

Association of Vitamin D Deficiency with ST-Elevation Myocardial Infarction: Evidence from Statistical Modeling and Machine Learning Classification

Cheragh Hussain

Hayatabad Medical Complex

Farida Ahmad

faridaahmad397@gmail.com

Khyber Girls Medical College

Humal Anwar

Khyber Girls Medical College Peshawar

Hafsa Gul

Khyber Girls Medical College Peshawar

Noor Afsar

Khyber Girls Medical College Peshawar

Umama Muskan

Khyber Girls Medical College Peshawar

Rabia Tahir

Khyber Girls Medical College Peshawar

Sadaf Amin

Khyber Girls Medical College Peshawar

Abdur Rehman

Khyber Girls Medical College Peshawar

Research Article

Keywords: Acute coronary syndrome, STEMI, HDL cholesterol, vitamin D deficiency, machine learning, Random Forest

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10993767/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Acute coronary syndrome (ACS), including ST-elevation myocardial infarction (STEMI) and non-ST-elevation myocardial infarction/unstable angina pectoris (NSTEMI/USAP) is a major cause of cardiovascular morbidity and mortality.

Objective

To determine clinical and biochemical predictors of ACS type and how well they predict using statistical and machine learning methods.

Methodology

A Case-Control study was performed at the Hayatabad Medical Complex, Peshawar, between November 2024 and March 2025 with 100 patients participating in the study with an equal number of STEMI and NSTEMI/USAP patients. IBM SPSS software version 27 was used to analyze the data. The Shapiro-Wilk test was used to determine the normality of the data; the Mann-Whitney U test and Chi-square test were used to compare the groups. Binary logistic regression was used to screen the predictors. Supervised machine learning models (random forest, support vector machine, k nearest neighbors) were trained to classify ACS type.

Results

There were statistically significant group differences in oxygen saturation and HDL. The main predictors were vitamin D levels, HDL, oxygen saturation and dyslipidemia. Machine learning models exhibited moderate performance with Logistic Regression and Support Vector Machine having the highest accuracy and Random Forest having an AUC of 0.792. Furthermore, vitamin D deficiency was observed more frequently among STEMI patients compared to NSTEMI/USAP patients.

Conclusion

Machine learning models moderately predicted ACS classification, with vitamin D deficiency, dyslipidemia, and oxygen saturation identified as key predictors. Vitamin D deficiency was more common in STEMI patients, suggesting a role in ACS severity.

Trial registration

not applicable

INTRODUCTION

Vitamin D insufficiency is a prevalent global health issue. The most common causes of vitamin D insufficiency include limited sun exposure, malabsorption, and liver and renal diseases(1). Vitamin D deficiency is a contributing factor to numerous health issues such as hypertension, diabetes mellitus, coronary artery disease, peripheral artery diseases, cerebrovascular accidents, stroke, rickets, osteomalacia, autoimmune diseases, and cancer(2, 3). On average, 50% of the world's inhabitants are vitamin D deficient and an indirect risk factor for the progression of cardiovascular problems (4). A review of publications from Pakistan between 2008 and 2018 revealed an alarming occurrence of vitamin D deficiency in the country. Nearly half of the population was found to be vitamin D deficient, with a higher incidence observed in women than men. The reason for women's vitamin D deficiency is their limited sun exposure due to their lifestyle and cultural norms(5).

Vitamin D, a pleiotropic nutrient, is crucial for cardiovascular health. Vascular smooth muscles, endothelium, myocytes, and many other tissues within the body contain vitamin D receptors (VDRs), where vitamin D acts on these receptors to prevent the evolution of atherosclerosis, calcifications of the arteries, and heart muscle hypertrophy (6). In addition Vitamin D has anti-inflammatory effects. It reduces inflammation by modulating cytokine production and immune response. Vitamin D help regulate cholesterol levels, reducing the risk of plaque buildup in arteries(7). Cardiac myocytes and fibroblasts express their enzymes, 25-hydroxylase and 1 α -hydroxylase. This implies that 25-hydroxycholecalciferol [25(OH)D] is locally converted to 1,25-dihydroxyvitamin D₃(calcitriol). Since 1 α -Hydroxylase in extra-renal tissues, including the heart, is mostly dependent on substrate availability, the amount of circulating 25(OH)D may play a substantial role in determining the effects of vitamin D in the myocardium(8). Vitamin D suppresses renin gene expression through activation of the vitamin D receptor (VDR), thereby acting as a negative endocrine regulator of the Renin-Angiotensin-Aldosterone system (RASS). Increased renin and angiotensin II levels caused increased RASS activity in vitamin D deficiency or VDR-null conditions result in RASS over activation, leading to hypertension and cardiac hypertrophy. Therefore vitamin D deficiency disrupts upstream RASS regulation, resulting in functional RASS driven fluids and electrolyte imbalance(9, 10).

Multiple studies have reported a significant association between vitamin D deficiency and STEMI patients(6, 11). However, these findings remains inconsistent and highly region specific. Vitamin D deficiency is highly prevalent in South Asia, with Pakistan being one of the most affected countries (12). In addition, South Asians have a disproportionately higher burden of coronary heart disease compared to European populations, indicating increased susceptibility to atherosclerotic events (13). Therefore, the present study aimed to evaluate the relationship between vitamin D levels and acute coronary syndrome, specifically STEMI and NSTEMI/USAP, while integrating both statistical and machine learning approaches to enhance disease classification. In addition, this study seeks to develop and compare

predictive models using clinical, demographic, and biochemical variables to differentiate ACS types and identify the most significant predictors through advanced machine learning techniques.

