Official Journal of Erciyes University Faculty of Medicine

DOI: 10.14744/cpr.2023.09719
J Clin Pract Res 2023;45(6):632–9

# To Determine LDL Phenotypes Using Lipids, Lipoproteins, Apoproteins, and sdLDL Through Association Rule Mining

Mehtap Atak,[1] Mehmet Kıvrak,[2] Hatice Sevim Nalkıran,[3] Hüseyin Avni Uydu,[4] Ömer Şatıroğlu[5]

[1]Department of Biochemistry, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye
[2]Department of Biostatistics and Medical Informatics, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye
[3]Department of Medical Biology, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye
[4]Department of Medical Biochemistry, Samsun University Faculty of Medicine, Samsun, Türkiye
[5]Department of Cardiology, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye

**Cite this article as:**
Atak M, Kıvrak M, Nalkıran HS, Uydu HA, Şatıroğlu Ö. To Determine LDL Phenotypes Using Lipids, Lipoproteins, Apoproteins, and sdLDL Through Association Rule Mining. J Clin Pract Res 2023; 45(6): 632–9.

**Address for correspondence:**
Mehtap Atak.
Department of Biochemistry, Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye
**Phone:** +90 464 212 30 09 - 3319
**E-mail:**
mehtap.atak@erdogan.edu.tr

**Submitted:** 14.08.2023
**Revised:** 01.11.2023
**Accepted:** 06.12.2023
**Available Online:** 13.12.2023

Erciyes University Faculty of Medicine Publications - Available online at www.jcpres.com



## ABSTRACT

**Objective:** The atherogenic lipoprotein phenotype is closely associated with the risk assessment of Coronary Artery Disease (CAD) and the monitoring of treatment processes. Particularly, high levels of small dense low-density lipoprotein (sdLDL) and low levels of large buoyant low-density lipoprotein (lbLDL) are critical in determining Pattern B. This study aims to determine the lipid phenotype using the Association Rule Mining (ARM) method, based on concentrations of lipids, lipoproteins, apoproteins, and sdLDL.

**Materials and Methods:** This retrospective case-control study utilized analytical research methods. Numerical variables were expressed as mean, standard deviation, median, and min-max values. Statistically significant differences were observed between the low-density lipoprotein (LDL) size categories in terms of triglycerides (TG), LDL, high-density lipoprotein (HDL), apolipoprotein B (ApoB), apolipoprotein E (ApoE), sdLDL, and lbLDL distributions. ARM was employed to detect the lipoprotein phenotype.

**Results:** Statistically significant differences were found between the LDL size categories in distributions of TG, LDL, HDL, ApoB, ApoE, sdLDL, and lbLDL ($p_{TG}$<0.001, $p_{LDL}$=0.03, $p_{HDL}$<0.001, $p_{ApoB}$=0.016, $p_{ApoE}$=0.004, $p_{sdLDL}$<0.001, and $p_{lbLDL}$<0.001). The ARM method revealed that the probability of phenotype B is 100% for sdLDL values in the range of 15.5–109 and lbLDL values in the range of 0–31.5.

**Conclusion:** This study introduces a contemporary approach for detecting lipoprotein phenotypes using ARM, further substantiating the strong correlation between atherogenic phenotypes and sdLDL.

**Keywords:** Lipoproteins, ldl phenotype, coronary artery disease, association rule mining.

## INTRODUCTION

Cardiovascular diseases (CVDs) are a significant cause of death globally. The World Health Organization stated that an estimated 17.9 million people died from CVDs in 2019, representing

32% of all global deaths.[1] The risk of CVD is closely associated with changes in the lipoprotein profile in blood plasma.[2] High levels of low-density lipoprotein cholesterol (LDL-C) in plasma are one of the major risk factors for the development of CVD. However, not only plasma LDL-C levels but also particle properties are of great importance in predicting disease risk.[3] Low-density lipoprotein (LDL) particles are divided into subclasses with different diameters and densities, physicochemical compositions, metabolic behaviors, and atherogenic potentials. These subclasses of LDL are determined by various analytical methods, such as density gradient ultracentrifugation, nuclear magnetic resonance, gradient gel electrophoresis, and polyacrylamide gel electrophoresis.[4] LDL consists of two subclasses differing in size and density: large and medium-sized LDL (lbLDL, pattern A) and smaller-sized LDL (sdLDL, pattern B).[5] SdLDL is strongly associated with CVD risk due to its easier penetration into arterial tissue, more efficient passage through the subendothelium, low affinity for the LDL receptor, and high oxidative sensitivity.[6] Additionally, the diameter of LDL particles is an important parameter in determining their atherogenic tendency, and is critical for predicting cardiovascular events. The LDL phenotype in plasma is typically determined by gradient gel electrophoresis. If the LDL particle diameter is 258-263 Å or more, it is classified as 'Type A'; below this threshold, it is classified as 'Type B'.[4] The 'Type A' phenotype, or pattern A, is usually characterized by large buoyant LDL (lbLDL) particles, while 'Type B' phenotype, or pattern B, is characterized by small density LDL (sdLDL) particles.[5]

