.01.2024, Accepted: 25.01.2024

Abstract

Cardiovascular diseases are a leading global cause of death, particularly in low to middle-income countries. Early and accurate diagnosis of Acute Coronary Syndrome (ACS) is vital, but limited access to healthcare hinders effective management. This study utilized machine learning to develop mathematical models for ACS risk detection. Data from 249 individuals with ACS or suspected heart disease were used to construct twelve models with different parameters and classifiers. Performance indicators, including accuracy, Matthews correlation coefficient, and precision, were employed for evaluation. The Random Forest classifier demonstrated superior performance, achieving 90.45% accuracy for internal validation and 86% for external validation. Critical criteria for ACS diagnosis were CK-MB, age, coronary artery disease, and Troponin T value. The models developed in this study significantly prevent potential deaths via rapid intervention and reduce healthcare expenditures by minimizing unnecessary human resources and repeat tests.

Keywords: Cardiovascular diseases; acute coronary syndrome; heart attack; machine learning; model performance.

1. Introduction

Cardiovascular diseases (CVDs) encompass conditions that cause damage to the heart and blood vessels, including cerebrovascular, coronary heart, and rheumatic heart diseases. The World Health Organization (WHO) identifies CVDs as the leading global cause of mortality [1], presenting a significant public health challenge. A 2016 study on the Global Burden of Non-Communicable Diseases reveals that 40% of non-communicable diseases in women and 50% in men can be attributed to cardiovascular and related diseases. In contrast, only 20% of female and 24% of male patients experience cardiovascular diseases alone [2].

Acute Coronary Syndrome (ACS), a specific type of cardiovascular disease, refers to heart tissue damage resulting from blockages in the coronary arteries, commonly known as a heart attack. ACS accounts for 4 out of 5 deaths related to CVDs, with 1 out of 3 deaths occurring in individuals below 70 years of age [3]. Chest pain serves as the prominent symptom of cardiovascular diseases and ACS, and it is challenging to distinguish between common chest pain and suspected ACS. Despite advances in treatment, readmission rates for ACS patients remain elevated [4].

Cardiovascular diseases (CVDs) claim the lives of approximately 18 million individuals annually, making them the leading cause of global mortality in 2019. Most of these deaths occur in middle- and low-income countries, including Central Asia, Eastern Europe, and other low-

income regions [5]. The economic impact of CVDs is substantial, with significant costs reported in various countries, such as Serbia, the Fiji Islands, sub-Saharan Africa, Turkey, Brazil, and India [6-11].

To ensure early diagnosis and treatment for individuals at high risk of CVDs, universal access to primary healthcare services is imperative within the next 25 years. Currently, many people in low-income countries need access to such services [12]. Artificial intelligence algorithms offer a promising solution by expanding primary healthcare and enabling early detection of ACS and other diseases such as COVID-19 without the need for costly tests or excessive human resources [13-15]. Mathematical models employing machine learning techniques have been developed to identify ACS risk using data from individuals who experienced ACS and those with suspected heart conditions but no ACS diagnosis.

Our research underscores the notable success achieved in early Acute Coronary Syndrome (ACS) diagnosis through the implementation of sophisticated mathematical machine-learning models. The overarching aim is to prevent avoidable deaths by facilitating timely intervention, and curbing unnecessary resource utilization, and redundant testing, thereby resulting in a tangible reduction in healthcare expenditures. Significantly, our model equips healthcare professionals with a streamlined approach to promptly assess patients with suspected ACS, eliminating the need for extensive and costly diagnostic procedures.

A groundbreaking aspect of our study is the introduction of an innovative approach that harnesses artificial intelligence algorithms for the early detection of ACS. Through the integration of intricate mathematical models and advanced machine learning techniques, our research endeavors to enhance early diagnosis of mortal diseases such as ACS, COVID-19, ebola virus, etc, mitigating the necessity for costly tests a pivotal advancement in optimizing healthcare resource allocation [13-15].

Central to our contributions is the development and validation of mathematical machine-learning models specifically tailored for the early diagnosis of ACS. Emphasizing the model's remarkable success rates, we highlight its capacity to empower healthcare practitioners to efficiently evaluate patients presenting with suspected ACS, leading to more targeted and cost-effective interventions.

In positioning our proposed model, we assert its pivotal role in averting preventable deaths through timely interventions, curbing unwarranted resource consumption, and ultimately reducing overall healthcare expenditures. This strategic alignment aligns with broader healthcare objectives, including the enhancement of efficiency, cost reduction, and improvement of patient outcomes.

2. Material and Methods

To create predictive models for the early detection of acute coroner syndrome, (Waikato Environment for Knowledge Analysis (WEKA) (https://www.cs.waikato.ac.nz/ml/weka) was utilized in this study [35]. The development process for these models is outlined in Figure 1. The initial step involved preprocessing the dataset, which included removing noisy data, normalizing the dataset, and determining the applicable range. Subsequently, the dataset was divided into a training set (80% of the data) and a test set (20% of the data). Various techniques were employed to identify the most suitable parameters for constructing accurate predictive machine-learning models. Twelve machine learning modeling algorithms were utilized to diagnose obesity based on blood analysis, and their performance was evaluated using measures derived from the confusion matrix for internal and external validation.