METHODOLOGY

Data source

This Case -Control study was conducted from November 2024 to March 2025 in the Cardiology Department of Hayatabad Medical Complex, Peshawar, Khyber Pakhtunkhwa, Pakistan. Ethical approval was obtained from the Institutional Research and Ethical Board (IREB) of Hayatabad Medical Complex, Peshawar, Pakistan (Approval No.2217). The sample size was calculated using WHO sample size calculator based on an expected frequency of cases (79.2%) and controls (41.5%), and estimating an anticipated Odds ratio of 5.5 with 95% confidence level and relative precision of 50%. The minimum required sample size was 82; however, to enhance statistical validity and reduce sampling error, a total of 100 patients were enrolled. Written informed consent was obtained from all participants in accordance with the ethical principles of the Declaration of Helsinki. Patients aged ≥ 18 years diagnosed with acute coronary syndrome including ST-elevation myocardial infarction (STEMI), unstable angina pectoris (USAP), and non-ST-elevation myocardial infarction (NSTEMI) were recruited using a non-probability convenience sampling technique. Participants were stratified into two equal groups: 50 patients with STEMI and 50 patients with non-ST-elevation acute coronary syndromes (USAP/NSTEMI). Patients with acute or chronic kidney disease, those who had received vitamin D supplementation within the preceding 3–6 months, and individuals with malabsorption syndromes (based on clinical history and medical record review) were excluded to minimize confounding factors. Ethical approval (No. 2217) was obtained from the institutional review board, and written informed consent was secured from all participants prior to data collection. Data were collected using a structured study Performa, which included demographic characteristics, baseline laboratory investigations (complete blood count, renal function tests, blood glucose, and lipid profile), and clinical parameters such as heart rate, blood pressure, oxygen saturation, and ECG findings. In addition, serum samples were analyzed for vitamin D and troponin levels. Vitamin D measurement was performed using a chemiluminescent electrochemiluminescence immunoassay (ECLIA) on a Cobas E601 analyzer under standardized laboratory conditions. Based on serum vitamin D concentrations, participants were categorized into three groups: Group A ($< 10\text{ng/mL}$), Group B ($10\text{--}30\text{ ng/mL}$), and Group C ($\geq 30\text{ ng/mL}$), allowing a structured comparison across deficiency levels for analytical assessment.

Statistical Analysis

The data were analyzed using IBM SPSS Statistics version 27. Descriptive statistics were calculated to describe how demographic, clinical and laboratory parameters were distributed among the study participants. To check the normality of the continuous variables, the Shapiro-Wilk test was used to determine the right statistical method to be used to compare groups. The intergroup analysis was conducted using non-parametric tests since most of the variables were not normally distributed.

Moreover, Chi-square tests were used to compare categorical variables, and logistic regression screening was performed to identify possible predictors in terms of ACS type. To further increase the predictive performance and complex non-linear relationship of variables, machine learning (ML) techniques were also utilized. Clinical, demographic, and biochemical predictors were used to create supervised ML models, such as Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) to classify ACS types (STEMI vs. NSTEMI/USAP) and evaluate their diagnostic performance. A p-value of < 0.05 was considered statistically significant.

RESULTS

This study included 50 patients in STEMI group and 50 in NSTEMI/USAP group.

Table 1
Comprehensive Descriptive Profile of Biochemical Parameters among Study Participants

Variable	Min	Max	Mean	Std. Deviation	Skewness	Kurtosis
Age	33	89	58.08	11.90	0.049	-0.442
Heart Rate	46	120	82.59	11.13	-0.226	1.335
Oxygen Saturation	86	100	95.81	2.69	-1.309	2.235
Troponin Levels	0.10	25.00	2.49	5.04	3.044	9.574
Hb	9.70	17.20	13.36	1.72	0.061	-0.676
Platelets	79	687	278.02	86.87	0.888	3.909
WBCs	4.64	17.70	8.57	2.26	1.269	2.503
Urea	10.00	147.00	36.28	22.25	2.785	9.824
Creatinine	0.14	7.59	1.06	0.76	6.729	56.943
Blood Glucose	8.37	586.00	142.84	91.86	2.311	6.695
Cholesterol	60	244	156.90	42.63	-0.132	-1.075
Triglycerides	52	482	146.77	61.59	3.444	16.141
HDL	4	70	46.84	14.52	-0.464	-0.201

Table 1 shows the descriptive statistics of the clinical and biochemical parameters of the study participants. The patients had a mean age of 58.08 ± 11.90 years, with ages ranging between 33 and 89 years of age, meaning that most of the patients were middle-aged to older individuals, a group with a high risk of developing cardiovascular diseases. The average heart rate was 82.59 ± 11.13 beats per minute, indicating a relatively stable hemodynamic condition among the patients at presentation. The relative levels of oxygen saturation were fairly high ($95.81 + 2.69$) though there was slight negative skew which means that the majority of patients were found to have oxygen saturation values that were

concentrated on the higher range. The biochemical indicators showed a significant level of variability. The levels of Troponin indicated a strong positive skewness (3.044) and a high level of kurtosis (9.574), indicating the existence of extreme values characteristic of myocardial injury conditions. On the same note, the indicators of renal function and the blood glucose levels also demonstrated positive skew, as there were some patients with significantly high levels. The parameters of lipid profile showed moderate variability. The mean cholesterol level was 156.90 ± 42.63 mg/dL, triglycerides averaged 146.77 ± 61.59 mg/dL, and HDL levels averaged 46.84 ± 14.52 mg/dL. On the whole, the descriptive statistics indicate that there is a great deal of heterogeneity in clinical and biochemical parameters in ACS patients, especially in those parameters that represent cardiac injury and metabolic status.