Data mining can be summarized as a collection of methods used to extract information from data, focusing on the process of discovering previously unknown, hidden patterns. It involves transforming data into qualified information by utilizing statistical analysis methods and artificial intelligence algorithms.[7] Data Mining Methods are instrumental in diagnosing various diseases such as cardiovascular diseases, diabetes, obesity, and cancer. These methods support the process through Machine Learning, developing solutions with self-learning methods.[8]

Association Rule Mining (ARM) aims to extract statistically significant relational patterns from large databases.[9] Increasingly used in medical literature, ARM has become a key method for identifying factors associated with diseases.[10] The Apriori algorithm is the most commonly used in the ARM model, but the Frequent Pattern Growth (FP-Growth) and Eclat Algorithms are also employed. Support and confidence are two crucial measures in creating strong, meaningful rules.[11] The size of the LDL diameter, an important parameter indicating the atherogenic tendencies of lipoproteins, is also significant for predicting cardiovascular events.

**Table 1.** Multiple normality test

| Test | HZ | p | Multiple normality test |
|---|---|---|---|
| Henze-Zirkler | 2.2659 | <0.001 | No |

As shown in Table 1, the data were found to be not multi-normally distributed at the 95% confidence interval; p<0.001.

In this study, we analyzed the concentrations of lipids, lipoproteins, apoproteins, and sdLDL using ARM to determine the LDL phenotype.

## MATERIALS AND METHODS

This is an observational study within the framework of quantitative research. We employed a case-control analytical research method for this retrospective study. All ARM inferences were generated using the 'arules' package in the R programming language.[12] RStudio version 1.1.456 was used for the analysis.[13] The Statistical Package for the Social Sciences (SPSS) 22.0 Package Program was utilized for statistical analysis.

### Dataset

The acquisition of the dataset comprises three steps. The first stage involves collecting serum samples, the second stage pertains to the results obtained through biochemical and clinical analyses, and the third stage focuses on the outputs obtained from Low-Density Lipoprotein (LDL) subfraction analysis. This research adhered to the Declaration of Helsinki and received approval from the Non-Interventional Clinical Research Ethics Committee (Recep Tayyip Erdoğan University Faculty of Medicine, Rize, Türkiye; Decision Number: 2017/163; Date: 27.10.2017). Each participant provided written, informed consent prior to registration. The study included 516 patients who visited the Recep Tayyip Erdoğan University (RTEU) Medical Faculty Teaching & Research Hospital Cardiology Outpatient Clinic, were diagnosed with Coronary Artery Disease (CAD), and met the criteria for elective conventional coronary angiography. Subsequently, fasting blood samples were collected in pre-cooled Ethylenediaminetetraacetic acid (EDTA) tubes at baseline from each patient. After centrifugation at 3,500 rpm for 20 minutes at 4 °C, all fresh plasma aliquots were immediately analyzed in the clinical biochemistry laboratory. In this study, triglyceride (TG) and total cholesterol (TC) concentrations were measured using an autoanalyzer (Abbott Architect C16000). High-Density Lipoprotein Cholesterol (HDL-C) levels were determined by the dextran sulfate-$Mg^{+2}$ precipitation method. Total LDL-C levels were calculated using Friedewald's formula [LDL-C=TC-(TG/5+HDL-C)]. Apolipoprotein (ApoA, ApoB, ApoE) concentrations were determined using the nephelometric method (Siemens BN 2). Finally, LDL subfractions were measured using the Lipoprint system according to the manufactur-