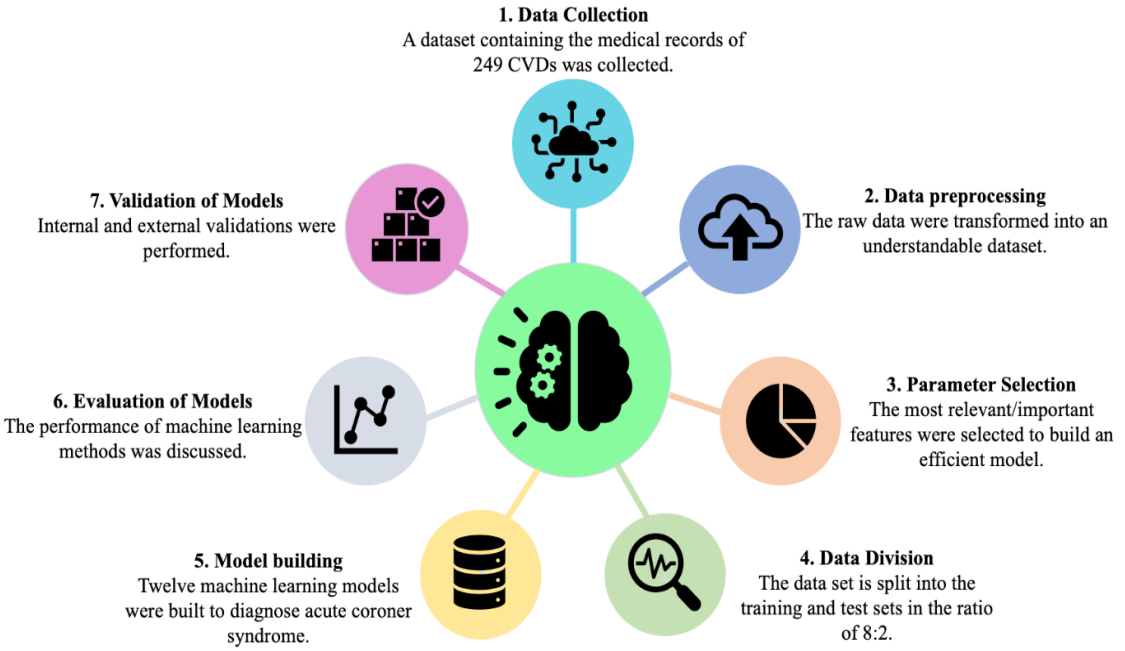

Fig. 1. The model development process for early diagnoses of acute coroner syndrome.

2.1. Details of the dataset

The study comprised 249 participants, some experiencing ST-elevation myocardial infarction or myocardial infarction without ST elevation, collectively called ACS. Among the participants, 191 individuals had a history of heart attack, while the remaining 58 did not exhibit signs of ACS. Detailed patient information, including age, gender, blood test results, and medical conditions like smoking, hypertension, and diabetes, which contribute to ACS risk, can be found in Table 1. The dataset provides further insights and is presented in Table 1.

Table 1. Features and risk factors of acute coronary syndrome

Serial Number	Group	Feature Names	Features Descriptions
1	Patient record data	Age	Age in years
2	Patient record data	Sex	1 = female; 2 = male
3	Patient record data	Smoke	1 = smoker; 0 = non smoker
4	Patient record data	Hypertension	History of hypertension
5	Patient record data	Diabetes	1 = history of diabetes; 0 = no such history
6	Patient record data	Troponin T	Troponin T value in blood test
7	Patient record data	CK-MB	CK-MB value in blood test
8	Patient record data	Hyperlipidemia	1 = high cholestrol; 0 = no high cholestrol
9	Patient record data	Heart failure	1 = History of heart failure; 0 = no such history
10	Patient record data	Coronary artery disease	1 = History of having a previous heart attack; 0 = no such history
11	Patient record data	Coronary Artery Graft	1 = History of having coronary angiography; 0 = no such history
12	Patient record data	Coronary by-pass	1 = History of having coronary artery by-pass surgery; 0 = no such history
13	Patient record data	Chronic renal failure	1 = History of chronic renal failure; 0 = no such history

The details of the dataset used in the study are as follows:

a. *Patient's age:* Age is a significant risk factor for coronary artery disease, and the risk of acute coronary syndrome (ACS) and mortality increases steadily. After accounting for other risk factors, the risk of ACS-related death rises by two to three times per decade [16]. Most ACS-related deaths happen in individuals aged 65 or older [17], with over half of these deaths occurring in those aged 70 or above.

b. *Gender:* Research has indicated that being male is an independent risk factor for coronary artery disease [18]. On average, men develop coronary artery disease 7-10 years earlier than women [19]. This disparity is attributed to the protective influence of estrogen, which may explain the heightened risk in women who undergo premature menopause [20].

c. *Smoke:* Smoking is a substantial and separate predictor for the onset of coronary artery disease, and quitting smoking reduces the risk of acute coronary syndrome (ACS) [21]. Smoking elevates blood pressure and heart rate, resulting in heightened peripheral vascular resistance and the release of catecholamines. It diminishes flow-mediated dilation in the coronary artery, promotes blood clot formation, and lowers HDL cholesterol levels. Moreover, smoking directly damages the endothelial lining, contributing to the development of atherosclerosis [17].

d. *Hypertension*: Numerous observational studies have established a robust association between elevated blood pressure and the likelihood of coronary artery disease [18]. Each increment of 20 mm/Hg in systolic blood pressure or 10 mm/Hg in diastolic blood pressure doubles the risk of ACS-related mortality and stroke [22].

e. *Diabetes:* Both type 1 and type 2 diabetes are linked to an increased risk of acute coronary syndrome (ACS) [23]. Individuals with diabetes who have a history of ACS face a greater risk of mortality compared to those without diabetes. Diabetes amplifies the risk of ACS by two to four times and, when combined with high cholesterol, serves as a significant predictor of coronary disease. Approximately 80% of patients with diabetes develop coronary atherosclerosis [21, 23].

f. *Troponin T:* Cardiac troponins play a crucial role in regulating the calcium-dependent interaction between actin and myosin in the heart [24]. Numerous studies have demonstrated a correlation between troponin levels and the risk of mortality in acute coronary syndrome (ACS) [25]. Troponin T, primarily present in the myocardium, exhibits high clinical sensitivity, enabling the detection of even minor increases [26, 27].

g. *CK-MB:* Creatine is a vital protein present in muscle cells, with notable concentrations in skeletal and cardiac muscles. In conditions that compromise muscle cell integrity, creatine kinase can be released into the bloodstream, leading to a substantial elevation in blood creatine kinase levels [28]. Among the three isoenzymes of creatine kinase, CK-MB was incorporated into the model due to its predominant occurrence in cardiac muscle cells and its tendency to increase in the blood during ischemic heart disease.

h. *Hyperlipidemia:* Hyperlipidemia denotes elevated levels of fats in the bloodstream. Numerous studies have established a robust association between low-density lipoprotein (LDL) cholesterol and the progression of atherosclerotic vascular disease. Additionally, reduced levels of high-density lipoprotein (HDL) cholesterol serve as a risk factor for heightened coronary artery disease risk [29].

i. *Congestive heart failure:* Heart failure can arise due to multiple factors, with the most prevalent causes being ischemic heart disease and acute coronary syndrome (ACS) induced

by atherosclerotic vascular disease. Malnutrition and myocardial necrosis resulting from ACS can result in impaired contraction function of the heart muscle [30].

j. *Coronary artery disease:* The coronary arteries play a critical role in providing the heart with essential oxygen and nutrients. Diseases affecting these arteries have a direct impact on heart function and efficiency. The buildup of substances, primarily cholesterol, results in the narrowing and blockage of the coronary arteries, giving rise to atheromatous plaques that progressively restrict the vessel's lumen. This narrowing hampers blood flow to the heart, leading to issues associated with coronary artery disease, often presenting as acute coronary syndrome (ACS) [31].

k. *Coronary artery imaging and coronary bypass:* For patients experiencing acute coronary syndrome (ACS), coronary angiography and angioplasty are employed to reinstate blood flow in the coronary artery obstructed by atheromatous plaque. In instances where coronary angioplasty fails to achieve revascularization, coronary bypass surgery is the preferred approach. This surgical intervention enhances symptoms and enhances the disease prognosis, leading to improved outcomes [32].

l. *Chronic kidney disease:* Chronic kidney disease is a condition marked by a decline in kidney function, resulting in the loss of nephrons and a gradual decrease in glomerular filtration rate (GFR) over time [33]. While chronic kidney disease itself is not a direct risk factor for acute coronary syndrome (ACS), it can lead to false-positive Troponin T levels in individuals with this condition [26].

