

Estimation of Atherogenic Index of Plasma Using HDL-C: A Machine Learning–Supported Approach

R. F. Famutimi

rfamutimi@unimed.edu.ng

University of Medical Sciences

M. A. Adegbenro

University of Medical Sciences

P. A Akpojotor

University of Medical Sciences

D. Akinnifesi

University of Medical Sciences

Research Article

Keywords: Atherogenic index of plasma. High density lipoprotein cholesterol, Cardiovascular risk assessment. Machine learning. Lipid profile, Coronary artery disease

Posted Date: May 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9431867/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Estimation of Atherogenic Index of Plasma Using HDL-C: A Machine Learning–Supported Approach

R. F. Famutimi^{1*}, M.A. Adegbenro², P. A Akpojotor¹, D. Akinnifesi²

*Email of Corresponding Author: rfamutimi@unimed.edu.ng

¹Mathematical and Computer Sciences Department, University of Medical Sciences Ondo City, Nigeria

²Medical Laboratory Department, University of Medical Sciences Ondo City, Nigeria

Abstract

Coronary artery disease (CAD) remains a leading cause of morbidity and mortality worldwide, necessitating reliable and accessible biomarkers for early risk assessment. The atherogenic index of plasma (AIP), defined as the logarithmic ratio of triglycerides to high-density lipoprotein cholesterol (HDL-C), has emerged as a strong predictor of cardiovascular risk. However, direct computation of AIP requires the availability of multiple lipid parameters, which may limit its use in resource-constrained settings. In this study, machine learning techniques were employed to explore relationships between AIP and conventional lipid profile parameters using clinical data obtained from adult diabetic and non-diabetic patients at a tertiary healthcare facility in southwestern Nigeria. Correlation analysis identified HDL-C as the lipid parameter most strongly associated with AIP ($r = -0.8277$). Guided by this relationship, a single-parameter linear regression model was developed to estimate AIP using HDL-C alone. Model performance was evaluated using an 80:20 training–testing split, yielding a mean absolute error of 0.0851 (8.51%). For benchmarking purposes, a random forest regression model incorporating multiple lipid features was also trained, achieving superior predictive performance ($R^2 = 0.943$; $MAE = 0.040$). While the random forest model demonstrated higher accuracy, the HDL-C-based linear model offers a clinically interpretable and computationally simple alternative for estimating AIP where full lipid panels may not be readily available. This machine learning–supported approach highlights the potential for simplified cardiovascular risk estimation and supports broader application of AIP-based screening in preventive cardiology.

Keywords: Atherogenic index of plasma. High density lipoprotein cholesterol, Cardiovascular risk assessment. Machine learning. Lipid profile, Coronary artery disease.

Back ground

Cardiovascular diseases (CVDs), particularly coronary artery disease (CAD), remain the leading cause of morbidity and mortality worldwide, posing a significant burden on global public health (World Health Organization 2020). The development of CAD is closely linked to atherosclerosis, a progressive pathological process driven by dyslipidaemia, inflammation, endothelial dysfunction, and oxidative stress (Libby 2021). Early identification of individuals at increased cardiovascular risk is therefore essential for effective prevention and timely clinical intervention.

Plasma lipid abnormalities play a central role in atherosclerotic progression, making lipid-based indices valuable tools for cardiovascular risk assessment. Among these indices, the atherogenic index of plasma (AIP), defined as the logarithmic ratio of triglycerides to high-density lipoprotein cholesterol (HDL-C), has emerged as a robust marker of atherogenic risk (Dobiášová and Frohlich 2001). Elevated AIP values have been associated with an increased risk of coronary artery disease, myocardial infarction, and adverse cardiovascular outcomes across diverse populations (Majeed *et al.* 2019; Ya Li *et al.* 2024).

Despite its clinical relevance, routine computation of AIP requires the availability of both triglyceride and HDL-C measurements, which may not always be feasible in routine clinical practice or in resource-limited settings. This limitation highlights the need for alternative approaches that reduce data requirements while maintaining reasonable predictive performance.

Recent advances in machine learning have provided powerful tools for uncovering complex relationships within biomedical data and improving cardiovascular risk prediction models (Chawla et al. 2017; Oussaid et al. 2021). In particular, machine learning methods have been shown to outperform conventional statistical approaches by capturing non-linear interactions among lipid parameters and clinical variables (Huang et al. 2020; Shah et al. 2025). However, many of these models prioritize predictive accuracy at the expense of interpretability and clinical simplicity.

This study adopts a machine learning-supported approach to identify the dominant lipid determinant of AIP and subsequently develops an interpretable single-parameter regression model for estimating AIP using HDL-C alone. By balancing data-driven analysis with clinical practicality, this work seeks to provide an accessible method for preliminary cardiovascular risk estimation, particularly in settings where full lipid profiling is constrained.

Existing Literature

Coronary artery disease (CAD) remains the most prevalent form of cardiovascular disease worldwide and is strongly driven by atherosclerotic processes involving lipid abnormalities (Ya Li et al. 2024; Zhao et al. 2025).

The disease is primarily driven by atherosclerotic processes involving lipid accumulation, chronic inflammation, endothelial dysfunction, and oxidative stress. As a result, the identification of reliable lipid-based markers for early cardiovascular risk assessment has been a major focus of clinical research.

66 The atherogenic index of plasma reflects the balance between atherogenic and anti-atherogenic
67 lipoproteins and has been shown to increase markedly in individuals with reduced HDL-C levels
68 (Dobiášová and Frohlich 2001; Majeed et al. 2019). Previous studies have demonstrated a strong
69 association between elevated AIP values and the severity of coronary artery disease, as well as adverse
70 cardiovascular outcomes. AIP has been shown to outperform traditional lipid parameters in predicting
71 cardiovascular risk in both general and high-risk populations, including individuals with diabetes and
72 metabolic syndrome. Its clinical relevance is further supported by evidence linking AIP to the
73 presence of small dense low-density lipoprotein particles, which are known to be highly atherogenic.

