Original Paper

A Deep Neural Network for Estimating Low-Density Lipoprotein Cholesterol From Electronic Health Records: Real-Time Routine Clinical Application

Sangwon Hwang[1], PhD; Chanwoo Gwon[2], MS; Dong Min Seo[3], BS; Jooyoung Cho[4], MD, PhD; Jang-Young Kim[5], MD, PhD; Young Uh[4], MD, PhD

[1]Artificial Intelligence Bigdata Medical Center, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
[2]Wonju Industry-Academic Cooperation Foundation, Yonsei University Mirae Campus, Wonju, Republic of Korea
[3]Department of Medical Information, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
[4]Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
[5]Department of Internal Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

Corresponding Author:
Young Uh, MD, PhD
Department of Laboratory Medicine
Yonsei University Wonju College of Medicine
20, Ilsan- ro, Wonju, Gangwon-do
Wonju, 26426
Republic of Korea
Phone: 82 33 741 1592
Fax: 82 33 731 0506
Email: u931018@yonsei.ac.kr

Abstract

Background: Previously, we constructed a deep neural network (DNN) model to estimate low-density lipoprotein cholesterol (LDL-C).

Objective: To routinely provide estimated LDL-C levels, we applied the aforementioned DNN model to an electronic health record (EHR) system in real time (deep LDL-EHR).

Methods: The Korea National Health and Nutrition Examination Survey and the Wonju Severance Christian Hospital (WSCH) datasets were used as training and testing datasets, respectively. We measured our proposed model's performance by using 5 indices, including bias, root mean-square error, P10-P30, concordance, and correlation coefficient. For transfer learning (TL), we pretrained the DNN model using a training dataset and fine-tuned it using 30% of the testing dataset.

Results: Based on 5 accuracy criteria, deep LDL-EHR generated inaccurate results compared with other methods for LDL-C estimation. By comparing the training and testing datasets, we found an overfitting problem. We then revised the DNN model using the TL algorithms and randomly selected subdata from the WSCH dataset. Therefore, the revised model (DNN+TL) exhibited the best performance among all methods.

Conclusions: Our DNN+TL is expected to be suitable for routine real-time clinical application for LDL-C estimation in a clinical laboratory.

(JMIR Med Inform 2021;9(8):e29331) doi: 10.2196/29331

KEYWORDS

low-density lipoprotein cholesterol; deep neural network; transfer learning; real-time clinical application

Introduction

Low-density lipoprotein cholesterol (LDL-C) is a major marker of cardiovascular disease (CVD) because of its role in the pathophysiology of atherosclerosis [1]. The contemporary reference measurement procedure for LDL-C is ultracentrifugation [2]. However, owing to the difficulty in applying this in a clinical setting, LDL-C levels have mostly been estimated by other means [3-6].

Friedewald et al [3] observed that most plasma samples are comprised of chylomicrons and that most triglycerides (TGs) in plasma are present in very low-density lipoprotein cholesterol (VLDL-C) at a ratio of 5:1, while the chylomicrons are undetectable. This observation led to the 1972 Friedewald (FW) equation, which is used to estimate LDL-C [3]. Martin et al [4] showed in 2014 that VLDL-C levels estimated by simply dividing the TG level by 5 may inaccurately predict LDL-C levels, specifically in hypertriglyceridemia. They divided subjects according to the levels of TG and non–high-density lipoprotein cholesterol (non-HDL-C), yielding 180 groups (clusters) [4]. For those, 180 equations were established and integrated into the novel estimation method. More recently, Sampson et al [5] used the interaction between TG and non-HDL-C and a correction factor (TG^2) to estimate LDL-C, resulting in the National Institutes of Health (NIH) method.

Deep learning techniques, specifically deep neural networks (DNNs), provide multilayer stacks of simple networks (eg, perceptrons or modules) with nonlinear functions applied between each layer [7]. The numerous perceptrons and the nonlinearity between them allow researchers to represent complex real data in a way that solves a variety of challenging tasks such as classification and regression. We previously established a deep learning model to estimate LDL-C, including 180 perceptrons [6], motivated by the model of Martin et al [4]. This yielded accurate results for LDL-C estimation.

Additionally, DNNs are easy to apply in clinical settings and hospital databases. Several studies have adopted linear regression to estimate LDL-C using fewer than 5 trained weights (parameters) [8,9]. With such a low number, it is possible to adapt the linear model–based LDL estimator to a hospital database without having to rebuild the system. With the DNN proposed by Lee et al [6], approximately 4600 trained weights were established as a matrix. Although it had many weights, it was applicable to clinical settings and hospital databases using matrix calculation. Moreover, if the independent DNN application server is present, it is easy to apply and upgrade without rebuilding the system.

Transfer learning (TL) is a method of transferring knowledge from a previously trained task to a new but related one [10]. In a clinical setting, it is enormously difficult to collect real patient data and preprocess them to analyzable forms (structured data). Moreover, for these analyses, a great deal of effort is needed to resolve ethical issues and receive board approval for data collection. The difficulty of preparing an analyzable dataset presents an enormous obstacle for training because it typically requires an enormous dataset to train numerous perceptrons [7]. However, TL adopts a pretrained model learned from publicly available or large-scale datasets. Hence, it is considered to be a powerful method when it comes to small-scale dataset training requirements.

Over the past decade, enormous volumes of medical data have been stored in electronic health records (EHRs) (ie, electronic medical records [EMRs]) from which many studies have compiled patient information for secondary use for health care tasks and medical decisions (eg, disease prediction). Shickel et al [11] reviewed the current research that applied deep learning to EHRs. Although there have been many studies that constructed models using data obtained from EHR data, very few were found to have performed real-time clinical applications of the established model [12]. This study aimed to remedy this by applying previously constructed models to an EHR system. Hence, we performed the following 3 tasks for this study. First, we applied the DNN model from Lee et al [6] to the Wonju Severance Christian Hospital (WSCH) EHR system to generate real-time results for estimated LDL-C (deep LDL-EHR; Figure 1). Second, we measured performance based on several accuracy indices for the estimated LDL-C levels provided by the real-time application of our DNN model (deep LDL-EHR) and compared them to those of other LDL estimation methods. Third, we revised the DNN model by using TL, a multitask learning algorithm (Figure 2).

Figure 1. Overall workflow of deep LDL-EHR: Steps 3, 7, and 8 provide input- or output-value transfers between 2 platforms; the (Tomcat)[a] web server was established using Apache Tomcat [13] on a JAVA server page and servlet application; the (Flask)[b] web server was established using the Flask framework [14], a lightweight web application framework based on TensorFlow and Keras in Python. DNN: deep neural network; EMR: electronic medical record; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; TC: total cholesterol; TG: triglyceride.