2.2. Data preprocessing

The preprocessing and data cleaning stages are crucial for the success and accuracy of machine learning modeling. Some collected data may require correction or may be missing due to noise [34]. During preprocessing, the noise was removed from the data set. Rows containing invalid data were deleted, and missing values were updated by taking the average of corresponding data groups. Duplicates were also removed. Additionally, names and last names were removed to avoid possible confusion. Normalization was applied by rescaling data between 0 and 1 to improve accuracy.

2.3. Data Division

The data was split into two sets: 80% for training and 20% for testing. The training and test data were independently separated to avoid a false accuracy rate, which can occur when using exact data for training and testing.

2.4. Parameter Selection

Real-time intrusion detection is nearly impossible due to the large amount of data flowing over the network. Feature selection can reduce the calculation time and complexity of the model. Feature selection is selecting the relevant and vital features by removing the most irrelevant and redundant features from the dataset to build an effective and efficient model. Two attribute selection techniques were used in this study. The CfsSubsetEval evaluator, combined with the Best First search method, is a correlation-based attribute selection method that determines the prediction power of each attribute. The CorrelationAttributeEval evaluator works with the Ranker search method and selects the attributes that will create the best models based on Pearson’s correlation logic. Feature selection reduces calculation time and model complexity by selecting relevant and vital features and removing irrelevant and redundant ones from the dataset.

2.5. Algorithms

In this study, four classifiers, namely Bayes Net, Logistic, IBk, and Random Forest were chosen and applied to the dataset. The selection of Bayes Net, Logistic Regression, IBk (k-Nearest Neighbors), and Random Forest classifiers was based on a strategic consideration of the dataset's characteristics. Bayes Net and Logistic Regression were chosen for their interpretability, IBk for its effectiveness in handling diverse data distributions, and Random Forest for its robustness and ability to capture complex relationships. These choices aim to provide a comprehensive evaluation, ensuring a balanced exploration of methodological approaches in addressing the study objectives. A summary of these classifiers is provided below.

2.5.1. Bayes Net

The Bayes Net classifier is a statistical classifier that uses multiple search algorithms and quality measures based on the Bayes network classifier [36]. It is based on Bayes' Theorem [37], which calculates the conditional probability of E_j, given A, from the probabilities of $E_1, E_2, \ldots, E_k$ and the conditional probabilities of A was given E_i, i=1,2, …, k is calculated by

$$P(E_j|A) = \frac{P(A \cap E_j)}{P(A)} = \frac{P(E_j)P(A|E_j)}{\sum_{i=1}^{k} P(E_i)P(A|E_i)}. \quad (1)$$

2.5.2. Logistic

The Logistic regression classifier is a popular classification method with extensive usage [38]. It is easy to implement and has exhibited excellent performance across various issues, including spam prediction. The classification model is appropriate for estimating discrete probabilities, such as outcomes in the form of yes, no, win, or lose. Additionally, this classifier is straightforward to execute and has demonstrated competent performance across various issues [39].

2.5.3. IBk

The IBk algorithm predicts the outcome of a test pattern in real-time without requiring the construction of a model during classification. To make predictions, the algorithm measures the distance between each test sample and k neighboring instances in the training data, selecting a particular distance measure to make an estimation. The classification function then incorporates the results obtained and the similarity function to determine which instances to include in the description concept [40].

2.5.4. Random forest

The random forest classifier combines decision trees to increase the classifier value. The algorithm generates multiple decision trees during classification and combines them to form a decision forest. Each classifier is created by randomly sampling a vector from the input [41]. The algorithm replaces the original sets to form training sets, and a new subset and random attribute selection create the tree. The node is then split using the best splitting on the randomly chosen attributes [42].

2.6.Evaluation of the Models

In this section, the metrics utilized in this study are described. These metrics are based on the four values (TP, FN, TN, FP) from the confusion matrix [56, 57]. As shown in Figure 2, the confusion matrix has labels for both positive and negative classes in both actual and predicted outcomes. When data has a positive label in the actual class and a positive label in the predicted class, it is considered a "True Positive (TP)," while a positive label in the predicted class but a negative label in the actual class is considered a "False Negative (FN)." Conversely, when data has a negative label in the actual class and a positive label in the predicted class, it is considered a "False Positive (FP)," while a negative label in both actual and predicted classes is considered a "True Negative (TN)."

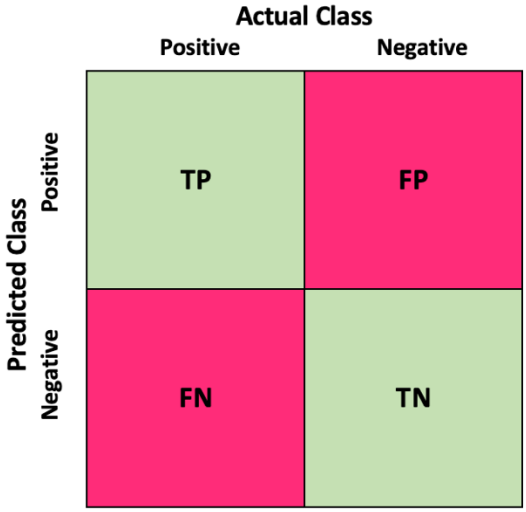

Fig. 2. Confusion matrix for binary classification

(a) Precision refers to the ratio of the truly positive or negative results to the total number of results.

$$\text{Precision} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{2}$$

(b) Recall is the ratio of the data that a machine learning model correctly identifies as belonging to a class of interest.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

(c) MCC evaluates the correlation between actual data and the data predicted by the model.

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{TP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{4}$$

(d) ROC Area: It is one of the most critical metrics for machine learning models. It shows the false positive and true positive rates, providing information about the model's overall performance.

