

## Research Paper

## Application of artificial neural networks to accurately predict non-HDL-C levels based on LDL-C and triglycerides

Anna Badura<sup>1</sup> , Urszula Marzec-Wróblewska<sup>2</sup> 

Department of Biopharmacy, Faculty of Pharmacy, Ludwik Rydygier Collegium Medicum in Bydgoszcz,  
Nicolaus Copernicus University in Toruń, Poland

## ARTICLE INFO

## Article history

Received: October 13, 2025

Accepted: February 26, 2026

Available online: May 25, 2026

## Keywords

Non-HDL cholesterol

LDL-C

Triglycerides

Friedewald formula

Artificial neural networks (ANN)

Cardiovascular risk assessment

## Doi

<https://doi.org/10.31648/paom/218622>

## User license

This work is licensed under  
a [Creative Commons Attribution –  
NonCommercial – NoDerivatives  
4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



## ABSTRACT

**Introduction:** While LDL-C concentration remains a key indicator in lipid profile assessment and cardiovascular risk prediction, non-HDL cholesterol (non-HDL-C) has been acknowledged as a superior marker, particularly among people with elevated levels of triglycerides (TG). However, non-HDL-C is not always calculated or reported in laboratory panels.

**Aim:** The main purpose of this study is to evaluate the effectiveness of artificial neural networks (ANNs) in predicting non-HDL-C levels based on standard lipid parameters, i.e. LDL-C and TG.

**Material and methods:** Based on 26,914 anonymized lipid profiles (TC, TG, LDL-C, non-HDL-C), an MLP 2–4–1 neural network was trained to estimate non-HDL-C values.

**Results and discussion:** The non-HDL-C values predicted by the model showed a strong correlation with laboratory results ( $R^2 = 0.944$  overall). Correlation was slightly higher for  $TG < 400$  mg/dL ( $R^2 = 0.945$ ) and remained relatively high in the hypertriglyceridemic subgroup ( $R^2 = 0.896$ ).

**Conclusions:** This study demonstrates that artificial neural networks are able to accurately estimate non-HDL-C values using the results of a lipid parameters survey, even in the cases of severe hypertriglyceridemia. Given its high prediction accuracy and its simplicity, the model is promising for the future as a practical decision-support tool in cardiovascular risk assessment, especially where non-HDL-C is unavailable or unreported.

## 1. INTRODUCTION

The European Society of Cardiology recommends the use of the SCORE2 and SCORE2-OP tables for assessing cardiovascular risk. Unlike the previously used SCORE tables, these tables are based on non-HDL cholesterol (non-HDL-C) rather than total cholesterol (TC).<sup>1</sup> While the differences between these rating systems are quite fundamental – SCORE assesses the 10-year risk of death from cardiovascular disease, whereas SCORE2 assesses the 10-year risk of a first major cardiovascular event – in both cases, the design of the system takes into account a single lipid parameter. For the SCORE tables, this is total cholesterol (TC), and for the SCORE2 tables, it is non-HDL cholesterol (non-HDL-C). Meanwhile, for appropriate patient treatment, low-density lipoprotein cholesterol (LDL-C) remains the key target according to the European Society of Cardiology/European Atherosclerosis Society guidelines.<sup>2</sup> The principle that lower LDL-C levels are associated with better outcomes has been confirmed in numerous studies.<sup>3,4</sup> From the perspective of a practicing physician, there is an increase in the number of individual lipid parameters. In developed countries, this is not a significant issue, but are all of them (TC, HDL-C, TG, LDL-C, non-HDL-C) always necessary? After all, some of these parameters are mathematically related, with an acceptable level of error. LDL-C can be calculated in several ways. The most commonly used methods are the Friedewald formula ( $LDL = TC - HDL - TG/5$ ) for values expressed in mg/dL<sup>1,5</sup> and the Sampson-NIH formula proposed in 2020.<sup>6</sup> The non-HDL-C<sub>LAB</sub> values can be calculated in the laboratory from a lipid profile ( $non-HDL = TC - HDL$ ).<sup>7</sup> Romaszko, Gromadziński, Buciński<sup>8</sup> demonstrated that a simple mathematical transformation of the Friedewald formula ( $non-HDL = LDL + TG/5$ ) allows for the calculation of non-HDL-C with approximately the same accuracy as the original Friedewald formula allowing for the calculation of LDL-C. It is tempting, therefore, to explore the relationship between non-HDL-C and LDL-C values using more modern tools. In the 21st century, such tools include neural networks, which are used to build self-learning Artificial Intelligence (AI).

Most equations used to estimate serum lipid levels are based on the assumption of linear relationships between variables. In contrast, artificial neural networks (ANNs) have the ability to model complex, nonlinear relationships, which can lead to improved accuracy in predicting lipid parameters.

## 2. AIM

The main purpose of this study is to evaluate the effectiveness of ANNs in accurately predicting non-HDL-C cholesterol levels based on LDL-C and triglyceride (TG) values. The analysis includes a comparison of the predictive accuracy of ANN models with traditional computational methods, as well as an assessment of the potential benefits of applying artificial intelligence to laboratory diagnostics.

## 3. MATERIAL AND METHODS

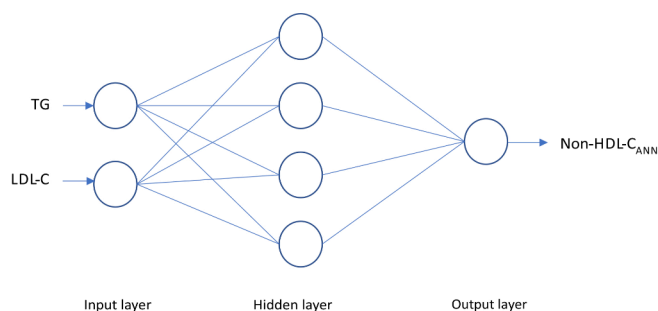
### 3.1. Acquisition of laboratory data

This study utilized publicly available laboratory data described in detail by Romaszko, Gromadziński, Buciński.<sup>8</sup> The dataset consisted of 26,914 anonymized laboratory test records containing complete lipid profiles, including TC, TG, LDL-C, and non-HDL-C. All lipid parameters were measured in certified clinical laboratories – the “Diagnostyka” network and the Voivodeship Specialist Hospital in Olsztyn (Poland). These laboratories determine LDL-C and other lipid parameters using the colorimetric-enzymatic method (Roche/Hitachi cobas 6000). Quality control of the tests was performed daily, and the permissible error for LDL-C measurement was 10%. The analyzed data were collected from January 1, 2021, to May 8, 2023, and were made available to researchers on May 10, 2023, for scientific purposes only. Based on that dataset, an attempt was made to construct an artificial neural network to predict the concentration of non-HDL-C based on LDL-C and TG.

### 3.2. ANN analysis

The collected numerical data regarding the lipid profile were analyzed using Statistica software (StatSoft, Inc., Tulsa, USA). In the first step, the dataset containing lipid profile data was randomly divided into three subsets: training (70% of cases), testing (15% of cases), and validation (15% of cases). These proportions were chosen to ensure an adequate number of training data and independent sets for validation and testing. The data randomization was performed prior to the division into subsets, ensuring their random and representative distribution.