74 High-density lipoprotein cholesterol (HDL-C) plays a protective role in cardiovascular health by
75 facilitating reverse cholesterol transport and exerting anti-inflammatory and antioxidant effects. Low
76 HDL-C levels have consistently been associated with increased cardiovascular risk, particularly when
77 accompanied by elevated triglyceride levels. Several studies have highlighted the inverse relationship
78 between HDL-C and AIP, emphasizing HDL-C as a key determinant of atherogenic risk. This
79 relationship underpins the potential utility of HDL-C as a surrogate marker in estimating AIP,
80 especially in settings where triglyceride measurements may be unavailable.

81 Advances in machine learning have enabled the development of data-driven models that improve
82 cardiovascular risk prediction by capturing complex relationships among lipid parameters (Chawla et
83 al. 2017; Oussaid et al. 2021). Machine learning models have demonstrated superior predictive
84 performance compared with conventional statistical methods by capturing complex, non-linear
85 interactions among lipid parameters and clinical variables. Several studies have demonstrated that
86 ensemble and machine learning-based approaches outperform traditional statistical methods in
87 cardiovascular risk stratification (Huang et al. 2020; Shah et al. 2025; Yadegar et al. 2025). These

approaches have shown particular promise in integrating lipid ratios, such as triglyceride-to-HDL-C and AIP, into predictive models.

Despite their predictive strengths, many machine learning-based models rely on multiple input variables and complex architectures, which may limit their clinical applicability and transparency. In routine clinical practice, especially in resource-constrained environments, there is a need for models that balance predictive accuracy with simplicity and interpretability. Consequently, research efforts have increasingly explored simplified or surrogate lipid markers capable of maintaining clinical relevance while reducing data requirements.

In this context, the present study builds upon existing literature by employing machine learning techniques to identify the most influential lipid determinant of AIP and subsequently developing an interpretable single-parameter regression model based on HDL-C. By bridging advanced data-driven analysis with clinical practicality, this work seeks to contribute a viable alternative approach for preliminary cardiovascular risk estimation.

Method

Study Design and Data Source

This study employed a retrospective observational design using lipid profile data as follows: total cholesterol (TC), triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and body mass index (BMI). obtained from adult male and female diabetic and non-diabetic patients attending a tertiary referral center, the University of Medical Sciences Teaching Hospital, Ondo City, and Southwestern Nigeria. Ethical approval for the use of

108 anonymized patient data was obtained prior to data collection. Only records with complete lipid
109 profile measurements were included in the analysis. Ninety records were obtained

110 **The baseline lipid characteristics and atherogenic index of plasma values of the study**
111 **participants are summarized in Table 1**

112

113

Table 1.

SNO	BMI	TC	TG	HDL	LDL	AIP
1	23.19	159.1	75.51	35.13	108.87	-0.03
2	21.87	135.45	62.43	12.63	110.33	0.33
3	20.82	142.73	137.19	32.37	82.92	0.27
4	25.08	119.7	97.94	14.21	85.9	0.48
5	28.48	196.36	112.9	14.6	159.18	0.53
6	27.36	129.4	96.07	4.34	105.85	0.99
7	25.35	184.23	82.99	16.18	151.45	0.35
8	23.32	159.7	66.17	12.24	134.23	0.37
9	23.67	123.33	86.73	48.94	57.04	-0.11
10	23.34	123.33	125.98	44.5	53.63	0.09
11	22.48	146.36	64.3	57.23	76.27	-0.31
12	25.48	136.66	64.3	56.84	66.96	-0.31
13	18.38	153.33	79.25	74.1	63.38	-0.33
14	20.52	193.94	77.38	49.33	129.13	-0.16
15	25.35	221.81	125.98	110.51	86.1	-0.3

16	20.29	208.48	80.32	93.3	99.12	-0.42
17	18.1	142.72	210	46.8	53.92	0.29
18	32.46	88.18	75.51	24.42	48.66	0.13
19	22.47	156.36	137.19	21.71	107.21	0.44
20	20.98	101.51	84.86	31.18	53.36	0.07
21	26.13	170.23	134.54	44.21	99.11	0.12
22	24.54	112.42	92.33	15.79	78.16	0.41
23	25.22	134.24	105.42	36.7	76.46	0.1
23	20.52	152.72	64.3	33.55	106.31	-0.08
24	27.12	210.5	71.77	94.54	101.61	-0.48
25	28.44	175.34	137.34	21.34	126.53	0.45
26	25.81	158.14	134.58	54.07	77.15	0.04
27	26.85	161.4	158.69	55.2	74.46	0.1
28	28.84	104.42	130.08	33.62	44.78	0.23
29	23.94	124.19	120.56	35.67	64.41	0.17
30	23.67	150.93	105.42	32.73	97.12	0.15
31	21.22	153.49	171.96	33.2	85.9	0.35
32	24.17	162.56	131.59	43.26	92.98	0.12
33	25.65	120.7	107.29	40.35	58.89	0.06
34	21.15	125.64	134.21	38.47	60.33	0.18
35	27.39	121.86	118.5	44.24	53.92	0.07
36	37.78	132.32	116.63	44.72	64.27	0.06
37	30.83	143.95	122.24	46.8	72.7	0.06