Table 2
Summary of Normality Test (Shapiro-Wilk) for Study Variables

Variable	STEMI (p-value)	NSTEMI/USAP (p-value)	Normality Decision
Age	0.757	0.264	Normal
Heart rate	0.005	0.492	Non-normal
Oxygen saturation	< 0.001	< 0.001	Non-normal
Troponin levels	< 0.001	< 0.001	Non-normal
Hb	0.433	0.633	Normal
Platelets	0.002	0.763	Non-normal
WBCs	0.001	0.002	Non-normal
Urea	< 0.001	< 0.001	Non-normal
Creatinine	< 0.001	0.137	Non-normal
Blood glucose	< 0.001	< 0.001	Non-normal
Cholesterol	0.001	0.472	Non-normal
Triglycerides	< 0.001	< 0.001	Non-normal
HDL	0.227	0.015	Non-normal

Table 2 summarizes the findings of the Shapiro-Wilk normality test when performed separately on STEMI and NSTEMI / USAP populations. The findings denote that the number of variables that were normally distributed was very small. In particular, the age and Haemoglobin(Hb) were normally distributed in both groups ($p > 0.05$). But most of the variables such as heart rate, oxygen saturation, troponin levels, renal function tests, blood glucose, cholesterol, triglycerides, HDL revealed a significant aberration of normality ($p < 0.05$). These results suggest that most clinical and biochemical variables were not normally distributed. Based on this pattern of distribution, non-parametric statistical techniques were deemed more suitable to compare continuous variables in cases of STEMI and NSTEMI/USAP. As a result, Mann-Whitney U test was used to compare groups.

Table 3
Mann-Whitney U Test Comparing STEMI and NSTEMI+USAP Patients

Variable	Mean Rank (STEMI)	Mean Rank (NSTEMI)	Mann–Whitney U	Z	p-value
Age	48.70	52.30	1160.0	-0.621	0.535
Heart Rate	50.46	50.54	1248.0	-0.014	0.989
Oxygen Saturation	41.37	59.63	793.50	-3.189	0.001
Troponin Levels	52.89	48.11	1130.5	-0.856	0.392
Urea	54.32	46.68	1059.000	-1.317	0.188
Creatinine	55.85	45.15	982.500	-1.844	0.065
Platelets	46.12	54.88	1031.000	-1.510	0.131
Hb	47.62	53.38	1106.000	-0.993	0.321
WBCs	49.85	51.15	1217.500	-0.224	0.823
Blood Glucose	51.83	49.17	1183.5	-0.459	0.646
Cholesterol	46.22	54.78	1036.0	-1.477	0.140
Triglycerides	52.56	48.44	1147.0	-0.711	0.477
HDL	35.06	65.94	478.0	-5.329	< 0.001

Table 3 shows the Mann–Whitney U test comparing STEMI and NSTEMI+USAP patients. The results indicate that most variables, including age, heart rate, troponin levels, urea, creatinine, platelet count, hemoglobin, white blood cell count, blood glucose, cholesterol, and triglycerides, did not show statistically significant differences between the two groups ($p > 0.05$), suggesting similar distributions in both conditions. However, oxygen saturation showed a statistically significant difference ($p = 0.001$), with higher mean ranks in NSTEMI+USAP patients compared to STEMI patients. Similarly, HDL levels were also significantly different between the groups ($p < 0.001$), with NSTEMI+USAP patients demonstrating markedly higher mean ranks than STEMI patients. These findings suggest that, among the studied variables, only oxygen saturation and HDL levels differ significantly between STEMI and NSTEMI+USAP patients, while all other parameters remain statistically comparable.

Table 4
Association of Acute Coronary Syndrome Type with Hypertension and Diabetes

Association	χ^2 (df)	p-value	Fisher's Exact p-value
STEMI/NSTEMI+USAP × Diabetic Mellitus	0.372 (1)	0.542	0.685
STEMI/NSTEMI+USAP × Hypertension	0.360 (1)	0.548	0.689

Figure 1 graphically shows the prevalence of hypertension and diabetes mellitus in the two ACS groups and Table 4 analyzes the relationship between the type of acute coronary syndrome and two key cardiovascular risk factors hypertension and diabetes mellitus. The Chi-square analysis revealed no statistically significant relationship between type of ACS and diabetes mellitus (0.372, 0.542) or with hypertension (0.360, 0.548). The absence of significant associations ($p = 0.685$ in diabetes and $p = 0.689$ in hypertension) was also confirmed by Fisher precise test. These results indicate that among the patients with STEMI and NSTEMI/USAP, the prevalence of hypertension and diabetes mellitus was relatively similar. Although both conditions are established cardiovascular risk factors, they might affect the overall development of coronary artery disease, but not the specific clinical presentation of acute coronary syndrome.

Table 5 Predictors Associated with Acute Coronary Syndrome Type Based on Logistic Regression Screening			
Predictor Variable	Score χ^2	df	p-value
Dyslipidemia	6.857	1	0.009
Vitamin D Levels	12.236	2	0.002
HDL	26.506	1	< 0.001
Oxygen Saturation	9.146	1	0.002
Creatinine	3.643	1	0.056
Cholesterol	3.683	1	0.055

Table 5 shows the results of the logistic regression screening analysis that is used to identify the possible predictors in relation to the type of acute coronary syndrome. A number of variables were found to have statistically significant relationships. The association between HDL levels was strongly correlated (Score $\chi^2 = 26.506$, $p < 0.001$), which is in line with previous conclusions that a lower level of HDL could be associated with more serious cardiac events. Others among the significant predictors were vitamin D levels ($p = 0.002$), oxygen saturation ($p = 0.002$) and dyslipidemia ($p = 0.009$). These results underscore the significance of metabolic, echocardiographic, and physiological parameters in the type of ACS. In turn, creatinine ($p = 0.056$) and cholesterol ($p = 0.055$) were close to reaching the conventional level of $p < 0.05$, meaning that their impact could be relatively smaller in this dataset. On the whole, the regression screening indicates that lipid metabolism, parameters of cardiac functioning, oxygenation, and the level of vitamin D are the important factors that can be associated with the clinical manifestation of acute coronary syndrome.