A regression model was then developed, trained, and validated using a multilayer perceptron (MLP) neural network, the structure of which corresponds to the classic MLP model. It consists of three main layers: input, hidden, and output. This is a feedforward model, meaning the signal flows in one direction – from the input layer, through the hidden neurons, to the output layer – ensuring efficient and precise data processing.<sup>9</sup> The neurons of the input layer are responsible for the input variables (LDL-C and TG), which are fed into the model. The hidden layer contains four neurons, and the output layer consists of a single neuron that generates the predicted non-HDL-C values (mg/dL) (Figure 1). Each input of the artificial neuron is associated with weights that determine its significance. The input signals are multiplied by these weights and then transformed by an activation function, which regulates the flow of information and defines how the data is processed. It is the activation functions, based on mathematical relationships, that decide which signals will be passed on. Initially, the network operates with random weights, which gradually adjust during the learning process, optimizing the model's performance.<sup>10,11</sup> In the discussed ANN model, the BFGS (Broyden-Fletcher-Goldfarb-Shanno) learning algorithm was used, which optimizes the network weights, allowing for the efficient finding of the error function minimum. The effectiveness of the network was



**Figure 1. Structure of the MLP 2-4-1 neural network for predicting non-HDL-C<sub>ANN</sub> values based on LDL and TG. The network consists of an input layer with two neurons (LDL-C, TG), one hidden layer, and an output layer generating the predicted non-HDL-C<sub>ANN</sub> value.**

assessed using the mean squared error, which measures the difference between the values predicted by the model and the actual observed values in the data. Detailed information on the parameters used in the model is presented in Table 1.

## 4. RESULTS

The literature documents that high TG concentrations can affect the calculation of LDL-C using the Friedewald formula.<sup>12</sup> To exclude any potential impact of this factor on our results, we conducted the study on the entire dataset ( $n = 26,914$ ), and then, when preparing the statistical results, we performed an additional analysis of the subset of cases where TG concentrations were below 400 mg/dL ( $n = 26,516$ ) and where TG was greater than or equal to 400 mg/dL ( $n = 398$ ). The results are presented in Table 2.

The non-HDL-C<sub>ANN</sub> values predicted by the neural network model showed a strong correlation (Pearson correlation) with the non-HDL-C<sub>LAB</sub> values calculated from a lipid profile using the laboratory method (non-HDL-C = TC – HDL). For cases where TG < 400 mg/dL ( $n = 26,516$ ), the correlation is slightly stronger, with an R<sup>2</sup> value of 0.945. The number of cases where TG ≥ 400 mg/dL was significantly smaller ( $n = 398$ ), but the correlation remained high, with an R<sup>2</sup> value of 0.896 (Figure 2). Such a high coefficient of determination

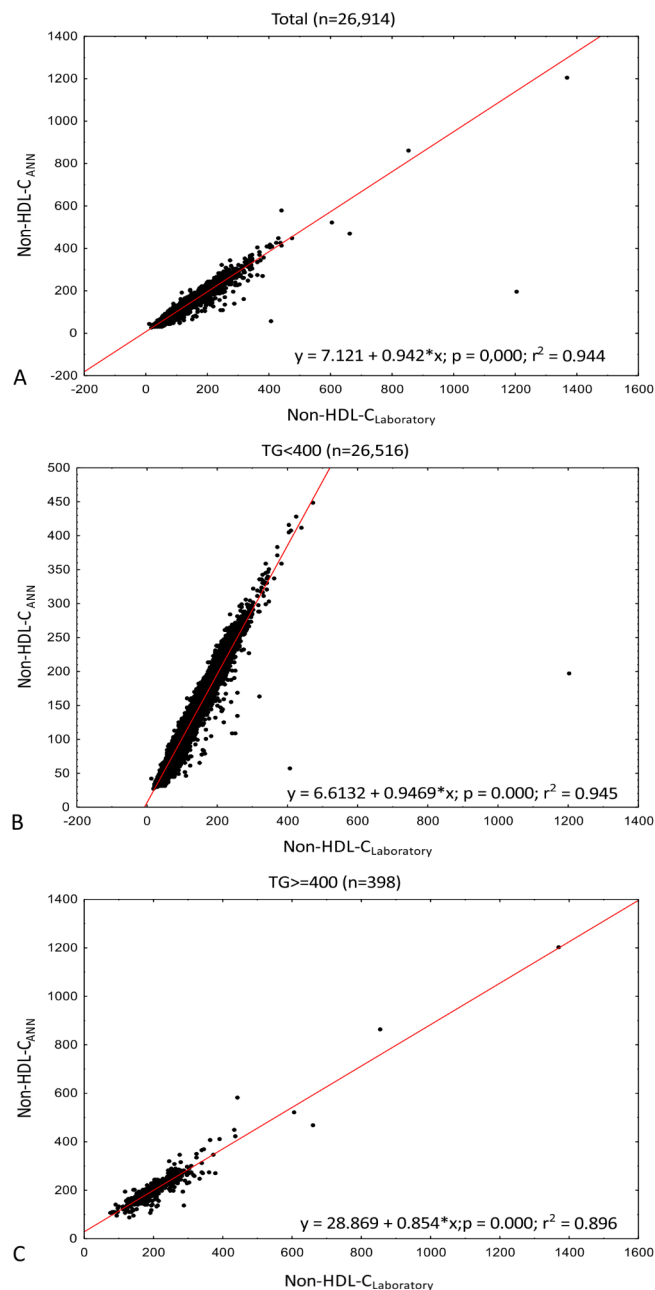
**Table 1. Summary of the parameters used in the ANN model.**