38	32.85	171.86	103.55	39.03	112.12	0.06
39	27.89	169.77	127.85	48.75	95.45	0.06
40	28.73	150	151.21	38.4	81.36	0.24
41	27.38	156.98	107.29	52.01	83.51	-0.05
42	35.96	145.12	111.03	46.67	76.24	0.02
43	29.36	162.56	97.94	74.51	68.46	-0.24
44	32.79	166.28	120.54	62.98	79.19	-0.08
45	23.18	181.4	135.32	40.97	113.37	0.16
46	23.67	255.64	200.34	20.65	194.92	0.63
47	21.72	167.21	73.64	67.7	84.78	-0.32
48	23.51	169.54	153.23	35.63	103.26	0.27
49	31.61	159.3	82.1	57.98	84.9	-0.21
50	24.34	174.74	153.98	42.78	101.16	0.2
51	27.7	152.63	148.38	47.29	75.66	0.14
52	35.06	210.54	124.73	53.96	131.63	0
53	33.2	182.69	106.67	55.04	106.32	-0.07
54	28.93	127.37	93.76	42.3	66.32	-0.01
55	24.34	175	127.35	36.73	112.8	0.18
56	23.05	170	372.03	28.42	67.17	0.76
57	22.57	164.67	153.98	39.1	94.77	0.24
58	28.28	153.45	153.98	39.1	83.55	0.24
59	24.93	194.74	217.2	39.51	111.79	0.38
60	30.82	160.26	136.77	46.66	86.25	0.11

61	33.71	182.9	102.36	19.93	142.5	0.35
62	25.22	120.8	76.56	43.26	62.23	-0.11
63	31.22	175	153.98	92.98	51.22	-0.14
64	26.85	161.58	151.83	36.6	94.61	0.26
65	28.48	178.69	159.14	30.28	116.58	0.36
66	27.61	119.47	113.12	54.44	42.41	-0.04
67	28.26	228.95	148.38	75.97	123.3	-0.07
68	29.21	183.32	170.35	36.06	113.19	0.31
69	26.64	169.74	114.12	76.94	69.98	-0.19
70	19.97	172.61	145.32	47.15	96.4	0.13
71	23.1	163.45	164	56.34	74.31	0.1
72	23.78	160.34	156	55.23	73.91	0.09
73	26.95	160.32	146	52.86	78.26	0.08
74	26.22	179.37	164	57.67	88.9	0.09
75	29.86	173.66	107.47	50.4	101.77	-0.03
76	23.18	164.45	126.64	39.83	99.29	0.14
77	30.22	152.33	129.56	35.23	91.19	0.21
78	24.11	163.45	115.5	48.34	92.01	0.02
79	31.25	205.63	138.3	28.45	149.52	0.33
80	21.72	173.21	120.43	47.56	101.56	0.04
81	24.77	171.3	139.45	42.45	100.96	0.16
82	32.79	163.33	151.23	33.54	99.54	0.29
83	23.88	206.14	131	55.34	124.6	0.01

84	32.04	175.52	157.34	34.34	109.71	0.3
85	25.39	163.2	159.03	52.32	79.07	0.12
86	26.72	141.3	108.45	38.45	81.16	0.09
87	32.39	134.53	114.34	32.98	78.68	0.18
88	25.65	149.23	155.23	35.32	82.86	0.28
89	24.38	204.32	157.34	26.45	146.4	0.41
90	26.09	174.37	114.52	29.34	122.13	0.23

Values are expressed in mmol/L.

Lipid Parameters and Atherogenic Index of Plasma

The atherogenic index of plasma (AIP) was calculated using the established formula (Equation 1):

$$AIP = \log_{10} \left(\frac{TG}{HDL-C} \right) \quad \text{Equation 1}$$

Where triglycerides and HDL-C were measured in mmol/L. AIP values were included as provided in the data set and used as the reference outcome variable for model development and evaluation.

Correlation Analysis

To identify the lipid parameter most strongly associated with AIP, correlation analysis was performed using the Python programming language within the Spyder integrated development environment. The Pearson correlation coefficient was computed between AIP and each lipid parameter using the SciPy library. Correlation coefficients ranged from -1 to $+1$, with values closer to -1 or $+1$ indicating

stronger negative or positive relationships, respectively. The lipid parameter exhibiting the strongest correlation with AIP was selected for subsequent model development.

Data Partitioning

Following correlation analysis, the data set was randomly partitioned into training and testing subsets. Eighty percent (80%) of the data were used for model training, while the remaining twenty percent (20%) were reserved for independent model testing. This split was chosen to ensure sufficient data for model learning while preserving an unseen subset for performance evaluation.

Model Development

Two predictive modeling approaches were implemented. First, a linear regression model was constructed using the lipid parameter most strongly associated with AIP as the sole predictor. This single-parameter model was developed to enhance interpretability and clinical usability. Model training was performed on the training data set using the scikit-learn library in Python.

For comparative benchmarking, a random forest regression model was also developed using multiple lipid parameters as predictors. The random forest model was trained using default hyper parameters to evaluate the potential performance gains achievable through ensemble learning methods.

Model Evaluation

Model performance was assessed using the testing dataset. The primary evaluation metric was the mean absolute error (MAE), which quantifies the average absolute difference between predicted and observed AIP values. Additionally, the coefficient of determination (R^2) was computed for the random forest model to assess the proportion of variance in AIP explained by the predictors. Graphical

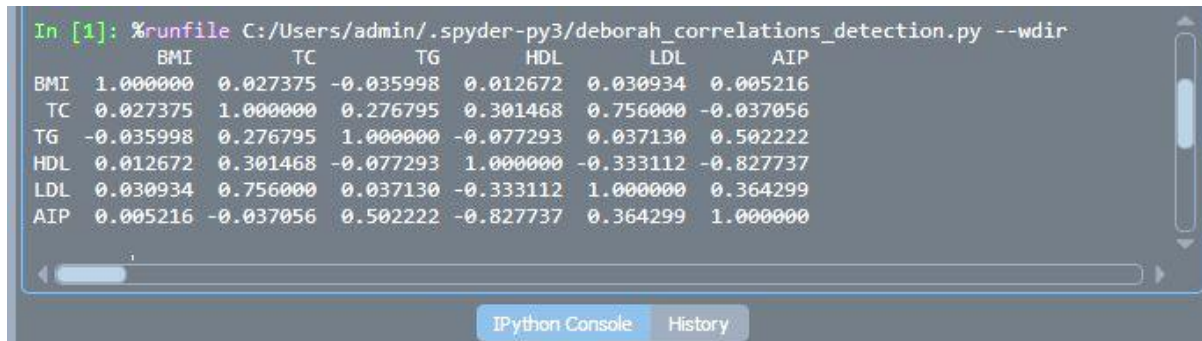
analyses, including scatter plots of expected versus predicted values and error distributions, were generated to visually assess predictive accuracy and bias

