International Journal of Recent Technology and Applied Science
p-ISSN: 2721-2017, e-ISSN: 2721-7280

Original Research Paper

Explaining Cholesterol-Related Coronary Artery Disease Risk Using Machine Learning and SHAP

Eka Pandu Cynthia[1*], Suzani Binti Mohamad Samuri[1], Wang Shir Li[1], Alabbas Hussein Saeed[2], Inggih Permana[3], Febi Yanto[3]

[1] *Department of Artificial Intelligence, Faculty of Computing and Meta Technology, Sultan Idris Education University. Perak, Malaysia.*
[2] *Department of General Practitioners, Faculty of Medicine, Hasanuddin University. Makassar, Indonesia.*
[3] *Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Sultan Syarif Kasim Riau. Pekanbaru, Indonesia.*

Article History
Received:
22.10.2025

Revised:
17.11.2025

Accepted:
29.01.2026

***Corresponding Author:**
Eka Pandu Cynthia
Email:
eka.cynthia@gmail.com

Abstract: Coronary Artery Disease (CAD) remains a leading cause of global mortality, with dyslipidemia recognized as a major modifiable risk factor. This study investigates the relationship between serum lipid parameters and CAD using the Z-Alizadeh Sani clinical dataset comprising 303 patients with 55 clinical, biochemical, and electrocardiographic attributes. Logistic Regression (LR) and Random Forest (RF) models were developed to predict CAD status, supported by a standardized preprocessing pipeline, multi-split train–test evaluation (70/30, 80/20, 90/10), and performance assessment using Accuracy, Precision, Recall, F1-Score, and AUC-ROC. SHapley Additive exPlanations (SHAP) were employed to enhance model interpretability and quantify the contribution of lipid-related and clinical features to individual predictions. The RF model consistently outperformed LR across all split configurations, achieving a maximum AUC of 0.96, while LR attained an AUC of 0.90. SHAP analysis revealed that total cholesterol (CHOL) and low-density lipoprotein (LDL) were strong positive predictors of CAD, whereas high-density lipoprotein (HDL) exhibited a protective effect, in line with established cardiovascular pathophysiology. These findings demonstrate that integrating explainable machine learning with routine clinical lipid profiles can provide accurate and transparent decision support for early CAD risk stratification.

Keywords: Coronary Artery Disease, Dyslipidemia, Logistic Regression, Random Forest, SHAP Explainability.