Machine Learning Technique

To further improve the predictive analysis, machine learning methods were used to classify the type of acute coronary syndrome (ACS) into STEMI and NSTEMI/USAP. Compared to more traditional statistical methods, machine learning algorithms do not assume strictness of relationships between variables and can capture complex and non-linear relationships between variables. Thus, these methods were deemed suitable in the modeling of the clinical and biochemical predictors of the type of ACS. A series of supervised machine learning models were applied, such as Logistic Regression, Random Forest, Support Vector Machine, and K-Nearest Neighbors. The clinical and laboratory variables measured on the study participants were used to train these models to predict the occurrence of STEMI versus NSTEMI/USAP. Standard classification measures such as accuracy, precision, recall and F1-score were used to evaluate model performance.

Table 6
Performance Comparison of Machine Learning Models for ECG Prediction

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.667	0.715	0.667	0.662
Random Forest	0.600	0.622	0.600	0.600
Support Vector Machine (SVM)	0.667	0.715	0.667	0.662
K-Nearest Neighbors (KNN)	0.667	0.667	0.667	0.667

Table 6 and Fig. 2 give a comparative analysis of the four machine learning models Logistic Regression, Random Forest, Support Vector Machine and K-Nearest Neighbors in ECG prediction based on accuracy, precision, recall and F1-score. The findings indicate that Logistic Regression and Support Vector machine show the highest predictive ability with an accuracy of 0.667, precision of 0.715, recall of 0.667 and an F1-score of 0.662, which implies that they have reliable and consistent classification performance. K-Nearest Neighbors model also demonstrates balance in results, with equal scores of 0.667 in all the evaluation measures, which represents a stable result of prediction. Conversely, Random Forest has a relatively lower level of effectiveness with an accuracy of 0.600, a precision of 0.622, a recall of 0.600 and an F1-score of 0.600. Overall, the results indicate that the Logistic Regression and Support Vector Machine models are the most dependable models of ECG prediction in this paper.

Table 7
Confusion Matrix for Classification of STEMI and NSTEMI/USAP Based on ECG Findings

Class	Meaning	Clinical Risk	Predicted 0	Predicted 1
0	STEMI	High risk (emergency)	11 (TP)	6 (FN)
1	NSTEMI + USAP	Lower risk	4 (FP)	9 (TN)

Table 8
Classification Performance Metrics for STEMI and NSTEMI/USAP Prediction

Metric	Class 0 (STEMI)	Class 1 (NSTEMI+USAP)	Overall
Precision	0.73	0.60	0.68 (weighted)
Recall (Sensitivity)	0.65	0.69	0.67 (accuracy)
F1-score	0.69	0.64	0.67
Support	17	13	30

Table 7, Table 8 and Fig. 3 summarize the performance of the RF model in classifying ECG findings into STEMI and NSTEMI/USAP. Based on the confusion matrix, out of 30 test cases, the model correctly identified 11 STEMI cases (true positives) and 9 NSTEMI/USAP cases (true negatives) and 6 STEMI cases were misclassified as NSTEMI/USAP (false negatives) and 4 NSTEMI/USAP cases were incorrectly predicted as STEMI (false positives). The model had a total accuracy of 0.67 with a weighted precision of 0.68 and an overall F1-score of 0.67. In the case of STEMI (Class 0), which is a high risk emergency situation, the model achieved a precision of 0.73 with a sensitivity (recall) of 0.65, meaning that about 65% of the true STEMI cases were correctly identified. In the case of NSTEMI/USAP (Class 1), the model had a precision of 0.60 with a recall of 0.69. These findings indicate that RF classifier has moderate performance in terms of distinguishing between STEMI and NSTEMI/USAP patients based on the ECG results as further demonstrated by the confusion matrix in Fig. 3.

Figure 4 illustrates the receiver operating characteristic (ROC) curve of the Random Forest model for differentiating STEMI from NSTEMI/USAP.

The ROC curve of the STEMI detection with the help of the Random Forests model shows a good level of discriminating ability with an area under the curve (AUC) of 0.792. The close value of AUC of about 0.80 indicates that the model has a high ability to differentiate the cases of STEMI and NSTEMI/USAP significantly exceeding random classification (AUC = 0.50). This implies that there is about a 79.2% chance that the model will rank a randomly selected STEMI case higher than a randomly selected NSTEMI/USAP case. The fact that the curve moves above the diagonal reference line is another indication of consistent sensitivity-specificity trade-off at different threshold levels. This can have clinical implications in that the model has good diagnostic potential to identify high-risk STEMI patients based on ECG findings although optimization can further enhance its sensitivity to detect emergency patients.

Table 9
Top 10 Most Important Features Identified by
Random Forest (Gini Importance)

Rank	Feature	Importance
1	Vitamin_D_levels	0.234579
2	Blood_pressure	0.120016
3	Hypertension	0.101669
4	Gender	0.099319
5	Diabetic_mellitus	0.099240
6	BMI	0.078184
7	Heart_Rate	0.078038
8	Age	0.066290
9	Dyslipidemia	0.066017
10	Oxygen_saturation	0.056647

Table 9 and Fig. 5 show the most important 10 predictors as identified by the Random Forest model using Gini importance to classify ECG findings (STEMI, NSTEMI+USAP) as the dependent variable. The outcome shows that the most effective predictor is Vitamin D levels (0.2346) followed by key cardiovascular and metabolic variables in the outcome such as blood pressure (0.1200), hypertension (0.1017), and diabetic mellitus (0.0992). All of these variables are indicative that both cardio-metabolic and nutritional status is a significant factor when it comes to the ECG-based diagnostic categories. Furthermore, the demographic and physiological variables, such as gender, BMI, heart rate, and age have moderate significance, implying their encouraging role in enhancing the accuracy of the classification. Conversely, dyslipidemia and oxygen saturation are less but still useful predictive variables. On the whole, the results obtained show that the Random Forest model of predicting ECG findings (dependent variable) is mainly driven by a combination of nutritional, metabolic and cardiovascular risk factors, which together increase the ability of the model to discriminate between the STEMI, NSTEMI+USAP cases.