Results

Correlation Analysis of Lipid Parameters

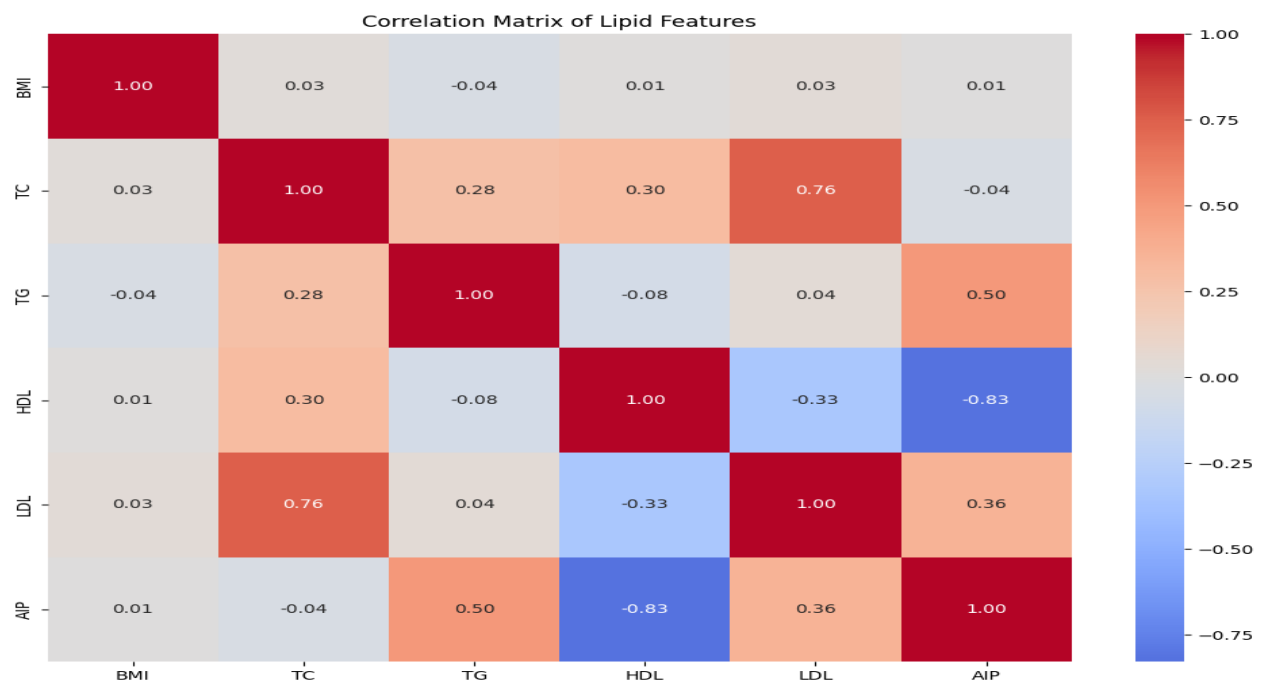
Correlation analysis was performed to examine the relationships between the atherogenic index of plasma (AIP) and individual lipid parameters. Among all variables analyzed, high-density lipoprotein cholesterol (HDL-C) demonstrated the strongest association with AIP, exhibiting a strong negative correlation ($r = -0.8277$). This finding indicates that increasing HDL-C levels are associated with decreasing AIP values. Moderate correlations were observed between AIP and triglycerides, consistent with established clinical understanding of lipid-related cardiovascular risk. Additionally, a strong positive correlation was identified between total cholesterol and low-density lipoprotein cholesterol, reflecting expected lipid profile relationships. Overall, these correlation patterns confirmed HDL-C as the lipid parameter most strongly associated with the atherogenic index of plasma (Figure 1; Figure 2). Figure 3 shows the feature importance of lipid parameters in predicting the atherogenic index of plasma (AIP) derived from machine learning analysis, highlighting HDL-C as the most influential predictor.

162



163

164 Figure 1: Correlation matrix showing the relationships between the atherogenic index of plasma
 165 (AIP) and lipid parameters