(e) Accuracy: The Accuracy value is calculated as the ratio of correctly predicted values to the total dataset in the model.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \quad (5)$$

(f) Kappa statistic obtains a value between 0 and 1 in the calculation. The statistics obtained between 0.00-0.20 are low, 0.20-0.40 are below average, 0.40-0.60 are average, 0.60-0.80 are strong, and 0.80-1.00 are excellent [43]. The kappa value is calculated as follows:

$$\text{Kappa statistic} = \frac{\text{observed aggrement} - \text{expected aggrement}}{1 - \text{expected aggrement}}. \quad (6)$$

2.7. Validation of the Models

Validation is a crucial statistical resampling technique that plays a pivotal role in objectively and accurately assess the performance of a machine learning model when applied to unseen data. We conducted internal and external validations to ensure a comprehensive and unbiased evaluation of the methods employed in our modeling process. These validations serve as robust and reliable mechanisms to gauge the effectiveness and reliability of our chosen approaches, enabling us to obtain a more thorough understanding of the model's performance and its potential for generalization to new, unseen data. By employing these validation techniques, we aim to minimize any potential biases or overfitting issues, allowing us to confidently ascertain the true efficacy of our machine learning methods and their suitability for accurately diagnosing obesity based on the parameters derived from blood test results.

3. Results

In our study, we embarked on the development of a diverse set of twelve machine learning models, which were created by employing a combination of four distinct classification algorithms and three feature selection options. This approach allowed us to explore various avenues and harness the potential of different techniques to enhance the predictability and reliability of our models. To comprehensively evaluate the performance and effectiveness of these models, we conducted both internal and external validations, each serving a unique purpose in assessing their capabilities.

Internal validation, a crucial component of our evaluation framework, involved the utilization of a 10-fold cross-validation technique. This technique provided a robust and objective means of gauging the models' performance by partitioning the available data into ten equally sized subsets. By iterative training and evaluating the models on different combinations of training and validation sets, we were able to obtain a comprehensive understanding of their generalization ability and reliability in different scenarios.

Furthermore, external validation was conducted to assess the predictive power of the models on unseen data. This validation process involved evaluating the performance of the trained models on a dedicated test set, which was kept completely separate from the training data. By subjecting the models to this rigorous evaluation on an independent dataset, we were able to gauge their real-world applicability and assess their ability to accurately predict obesity based on the parameters derived from blood test results.

Through this extensive and rigorous validation process, we aimed to ensure that our machine learning models not only demonstrated strong performance during internal validation but also

exhibited a high degree of reliability and predictive capability when confronted with new, unseen data. By adopting these comprehensive validation procedures, we gained confidence in the efficacy and generalizability of our models, enabling us to make informed decisions about their suitability for diagnosing obesity based on blood analysis.

3.1. Internal Validation Performance

Among the various machine learning models and feature selection techniques evaluated in our study, the Random Forest classifier with CfsSubsetEval feature selection emerged as the top-performing model in terms of accuracy. It achieved an impressive accuracy rate of 90.4523%, accompanied by notable metrics such as an MCC (Matthews Correlation Coefficient) of 0.726, ROC (Receiver Operating Characteristic) of 0.942, Precision of 0.903, Recall of 0.905, and Kappa Statistic of 0.7251. These results highlight the model's ability to make accurate predictions and capture the complexity of the underlying data.

Furthermore, the Random Forest classifier also exhibited superior performance in the training data analysis, further validating its efficacy. This classifier utilizes an ensemble approach, building numerous decision trees on subsets of the data and combining their outcomes to enhance prediction accuracy. The strength of this approach lies in its ability to mitigate the limitations of individual decision trees and capture diverse patterns in the data.

However, it is important to acknowledge that the Random Forest classifier has some drawbacks. One notable drawback is its computational requirements and resource-intensive nature, which necessitates substantial computational power to achieve optimal accuracy. Additionally, compared to other models, training the Random Forest classifier may take a longer time due to its ensemble nature and the need to construct multiple decision trees.

Table 2. Internal Validation Performance

Algorithm	Feature Selection	Metric					
		ACC	MCC	ROC	Precision	Recall	Kappa Statistics
BayesNet	No attribute selection	87.43	0.655	0.918	0.878	0.874	0.6544
	CfsSubEval_BestFirst	87.43	0.655	0.918	0.878	0.874	0.654
	CorAttEval_Ranker	87.43	0.655	0.918	0.878	0.874	0.6544
Logistic	No attribute selection	87.43	0.644	0.937	0.873	0.874	0.6439
	CfsSubEval_BestFirst	87.93	0.678	0.939	0.886	0.879	0.6755
	CorAttEval_Ranker	89.94	0.699	0.935	0.893	0.889	0.6982
IBk	No attribute selection	77.88	0.378	0.782	0.779	0.779	0.3779
	CfsSubEval_BestFirst	79.39	0.425	0.872	0.796	0.794	0.4247
	CorAttEval_Ranker	81.40	0.481	0.856	0.816	0.814	0.2088
Random Forest	No attribute selection	90.45	0.720	0.940	0.902	0.905	0.7163
	CfsSubEval_BestFirst	90.45	0.726	0.942	0.903	0.905	0.7251
	CorAttEval_Ranker	90.45	0.726	0.939	0.903	0.905	0.7251

ACC: Accuracy of classification, MCC: Matthews correlation coefficient, ROC: area under the receiver operating characteristic curve

In terms of computational efficiency, our study revealed varying processing times for different models. Specifically, Bayes Net and IBk models completed their training in 0 seconds, while Logistic took 0.01 seconds. On the other hand, the Random Forest classifier required relatively more time, taking approximately 0.04 seconds for training. These time differences can be important considerations when deploying the models in real-world applications that require prompt responses.

3.2. External Validation Performance

In Table 3, we present the performance evaluation of twelve different models on the external validation set, providing insightful statistical results that shed light on the comparative effectiveness of each model. Among these models, the Random Forest classifier with CorAttEval_Ranker feature selection stands out as the top-performing model, demonstrating its exceptional capability in accurately predicting outcomes.

The success rate achieved by the Random Forest model, reaching an impressive 86%, showcases its remarkable predictive power. This high success rate indicates the model's ability to correctly classify instances and accurately predict the unseen data in the external validation set.

Furthermore, the Random Forest model outperforms the other models in several key metrics, highlighting its superiority. Specifically, it exhibits higher values for the MCC (Matthews Correlation Coefficient) at 0.583, Precision at 0.860, and Recall at 0.860 compared to the alternative models. These metrics are essential indicators of the model's performance, demonstrating its ability to strike a balance between true positive and true negative rates, as well as its precision in correctly classifying instances belonging to the positive class.