DISCUSSION

The current research measured the association of vitamin D deficiency and other clinical and biochemical features of the patients diagnosed with ACS including STEMI and NSTEMI/USAP. The study results emphasize the possibility that vitamin D deficiency may contribute significantly to the pathophysiology of STEMI, a more severe form of myocardial infarction characterized by complete coronary artery blockage, compared to NSTEMI, which typically involves partial blockage. Identifying a stronger connection between lower levels of vitamin D and STEMI opens avenues for further exploration

into how improving vitamin D levels might mitigate the probability or severity of acute coronary syndromes. Alongside standard statistical techniques, machine learning algorithms were applied to assess their value for ACS classification and prediction. The results showed that vitamin D deficiency and low HDL levels were significantly linked to STEMI presentation.

Vitamin D exerts significant anti-atherosclerotic effects by modulating the immune system, acting through the vitamin D receptor (VDR) and activating enzymes (CYP27A1, CYP27B1) in immune cells locally, which result in suppression of inflammation and stabilization of atherosclerotic plaques. It regulates innate immunity by promoting monocyte-to-macrophage differentiation, inhibiting toll-like receptor expression, and reducing pro-inflammatory cytokine production, thereby decreasing inflammatory activity within arterial walls. In adaptive immunity, it enhances Th2 and Treg responses while suppressing B-cell proliferation and Th1/Th17 responses, helping to reduce vascular inflammation and improve plaque stability. Additionally, vitamin D–VDR signalling promotes cholesterol efflux through ABCA1 and downregulates macrophage scavenger receptors, which limits LDL uptake and foam cell generation. This reduces the development of atherosclerotic plaque and improves lipid management inside the arterial wall ((7), (10). By increasing nitric oxide (NO) generation through the eNOS pathway, vitamin D enhances endothelial function and supports vascular integrity and vasodilation, while its anti-inflammatory and antioxidant properties reduce pro-inflammatory signalling and oxidative stress. Furthermore, it increases thrombomodulin and antithrombin, lowers tissue factor, platelet activity, and PAI-1 levels, and inhibits NF- κ B to reduce pro-thrombotic cytokines. It also inhibits adhesion molecule expression, reducing leukocyte migration and further limiting progression of atherosclerosis, thereby reducing thrombosis. Overall, our study findings are consistent with the molecular mechanisms of vitamin D reported in the literature, supporting its role in antithrombotic and anti-atherosclerotic pathways ((14, 15)

The results of our study are in line with several previous studies that found a negative correlation between serum vitamin D levels and the severity of coronary artery disease. A comparable study conducted in 2021 at Islamabad Medical and Dental College found vitamin D deficiency in 79.2% of STEMI patients and 41.5% of patients without STEMI(11). A similar study was done by Safaie et al., which included 88 patients, out of which low levels of vitamin D were recorded in 59.1%, with a greater frequency in the STEMI group, and the frequency of vitamin D insufficiency was 19.3% (it was 15% in the STEMI group and 22.9% in the NSTEMI) (6).

The study of Balasubramanian et al. emphasizes that vitamin D is essential for cardiovascular health and also has an impact on the severity of myocardial infarction. The high prevalence of vitamin D deficiency directs that supplementation could be valuable in reducing cardiovascular risk. The results of our study were consistent with the significant correlation found by binary logistic regression analysis between vitamin D deficiency and the incidence of STEMI elsewhere (16).

However some studies did not show any significant association At Zagazig University in Egypt, A cross-sectional study (n = 71) was performed that reported low vitamin D levels in ACS patients, but no

statistically significant difference was seen between ACS subgroups (STEMI, NSTEMI, and USA) ($p > 0.05$)(17). Similarly, the MINI-Finland Health Survey of 6,219 adults aged over 30 without prior heart issues found no link between vitamin D deficiency and coronary heart disease mortality after controlling for traditional cardiovascular risk factors and seasonal variation, during a 27-year follow-up period (18). Our study population has a high prevalence of low vitamin D levels, which implies that vitamin D supplementation may significantly lower the risk of cardiovascular events. Thompson et al. (2023) found that vitamin D supplementation lowers the risk of cardiovascular events in patients with vitamin D deficiency(19).

Our study, like other studies, has some limitations, such as: relatively small sample size (i.e 100). In a small sample size of 100, the results might not be generalized. Our study is a monocentric study conducted in the Cardiology Department of Hayatabad Medical Complex, Peshawar, so there is a chance of selection bias and lack of diversity. As our study lacks follow-up so we were unable to observe and evaluate the long-term changes associated with vitamin D deficiency. Also the cross-sectional nature of the comparison limits the ability to establish a causal relationship.