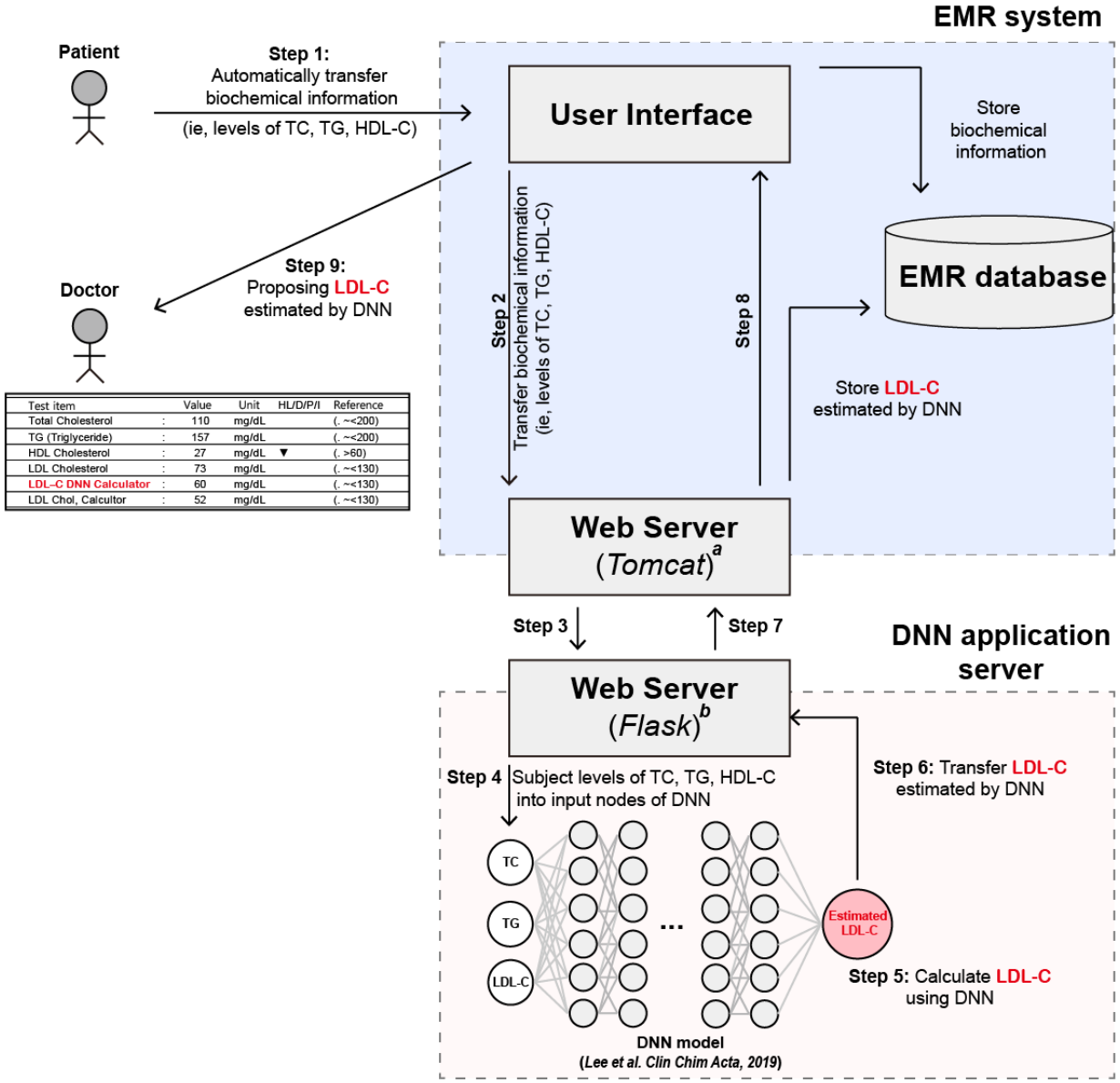
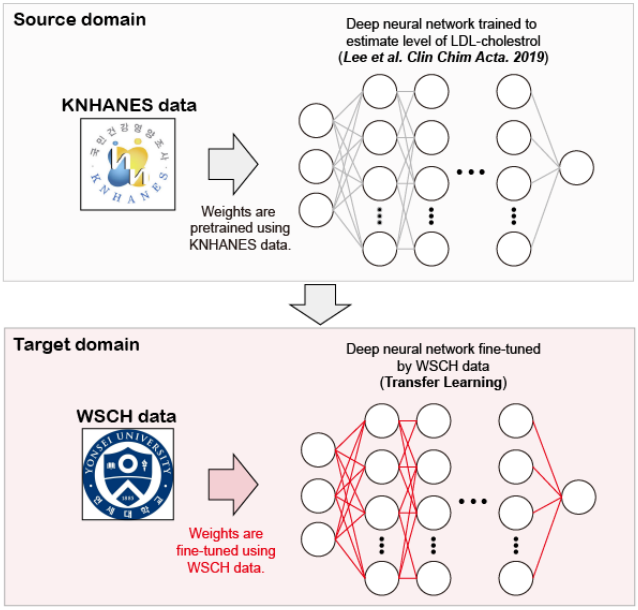


Figure 2. Transfer learning: For the task in the source domain, the deep neural network (DNN) model has the same structure and data as those previously trained by Lee et al [6], while ours is trained and saved on the DNN application server. For the task in the target domain, the DNN model saved in the DNN application server is loaded and retrained (fine-tuned) using Wonju Severance Christian Hospital (WSCH) data (30% randomly selected subjects) on a local computer. KNHANES: Korea National Health and Nutrition Examination Survey; LDL: low-density lipoprotein.



Methods

Application of Our DNN Model in a Clinical Laboratory

Experts in various fields (ie, clinical pathologists, database administrators, cardiologists, and computer scientists) have collaborated to construct a deep LDL-EHR model that we are using to provide LDL-C estimations for hospital patients. The application of our DNN model (ie, the deep LDL-EHR) in a clinical laboratory consists of 2 main subsystems: the EMR and a DNN application server. The EMR system is responsible for receiving and storing patient medical data (eg, levels of total cholesterol [TC], HDL-C, and TG) and transferring them to the DNN application server. The following core components are part of the EMR system: a user interface that receives data from users and stores them in the EMR database; a web server that hosts the application that permits users to see laboratory results and estimates via a web browser; a database that stores all data, including laboratory markers (input data) and results estimated by deep learning; and a physical server that runs these software components. The web service was developed using JAVA Server Pages (JSP) and a servlet application [15], and the user interface is based on the hypertext markup language, cascade style sheets, and JavaScript [16]. The web server was established in Apache Tomcat [13] based on JSP and servlets. We used a Sybase relational database management system for its construction [17].