The exceptional performance of the Random Forest model with CorAttEval Ranker feature selection suggests its suitability for accurately predicting outcomes in this specific context. Its robustness and ability to capture the relevant patterns and relationships in the data set it apart from the other models considered in this study.

Table 3. External Validation Performance

Algorithm	Feature Selection	Metric					
		ACC	MCC	ROC	Precision	Recall	Kappa Statistics
BayesNet	No attribute selection	82.00	0.448	0.865	0.808	0.820	0.4246
	CfsSubEval_BestFirst	82.00	0.448	0.865	0.808	0.820	0.4246
	CorAttEval_Ranker	82.00	0.448	0.865	0.808	0.820	0.4246
Logistic	No attribute selection	84.00	0.539	0.919	0.833	0.840	0.5349
	CfsSubEval_BestFirst	84.00	0.521	0.925	0.831	0.840	0.5050
	CorAttEval_Ranker	84.00	0.521	0.925	0.831	0.840	0.5050
IBk	No attribute selection	78.00	0.380	0.696	0.774	0.780	0.3792
	CfsSubEval_BestFirst	82.00	0.437	0.784	0.817	0.820	0.3836
	CorAttEval_Ranker	76.00	0.304	0.817	0.747	0.760	0.3023
Random Forest	No attribute selection	80.00	0.393	0.893	0.783	0.800	0.3812
	CfsSubEval_BestFirst	84.00	0.513	0.883	0.839	0.840	0.4709
	CorAttEval_Ranker	86.00	0.583	0.893	0.860	0.860	0.5224

ACC: Accuracy of classification, MCC: Matthews correlation coefficient, ROC: area under the receiver operating characteristic curve

By carefully considering these statistical findings, we can confidently assert that the Random Forest model with CorAttEval_Ranker feature selection exhibits the highest performance among the twelve models evaluated for the external validation set. Its impressive success rate and superior values in metrics such as MCC, Precision, and Recall showcase its effectiveness and reinforce its potential as a reliable predictive model for the given problem domain.

Our investigation identified the Troponin T value as the most influential criterion among the thirteen parameters used to build our predictive model. This finding aligns with existing research, which highlights Troponin as a primary biomarker that provides crucial insights into the development of Acute Coronary Syndrome (ACS) [44, 45]. Consequently, our prediction

model's reliance on the Troponin T value reinforces its ability to detect ACS based on this essential biomarker effectively.

By taking into account these factors, including the Random Forest classifier's outstanding performance, its computational requirements, and the significance of the Troponin T value in our predictive model, we can assert that our approach provides a robust and accurate means of diagnosing ACS based on relevant criteria obtained from blood tests.

4. Discussion

The results obtained from the evaluation of the models in Tables 2 and 3 reveal distinct levels of prediction power exhibited by each model. In terms of internal validation, it is evident that the Random Forest (RF) model with CorAttEval_Ranker evaluator stands out as the model with the highest prediction power. The RF model achieved an impressive accuracy rate (ACC) of 90.45%, indicating its suitability and efficacy in predicting early diagnoses of acute coronary syndrome (ACS). This finding emphasizes the RF model's potential in accurately identifying instances of ACS at an early stage, enabling timely intervention and treatment.
Similarly, when considering external validation, the Random Forest model with CorAttEval_Ranker evaluator once again emerges as the frontrunner, boasting a notable percentage rate of 86.00%. This result further supports the notion that the RF model, with its chosen feature selection technique, excels in making accurate predictions on unseen data, reinforcing its reliability and generalizability in real-world scenarios. The findings from our study strongly indicate that the Random Forest algorithm consistently outperforms the other models examined across both internal and external validation. Numerous studies have acknowledged the suitability of Random Forest and similar ensemble models in handling medical data, especially in cardiovascular-related prediction tasks. The robust performance of RF, as demonstrated in our study, reinforces the utility of these models for early ACS diagnosis. However, it is essential to acknowledge the variability in model performance across different datasets and study contexts, as mentioned in our discussion [58,59].

Table 4: The selected parameters

Parameter Selection	Number of Parameters	Parameters
CfsSubsetEval, BestFirst	4	TroponinT CK-MB Age CABG
CorrelationAttributeEval, Ranker	6	Age CK-MB CABG HT TroponinT KAG

CfsSubsetEval: Correlation-based attribute evaluator, CorrelationAttributeEval: Correlation-based attribute evaluator, BestFirst: Search method, Ranker: Search method

Additionally, the emphasis on parameter tuning in our study resonates with the literature's acknowledgment of the critical role that well-calibrated parameters play in enhancing the predictive performance of machine learning models [60, 61]. Table 4 provides an overview of the parameters employed in this study, underscoring their significance in the predictive performance of the Random Forest model with the CorAttEval_Ranker evaluator. By fine-tuning and selecting the appropriate parameters, researchers can enhance the model's ability to capture relevant patterns and accurately predict the desired outcomes.

Cardiac troponins, which were initially described in 1965, emerged as a reliable method for measuring their levels in the blood. However, it was only in the late 1990s that the methodology for assessing cardiac troponins was fully developed. These specific biomarkers indicate cardiac muscle damage and play a crucial role in diagnosing acute coronary syndrome (ACS) [46, 47]. Recent advancements have demonstrated the remarkable sensitivity of troponin detection, even at minimal levels associated with cardiac muscle damage. This heightened sensitivity has contributed to the widespread acceptance of troponin as a standard biomarker for diagnosing ACS, endorsed by esteemed organizations such as the European Society of Cardiology (ESC) and the American College of Cardiology (ACC) since 2000. The specificity and sensitivity of troponin to cardiac muscle damage make it an invaluable tool in diagnosing and monitoring unstable angina pectoris (UAP), a condition recognized by the ACC and the American Heart Association (AHA) [48]. CK-MB is one of the three iso-enzymes of creatine kinase in the heart muscle and 3% of skeletal muscle. It has been reported that, compared to other iso-enzymes, CK-MB is more specific to the myocardium. CK-MB levels rise between 4 and 12 hours after cardiac injury and peak in the blood at 24 hours [49, 50]. Biomarkers associated with troponin, glucose, CK-MB, cholesterol along with age are also used to diagnose COVID-19 with machine learning [13,62,63].

Age is recognized as a significant risk factor for the development of coronary artery disease (CAD). Specifically, individuals who are 45 years or older (for men) and 55 years or older (for women) are considered to be potentially at higher risk for CAD. This association between age and CAD persists even after accounting for other known risk factors. Notably, research has indicated that the probability of developing vascular disease doubles with each passing decade of age, further highlighting the impact of aging on CAD risk [18].