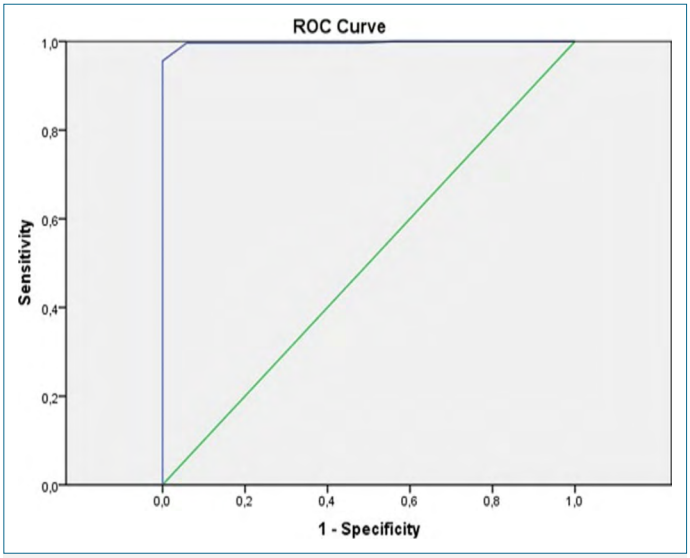


**Figure 1.** The cut-off point of LDL particles: The area under the ROC curve was 0.997 with a standard error of 0.002 (p<0.001). The cut-off value is the one that maximizes the sensitivity/(1-specificity) ratio).

er's instructions (Quantimetrix Inc., Redondo Beach, California). This method separates plasma lipoproteins in a non-denaturing gel gradient polyacrylamide. In the Lipoprint system, lipoproteins are separated based on their varying net surface charges and the size of the LDL particle. Seven types of LDL are identified by this method, classified as lb_LDL and sd_LDL based on LDL particle size. Additionally, LDL particles are divided into 'large' and 'small' types; the cut-off point for classification was a particle size of 264 Å as determined by Receiver Operating Characteristic (ROC) curve analysis. The area under the ROC curve was found to be 0.997. A particle diameter larger than 264 Å is defined as phenotype A, while a diameter smaller than 264 Å is defined as phenotype B. Individuals with phenotype B are considered to be in a higher risk group for CAD.[14] The cut-off point for LDL particles is presented in Figure 1.

In our dataset, the LDL size variable is used as the output variable. The input variables include TG, TC, LDL, HDL, ApoA, ApoB, ApoE, sd_LDL, and lb_LDL. Association Mining is carried out in two steps: Frequent Itemset (FI) discovery and Association Rule (AR) generation.[15] A randomly selected sample may contain statistical errors, particularly in support and confidence calculations. The error is calculated either at the FI discovery step or the AR generation step and is compared with the corresponding value of the universe, either absolutely or relatively. The classification of sampling size estimation techniques depends on the type of error and the step where the error emerges.[16] The formula for absolute frequent itemset ($fI_{abs}$) is given as:

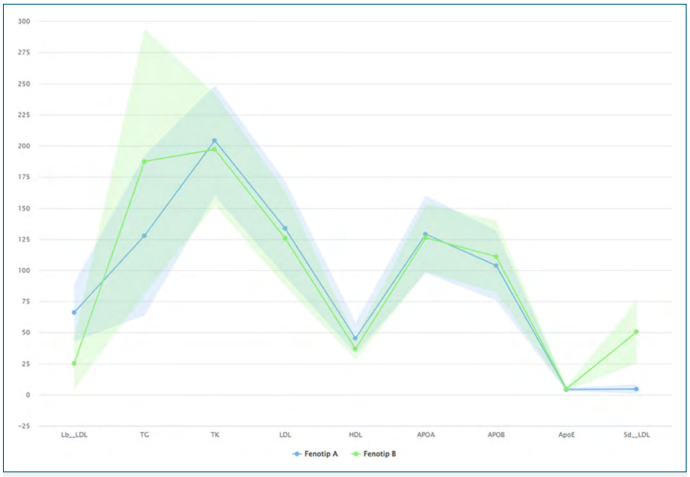


**Figure 2.** Using the specified formula, the minimum sample size in association rule mining was calculated as 395.

$$fI_{abs} = \frac{-2\ln(1-\gamma)}{\vartheta\delta^2} \quad (1)$$

Where:

γ: minimum confidence of AR,

ϑ: minimum support of FI,

δ: failure probability in FI discovery/AR generation step.[17]

Using this formula, the minimum sample size required for Association Rule Mining was calculated to be 395. The classification performance is depicted in a deviation plot in Figure 2.

## Data Preprocessing

Initially, Henze-Zirkler's Multiple Normality Test was conducted. The data were found not to be multi-normally distributed at the 95% confidence interval (p<0.001) (Table 1). This finding aligns with Henze and Zirkler's[18] work on multivariate normal distribution, as shown in the following setting:

$$Hz = \frac{1}{\sqrt{2}}\left[\frac{2d+1)n}{4}\right]^{1/(d+4)} \quad (2)$$

Missing values were imputed in the dataset using the Random Forest Method. Additionally, 31 outliers with a random distribution were identified and removed from the dataset, as they were determined to have no effect on the analysis results (Fig. 3).