Model	Learning algorithm	Activation functions		
		Input layer	Hidden layer	Output layer
MLP 2–4–1	BFGS 90	linear	exponential	logistic

**Table 2. Comparison of the mean values of non-HDL-C<sub>LAB</sub> calculated from a lipid profile using the laboratory method and non-HDL-C<sub>ANN</sub> predicted using ANN, as well as the percentage of erroneous measurements.**

	Non-HDL-C [mg/dL]		Error (%) mean (± SD; median)
	non-HDL-C <sub>LAB</sub> mean (± SD; median)	non-HDL-C <sub>ANN</sub> mean (± SD; median)	
<b>Total (n = 26,914)</b>	124.39 (48.06; 118.00)	124.45 (46.65; 118.21)	–0.70 (7.93; 0.05)
<b>Underestimate &lt;10%</b>		39.62% (n = 10,663)	
<b>Error of the absolute &gt;25%</b>		0.394% (n = 106)	
<b>Overestimate &lt;10%</b>		45.34% (n = 12,202)	
<b>TG &lt; 400 mg/dL (n = 26,516)</b>	123.18 (45.63; 117.52)	123.19 (44.63; 117.38)	–0.69 (7.80; 0.06)
<b>Underestimate &lt;10%</b>		39.87% (n = 10,573)	
<b>Error of the absolute &gt;25%</b>		0.356% (n = 92)	
<b>Overestimate &lt;10%</b>		45.57% (n = 12,082)	
<b>TG ≥ 400 mg/dL (n = 398)</b>	209.430 (94.863; 194.00)	207.820 (85.641; 195.441)	–1.203 (13.976; –1.186)
<b>Underestimate &lt;10%</b>		31.658% (n = 126)	
<b>Error of the absolute &gt;25%</b>		3.517% (n = 14)	
<b>Overestimate &lt;10%</b>		30.151% (n = 120)	

SD – standard deviation



**Figure 2.** Pearson correlation for non-HDL-C<sub>LAB</sub> calculated from a lipid profile using the laboratory method and non-HDL-C<sub>ANN</sub> for all cases (A), for cases where TG < 400 mg/dL (B), and for cases where TG ≥ 400 mg/dL (C).

indicates a strong accuracy of the model in predicting non-HDL-C<sub>ANN</sub> values, even in hypertriglyceridemic trials.