The DNN application server hosts the DNN application, which is built upon a Python environment running separately from the EMR system. It is responsible for performing the estimation of LDL-C values based on the received data (TC, HDL-C, TG) from the EMR system and for transferring the estimated values of LDL-C back to the EMR system (Figure 1). This application server is comprised of several core components, including a flask-based web server [14] built using the flask framework (ie, a lightweight web application framework on Python), which receives data from the EMR system and transfers estimated LDL-C values back to the EMR system. It is also comprised of an application that calculates LDL-C values using the data received from the EMR system, a TensorFlow [18] framework that provides various Python application programming interfaces (APIs) that execute high-performance DNN analysis, a Keras [19] neural network library installed atop a Microsoft cognitive toolkit, TensorFlow, and Theano, which provides high-level easy-to-use APIs for creating neural networks. Although the 2 libraries are technically separate, TensorFlow and Keras are typically used in a unified manner.

Note that the optimization of weights or parameters is performed on a local computer and is saved in the form of a matrix; the DNN application server processes only the matrix operations using previously trained weights in the local computer.

Data Collection

From July 2020 to December 2020, we obtained 11,125 estimated LDL results from a real-time system. Because these results were obtained from inpatients and outpatients from all departments (eg, cardiology, gastroenterology, endocrinology, oncology, and health check-up centers) in real time, we could not trace whether examinations were performed before or after fasting. The TC, TG, HDL-C, and LDL-C data were analyzed using the modular Diagnostic de Performance Énergétique system (Roche Diagnostics, Basel, Switzerland).

We collected 2009-2015 Korea National Health and Nutrition Examination Survey (KNHANES) datasets to replicate the DNN model of Lee et al [6] Note that results in Multimedia Appendix 1 refer to the DNN model of Lee et al [6], and those in Figure 4 refer to the replicated DNN model. Subjects missing TC, HDL-C, TG, and LDL-C data were excluded. Therefore, data for 15,074 subjects were analyzed for this study, nearly the same as the number used in the previous study [6]. All participants were tested for lipid profiles after at least 12 hours of fasting.

Lipid profiles (ie, TC, HDL-C, TG, and LDL-C) were measured using the Hitachi 7600 analyzer (Hitachi, Tokyo, Japan).

Other LDL-C Estimation Methods

There have been numerous studies on the estimation of LDL-C, and they largely used linear regression methods [20,21]. Among them, we empirically selected some representative methods, including FW, Novel, and NIH methods [3-5]. The FW method estimates LDL-C by subtracting levels of HDL-C and TG/5 from TC. The Novel method integrates clustering and linear regression, initially arranging a sample into one of 180 subgroups previously determined by TG and non-HDL-C levels. Afterward, a case of 180 linear regression equations is applied to the sample. The NIH method uses TC, HDL-C, TG, and their combinations, including the square of TG (TG^2) and a multiplication value between TG and non-HDL-C. The source code for these equations is available at our GitHub homepage [22].

DNN and TL

The DNN model included 6 hidden layers with 30 hidden nodes in each. We used a rectified linear unit as an activation function to implement nonlinearity between the hidden layers. The details of this model are described in the study by Lee et al [6].

We used TL [10] to upgrade this DNN model [6]. TL includes a source domain that is typically a large-scale dataset alongside a small-scale target domain that contains more specific data compared with those of the source domain [10]. As described in Figure 2, from the source task (ie, KNHANES dataset), we extracted the desired information (ie, trained weights). From the target task (ie, subset of the WSCH dataset), we retrained (fine-tuned) the DNN. The source code for the DNN+TL is available at our GitHub homepage [22].

Performance Measurement

To assess and compare the accuracy of each LDL-C estimation method, we measured the following 5 indices: bias (estimated LDL-C [eLDL-C] – measured LDL-C [mLDL-C]), root mean square error (RMSE), P10 to P30, concordance, and correlation coefficient.

Jeong et al [23] implemented the one-sample *t* test to compare the average bias between true and estimated values from a regression task. Motivated by this, we used the one-sample *t* test to measure the degree of average bias of each estimation method differing from zero.

Numerous studies have implemented RMSE to measure the degree of accuracy for LDL-C estimation methods [4-6,23]. Hence, we decided to use the RMSE for the estimation accuracies of each method as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{eLDL-C} - \text{mLDL-C})^2}$$

P30 has been implemented to measure the clinical accuracy of estimation methods for glomerular filtration rate [23]. This study used P10 and P30, and we expanded these indices as *Pn* (*n* = 10, 15, 20, 25, and 30), measured as the ratio of samples from which LDL-C was estimated using each method within mLDL-C ± n% divided by all samples.

$$Pn = \frac{\text{\# of samples with eLDL-C within mLDL} \pm n\% \text{ of mLDL-C}}{\text{\# of all samples}}$$

In studies that provided the estimation method for LDL-C [4,5], concordance has been used to examine the classification accuracy between mLDL-C and eLDL-C. In detail, both mLDL-C and eLDL-C values are categorized as 6 subgroups based on the National Cholesterol Education Program (NCEP) Adult Treatment III guideline cutoffs that other studies used [24,25]. Concordance was measured as follows:

$$\text{Concordance} = \frac{\text{\# of B}\cap\text{A}}{\text{\# of A}}$$

where *A* are samples with mLDL-C within a specific range and *B* are samples with eLDL-C in the same interval as mLDL-C.

Several methods of correlation have been used to measure the degree of consistency between true and estimated values (ie, mLDL-C and eLDL-C) [5,23]. Specifically, we used Pearson correlation coefficient, a normalized measurement of the covariance of 2 lists of values (ie, mLDL-C and eLDL-C) divided by the product of their standard deviation.

Jacob and Speed [26] suggested that the selected features and their predictive performances should be examined based on a random sampling perspective for generalization. In other words, the samples selected for the training model (ie, DNN+TL) greatly affect its performance. Therefore, we performed the following tasks considering the random sampling perspective. In step 1, we made a pair of random sample datasets, including training and testing, which were randomly divided at a ratio of 0.3 and 0.7, respectively. In step 2, we established a DNN+TL model using the randomly selected training set and measured the *t* value and RMSE of the DNN+TL model for the testing set. We also measured the *t* value and RMSE of other models (ie, FW, Novel, NIH, and DNN) for the testing set. In step 3, we iterated Steps 1 to 2 at 1000 times, and 2 matrices consisting of 5 columns (5 LDL-C estimation methods) and 1000 rows (# of iterations) were generated, including the *t* value and RMSE. We compared 2 indices (ie, *t* value and RMSE) among the 5 methods based on one-way analysis of variance and performed multiple comparisons using the Bonferroni post hoc test.

Variance Importance

We implemented permutation importance [27] and Shapley addictive explanations (SHAP) [28] to identify the contribution of each feature (ie, TC, HDL-C, and TL) to the final output of the DNN model. Permutation importance is a heuristic method used to measure normalized feature importance by measuring the decrease in a model's performance when a feature is permuted [27]. SHAP is an addictive feature attribution method used to determine feature importance by measuring a weighted average value of all possible differences between 2 sets of outputs that are resulted from models with and without the feature [28]. The permutation importance was measured using the *permutation_importance* function in the sklearn package [29], and the SHAP was calculated using the *DeepExplainer* function in the SHAP package [28].