## Association Rule Mining

ARM aims to uncover meaningful relationship structures by generating rules from pattern structures within a specific dataset. Support and confidence are two important statistical criteria for interpreting these ratios. Originally utilized in

**Table 2.** Descriptive statistics and significance between groups

| Variable [Mean±SD/Median (Min–Max] | Phenotype A (n=283) | Phenotype B (n=163) | U statistics | p |
|---|---|---|---|---|
| TG | 127.8±64.1/112 (37–423) | 187.4±106.5/156 (43–674) | -6.5956 | <**0.001** |
| TC | 204.3±44.3/203 (109–376) | 197.1±44.7/192 (98–322) | 1.8195 | 0.069 |
| LDL | 133.6±38.1/132 (62–301) | 125.7±38.01/122.7 (54.9–236) | 2.1719 | **0.030** |
| HDL | 45.3±12.5/43.8 (18–89) | 36.9±8.7/36 (20.3–65.8) | 7.2168 | <**0.001** |
| ApoA | 129.1±30.9/126 (72.9–281) | 126.1±27.4/123 (71.1–203) | 0.6968 | 0.486 |
| ApoB | 103.8±28.1/104 (35–188) | 111.1±28.8/108.5 (49.3–223) | -2.4128 | **0.016** |
| ApoE | 4.2±1.3/4.1 (1.6–14.1) | 4.75±1.6/4.35 (2.1–11.8) | -2.8618 | **0.004** |
| sdLDL | 4.5±3.6/4 (0–23) | 50.6±25.3/51 (6–109) | -17.115 | <**0.001** |
| lbLDL | 66.2±23.3/65 (12–155) | 25.1±20.5/19 (0–121) | 14.115 | <**0.001** |

Numerical variables are expressed as mean, standard deviation, median, and minimum–maximum values. Due to the dataset's inability to satisfy multiple normality assumptions, the Mann-Whitney U test, a non-parametric test, was used for comparing groups in quantitative data. The Type 1 error (α) was set at 0.05. ApoA: Apolipoprotein A; ApoB: Apolipoprotein B; ApoE: Apolipoprotein E; HDL: High density lipoprotein; lb_LDL: Large-buoyant low density lipoprotein; SdLDL: Small-density low density lipoprotein; TC: Total cholesterol; TG: Triglyceride.

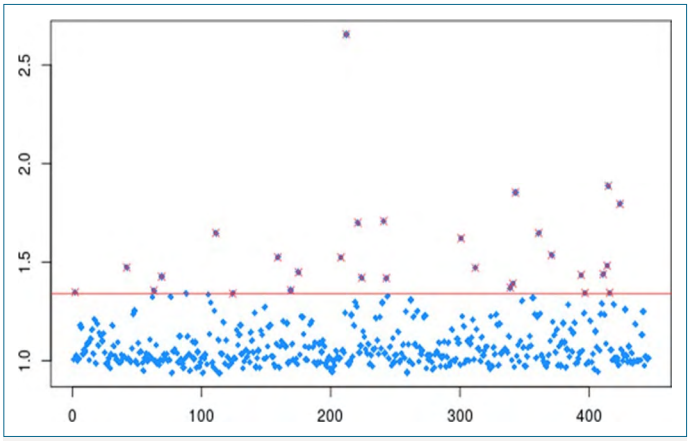


**Figure 3.** Outlier distribution: The x-axis represents the local outlier factor, while the y-axis shows the observation points. A total of 31 rows containing outliers or excessive values were identified in the dataset and subsequently removed from the dataset based on random selection.

marketing, ARM is now employed in medicine, particularly for identifying disease risk factors. The general formulation in ARM is X ⇒ Y, where X represents the rule's premise and Y its conclusion.[10] The most commonly used method in ARM is the Apriori Algorithm.

### Apriori Algorithm

The Apriori Algorithm is a method used to identify frequently recurring significant items from a given database through multiple iterative scanning operations. By establishing minimum support and confidence values, items below the minimum support value are disregarded. Subsequently, triple association rules are formulated, and those falling below the minimum support value are excluded from these rules.[19]

### Support

In a dataset (XY), the support value is expressed as the percentage of rows that include both X and Y values. It is represented as:

Support (X⇒Y) = P(XUY).