CONCLUSION

The study has assessed the clinical, biochemical, and metabolic features in relation to various types of acute coronary syndrome, and has examined how machine learning methods can be used to predict the classification of acute coronary syndrome. The findings indicated that the level of oxygen saturation and the level of HDL differed significantly between STEMI and NSTEMI/USAP patients, which could be considered to play the role of the severity of the presentation of ACS. Additional screening with logistic regression identified a few significant predictors, such as, vitamin D levels, dyslipidemia, and oxygen saturation, which demonstrated the combined effects of cardiovascular, metabolic, and physiological factors in the development of ACS. Machine learning models were shown to have moderate predictive capacity, with Logistic Regression and Support Vector Machine showing the highest overall performance and random forest showing good discriminatory ability with a AUC of 0.792 to detect STEMI. The analysis of feature importance also established that the levels of vitamin D and the crucial cardio metabolic risks factors are significant predictors of ACS outcomes. Moreover, the prevalence of vitamin D deficiency was found more frequently among STEMI patients than among NSTEMI/USAP patients, indicating that ACS patients with vitamin D deficiency could be having a higher risk of developing STEMI. In the general, the combination of conventional statistical methods with machine learning techniques presents a complete framework of determining important predictors and enhancing diagnostic classification of acute coronary syndrome. The findings can be used to assist clinicians in early risk stratifying and decision making which will ultimately aid in better patient managements and cardiovascular outcomes.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Institutional Research and Ethical Board (IREB) of Hayatabad Medical Complex, Peshawar, Pakistan (Approval No.2217).

Consent for publication

not applicable

Funding

Nil

Author Contribution

Authors' Contributions Cheragh Hussain contributed to the conception and design of the study, provided expert clinical oversight in cardiology, supervised patient selection and clinical data acquisition, interpreted cardiology-specific findings, and critically revised the manuscript for important intellectual content. He approved the final submitted version and agrees to be accountable for all aspects of the work. Farida Ahmad contributed to the conception and design of the study, supervised the overall research methodology, coordinated the research team, interpreted the data, drafted and substantively revised the manuscript, and managed submission correspondence as co-corresponding author. She approved the final submitted version and agrees to be accountable for all aspects of the work. Humal Anwar contributed to data acquisition, assisted in drafting the manuscript, participated in interpretation of results, and served as co-corresponding author managing editorial communications. She approved the final submitted version and agrees to be accountable for all aspects of the work. Hafsa Gul contributed to data acquisition, assisted in literature searching and review, and participated in drafting sections of the manuscript. She approved the final submitted version and agrees to be accountable for all aspects of the work. Noor Afsar contributed to data acquisition, assisted in literature searching and review, and participated in drafting sections of the manuscript. She approved the final submitted version and agrees to be accountable for all aspects of the work. Umama Muskan contributed to data acquisition, assisted in literature searching and review, and participated in drafting sections of the manuscript. She approved the final submitted version and agrees to be accountable for all aspects of the work. Rabia Tahir contributed to data acquisition, assisted in literature searching and review, and participated in drafting sections of the manuscript. She approved the final submitted version and agrees to be accountable for all aspects of the work. Sadaf Amin contributed to data acquisition, assisted in literature searching and review, and participated in drafting sections of the manuscript. She approved the final submitted version and agrees to be accountable for all aspects of the work. Abdur Rehman (Department of Research) contributed to the statistical analysis and interpretation of data, including statistical modeling and machine learning classification procedures. He approved the final submitted version and agrees to be accountable for all aspects of the work. All authors have read and approved the final manuscript. All authors have agreed to be personally accountable for their own contributions and to

ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated, resolved, and the resolution documented in the literature.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests; Nil

Acknowledgements; We acknowledge department of Cardiology HMC

References

1. Saeed MK, Khan A, Chachar AZK, Khanum A, Khan H. Vitamin D Deficiency in Acute Myocardial Infarction and its Association with Diabetes. *Annals King Edw Med Univ.* 2022;28(1):13–8.
2. Kheiri B, Abdalla A, Osman M, Ahmed S, Hassan M, Bachuwa G. Vitamin D deficiency and risk of cardiovascular diseases: a narrative review. *Clin Hypertens.* 2018;24(1):9.
3. Basit S. Vitamin D in health and disease: a literature review. *Br J Biomed Sci.* 2013;70(4):161–72.
4. Zhernakova N, Bunova S, Agarkov N, Lebedev D, Aksenov V. Vitamin D deficiency as an independent predictor of myocardial infarction in the elderly. *Arch Razi Inst.* 2021;76(4):1069.
5. Syed F, Latif MSZ, Ahmed I, Bibi S, Ullah S, Khalid N. Vitamin D deficiency in Pakistani population: critical overview from 2008 to 2018. *Nutr Food Sci.* 2020;50(1):105–15.
6. Safaie N, Rezaee H, Dvati BS, Entezari-Maleki T. Vitamin D deficiency predicts the ST elevation type of myocardial infarction in patients with acute coronary syndrome. *Iran J Pharm research: IJPR.* 2018;17(Suppl):73.
7. Fenercioglu AK. The anti-inflammatory roles of vitamin D for improving human health. *Curr Issues Mol Biol.* 2024;46(12):13514–25.
8. Yilmaz A, Eren E, Ellidag HY, Yuksel IO, Yilmaz N, Arslan S, et al. High prevalence of severe vitamin D deficiency in patients with acute myocardial infarction. *LaboratoriumsMedizin.* 2015;39(3):175–81.
9. Ajabshir S, Asif A, Nayer A. The effects of vitamin D on the renin-angiotensin system. *J nephropathology.* 2014;3(2):41.
10. Condoleo V, Pelaia C, Armentaro G, Severini G, Clausi E, Cassano V, et al. Role of vitamin D in cardiovascular diseases. *Endocrines.* 2021;2(4):417–26.
11. Syed Ikramullah AKMUABM. Frequency of Vitamin D Deficiency in Patients with Acute ST Elevation MI. Volume 16. *PAKISTAN JOURNAL OF MEDICAL AND HEALTH SCIENCES*; 2022. 5.
12. Siddiquee MH, Bhattacharjee B, Siddiqi UR, MeshbahurRahman M. High prevalence of vitamin D deficiency among the South Asian adults: a systematic review and meta-analysis. *BMC Public Health.* 2021;21(1):1823.