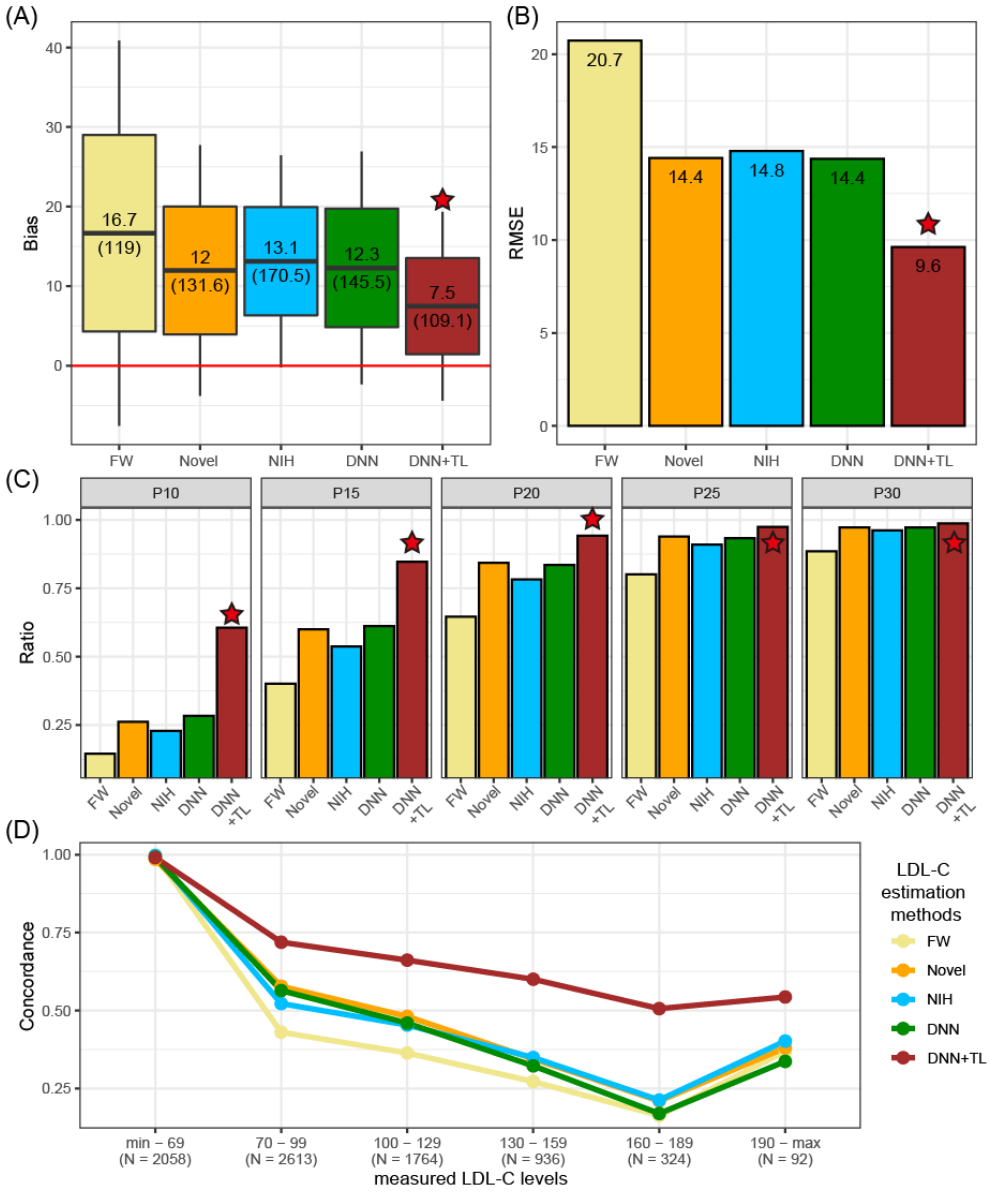
Statistics

Statistical analyses were performed using the *R* programming language (v.3.6.4). For a comparison of continuous variables based on 2 groups, we used the *t* test and the Mann Whitney *U* test. For categorical variables, we used the Chi-squared test, and a P value of <.05 was considered to be statistically significant.

Results

From the real-time application (Figure 1), we obtained 11,125 LDL values estimated using the DNN model. The distribution of bias (box plot) and RMSE (bar plot) of each LDL estimation method are illustrated in Multimedia Appendix 1. The estimated LDL-C values using the Novel method differed least from zero, and the values using the FW equation method were biased the most from zero. The eLDL-C levels using the DNN application system had, from among the 4 methods, the second most biased distribution from zero among the difference values between eLDL-C and mLDL-C (Multimedia Appendix 1). When comparing the RMSE of each method, the FW method resulted in the highest RMSE, followed by the DNN application system. In all the P10 to P30, the FW method showed the lowest ratio, and the DNN application system showed the second lowest ratio (Figure 3C; Multimedia Appendix 1). We compared concordances between groups stratified by mLDL-C and eLDL-C levels obtained from the 4 methods (Figure 3D). Therefore, the novel method showed the highest concordance from 70 to 129 of the mLDL-C levels, and the NIH method showed the highest concordance from 130 to the maximum mLDL-C levels (Multimedia Appendix 1). Collectively, the DNN application generated inaccurate results compared with the others.

Figure 3. Performance of 5 LDL estimation methods: (A) upper and lower numbers indicate the average and one-sample *t* value, respectively, while the black bars, upper or lower margins, and maximum or minimum lines for each boxplot indicate 1 SD and 1.96 SDs, respectively; (B) numbers in bar plots indicate real values of RMSE; (C) P10 to P30; (D) concordance of each LDL-C estimation method. Stars in each plot indicate the model with the best performance. Note that the deep neural network (DNN) method was the replicated model for the DNN model. FW: Friedewald equation; NIH: National Institutes of Health; RMSE: root mean square error; TL: transfer learning.



We compared the lipid profiles of the KNHANES dataset with those of the WSCH dataset (Table 1). All 4 variables differed significantly between the 2 datasets. We concluded that differential characteristics between the training set (KNHANES) and the testing set (WSCH) triggered inaccurate results from the DNN application system. In other words, an overfitting problem existed in the deep LDL-EHR model. To overcome this limitation, we adopted the TL method [10]. Using the 2009-2015 KNHANES datasets, we trained the DNN model using the same structure and hyperparameters as those of the model proposed by Lee et al [6], yielding a pretrained DNN model. Next, we randomly selected 30% of the WSCH dataset, which was used to fine-tune the pretrained DNN model (Figure 2).