166

167 Figure 2: The correlations matrix among the LIPIDS

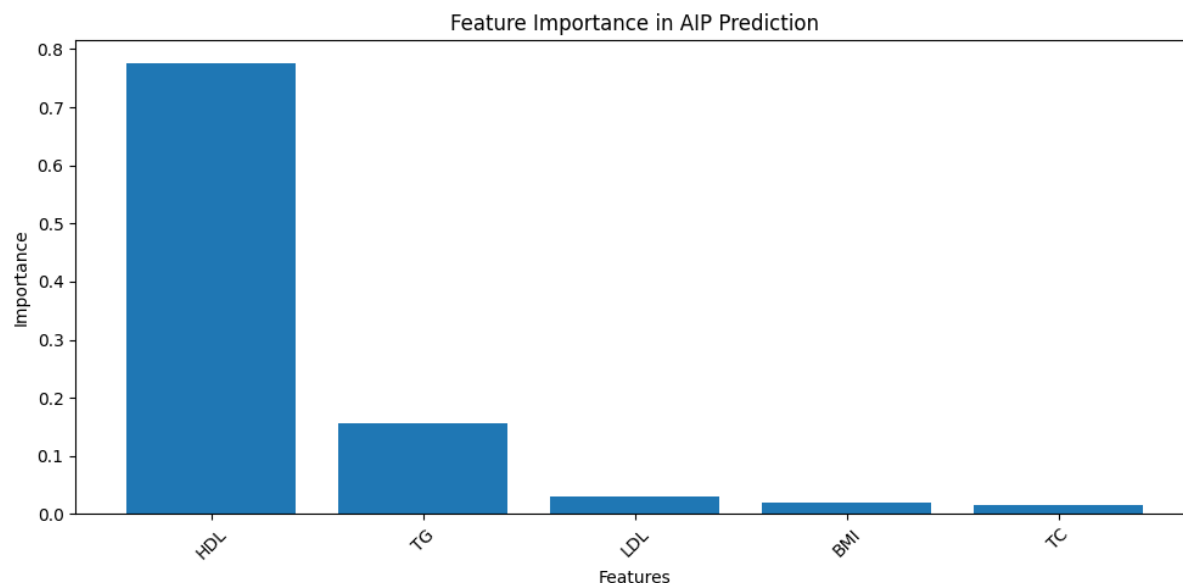
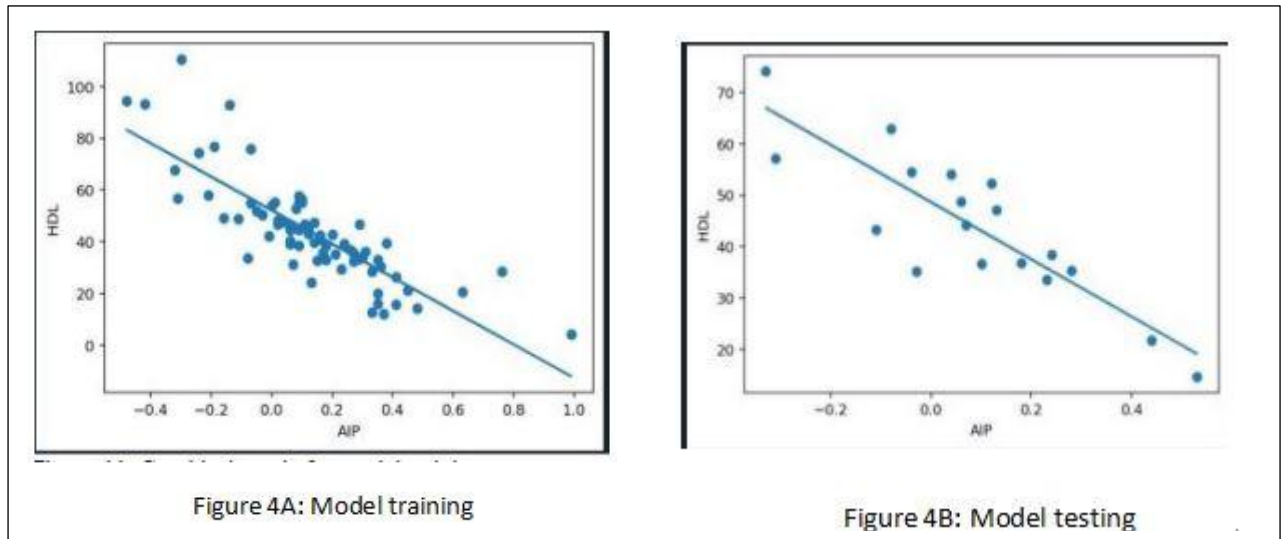


Figure 3 showing the feature importance of lipid parameters in predicting the atherogenic index of plasma (AIP) derived from machine learning analysis, highlighting HDL-C as the most influential predictor

Training and Testing Data Distribution

The model training is shown in Figure 4A while Figure 4B shows the testing. When evaluated on the testing data set, the linear regression model achieved a mean absolute error (MAE) of 0.0851, corresponding to an average prediction deviation of 8.51%. The relationship between the observed and predicted atherogenic index of plasma (AIP) values is illustrated in Figure 5A, showing reasonable agreement along the identity line. The corresponding error distribution (Figure 5B) demonstrates a balanced spread of positive and negative errors, indicating the absence of systematic prediction bias, although reduced accuracy was observed at extreme AIP values. Table 2 shows Regression Model validation result.



After training the Model , the estimation of AIP is shown in equation 2.

$$AIP = \frac{HDL - 52.1565166934}{64.8330629390} \quad \text{..... (equation 2)}$$

Table 2: Regression Model validation result.

EXPECTED	PREDICTED	ERROR	ABSOLUTE
AIP	AIP		ERROR
0.24	0.21	0.03	0.03
0.1	0.24	-0.14	0.14
0.18	0.238	-0.058	0.058
0.28	0.26	0.02	0.02
-0.03	0.263	-0.293	0.293

	0.04	-0.029	0.069	0.069
	0.06	0.053	0.007	0.007
	-0.04	-0.035	-0.005	0.005
	-0.31	-0.167	-0.143	0.143
	-0.08	-0.166	0.086	0.086
TOTAL ABSOLUTE ERROR			0.851	
MEAN ABSOLUTE ERROR			0.0851	
PCNT ERROR			8.51%	