Hypertension, commonly referred to as high blood pressure, represents another silent yet prevalent risk factor for cardiovascular diseases. Numerous studies have demonstrated a strong association between hypertension and cardiovascular conditions, encompassing both elevated systolic and diastolic blood pressure readings. For instance, one study revealed that patients with hypertension exhibited a 63.3% risk of developing coronary artery disease, whereas individuals with normal blood pressure showed a comparatively lower risk of 46.1% [51].

Acute coronary syndrome (ACS) can manifest as a consequence of lesions or blockages occurring in vessels subjected to revascularization through angioplasty, as well as in native or graft vessels used post-surgery, in patients who have previously undergone coronary angiography (CAG) or coronary artery bypass graft surgery (CABG) [52, 53].

Notably, patients with a history of coronary angioplasty or coronary artery bypass graft surgery and subsequently experiencing ACS tend to have a more favorable prognosis compared to individuals with ACS who have not undergone prior interventions. Research findings indicate that the prognosis for patients with prior CAG or CABG who develop ACS is generally more favorable, while those without a history of such interventions tend to have a poorer prognosis [54, 55].

Several limitations should be also considered in the interpretation of our study findings. First, the dataset's representativeness may affect model generalizability, emphasizing the need for diverse datasets. Second, sensitivity to algorithm and feature selection choices underscores considerations for model applicability across datasets. Third, temporal considerations highlight the need for periodic model updates to align with evolving healthcare landscapes. Finally, the clinical applicability of models, including integration into healthcare practices and real-world validation, requires further investigation, presenting avenues for future research.

In summary, our comprehensive analysis highlights the superiority of the Random Forest algorithm, particularly when combined with the CorAttEval_Ranker feature selection method, in terms of prediction power for diagnosing ACS. These findings emphasize the importance of selecting appropriate algorithms and parameters to maximize the accuracy and reliability of machine-learning models in specific contexts.

5. Conclusion

In this study, our primary objective was to develop effective machine-learning models for the early diagnosis of acute coronary syndrome (ACS). We constructed twelve distinct machine learning models, leveraging 13 relevant features associated with ACS. Through our analysis, we aimed to identify the most accurate model capable of predicting ACS at its early stages. Our findings demonstrated that the Random Forest algorithm outperformed the other models we developed, showcasing its superior predictive capabilities. This highlights the effectiveness of the Random Forest algorithm in accurately identifying ACS cases early on, potentially enabling timely interventions and treatments.

Furthermore, our analysis revealed that several key parameters play a crucial role in the early diagnosis of ACS. Notably, Troponin T, CK-MB, age, and the presence of coronary artery disease emerged as vital factors for accurate prediction. These findings provide valuable insights into the risk factors and diagnostic indicators associated with acute coronary artery disease. Importantly, the proposed methodology and models developed in this study hold promise for real-world applications beyond ACS diagnosis. By applying similar approaches and feature selection techniques, our models can be adapted to analyze more extensive datasets and explore risk factors for various other diseases. This versatility underscores the potential of our method to contribute to medical research and improve diagnostic practices across different healthcare domains.

In summary, our study contributes to the field of early ACS diagnosis by presenting a robust and accurate machine-learning model. The superior performance of the Random Forest algorithm and the identified key parameters pave the way for improved early detection of ACS. Furthermore, our methodology and models offer potential for broader applications in medical diagnostics, facilitating the identification of risk factors and enhancing our understanding of various diseases when applied to larger datasets.

Ethical Approval

The study was carried out following the principles laid down by the Declaration of Helsinki and received approval from the ethics committee of Karamanoglu Mehmetbey University School of Medicine. Since the research was retrospective in nature, informed consent was not deemed necessary.

Author Contributions

Umut Utku Tiryaki: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing.
Gul Karaduman: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing - Review & Editing, Supervision.
Sare Nur Cuhadar: Conceptualization, Methodology, Investigation.
Ahmet Uyanık: Conceptualization, Methodology, Investigation.
Habibe Durmaz: Conceptualization, Methodology, Investigation, Writing - Review & Editing.

References

[1] Wilkins, E., Wilson, L., Wickramasinghe, K., Bhatnagar, P., Leal, J., Luengo-Fernandez, R., Burns, R., Rayner, M., Townsend, N., European Cardiovascular Disease Statistics 2017. *European Heart Network*, 2017.

[2] Thomas, H., Diamond, J., Vieco, A., Chaudhuri, S., Shinnar, E., Cromer, S., Perel, P., Mensah, G. A., Narula, J., Johnson, C. O., Roth, G. A., Moran, A. E., Global Atlas of Cardiovascular Disease 2000-2016: The Path to Prevention and Control. *Global heart*, 13(3), 143–163, 2018.

[3] World Health Organization, *Cardiovascular Diseases,* 2020.

[4] Şencan, I., Keskinkılıç, B., Ekinci, B., Öztemel, A., Sarıoğlu, G., Çobanoğlu, N., Türkiye Kalp ve Damar Hastalıkları Önleme ve Kontrol Programı Eylem Planı (2015-2020). T.C. Türkiye Halk Sağlığı Kurumu. *T.C. Sağlık Bakanlığı Yayın,* 988-1-63,2015.

[5] Benziger, C. P., Roth, G. A., & Moran, A. E., The Global Burden of Disease Study and the Preventable Burden of NCD. *Global heart*, *11*(4), 393–397, 2016.

[6] Lakic, D., Tasic, L. Kos, M., Economic burden of cardiovascular diseases in Serbia. *Vojnosanit Pregl*, 71(2),137 –143, 2014.

[7] Maharaj, J.C., Reddy, M., Young Stroke Mortality in Fiji Islands: An Economic Analysis of National Human Capital Resource Loss. *International Scholarly Research Notices*, 802785, 2012.

[8] Gaziano, T. A., Bitton, A., Anand, S., Weinstein, M. C., & International Society of Hypertension, The global cost of nonoptimal blood pressure. *Journal of hypertension*, *27*(7), 1472–1477, 2009.

[9] Balbay, Y., Gagnon-arpin, I., Malhan, S., Öksüz, M. E., Sutherland, G., Dobrescu, A., Villa, G., Ertuğrul, G., Habib, M., Modeling the burden of cardiovascular disease in Turkey. *Anatol J Cardiol*, 20(4), 235-240, 2018.

[10] Azambuja, M.I.R., Foppa, M., Maranhao, M.F.C., Achutti, A.C., Economic burden of severe cardiovascular diseases in Brazil: an estimate based on secondary data. *Arq Bras Cardiol*, 91(3),163 –171, 2008.

[11] Bloom, D., Cafiero, E., McGovern, M., Prettner, K., Stanciole, A., The Economic Impact of Non-Communicable Disease in China and India: Estimates, Projections, and Comparisons. *The Journal of the Economics of Ageing, 4*,100–111, 2013.