Table 1. General characteristics of and comparisons between the Korea National Health and Nutrition Examination Survey (KNHANES) and Wonju Severance Christian Hospital (WSCH) datasets.

Variable	KNHANES (n=15,074)	WSCH (n=11,125)	P value
Age (years), mean (SD)	45.5 (18.2)	59.4 (15.5)	<.001[a]
Age (years), median (IQR)	46 (32-60)	60 (51-70)	<.001[b]
Male, n (%)	7507 (49.8)	6435 (57.8)	<.001[c]
Total cholesterol (mg/dL), mean (SD)	188.8 (37.7)	156.4 (41.6)	<.001[a]
Total cholesterol (mg/dL), median (IQR)	186 (162-212)	152 (128-182)	<.001[b]
HDL[d] cholesterol (mg/dL), mean (SD)	48.7 (12.1)	50.2 (14.2)	<.001[a]
HDL cholesterol (mg/dL), median (IQR)	47.3 (40.1-55.7)	48 (40-58)	<.001[b]
Triglyceride (mg/dL), mean (SD)	160.2 (135.6)	139.7 (126.2)	<.001[a]
Triglyceride (mg/dL), median (IQR)	120 (76-211)	114 (83-163)	<.001[b]
Measured LDL[e] cholesterol (mg/dL), mean (SD)	112 (32.3)	94.8 (35.9)	<.001[a]
Measured LDL cholesterol (mg/dL), median (IQR)	109 (89-132)	90 (68-117)	<.001[b]

[a]Determined using a *t* test.

[b]Determined using a Mann-Whitney *U* test.

[c]Determined using a Chi-squared test.

[d]HDL: high-density lipoprotein.

[e]LDL: low-density lipoprotein.

We compared the performances of the 5 methods, including the aforementioned 4 and DNN+TL methods (Figure 3). Based on the bias and RMSE, the DNN+TL was biased least from zero (mean 7.5; t_{7786}=109.1) and had the lowest RMSE (Figures 3A and 3B). In all of P10 to P30, the DNN+TL method had the highest ratio among the other methods. Particularly in P10, the superior performance of the DNN+TL method was notable (Figure 3C). Regarding the concordance of the LDL-C estimation methods, the DNN+TL method had the highest ratio through most of the LDL-C range except for a section of LDL-C from the minimum to 69 mg/dL (Figure 3D).

We illustrated correlation plots describing the distribution of eLDL-C values and the matched LDL-C levels estimated by the 5 methods, including FW, Novel, and DNN (Figure 4). In DNN+TL, the LDL-C level is the most accurately estimated among the other 4 methods based on the Pearson correlation coefficient (Figure 4).

For the 5 LDL-C estimation methods, we generated distributions of *t* values and RMSE, separately, by iterating the random selection of training set at 1000 times (Figure 5). As a result, DNN+TL exhibited the best performance for both bias from zero (*t* value, Bonferroni-corrected *P*<.001 for DNN+TL vs other methods) and absolute error (RMSE, Bonferroni-corrected *P*<.001).

For input features (ie, TC, HDL-C, and TG) and their deep learning models (ie, DNN and DNN+TL), we measured the variance (global) importance by using permutation importance and SHAP (Figure 6). In both DNN and DNN+TL, TC was the best crucial feature based on 2 indices of the variance importance. Moreover, TG and HDL-C comprised the second-most important variable based on permutation importance and SHAP, respectively (Figure 6A). In DNN+TL, the second important feature was TG, based on all indices of the variance importance (Figure 6B). Moreover, we illustrated the distribution of the ratio of TG to VLDL-C in relation to TG levels (Multimedia Appendix 2). VLDL-C, as analyzed in our study, is not a measured value, but is instead the result calculated by subtracting the values of HDL-C and eLDL-C (by the 5 methods) from TC. We found that the TG to VLDL-C ratio estimated by 3 models had large variance at high TG levels (Multimedia Appendix 2), which was similar with the results in the study by Martin et al [4]. The distribution of the TG to VLDL-C ratio estimated by the DNN+TL model looked like a mixture between the ratios by mLDL-C and DNN (Multimedia Appendix 2), indicating that the DNN+TL had fine-tuned the

previous DNN model [6] to represent the characteristics of the WSCH dataset by importantly considering the TG variable (Figure 6).

Figure 4. Correlation plots and coefficients between measured low-density lipoprotein cholesterol (mLDL-C) and estimated LDL-C (eLDL-C) calculated by 5 methods. The points on the scatterplots indicate the individual samples. A star indicates the highest Pearson correlation coefficient. DNN: deep neural network; FW: Friedewald method; NIH: National Institutes of Health; TF: transfer learning.