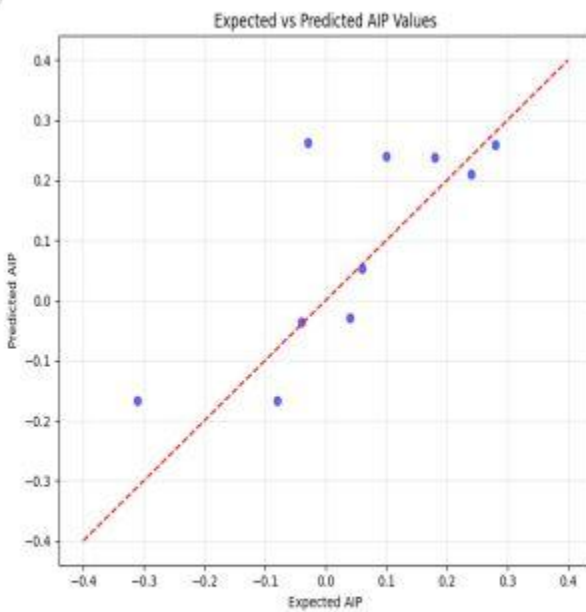


Figure 5A: Regression Expected AIP

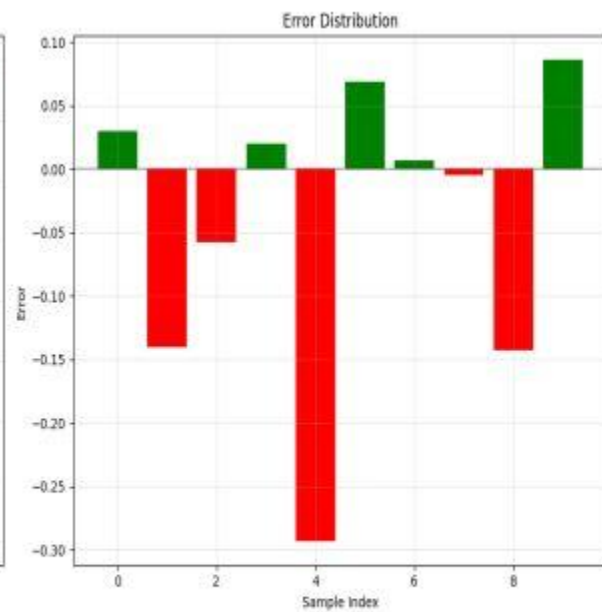


Figure 5B: Error distribution versus predicted AIP

Linear Regression Model Performance

A single-parameter linear regression model was developed using HDL-C as the sole predictor of AIP. Model training revealed a statistically meaningful inverse relationship between HDL-C and AIP, with a regression coefficient of -0.0104 , confirming that higher HDL-C levels correspond to

lower atherogenic risk. The model explained approximately 65% of the variance in AIP values ($R^2 = 0.6504$), indicating a moderately strong predictive relationship.

When evaluated on the testing data set, the linear regression model achieved a mean absolute error (MAE) of 0.0851, corresponding to an average prediction deviation of 8.51%. Scatter plots comparing observed and predicted AIP values demonstrated reasonable alignment along the identity line, although reduced accuracy was observed at extreme negative AIP values. Error distribution analysis revealed a balanced spread of positive and negative errors, suggesting the absence of systematic prediction bias.

Random Forest Regression Model Performance

For comparative analysis, a random forest regression model incorporating multiple lipid parameters was trained and evaluated (Figure 6). The random forest model demonstrated superior predictive performance relative to the linear regression approach, achieving a mean absolute error of 0.040 and explaining 94.3% of the variance in AIP values ($R^2 = 0.943$). Predicted AIP values closely aligned with observed values across the full range of measurements, as illustrated by tight clustering along the identity line in the scatter plot.

Error analysis further confirmed the robustness of the random forest model, with the majority of prediction errors falling within ± 0.1 units (Figure 7). These results indicate that while multi-parameter ensemble learning substantially enhances predictive accuracy, it does so at the cost of reduced interpretability compared with the single-parameter regression model.

For comparative analysis, a random forest regression model incorporating multiple lipid parameters was trained and evaluated. The model demonstrated superior predictive performance, achieving a

mean absolute error of 0.040 and explaining 94.3% of the variance in AIP values ($R^2 = 0.943$). As shown in Figure 4, the observed and predicted AIP values exhibited close alignment along the identity line, indicating strong predictive accuracy across the full range of measurements.