13. Wang J, Tillin T, Hughes AD, Chaturvedi N. Associations between family history and coronary artery calcium and coronary heart disease in British Europeans and South Asians. *Int J Cardiol.* 2020;300:39–42.
14. Ojaroodi AF, Jafarnezhad F, Eskandari Z, Keramat S, Stanek A. Recent updates and advances in the association between vitamin D deficiency and risk of thrombotic disease. *Nutrients.* 2024;17(1):90.
15. Khanolkar S, Hirani S, Mishra A, Vardhan S, Hirani S, Prasad R et al. Exploring the role of vitamin D in atherosclerosis and its impact on cardiovascular events: a comprehensive review. *Cureus.* 2023;15(7).
16. Balasubramanian A, Kunchala K, Shahbaz A, Kar A, Sankar J, Anand S et al. Association of vitamin D deficiency as an independent risk factor for myocardial infarction and its therapeutic implications: a systematic review. *Cureus.* 2025;17(1).
17. Shereef AS, Kandeel NT. Role of Vitamin D deficiency among patients with acute coronary syndrome. *J Indian Coll Cardiol.* 2019;9(2):95–9.
18. Kilkkinen A, Knekt P, Aro A, Rissanen H, Marniemi J, Heliövaara M, et al. Vitamin D status and the risk of cardiovascular disease death. *Am J Epidemiol.* 2009;170(8):1032–9.
19. Thompson B, Waterhouse M, English DR, McLeod DS, Armstrong BK, Baxter C et al. Vitamin D supplementation and major cardiovascular events: D-Health randomised controlled trial. *BMJ.* 2023;381.

Figures

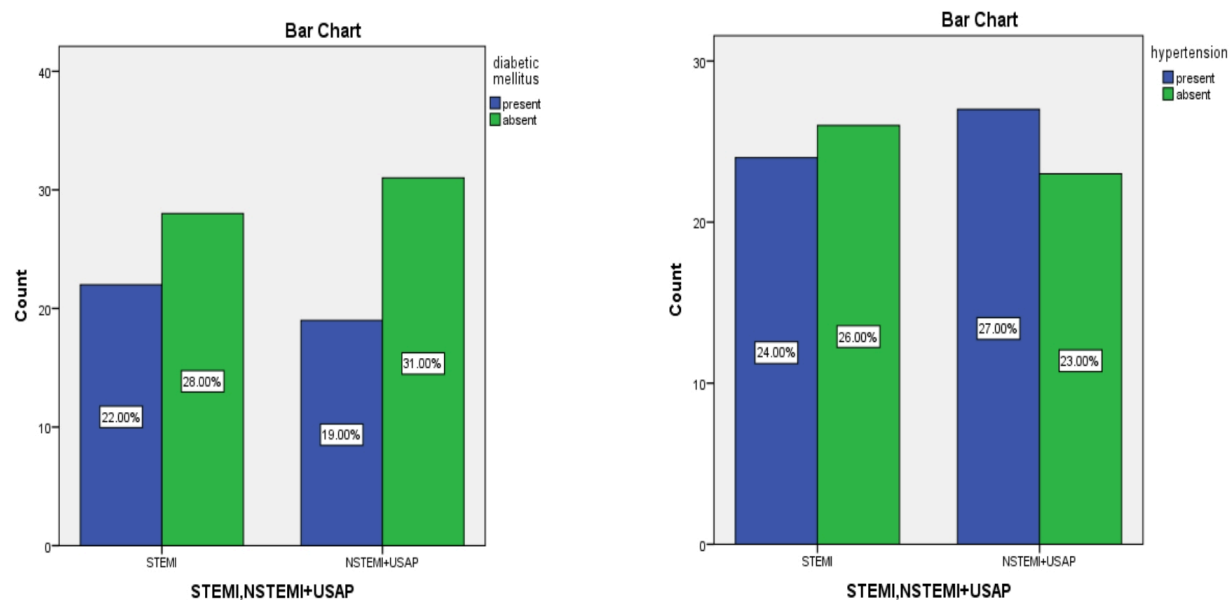


Figure 1

Association of Acute Coronary Syndrome Type with Hypertension and Diabetes Mellitus

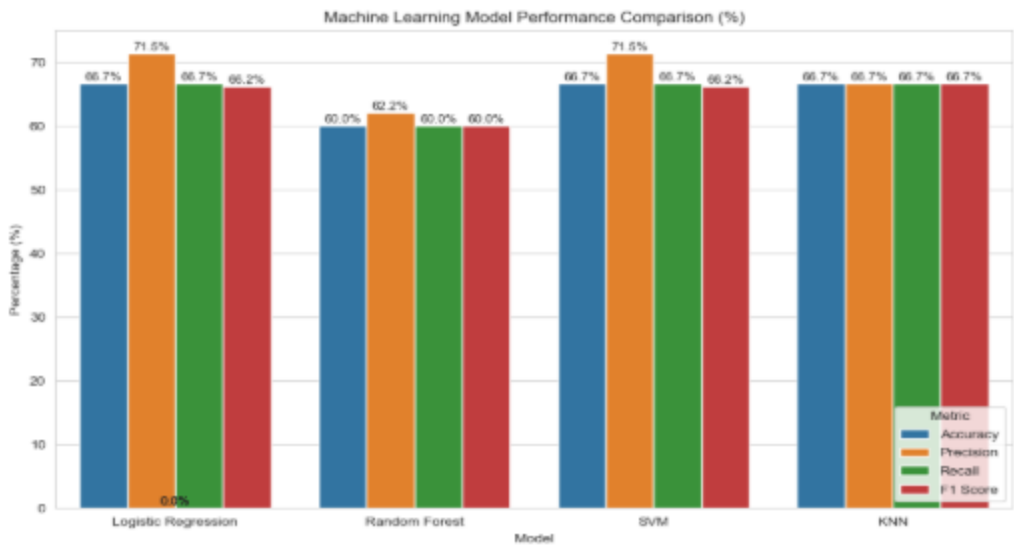


Figure 2

Performance Comparison of Machine Learning Models Based on Accuracy, Precision, Recall, and F1-Score (%)

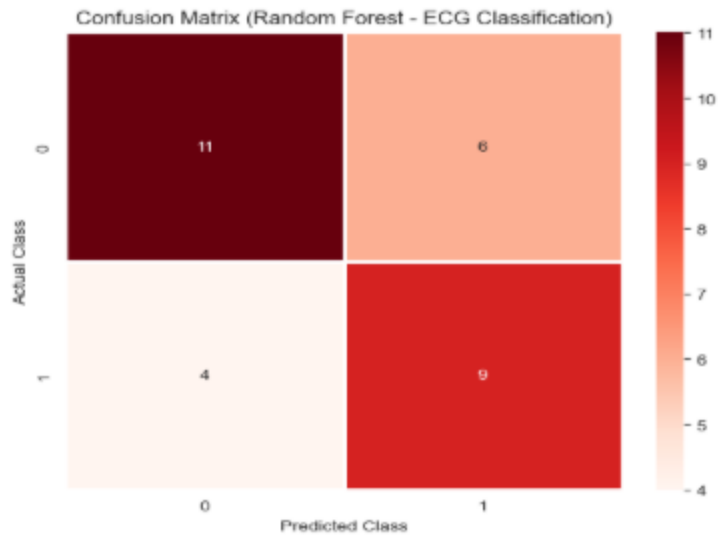


Figure 3

Confusion Matrix of Random Forest Classifier for ECG Findings Prediction

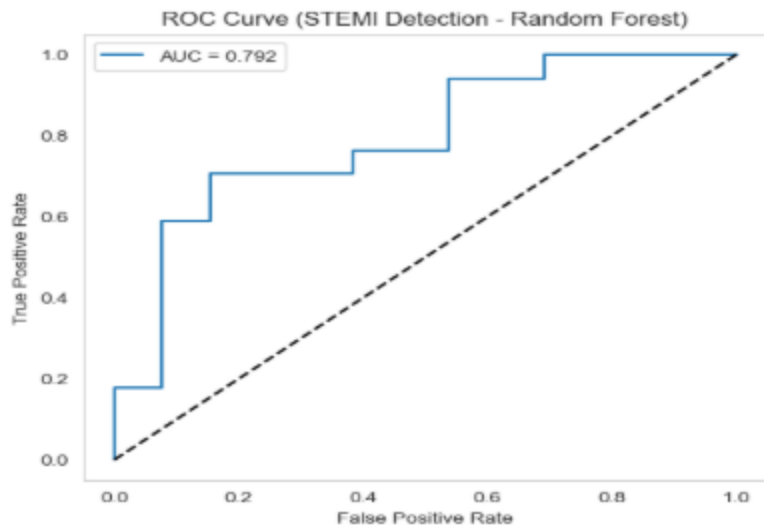


Figure 4

Receiver Operating Characteristic (ROC) Curve for STEMI Detection Using the Random Forest Model

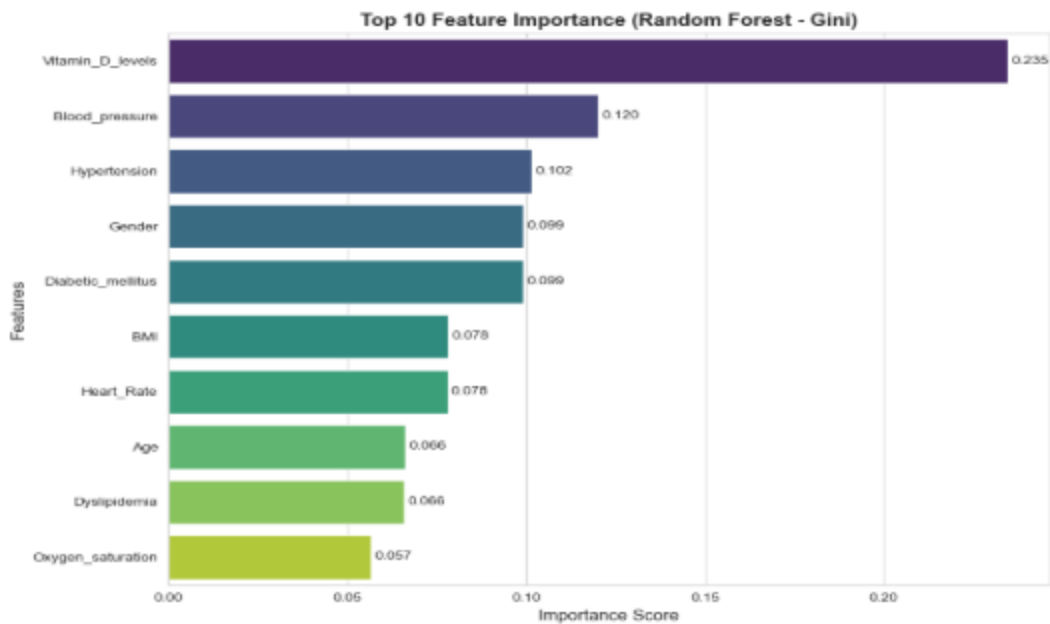


Figure 5

Top 10 Most Important Features Identified by the Random Forest Model Based on Gini Importance