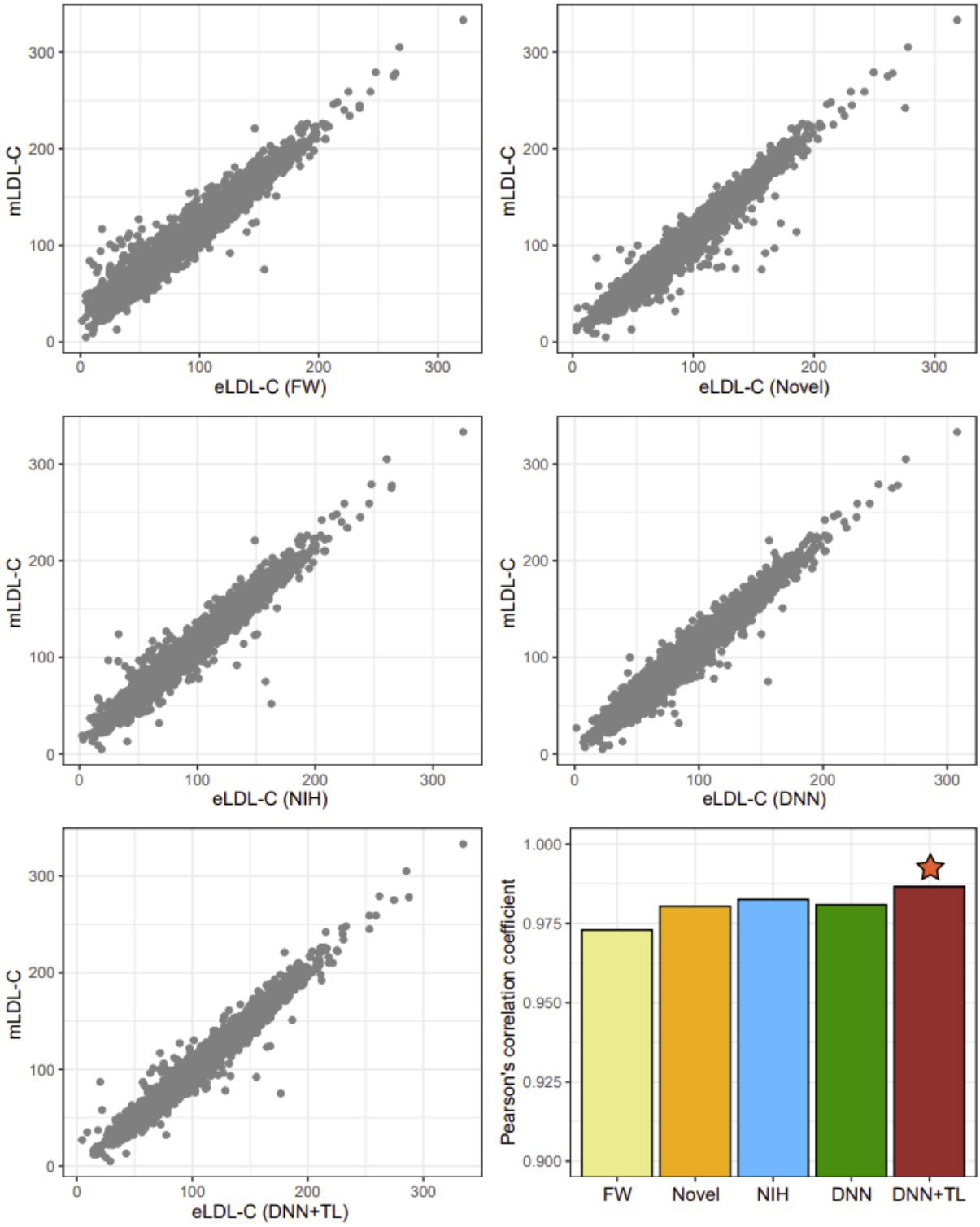


Figure 5. Comparison of performance based on a random sample perspective. A one-sample *t* test was used. DNN: deep neural network; FW: Friedewald method; NIH: National Institutes of Health; TL: transfer learning.

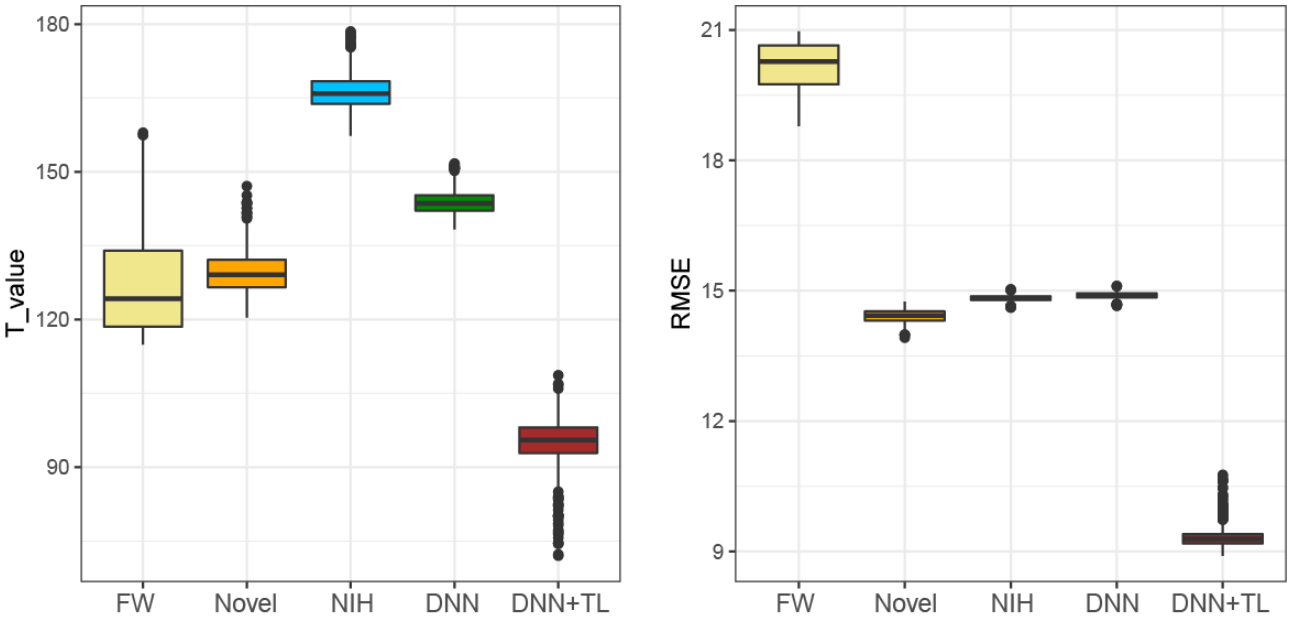
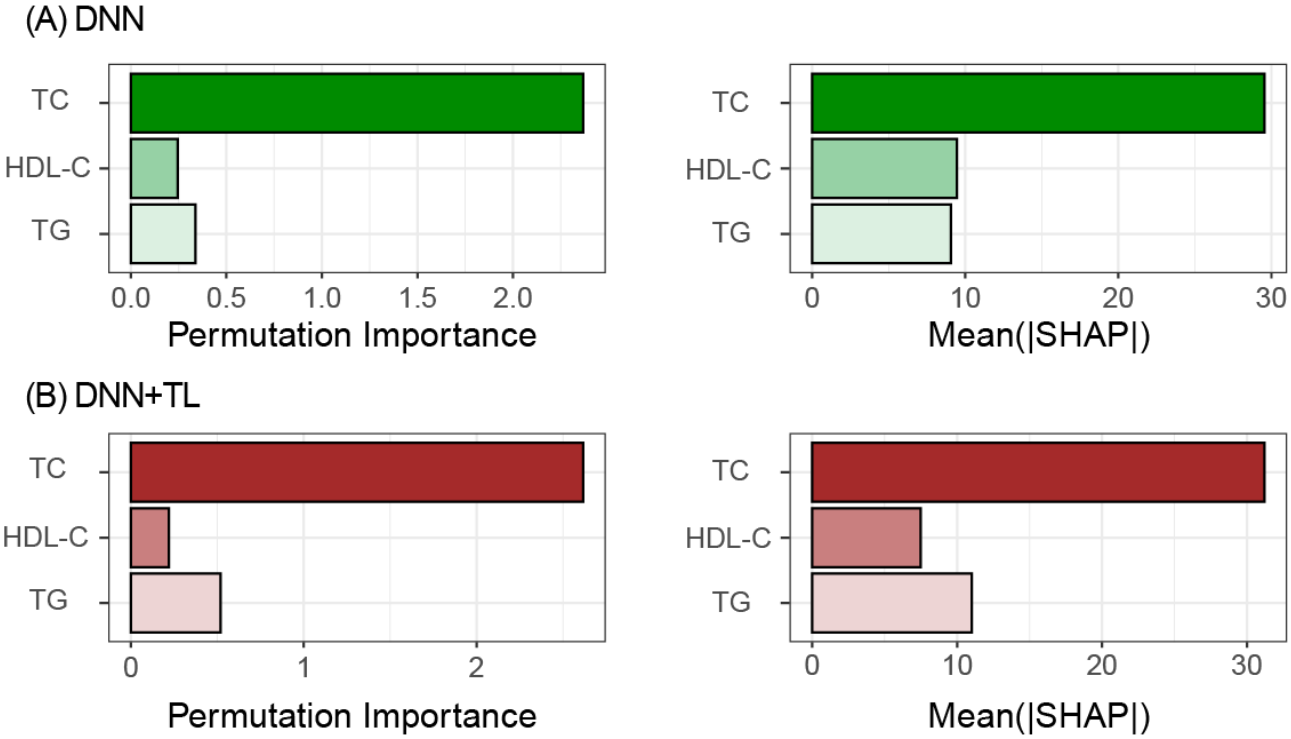


Figure 6. Variance importance based on permutation importance and Shapley addictive explanations (SHAP). DNN: deep neural network; HDL-C: high-density lipoprotein cholesterol; TC: total cholesterol; TG: triglyceride; TL: transfer learning.