### Confidence

Another essential measure for association rules is confidence. Confidence indicates the degree of association discovered. P(Y|X) represents the probability of including Y in a transaction that also includes X, and it is defined as follows:

Confidence (X⇒Y) = P (Y|X).[20]

Strong rules are established when support and confidence values are close to 1. The hyperparameter values for support and confidence used in the formation of the rules were determined to be 0.2 and 0.1, respectively.[11]

## RESULTS

### Statistical Analysis

Numerical variables were expressed as mean, standard deviation, median, and min-max values. As the dataset did not meet the assumptions of multivariate normality, the Mann-Whitney U test, a non-parametric test, was used for group comparisons in quantitative data. The sums of ranks in the Mann-Whitney test are denoted by $R_1$ and $R_2$. The statistic U is computed as the smaller of $U_1$ and $U_2$, where:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

The significance of the differences in distributions is determined from the values of n1, n2, and U using tables. The choice of tables depends on the values of n1 and n2. Two cases are considered.[21] Type 1 error (α) was set at 0.05. There is a statistically significant difference between the LDL size categories in terms of TG, LDL, HDL, ApoB, ApoE, sdLDL, and lbLDL distributions. However, no statistically significant difference was observed between the LDL size categories in ApoA and TC distributions. TG values were higher in the Phenotype B category (156) than in the Phenotype A category (112) (p<0.001). LDL values were higher in the Phenotype A category (132) than in the Phenotype B category (122.7) (p=0.030).

HDL levels were higher in the Phenotype A category (43.8) than in the Phenotype B category (36) (p<0.001). There was no statistically significant difference in ApoA distribution between LDL size categories (p=0.486), nor in TC (p=0.069). ApoB values were higher in the Phenotype B category (108.5) than in the Phenotype A category (104) (p=0.016). ApoE values were higher in the Phenotype B category (4.35) than in the Phenotype A category (4.1) (p=0.004). SdLDL values were higher in the Phenotype B category (51) compared to the Phenotype A category (4) (p<0.001). LbLDL values were higher in the Phenotype A category (65) than in the Phenotype B category (19) (p<0.001). Descriptive statistics and statistical significance between groups are presented in Table 2.

The continuous variables (TG, TC, LDL, HDL, ApoA, ApoB, ApoE, sdLDL, and lbLDL) were transformed into categorical variables for generating association rules to detect lipoprotein phenotypes using ARM.

Considering the support and confidence criteria (0.2/0.5), 13 rules were created in the analysis. The resulting rules are provided in Table 3 and Figure 4. Generally, in relational classification methods, the rule with the highest confidence is used for classification. In this context, three examples of rule from Table 3 are given below.

**Rule 1:** Individuals with sdLDL values between 15.5 and 109, and lbLDL values between 0 and 31.5, are 100% Phenotype B.

**Rule 4:** Individuals with an HDL of 18.37 and an sdLDL between 15.5 and 109 are 100% Phenotype B.

**Rule 11:** Individuals with an HDL value of 37.89, ApoE between 1.6 and 6.64, sdLDL between 0 and 15.5, and lbLDL between 31.5 and 155 are 100% Phenotype A.

The interpretation of other rules is provided in Table 3.

## DISCUSSION

In this study, ARM was employed to determine the impact of serum lipid, lipoprotein, and LDL subfractions on the formation of atherogenic (Pattern B) and anti-atherogenic (Pattern A) LDL phenotypes in CAD. Additionally, individuals were categorized based on Patterns A and B, and lipid parameters were analyzed within these groups. It was observed that serum lipid levels (TG, TC, and LDL-C) were higher in individuals with Pattern B. Furthermore, sdLDL levels were found to be lower in the Pattern A group compared to the Pattern B group, while lbLDL levels were higher in the Pattern A group compared to the Pattern B group.

Previous literature has indicated that serum sdLDL is associated with CVD risk, independent of LDL-C concentration. In this study, no difference in LDL-C concentration was observed between the groups. However, the high concentration of TG and sdLDL, combined with low levels of HDL-C, are considered significant markers of dyslipidemia and are characteristic of the atherogenic lipoprotein phenotype (Pattern B). Additionally, the Pattern B exhibited high levels of TG and sdLDL, along with low HDL-C levels. These findings align with the existing literature.[22]

SdLDL formation is closely associated with high serum TG levels (>120 mg/dL), low HDL-C levels, and increased hepatic lipase enzyme activity.[23] TGs and cholesterol esters are exchanged between lipoprotein particles via the cholesteryl ester transfer protein (CETP).[24] When serum TG levels are high and HDL levels are low, TGs in the Very Low-Density Lipoprotein (VLDL) structure are transferred to LDL, instead of HDL, via CETP. The newly formed LDL particle becomes rich in TGs. Following the hydrolysis of TGs in this particle by hepatic lipase, the residue is called sdLDL.[25,26] Our findings showed that TG levels were higher and HDL levels were lower in the Pattern B group, leading to a more dominant sdLDL particle in the plasma, which is consistent with previous research.[27]

LDL particle diameter is a key factor in calculating sdLDL concentration. Those with an LDL particle diameter greater than 264 Å are classified as lbLDL (Pattern A), and those with smaller diameters as sdLDL (Pattern B).[14] LDL particle diameter is critical in determining the pattern.[28,29] However, Munusuru et al.[30] stated that, in addition to particle diameter, factors such as VLDL, LDL, LDL subfractions (sdLDL, lbLDL), and HDL subfractions (HDL2, HDL3) play significant roles in determining the atherogenic lipoprotein phenotype. Similarly, in our study, concentrations of lipids, lipoproteins, apoproteins, and sdLDL were used to determine the LDL phenotype using ARM. It is essential to note that for identifying Pattern B, it is not only necessary to have high sdLDL levels but also to have low lbDL levels. Serum apoprotein (ApoA, ApoB, ApoE) concentrations within certain limits are also effective in determining the pattern. Likewise, in defining Pattern A, it was determined that, in addition to lbLDL, ApoA, ApoE, HDL, and sdLDL should be within specific ranges.

**Table 3.** Generated association rules

| Rule no | Association rules (X⇒Y) | Support | Confidence | Count |
|---|---|---|---|---|
| 1 | sdLDL=(15.5–109), lbLDL=(0–31.5) LDL_Size=Phenotype B | 0.289 | 1 | 105 |
| 4 | HDL=(18–37), sdLDL=(15.5–109) LDL_Size=Phenotype B | 0.207 | 1 | 75 |
| 11 | TG=(40–232), TC=(200–376), sdLDL=(0–15.5) LDL_Size=Phenotype A | 0.298 | 0.982 | 108 |

HDL: High density lipoprotein; lb_LDL: Large-buoyant low density lipoprotein; SdLDL: Small-density low density lipoprotein; TC: Total cholesterol; TG: Triglyceride.