[12] Dejaco, C., Singh, Y. P., Perel, P., Hutchings, A., Camellino, D., Mackie, S., Abril, A., Bachta, A., Balint, P., Barraclough, K., Bianconi, L., Buttgereit, F., Carsons, S., Ching, D., Cid, M., Cimmino, M., Diamantopoulos, A., Docken, W., Duftner, C., Fashanu, B., et., al. 2015 Recommendations for the management of polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Arthritis & rheumatology*, 67(10), 2569–2580, 2015.

[13] Huyut, M. T., Automatic detection of severely and mildly infected COVID-19 patients with supervised machine learning models. *IRBM*, 44(1), 100725, 2023.

[14] Huyut, M. T., & Huyut, Z., Effect of ferritin, INR, and D-dimer immunological parameters levels as predictors of COVID-19 mortality: A strong prediction with the decision trees. *Heliyon*, *9*(3),2023.

[15] Huyut, M. T., & Üstündağ, H.. Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model: a retrospective observational study. *Medical gas research*, 12(2), 60,2022.

[16] Kannel, W. B., Coronary heart disease risk factors in the elderly. *The American Journal of geriatric cardiology*, 11(2), 101–107, 2002.

[17] İkitimur, B., Karadağ, B., Öngen, Z., Yaşlılarda Koroner Arter Hastalığı. *Turkish Journal of Geriatrics*, 2,13-20, 2010.

[18] Savji, N., Rockman, C. B., Skolnick, A. H., Guo, Y., Adelman, M. A., Riles, T., Berger, J. S., Association between advanced age and vascular disease in different arterial

territories: a population database of over 3.6 million subjects. *Journal of the American College of Cardiology*, 61(16), 1736–1743, 2013.

[19] Onat, A., Kaya, A., Şimşek, T., Şimşek, B., Tusun, E., Karadeniz, Y., Can, G., Twenty-five years of the TARF study: The 2015 survey and temporal trends in mortality and loss to follow-up. *Turk Kardiyoloji Dernegi Arsivi,* 44(5),365–370, 2016.

[20] Yadav, P., Joseph, D., Joshi, P., Sakhi, P., Jha, R. and Gupta, J., Clinical Profile & Risk Factors in Acute Coronary Syndrome. *National Journal of Community Medicine*, 1, 150-151, 2010.

[21] National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) . Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report*. Circulation*, 106(25), 3143–3421, 2002.

[22] Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., Jr, Jones, D. W., Materson, B. J., Oparil, S., Wright, J. T., Jr, Roccella, E. J., Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. National Heart, Lung, and Blood Institute, & National High Blood Pressure Education Program Coordinating Committee. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension*, 42(6), 1206–1252, 2003

[23] Haffner, S. M., Lehto, S., Rönnemaa, T., Pyörälä, K., & Laakso, M., Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *The New England journal of medicine*, 339(4), 229–234,1998.

[24] Christenson, R. H., Apple, F. S., Morgan, D. L., Alonsozana, G. L., Mascotti, K., Olson, M., McCormack, R. T., Wians, F. H., Jr, Keffer, J. H., & Duh, S. H., Cardiac troponin I measurement with the ACCESS immunoassay system: analytical and clinical performance characteristics. *Clinical chemistry*, 44(1), 52–60, 1998.

[25] Wu, A. H., Feng, Y. J., Moore, R., Apple, F. S., McPherson, P. H., Buechler, K. F., & Bodor, G., Characterization of cardiac troponin subunit release into serum after acute myocardial infarction and comparison of assays for troponin T and I. American Association for Clinical Chemistry Subcommittee on cTnI Standardization. *Clinical chemistry*, 44, 1198–1208, 1998.

[26] Bhagavan, N. V., Medical Biochemistry, Chapter 21.3, Energy supply in muscle. *Canada, Acad. Pub*, *122*.

[27] Hillis, G. S., & Fox, K., Cardiac troponins in chest pain can help in risk stratification. *British Medical Journal*, 319(7223), 1451-2, 1999.

[28] Kocaman, S., Ratlarda deneysel olarak oluşturulacak kalp krizi ve hasarı modeli ile farklı tedavi yöntemlerinin karşılaştırmalı olarak test edilmesi. *PhD Thesis*, 2022.

[29] Raines, E. W., & Ross, R., Smooth muscle cells and the pathogenesis of the lesions of atherosclerosis. *British heart journal*, 69, S30–S37, 1993.

[30] Kumbasar, D., Kalp sağlığı, 2013.

[31] Gaziano T. A., Cardiovascular disease in the developing world and its cost-effective management. *Circulation*, 112(23), 3547–3553, 2005.

[32] Knuuti, J., Wijns, W., Saraste, A., Capodanno, D., Barbato, E., Funck-Brentano, C., Prescott, E., Storey, R. F., Deaton, C., Cuisset, T., Agewall, S., Dickstein, K., Edvardsen, T., Escaned, J., Gersh, B. J., Svitil, P., Gilard, M., Hasdai, D., Hatala, R., Mahfoud, F., 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: The Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *European heart journal*, 41(3), 407–477, 2020.