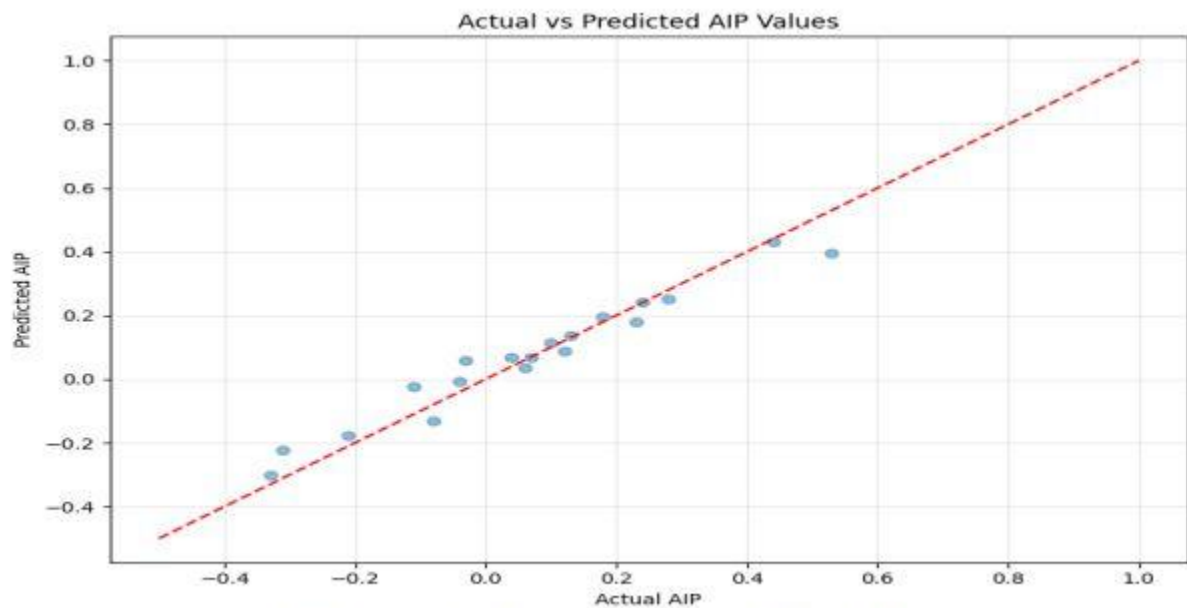


Figure 6: Random Forest Expected versus Predicted AIP values

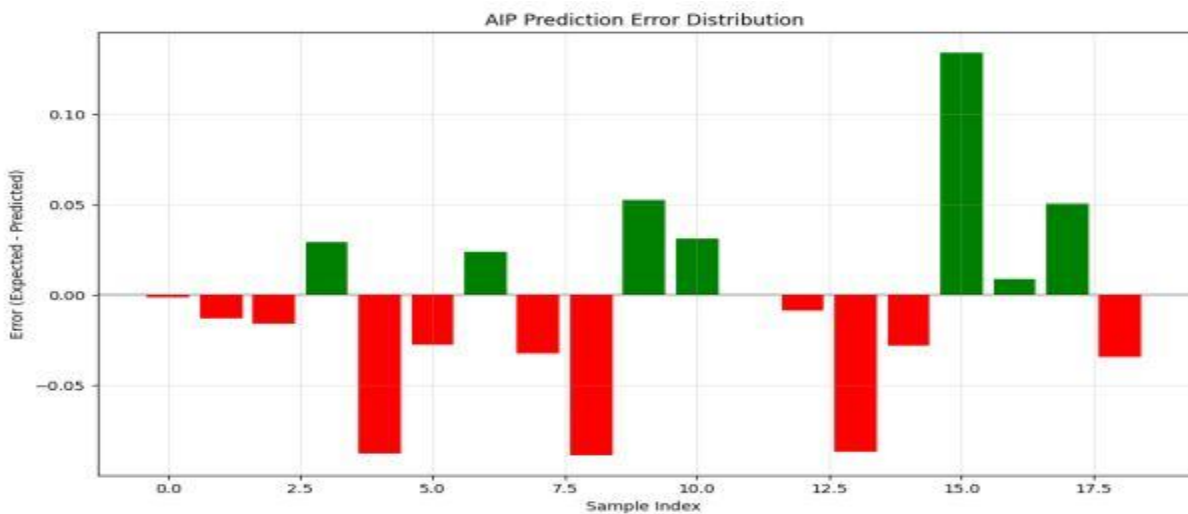


Figure 7: Random Forest Error distribution of expected versus predicted AIP

Discussion

This study investigated the relationship between conventional lipid profile parameters and the atherogenic index of plasma (AIP) using a machine learning–supported approach, with the aim of developing an interpretable model for cardiovascular risk estimation. The findings demonstrate a strong inverse relationship between high-density lipoprotein cholesterol (HDL-C) and AIP, highlighting HDL-C as the most influential single lipid determinant of atherogenic risk within the studied population. The observed negative correlation between HDL-C and AIP is consistent with established atherosclerotic mechanisms and prior reports demonstrating the protective role of HDL-C in cardiovascular health (Dobiášová and Frohlich 2001; Libby 2021).

HDL-C plays a critical role in reverse cholesterol transport and exerts anti-inflammatory and antioxidant effects that mitigate lipid accumulation in arterial walls. Consequently, reduced HDL-C levels are associated with increased atherogenic burden, which is reflected in elevated AIP values. The strength of the correlation observed in this study reinforces the clinical relevance of HDL-C as a key protective lipid parameter.

The single-parameter linear regression model developed using HDL-C alone demonstrated moderate predictive performance, explaining approximately 65% of the variance in AIP values and achieving a mean absolute error of 0.0851. While this performance is lower than that of the multi-parameter random forest model, the regression model offers clear advantages in interpretability and ease of implementation. In routine clinical practice, especially in resource-limited settings, simple and transparent models are often preferable, as they facilitate rapid risk estimation without requiring extensive computational resources or comprehensive lipid measurements.

In contrast, the random forest regression model achieved substantially higher predictive accuracy, accounting for over 94% of the variance in AIP values. This improvement reflects the ability of ensemble machine learning methods to capture complex, non-linear interactions among multiple

lipid parameters, which is consistent with prior studies reporting improved predictive performance of ensemble machine learning models in cardiovascular risk stratification (Huang et al. 2020; Shah et al. 2025). This improvement reflects the ability of ensemble machine learning methods to capture complex, non-linear interactions among multiple lipid parameters. However, such models typically rely on complete input data and are less transparent, which may limit their widespread clinical adoption despite their superior performance.

The comparative results of this study suggest a complementary role for both modeling approaches. The random forest model may serve as a high-accuracy analytical tool in research or well-resourced clinical environments, whereas the HDL-C–based regression model provides a pragmatic alternative for preliminary cardiovascular risk assessment when full lipid panels are unavailable. This dual-model perspective aligns with current efforts to balance predictive performance with clinical usability in cardiovascular risk stratification.

Several limitations should be acknowledged. The study was conducted using data from a single tertiary institution, which may limit generalizability to other populations. Additionally, the sample size, while adequate for exploratory modeling, may not capture the full spectrum of lipid variability. Reduced predictive accuracy at extreme AIP values was observed with the linear regression model, indicating that caution is warranted when applying the model in such cases. Future studies involving larger, multi-center datasets and external validation are necessary to further evaluate the robustness and generalizability of the proposed approach.

Overall, this study contributes to the growing body of literature on machine learning applications in cardiovascular risk assessment by demonstrating that data-driven methods can be effectively leveraged to develop simplified and clinically interpretable models. The findings support the potential utility of HDL-C–based estimation of AIP as a practical tool in preventive cardiology.