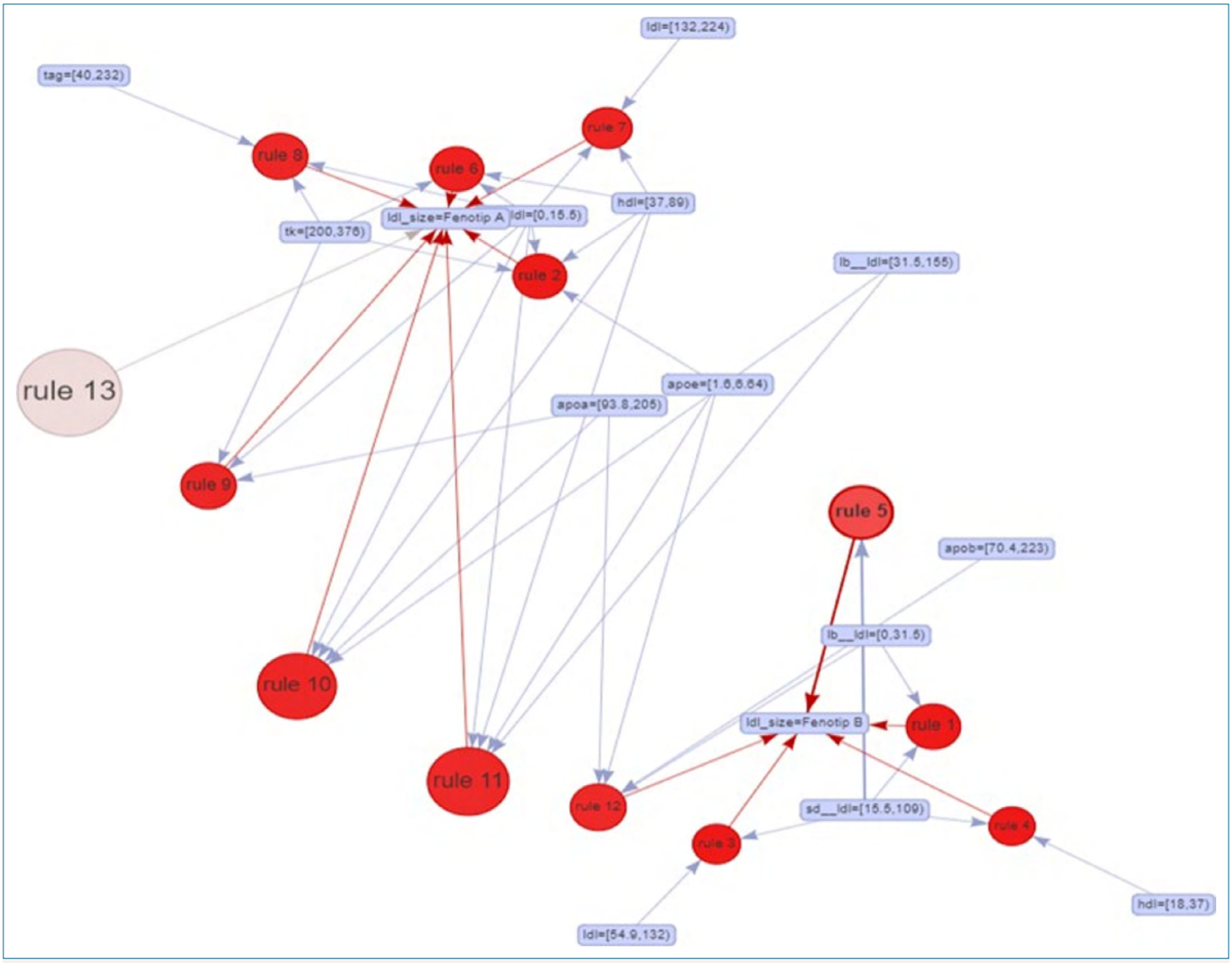


**Figure 4.** Generated association rules: LDL size was set as the response output variable, and Phenotype B was identified as the positive predictor class. These results were obtained using the Ameva discretization method, incorporating predictive variables with confidence values of 0.2 for support and 0.5. for classification. The classification-based associations rules method was utilized in the algorithm.

## CONCLUSION

In conclusion, the atherogenic lipoprotein phenotype is closely associated with the determination of CAD risk and the monitoring of treatment. This study presents a novel approach to defining lipoprotein phenotypes using ARM and corroborates prior research on the impact of sdLDL on lipoprotein phenotypes.

**Peer-review:** Externally peer-reviewed.

**Ethics Committee Approval:** The Recep Tayyip Erdoğan University Non-Interventional Clinical Research Ethics Committee granted approval for this study (date: 27.10.2017, number: 2017/163).

**Informed Consent:** Written informed consent was obtained from patients who participated in this study.

**Author Contributions:** Concept – MA, MK; Design – MA, MK; Supervision – MA, HAU; Resource – HAU, ÖŞ, MA; Materials – ÖŞ; Data Collection and/or Processing – ME, ÖŞ; Analysis and/or Interpretation – MA, ÖŞ, HAU; Literature Search – MA, MK, HSN; Writing – HSN, MA, MK; Critical Reviews – HSN, HAU, ÖŞ.