[33] Uzun, Ş., Kara, B., İşcan, B. Hemodiyalize giren kronik böbrek yetmezliği olan hastalarda uyku sorunları. *Türk Nefroloji Diyaliz ve Transplantasyon Dergisi*, 12(1): 61-6, 2003.
[34] Kelleci Çelik, F., Karaduman, G., In silico QSAR modeling to predict the safe use of antibiotics during pregnancy. *Drug and Chemical Toxicology*, 1-10, 2022.
[35] Frank, Eibe, Mark A. Hall, and Ian H. Witten. *The WEKA workbench*. Morgan Kaufmann, 2016.
[36] Tang, B., He, H., Baggenstoss, P. M., Kay, S., A Bayesian classification approach using class-specific features for text categorization. *IEEE Transactions on Knowledge and Data Engineering*, 28(6), 1602-1606, 2016 .
[37] Hogg, R.V., Tanis, E.A., Probability and Statistical Inference. *Upper Saddle River*, NJ: Prentice Hall, 1997.
[38] Badresiya, A., Vohra, S., Teraiya, J., Performance Analysis of Supervised Techniques for Review Spam Detection. *International Journal of Advanved Networking Applications*, 21–24, 2014.
[39] Visani, V., Jadeja, N., Modi, M., A Study on Different Machine Learning Techniques for Spam Review Detection. Conference: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017.
[40] Swetha, K., & Ranjana, R., Breast cancer prediction using machine learning and data mining. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 6(3), 610-5, 2020.
[41] Sridharan, K., & Komarasamy, G., Sentiment classification using random harmony forest and harmony gradient boosting machine. *Soft Computing*, 24(10), 7451-7458, 2020.
[42] Wang, Z., Chegdani, F., Yalamarti, N., Takabi, B., Tai, B., El Mansori, M., & Bukkapatnam, S., Acoustic Emission Characterization of Natural Fiber Reinforced Plastic Composite Machining Using a Random Forest Machine Learning Model. *Journal of Manufacturing Science and Engineering*, 142(3), 031003, 2020.
[43] Viera, A. J., Garrett, J. M., Understanding interobserver agreement: the kappa statistic. *Family medicine*, 37(5), 360-363, 2005.
[44] Gerhardt, W., Katus, H. A., Ravkilde, J., Hamm, C., Jørgensen, P. J., Peheim, E., Ljungdahl, L., Löfdahl, P., S-troponin T in suspected ischemic myocardial injury compared with mass and catalytic concentrations of S-creatine kinase isoenzyme MB. *Clinical chemistry*, 37(8), 1405–1411, 1991.
[45] Hamm, C.W., Goldmann, B.U., Heeschen, C., Kreymann, G., Berger, J., Meinertz, T., Emergency room triage of patients with acute chest pain by means of rapid testing for cardiac troponin T or troponin. *New England Journal of Medicine*, 337 (23), 1648-1653, 1997.
[46] Ebashi, S., Kodama, A., A new protein factor promoting aggregation of tropomyosin. *Journal of biochemistry*, 58(1), 107–108, 1965.
[47] Tahir, K., Pauley, E., Dai, X., Smith, S. C., Jr, Sweeney, C., & Stouffer, G. A., Mechanisms of ST Elevation Myocardial Infarction in Patients Hospitalized for Noncardiac Conditions. *The American journal of cardiology*, 123(9), 1393–1398,2019.
[48] Sheehan, P., Vasikaran, S.D. The evolving clinical role of cardiac troponins and new acute myocardial infarction guidelines: Implications for the clinical laboratory. *Clin Bichemist Rev*, 23,52-65,2001.
[49] Reichlin, T., Hochholzer, W., Bassetti, S., Steuer, S., Stelzig, C., Hartwiger, S., Biedert, S., Schaub, N., Buerge, C., Potocki, M., Noveanu, M., Breidthardt, T., Twerenbold, R., Winkler, K., Bingisser, R., Mueller, C., Early diagnosis of myocardial infarction with sensitive cardiac troponin assays. *The New England journal of medicine*, 361(9), 858–867,2009.

[50] Hornemann, T., Dorothea, R., Wallimann, T., Why is creatine kinase a dimer? Evidence for cooperativity between the two subunits. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology,* 1480.1,365-373, 2000.

[51] Hebert, P. R., Moser, M., Mayer, J., Glynn, R. J., & Hennekens, C. H., Recent evidence on drug therapy of mild to moderate hypertension and decreased risk of coronary heart disease. *Archives of internal medicine*, 153(5), 578–581,1996.

[52] Campeau, L., Lesperance, J., Bourassa, M.G., Natural history of saphenous vein aortocoronary bypass grafts. *Mod Concepts Cardiovasc Dis 1984*,53:59-63.

[53] Fitzgibbon, G. M., Kafka, H. P., Leach, A. J., Keon, W. J., Hooper, G. D., & Burton, J. R., Coronary bypass graft fate and patient outcome: angiographic follow-up of 5,065 grafts related to survival and reoperation in 1,388 patients during 25 years. *Journal of the American College of Cardiology*, 28(3), 616–626, 1996.

[54] Labinaz, M., Mathias, J., Pieper, K., Granger, C. B., Lincoff, A. M., Moliterno, D. J., Van de Werf, F., Simes, J., White, H. D., Simoons, M. L., Califf, R. M., Topol, E. J., Armstrong, P. W., Harrington, R. A.,Outcomes of patients with acute coronary syndromes and prior percutaneous coronary intervention: a pooled analysis of three randomized clinical trials. *European heart journal*, 26(2), 128–136, 2005.

[55] Lee, K. L., Woodlief, L. H., Topol, E. J., Weaver, W. D., Betriu, A., Col, J., Simoons, M., Aylward, P., Van de Werf, F., Califf, R. M., Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction. Results from an international trial of 41,021 patients. *Circulation,* 91(6), 1659–1668, 1995.

[56] Karaduman, G., Kelleci Çelik, F., 2D-Quantitative structure-activity relationship modeling for risk assessment of pharmacotherapy applied during pregnancy. *Journal of Applied Toxicology, 43*(10), 1436-1446, 2023.

[57] Erturan, A. M., Karaduman, G., Durmaz, H., Machine learning-based approach for efficient prediction of toxicity of chemical gases using feature selection. *Journal of Hazardous Materials*, *455*, 131616, 2023.

[58] Raihan, M., Islam, M. M., Ghosh, P., Shaj, S. A., Chowdhury, M. R., Mondal, S., & More, A., A comprehensive Analysis on risk prediction of acute coronary syndrome using machine learning approaches. In 2018 21st International Conference of Computer and Information Technology (ICCIT) (pp. 1-6). IEEE, 2018.

[59] Ke, J., Chen, Y., Wang, X., Wu, Z., Zhang, Q., Lian, Y., & Chen, F., Machine learning-based in-hospital mortality prediction models for patients with acute coronary syndrome. *The American journal of emergency medicine*, 53, 127-134, 2022.

[60] Huang, Z., Ge, Z., Dong, W., He, K., Duan, H., & Bath, P., Relational regularized risk prediction of acute coronary syndrome using electronic health records. *Information Sciences*, 465, 118-129, 2018.

[61] Bouzid, Z., Faramand, Z., Gregg, R. E., Frisch, S. O., Martin-Gill, C., Saba, S., ... & Al-Zaiti, S., In search of an optimal subset of ECG features to augment the diagnosis of acute coronary syndrome at the emergency department. *Journal of the American Heart Association*, 10(3), e017871, 2021.

[62] Huyut, M. T., & Ilkbahar, F., The effectiveness of blood routine parameters and some biomarkers as a potential diagnostic tool in the diagnosis and prognosis of Covid-19 disease. *International Immunopharmacology*, 98, 107838, 2021.

[63] Huyut, M. T., Huyut, Z., Ilkbahar, F., & Mertoğlu, C., What is the impact and efficacy of routine immunological, biochemical and hematological biomarkers as predictors of COVID-19 mortality?. *International Immunopharmacology*, 105, 108542, 2022.