Conclusion

This study demonstrated a strong inverse association between high-density lipoprotein cholesterol (HDL-C) and the atherogenic index of plasma (AIP), confirming HDL-C as a key determinant of atherogenic risk. By leveraging machine learning techniques to explore lipid parameter relationships, an interpretable single-parameter regression model was developed to estimate AIP using HDL-C alone. The model exhibited reasonable predictive performance, supporting previous evidence that simplified lipid-based indices remain clinically valuable in cardiovascular risk assessment (Majeed et al. 2019).

Although a multi-parameter random forest model achieved higher predictive accuracy, the HDL-C-based model offers advantages in terms of simplicity, transparency, and clinical applicability. This approach may be particularly useful in settings where complete lipid profiles are unavailable or limited. Future studies incorporating larger, multi-center datasets and external validation are recommended to further assess the generalizability and clinical utility of the proposed method.

List of abbreviations

AIP – Atherogenic Index of Plasma
CAD – Coronary Artery Disease
CVD – Cardiovascular Disease
HDL-C – High-Density Lipoprotein Cholesterol
LDL-C – Low-Density Lipoprotein Cholesterol
TG – Triglycerides
TC – Total Cholesterol
MAE – Mean Absolute Error

References.

1. Ya Li, Feng Y, Feng Y, Li S, Ma Y, Lin J, Wan J, Zhao M. The atherogenic index of plasma (AIP) is a predictor for the severity of coronary artery disease. *Scientific Reports*. 2024; <https://doi.org/https://doi.org/10.3389/fcvm.2023.1140215>

2. Zhao M, Xiao M, Zhang H, Tan Q, Ji J, Cheng Y, Lu F.
Relationship between plasma atherogenic index and incidence of cardiovascular diseases in Chinese middle-aged and elderly people.
Scientific Reports. 2025;15:8775.
<https://doi.org/10.1038/s41598-025-08775-x>
3. Chawla A, et al.
Machine learning in cardiovascular medicine: Current applications and future directions.
Journal of the American College of Cardiology. 2017;70(20):2549–2558.
<https://doi.org/10.1016/j.jacc.2017.09.016>
4. Dobiášová M, Frohlich J.
The atherogenic index of plasma (AIP) as a risk factor for coronary heart disease: A review.
Vascular Health and Risk Management. 2001;17(3):269–278.
5. Fihn SD, Blankenship JC, Alexander KP, et al.
ACC/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the management of patients with stable ischemic heart disease.
Journal of the American College of Cardiology. 2014;63(4):299–346.
<https://doi.org/10.1016/j.jacc.2014.02.561>
6. Huang J, Chen Y, Li J, et al.
Predicting cardiovascular disease risk using machine learning in a population of patients with diabetes.
Diabetes Care. 2020;43(1):173–179.
<https://doi.org/10.2337/dc19-0853>
7. Libby P.
The pathobiology of atherosclerosis: A series of articles.
Journal of the American College of Cardiology. 2021;77(24):3069–3081.
<https://doi.org/10.1016/j.jacc.2021.04.029>
8. Majeed A, et al.
Role of atherogenic index of plasma in predicting cardiovascular risk.
Journal of Public Health Research. 2019;8(1):125–131.
<https://doi.org/10.4081/jphr.2019.1554>
9. Oussaid S, et al.
Machine learning in cardiovascular disease: A review.
Journal of the American Heart Association. 2021;10(3):e021367.
<https://doi.org/10.1161/JAHA.121.021367>
10. World Health Organization.
Global health estimates: Leading causes of death.
Geneva: World Health Organization; 2020.
Available at: <https://www.who.int>
11. Liu Z, Zhang L, Wang L, et al.
The predictive value of cumulative atherogenic index of plasma (AIP) for cardiovascular outcomes: A prospective community-based cohort study.
Cardiovascular Diabetology. 2024;23:264.
<https://doi.org/10.1186/s12933-024-02164-5>
12. Shah P, Shukla M, Dholakia NH, et al.
Predicting cardiovascular risk with hybrid ensemble learning and explainable artificial intelligence.

- 341 *Scientific Reports*. 2025;15:17927.
342 <https://doi.org/10.1038/s41598-025-17927-x>
343 13. Yadegar A, Mohammadi F, Seifouri K, et al.
344 Surrogate markers of insulin resistance and coronary artery disease in type 2 diabetes: U-
345 shaped TyG association and insights from machine learning integration.
346 *Lipids in Health and Disease*. 2025;24:96.
347 <https://doi.org/10.1186/s12944-025-01996-9>

349 **Declarations:**

350 **Ethics Approval and Consent to Participate**

351 Ethical approval for this study was obtained from the Ethics Review Committee of the University of
352 Medical Sciences Teaching Hospital, Ondo City, Nigeria. The study utilized anonymized
353 retrospective clinical data, and all procedures were conducted in accordance with the ethical standards
354 of the institutional review committee and the principles of the Declaration of Helsinki. The
355 requirement for informed consent was waived due to the retrospective nature of the study and the use
356 of de-identified patient records.

357 **Clinical trial number:** Not Applicable

358 **Consent for publication:** Not Applicable

359 **Availability of Data and Materials**

360 The datasets used and/or analyzed during the current study are available from the corresponding
361 author on reasonable request, subject to institutional and ethical approvals.

362 **Competing Interests**

363 The authors have no competing interests.

364 **Funding**

365 This research received no specific grant from any funding agency in the public, commercial, or not-
366 for-profit sectors.

367 **Authors' Contributions**

368 R. F.F. conceived and designed the study.
369 M.A.A. contributed to data analysis, machine learning implementation, and manuscript drafting.
370 P.A.A. assisted with methodology development and critical revision of the manuscript.
371 D.A. contributed to data acquisition and clinical interpretation.

372 All authors read and approved the final manuscript.

373 **Acknowledgements**

374 The authors acknowledge the management and staff of the University of Medical Sciences Teaching
375 Hospital, Ondo City, Nigeria, for providing access to the clinical data used in this study.

376