**Conflict of Interest:** The authors have no conflict of interest to declare.

**Financial Disclosure:** The authors declared that this study has received no financial support.

## REFERENCES

1. WHO. Cardiovascular diseases (CVDs), 2021. Available from: URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed 11 Dec, 2023.
2. Ivanova EA, Myasoedova VA, Melnichenko AA, Grechko AV, Orekhov AN. Small dense low-density lipoprotein as biomarker for atherosclerotic diseases. Oxid Med Cell Longev 2017; 2017: 1273042. [CrossRef]
3. Vekic J, Zeljkovic A, Jelic-Ivanovic Z, Spasojevic-Kalimanovska V, Bogavac-Stanojevic N, Memon L, et al. Small, dense LDL cholesterol and apolipoprotein B: relationship with serum lipids and LDL size. Atherosclerosis 2009; 207(2): 496–501. [CrossRef]
4. Rizzo M, Berneis K. Should we measure routinely the LDL peak particle size? Int J Cardiol 2006; 107(2): 166–70. [CrossRef]
5. Diffenderfer MR, Schaefer EJ. The composition and metabolism of large and small LDL. Curr Opin Lipidol 2014; 25(3): 221–6. [CrossRef]
6. Uydu HA, Bostan M, Atak M, Yılmaz A, Demir A, Akçan B, et al. Cholesterol forms and traditional lipid profile for projection of atherogenic dyslipidemia: lipoprotein subfractions and erythrocyte membrane cholesterol. J Membr Biol 2014; 247(2): 127–34. [CrossRef]
7. Arslan AK, Colak C, Sarihan ME. Different medical data mining approaches based prediction of ischemic stroke. Comput Methods Programs Biomed 2016; 130: 87–92.130.
8. Dogan A, Birant D. Machine learning and data mining in manufacturing. Expert Systems with Applications 2021; 166(2): 114060. [CrossRef]
9. Han J, Kamber M, Pie J. Data mining: concepts and techniques. 3rd edition. Elsevier;2012.
10. Zhang WJ, Ma DL, Dong B. The automatic diagnosis system of breast cancer based on the improved Apriori algorithm. 2012 International Conference on Machine Learning and Cybernetics. IEEE; 2012. pp.63–6.
11. Akbaş KE, Kivrak M, Arslan AK, Çolak C. Assessment of association rules based on certainty factor: An application on heart data set, 2019 International artificial intelligence and data processing symposium (IDAP). IEEE; 2019. pp. 1–5. [CrossRef]
12. Hahsler M, Grün B, Hornik K. Arules- A computational environment for mining association rules and frequent item sets. J Statistical Software 2005; 14(15): 1–25. [CrossRef]
13. Campbell M. RStudio Projects, Learn RStudio IDE. Springer; 2019.pp. 39–48. [CrossRef]
14. Vega GL, Ma PTS, Cater NB. Effects of adding fenofibrate (200 mg/day) to simvastatin (10 mg/day) in patients with combined hyperlipidemia and metabolic syndrome. Am J Cardiol 2003; 91(8): 956–60. [CrossRef]
15. Agrawal R, Imieliński T, Swami, A. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data 1993; pp. 207–16. [CrossRef]
16. Riondato M, Upfal E. Efficient discovery of association rules and frequent itemsets through sampling with tight performance guarantees. ACM Transactions on Knowledge Discovery from Data (TKDD) 2014; 8(4): 1–32. [CrossRef]
17. Halıcı T, Ketenci UG. Comparison of sampling size estimation techniques for association rule mining. In: 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K). IEEE 2015; 1:195–202. [CrossRef]
18. Henze N, Zirkler B. A class of invariant and consistent tests for multivariate normality. Comm Statist Theory Methods 1990; 19(10): 3595–617. [CrossRef]
19. Rao S, Gupta P. Implementing improved algorithm over apriori data mining association rule algorithm. Available from: URL: https://www.ijcst.com/vol31/3/sanjeev.pdf. Accessed 11 Dec, 2023.
20. Han J, Pei J, Kamber M. Data mining: concepts and techniques. 3rd. Edition. Elsevier, 2011.

21. Tallarida R J, Murray RB. Mann-whitney test. Manual of pharmacologic calculations: with computer programs. Springer; 1987.p. 149–53. [CrossRef]
22. Duran EK, Aday AW, Cook NR. Triglyceride-rich lipoprotein cholesterol, small dense LDL cholesterol, and incident cardiovascular disease. J Am Coll Cardiol 2020; 75(17): 2122–35. [CrossRef]
23. Feingold KR, Grunfeld C. The effect of inflammation and infection on lipids and lipoproteins. Biology Med; 2022.
24. Oestereich F, Yousefpour N, Yang E, Phénix J, Nezhad ZS, Nitu A, et al. The cholesteryl ester transfer protein (CETP) raises cholesterol levels in the brain. J Lipid Res 2022; 63(9): 100260. [CrossRef]
25. Krauss RM, Wojnooski K, Orr J, Geaney JC, Pinto CA, Liu Y, et al. Changes in lipoprotein subfraction concentration and composition in healthy individuals treated with the CETP inhibitor anacetrapib. J Lipid Res 2012; 53(3): 540–7. [CrossRef]
26. Brousseau ME, Schaefer EJ, Wolfe ML, Bloedon LT, Digenio AG, Clark RW, et al. Effects of an inhibitor of cholesteryl ester transfer protein on HDL cholesterol. N Engl J Med 2004; 350(15): 1505–15. [CrossRef]
27. Hanak V, Munoz J, Teague J, Stanley A.Jr, Bittner V. Accuracy of the triglyceride to high-density lipoprotein cholesterol ratio for prediction of the low-density lipoprotein phenotype B. Am J Cardiol 2004; 94(2): 219–22. [CrossRef]
28. Krauss RM, Dreon DM. Low-density-lipoprotein subclasses and response to a low-fat diet in healthy men. Am J Clin Nutr 1995; 62(2): 478S–87S. [CrossRef]
29. Hirano T, Ito Y, Saegusa H, Yoshino G. A novel and simple method for quantification of small, dense LDL. J Lipid Res 2003; 44(11): 2193–201. [CrossRef]
30. Musunuru K, Orho-Melander M, Caulfield MP, Li S, Salameh WA, Reitz RE, et al. Ion mobility analysis of lipoprotein subfractions identifies three independent axes of cardiovascular risk. Arterioscler Thromb Vasc Biol 2009; 29(11): 1975–80.