Discussion

Principal Findings

We applied the DNN model for LDL-C estimation from EHR (deep LDL-EMR) data to generate real-time results. However, we found that our original deep LDL-EMR generated inaccurate results compared with other LDL estimation methods. We hypothesized that these inaccuracies may have been caused by the batch effect between the 2 different datasets. We therefore adopted a TL method to fine-tune the DNN model using local data-specific characteristics. Therefore, the DNN+TL method resulted in the most accurate results of all methods.

Approximately 15,000 subjects (KNHANES) were used to construct the DNN, and about 3300 WSCH LDL-C results were used for fine-tuning it. Martin et al [4] assigned approximately 900,000 subjects to develop the Novel method. Meeusen et al [25] enrolled 23,055 individuals from the Mayo Clinic and externally validated the Novel method. In 2020, Sampson et al [5] used approximately 9000 LDL-C test results to develop the NIH method while internally and externally validating it through approximately 9000 LDL-C results and those of another 4

databases. Our DNN model was established using approximately 18,000 LDL-C results obtained from 2 different institutions, and validation was established using approximately 77,000 LDL-C results, which was comparable to the validation in other studies.

In the study by Martin et al [4] (the Novel method), the median TG distribution was 115 (IQR 82-166). Research by Meeusen et al [25] resulted in a median TG distribution of 131 (IQR 89-196). In a study by Sampson et al [5] (NIH method), the median TG distribution was 149 (IQR 98-253). Our derivation dataset (KNHANES) had a median TG of 120 (IQR 76-211), and our validation dataset had a median TG of 114 (IQR 83-163). Although data from the Novel method had a TG distribution more similar to our validation dataset than the TG distribution from the NIH method, the performances obtained from these methods were almost identical. However, we found that our deep LDL-EHR model generated extremely accurate results for the derivation set and comparably inaccurate results for the testing dataset. In other words, an overfitting problem occurred in our deep LDL-EHR model. Therefore, we adopted a TL method to fine-tune (overall retainment with little change in trained parameters) the deep LDL-EHR (DNN+TL) model, yielding the best performance among all the methods.

Limitations and Future Work

The most important limitation of the present study is the referenced homogenous method used to measure LDL-C. Representative methods for estimating LDL-C [3-5] use the heterogeneous method of ultracentrifugation (eg, beta-quantification) [30,31]. Besides, we implemented the homogeneous precipitation-based (direct) method as the reference for establishing an LDL-C regression model. Nauck et al [30] suggested that the homogenous method satisfied the NCEP requirements and proposed accurate LDL-C results with a coefficient of variation less than 4% and a bias less than 4%. Moreover, the homogenous method seems to have better classified subjects into NCEP criteria than the FW method [30]. The homogenous method does not require the preliminary lipoprotein fractionation step (eg, ultracentrifugation). In other words, it is easy to use and often provides improved precision; therefore, it has gained rapid acceptance worldwide [31]. However, for high-risk CVD patients or groups, future studies should analyze both beta-quantifications and direct methods to provide more accurate and generalized estimates for decreasing CVD-related mortality.

In future studies, we plan to update the trained weights in the LDL-EHR model with optimized parameters using TL. Another study is needed to evaluate the performance of an updated version of the LDL-EHR (DNN+TL) model for the newly selected samples. Furthermore, as suggested by other studies [6,32], it is crucial to develop an LDL-C estimation method that considers demographic, medical, anthropometric, and laboratory phenotypes, such as age, obesity, chronic disease, and liver profiles.

Conclusion

We applied a real-time deep learning model to estimate LDL-C using EHR system data. However, we encountered several unforeseen problems. When applying the DNN model to real patients, our tool could not outperform the other LDL-C estimation methods (ie, Novel and NIH). We overcame this by upgrading our DNN using a TL algorithm (DNN+TL), resulting in superior LDL-C estimation performance compared with the other methods. Our study suggests that the revised version of our deep LDL-EHR (DNN+TL) may contribute to future accurate estimations for LDL-C in real clinical settings.

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI19C1035).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Performances of four LDL estimation methods.
[PDF File (Adobe PDF File), 369 KB-Multimedia Appendix 1]

Multimedia Appendix 2

The distribution of TG:VLDL-C in relation to TG.
[PDF File (Adobe PDF File), 288 KB-Multimedia Appendix 2]

References

1. Ference B, Ginsberg H, Graham I, Ray K, Packard C, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J 2017 Aug 21;38(32):2459-2472 [FREE Full text] [doi: 10.1093/eurheartj/ehx144] [Medline: 28444290]