## 5. DISCUSSION

Although LDL-C concentration still plays a key role in assessing the lipid profile, both in diagnosing and monitoring the treatment of lipid disorders and in predicting the risk of cardiovascular incidents, since 2021 non-HDL-C has been

recognized as being an equally important indicator to LDL-C. Traditional calculation formulas, such as Friedewald, Sampson-NIH, and Martin-Hopkins, are widely used in laboratory diagnostics to estimate LDL-C, non-HDL-C, and VLDL-C levels. Romaszko, Gromadziński, Buciński<sup>8</sup> demonstrated that non-HDL-C concentration can be accurately estimated based on LDL-C and TG values, without the need to measure TC and HDL-C levels. The equation they proposed may reduce testing costs by eliminating the need for a full lipid profile or, more importantly, serve as an auxiliary method in the absence of TC results (avoiding the need for a repeat blood draw) and still provide results that were closer to laboratory measurements than the classic Friedewald formula for LDL-C. However, the effectiveness of the calculation formulas is limited, particularly in cases of elevated triglyceride levels or other lipid disorders.<sup>5,6</sup> In such situations, the prediction error can be significant, negatively impacting the quality of diagnosis and therapeutic decision-making. In response to these limitations, models based on artificial intelligence, such as artificial neural networks (ANN), are gaining increasing importance due to their higher precision in predicting laboratory values.

In this study, we utilized artificial neural networks to predict non-HDL-C, like Romaszko, Gromadziński, Buciński<sup>8</sup>, based on only LDL-C and TG. Our results indicate that the use of ANN allows for even more accurate estimations, minimizing the mean prediction error compared to previous methods. A comparison of the results for non-HDL-C obtained using the ANN model and the transformed Friedewald formula method shows significant differences in prediction accuracy.<sup>8</sup> Although the calculation of non-HDL-C using the Romaszko, Gromadziński, Buciński<sup>8</sup> method is characterized by an error of 8.88%, which falls within the acceptable range of 6–12% defined by international and Polish laboratory diagnostic societies (NCEP, CLSI, PDL), the ANN model achieves a much smaller error (–0.70%). This result is well below the acceptable norms, suggesting the high applicability of ANN in routine clinical diagnostics. Additionally, the ANN model very accurately reflects laboratory values for non-HDL-C, showing minimal differences in means and medians. Importantly, the percentage of large errors (>25%) for the ANN model is ten times lower than for the Friedewald method (0.394% vs. 3.92%). Moreover, the prediction error level for patients with TG > 400 mg/dL remains similar to that for the entire group, suggesting that the ANN model allows for accurate prediction of non-HDL-C even with high TG concentrations.

Researchers who compared the effectiveness of machine learning methods with traditional formulas such as the Friedewald formula, the Martin-Hopkins equation, and the Sampson-NIH formula in estimating LDL-C levels have presented similar results.<sup>13–15</sup> It showed that machine learning-based methods outperform traditional formulas in terms of accuracy, particularly in cases of high triglyceride levels. Moreover, other researchers<sup>16,17</sup> have demonstrated not only the effectiveness of AI methods in predicting LDL-C levels but also applied Explainable AI (XAI) techniques to better understand the factors influencing prediction results. These

studies emphasized that the interpretability of AI models is crucial in medical applications because it enhances the trust of healthcare personnel in machine learning-based systems. Our study used neural networks to predict non-HDL-C, analogous to the studies mentioned above predicting LDL-C. The ANN model demonstrated high accuracy in predicting non-HDL-C concentrations based not only on just two but also directly measured TG and LDL-C values, maintaining strong agreement with reference laboratory results even at TG levels exceeding 400 mg/dL. To the best of our knowledge, this is the first such study.

## 6. CONCLUSION

- (1) The ANN model can be an effective method for predicting non-HDL-C, achieving results nearly identical to laboratory measurements and demonstrating high accuracy even with TG > 400 mg/dL.
- (2) This new method can be easily implemented into most laboratory reporting systems with little to no additional cost.
- (3) These findings indicate that in populations with elevated triglyceride concentrations – where traditional calculation methods for non-HDL-C are subject to considerable uncertainty – the application of the ANN model may provide a valuable complementary approach and help address an existing methodological gap.

### Ethics approval

The data used in this analysis were obtained from publicly available resources published by Romaszko, Gromadziński and Bucinski.<sup>8</sup> The data are retrospective and fully anonymized; therefore, approval from an ethics committee was not required.

The data that support the findings of this study are openly available at the: <https://www.frontiersin.org/articles/10.3389/fmed.2023.1247126/full#supplementary-material>. Further inquiries can be directed to the corresponding author.

### Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Funding

This work was supported by the maintenance of research capacities funds from Nicolaus Copernicus University in Toruń.

### Author contributions

Study design: AB

Data collection: AB

Statistical analysis: AB

Data interpretation: AB

Manuscript preparation: AB, UM-W

Literature search: AB, UM-W

Funds collection: AB, UM-W

## References

- 1 Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). *Rev Esp Cardiol (Engl Ed)*. 2022;75:429.
- 2 Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: Lipid modification to reduce cardiovascular risk. *Eur Heart J*. 2020;41:111–188.
- 3 Nanchen D, Tokgozoglul L, Komen JJ, Bardet A, Catapano AL, Ray KK. Contemporary LDL-cholesterol management in male and female patients at high-cardiovascular risk: Results from the European observational SANTORINI study. *Eur Heart J*. 2024.
- 4 Zeng S, Wang H, Li X, et al. Composition design of fullerene-based hybrid electron transport layer for efficient and stable wide-bandgap perovskite solar cells. *J Energy Chem*. 2025;102:172–178.
- 5 Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18:499–502.
- 6 Sampson M, Ling C, Sun Q, et al. A New Equation for Calculation of Low-Density Lipoprotein Cholesterol in Patients With Normolipidemia and/or Hypertriglyceridemia. *JAMA Cardiol*. 2020;5:540–548.
- 7 Virani SS. Non-HDL cholesterol as a metric of good quality of care: Opportunities and challenges. *Tex Heart Inst J*. 2011;38:160–162.
- 8 Romaszko J, Gromadziński L, Bucinski A. Friedewald formula may be used to calculate non-HDL-C from LDL-C and TG. *Front Med (Lausanne)*. 2023;10:1247126.
- 9 Kruse R, Mostaghim S, Borgelt C, Braune C, Steinbrecher M. Multi-layer perceptrons. *Computational intelligence: A methodological introduction*: Springer; 2022:53–124.
- 10 Weiss R, Karimijafarbigloo S, Roggenbuck D, Rödiger S. Applications of Neural Networks in Biomedical Data Analysis. *Biomedicines*. 2022;10.
- 11 Ramchoun H, Idrissi MAJ, Ghanou Y, Ettaouil M. Multilayer Perceptron: Architecture Optimization and Training. *Int J Interact Multim Artif Intell*. 2016;4:26–30.
- 12 Martin SS, Blaha MJ, Elshazly MB, et al. Comparison of a novel method vs the Friedewald equation for estimating low-density lipoprotein cholesterol levels from the standard lipid profile. *Jama*. 2013;310:2061–2068.
- 13 P PA, Kumari S, Rajasimman AS, Nayak S, Priyadarsini P. Machine learning predictive models of LDL-C in the population of eastern India and its comparison with

- directly measured and calculated LDL-C. *Ann Clin Biochem.* 2022;59:76–86.
- <sup>14</sup> Tsigalou C, Panopoulou M, Papadopoulos C, Karvelas A, Tsairidis D, Anagnostopoulos K. Estimation of low-density lipoprotein cholesterol by machine learning methods. *Clin Chim. Acta.* 2021;517:108–116.
- <sup>15</sup> Singh G, Hussain Y, Xu Z, et al. Comparing a novel machine learning method to the Friedewald formula and Martin-Hopkins equation for low-density lipoprotein estimation. *PLoS One.* 2020;15:0239934.
- <sup>16</sup> Sezer S, Oter A, Ersoz B, et al. Explainable artificial intelligence for LDL cholesterol prediction and classification. *Clin Biochem.* 2024:110791.
- <sup>17</sup> Öter A. Deep learning-based LDL-C level prediction and explainable AI interpretation. *Comput Biol Med.* 2025; 188:109905.