2. Miller W, Myers G, Sakurabayashi I, Bachmann L, Caudill S, Dziekonski A, et al. Seven direct methods for measuring HDL and LDL cholesterol compared with ultracentrifugation reference measurement procedures. Clin Chem 2010 Jun;56(6):977-986 [FREE Full text] [doi: 10.1373/clinchem.2009.142810] [Medline: 20378768]
3. Friedewald W, Levy R, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972 Jun;18(6):499-502. [Medline: 4337382]
4. Martin SS, Blaha MJ, Elshazly MB, Toth PP, Kwiterovich PO, Blumenthal RS, 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 Nov 20;310(19):2061-2068 [FREE Full text] [doi: 10.1001/jama.2013.280532] [Medline: 24240933]
5. Sampson M, Ling C, Sun Q, Harb R, Ashmaig M, Warnick R, et al. A New Equation for Calculation of Low-Density Lipoprotein Cholesterol in Patients With Normolipidemia and/or Hypertriglyceridemia. JAMA Cardiol 2020 May 01;5(5):540-548 [FREE Full text] [doi: 10.1001/jamacardio.2020.0013] [Medline: 32101259]
6. Lee T, Kim J, Uh Y, Lee H. Deep neural network for estimating low density lipoprotein cholesterol. Clin Chim Acta 2019 Feb;489:35-40. [doi: 10.1016/j.cca.2018.11.022] [Medline: 30448282]
7. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
8. Chen Y, Zhang X, Pan B, Jin X, Yao H, Chen B, et al. A modified formula for calculating low-density lipoprotein cholesterol values. Lipids Health Dis 2010 May 21;9:52 [FREE Full text] [doi: 10.1186/1476-511X-9-52] [Medline: 20487572]
9. de Cordova CMM, de Cordova MM. A new accurate, simple formula for LDL-cholesterol estimation based on directly measured blood lipids from a large cohort. Ann Clin Biochem 2013 Jan 29;50(Pt 1):13-19. [doi: 10.1258/acb.2012.011259] [Medline: 23108766]
10. Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/tkde.2009.191]
11. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604 [FREE Full text] [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
12. Kim YY, Oh SJ, Chun YS, Lee WK, Park HK. Gene expression assay and Watson for Oncology for optimization of treatment in ER-positive, HER2-negative breast cancer. PLoS One 2018 Jul 6;13(7):e0200100 [FREE Full text] [doi: 10.1371/journal.pone.0200100] [Medline: 29979736]
13. Vukotic A, Goodwill J. Apache Tomcat 7. Cham, Switzerland: Springer Publishing Company; 2011.
14. Ronacher A. Flask: web development, one drop at a time. URL: https://flask-doc.readthedocs.io/en/latest/ [accessed 2021-07-23]
15. Wolf D, Henley AJ. Java EE Web Application Primer: Building Bullhorn: A Messaging App with JSP, Servlets, JavaScript, Bootstrap and Oracle. New York, NY: Apress; 2017.
16. McGrath M. HTML, CSS & JavaScript in easy steps. Leamington Spa, UK: In Easy Steps Limited; 2020.
17. Hadjigeorgiou C. RDBMS vs NoSQL: Performance and scaling comparison. University of Edinburgh. 2013. URL: https://static.epcc.ed.ac.uk/dissertations/hpc-msc/2012-2013/RDBMS%20vs%20NoSQL%20-%20Performance%20and%20Scaling%20Comparison.pdf [accessed 2021-07-23]
18. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A System for Large-Scale Machine Learning. 2016 Presented at: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16); November 2-4, 2016; Savannah, GA.
19. Manaswi NK. Understanding and Working with Keras. In: Deep Learning with Applications Using Python. Berkeley, CA: Apress; 2018:31-43.
20. Wadhwa N, Krishnaswamy R. Comparison of LDL-Cholesterol Estimate using Various Formulae with Directly Measured LDL-Cholesterol in Indian Population. J Clin Diagn Res 2016 Dec;10(12):BC11-BC13 [FREE Full text] [doi: 10.7860/JCDR/2016/22272.9018] [Medline: 28208843]
21. Piani F, Cicero AF, Ventura F, Dormi A, Fogacci F, Patrono D, BLIP Study Group. Evaluation of twelve formulas for LDL-C estimation in a large, blinded, random Italian population. Int J Cardiol 2021 May 01;330:221-227. [doi: 10.1016/j.ijcard.2021.02.009] [Medline: 33581176]
22. WCH-AI-LAB: DNN-TL. GitHub. URL: https://github.com/WCH-AI-LAB/DNN-TL [accessed 2021-07-25]
23. Jeong T, Cho E, Lee W, Chun S, Hong K, Min W. Accuracy Assessment of Five Equations Used for Estimating the Glomerular Filtration Rate in Korean Adults. Ann Lab Med 2017 Sep 01;37(5):371-380 [FREE Full text] [doi: 10.3343/alm.2017.37.5.371] [Medline: 28643485]
24. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation,Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002 Dec 17;106(25):3143-3421. [Medline: 12485966]
25. Meeusen J, Lueke A, Jaffe A, Saenger AK. Validation of a proposed novel equation for estimating LDL cholesterol. Clin Chem 2014 Dec;60(12):1519-1523. [doi: 10.1373/clinchem.2014.227710] [Medline: 25336719]

26. Jacob L, Speed TP. The healthy ageing gene expression signature for Alzheimer's disease diagnosis: a random sampling perspective. Genome Biol 2018 Jul 25;19(1):97 [FREE Full text] [doi: 10.1186/s13059-018-1481-6] [Medline: 30045771]
27. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics 2010 May 15;26(10):1340-1347. [doi: 10.1093/bioinformatics/btq134] [Medline: 20385727]
28. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Cornell University. URL: https://arxiv.org/abs/1705.07874 [accessed 2021-07-23]
29. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825-2830.
30. Nauck M, Warnick GR, Rifai N. Methods for measurement of LDL-cholesterol: a critical assessment of direct measurement by homogeneous assays versus calculation. Clin Chem 2002 Feb;48(2):236-254. [Medline: 11805004]
31. Warnick GR, Kimberly MM, Waymack PP, Leary ET, Myers GL. Standardization of Measurements for Cholesterol, Triglycerides, and Major Lipoproteins. Laboratory Medicine 2008 Jul 17;39(8):481-490. [doi: 10.1309/6ul9rhjh1jffu4py]
32. Chakraborty M, Tudu B. Comparison of ANN models to predict LDL level in Diabetes Mellitus type 2. InInternational Conference on Systems in Medicine and Biology; 2011 Presented at: International Conference on Systems in Medicine and Biology; December 16-18, 2010; Kharagpur, India. [doi: 10.1109/icsmb.2010.5735410]

Abbreviations

API: application programming interface
CVD: cardiovascular disease
DNN: deep neural network
EHR: electronic health record
eLDL-C: estimated low-density lipoprotein cholesterol
EMR: electronic medical record
HDL-C: high-density lipoprotein cholesterol
JSP: JAVA Server Pages
KNHANES: Korea National Health and Nutrition Examination Survey
LDL-C: low-density lipoprotein cholesterol
mLDL-C: measured low-density lipoprotein cholesterol
NCEP: National Cholesterol Education Program
NIH: National Institutes of Health
RMSE: root mean square error
SHAP: Shapley addictive explanations
TC: total cholesterol
TG: triglyceride
TL: transfer learning
VLDL-C: very low-density lipoprotein cholesterol
WSCH: Wonju Severance Christian Hospital

Edited by G Eysenbach; submitted 02.04.21; peer-reviewed by T Lee, Z Ren, H Li, PP Zhao; comments to author 23.04.21; revised version received 18.06.21; accepted 05.07.21; published 03.08.21

Please cite as:
Hwang S, Gwon C, Seo DM, Cho J, Kim JY, Uh Y
A Deep Neural Network for Estimating Low-Density Lipoprotein Cholesterol From Electronic Health Records: Real-Time Routine Clinical Application
JMIR Med Inform 2021;9(8):e29331
URL: https://medinform.jmir.org/2021/8/e29331
doi: 10.2196/29331
PMID: