IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 9, 2020, accepted July 22, 2020, date of current version September 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015757

Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)

GIHUN JOO[1], YEONGJIN SONG[1], HYEONSEUNG IM[1], AND JUNBEOM PARK[2]
[1]Department of Computer Science, Kangwon National University, Chuncheon 24341, South Korea
[2]Department of Cardiology, College of Medicine, Ewha Womans University Medical Center, Seoul 07985, South Korea

Corresponding authors: Hyeonseung Im (hsim@kangwon.ac.kr) and Junbeom Park (parkjb@ewha.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Korea Government (MSIT) under Grant NRF-2019R1F1A1063272, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning under Grant NRF-2017R1E1A1A01078382.

ABSTRACT Machine learning (ML) and large-scale big data are key factors in developing an accurate prediction model for cardiovascular disease (CVD). Although the CVD risk often depends on the race and ethnicity, most previous studies considered only US or European populations for the CVD risk prediction. In this work, to complement previous researches, we analyzed the Korean National Health Insurance Service–National Health Sample Cohort (KNHSC) data and studied the characteristics of ML and big data for predicting the CVD risk. More specifically, we assessed the effectiveness of various ML methods in predicting the 2-year and 10-year risk of CVD such as atrial fibrillation, coronary artery disease, heart failure, and strokes. To develop prediction models, we considered the usual medical examination data, questionnaire survey results, comorbidities, and past medication information available in the KNHSC data. We developed various ML-based prediction models using logistic regression, deep neural networks, random forests, and LightGBM, and validated them using various metrics such as receiver operating characteristic curves, precision-recall curves, sensitivity, specificity, and F1 score. Experimental results showed that all ML models outperformed the baseline method derived from the ACC/AHA guidelines for estimating the 10-year CVD risk, demonstrating the usefulness of ML methods. In addition, in our analysis, whether we included the past medication information as a feature or not, the prediction accuracy of all ML models was comparable to each other. Since the use of medications by the physicians provided important information on the occurrence of diseases, when we included it as a feature, all prediction models achieved a slightly higher prediction accuracy.

INDEX TERMS Atrial fibrillation, cardiovascular disease, coronary artery disease, heart failure, stroke, Korean national health insurance data, machine learning.

I. INTRODUCTION

Representative cardiovascular disease (CVD) includes myocardial infarctions, atrial fibrillation, heart failure, and strokes. The occurrence of CVD is affected by various risk factors such as the race, ethnicity, age, sex, weight, height, body mass index, and a blood test result including the kidney function, liver function, and cholesterol levels [1]–[4]. These factors are often intertwined and affect the development of various diseases in a complicated way. Hence, prediction models based on conventional statistical methods often cannot reflect all the complex causal relationships between various risk factors [5], [6].

The recent standardization of medical big data and the systematization of national health examination data have made it possible to analyze previously unknown risk factors that may have a statistically significant association with the occurrence of disease, which may in turn allow us to trace back

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao.

various disease mechanisms. Moreover, a big data analysis is crucial in developing accurate prediction models for the occurrence of disease [7]. Traditionally, various statistical methods have been used to develop prediction models and to discover important risk factors. However, recently, artificial intelligence (AI) and big data have gained a lot of attention and are being increasingly used to develop prediction models for various diseases [5], [6], [8]–[12]. One possible limitation of AI-based prediction models is that since they are often black box models, it is challenging to analyze the causal relationship between risk factors and the occurrence of disease. Nevertheless, by learning complex patterns and regularities from big data, AI-based prediction models often improve the accuracy of predicting the occurrence of disease.

For the CVD risk prediction using machine learning (ML) and big data, most previous work mainly considered only US or European populations [5], [6], [8], even though the CVD risk often depends on the race and ethnicity. In this work, to complement previous researches, we developed various ML models based on logistic regression, deep neural networks, random forests [13], and LightGBM [14] to predict the risk of CVD using systematically organized large-scale nationwide health examination data in Korea [15], [16]. In particular, since most cardiovascular risk factors change constantly and a short-term risk prediction has higher clinical significance from the physicians' clinical perspective, we compared the results of the 2-year and, the usual, 10-year risk predictions of the proposed ML models. Validation results under various metrics showed that all proposed ML models outperformed the baseline method derived from the American Heart Association/American College of Cardiology (ACC/AHA) guidelines for an estimation of the 10-year risk for CVD [17].

In addition, retrospective medical data, on which most previous ML prediction models are based [5], [6], [8], often includes bias of physicians for the use of disease-related medications. To evaluate the effect of the bias, we considered the use of cardiovascular drugs as a feature variable. In our analysis, when the past medication information was exploited, all ML models achieved a slightly higher prediction accuracy, partly because the use of medications by the physicians provided important information on the occurrence of diseases.

The main contributions of the paper are summarized as follows:

- We studied the characteristics of ML and big data, in particular for a less studied Asian population, for the CVD risk prediction.
- We developed various ML-based prediction models and experimentally showed that they outperformed the standard baseline method, thus demonstrating the usefulness of ML methods in predicting CVD.
- We analyzed the effect of the short-term and long-term predictions by comparing the results of the 2-year and 10-year risk predictions.
- We also considered the past cardiovascular medication information to evaluate the physicians' bias in the use of disease-related medications.

The rest of the paper is organized as follows. Section 2 describes the study population and feature variables. It also presents the employed ML methods and experimental settings. Section 3 presents experimental results by comparing the performance of the proposed ML-based prediction models using various metrics. Section 4 discusses the experimental results and related work. Finally, Section 5 concludes.

II. METHODS

A. STUDY POPULATION

In this study, we developed ML-based prediction models for CVD such as atrial fibrillation (AF), coronary artery disease (CAD), heart failure (HF), and strokes by analyzing the Medical Check-up Cohort DB ver 1.0 provided by Korean National Health Insurance Service [15], [16] (NHIS-2016-2-263). This cohort database is a non-personally identifiable research DB consisting of general medical examination results from 2002 to 2013 (12 years) of about 510,000 qualified individuals who were between 40 and 79 years old as of 2002. The baseline year was set to 2003 to consider 10 years of follow-up, as suggested by the ACC/AHA guidelines. Among 310,210 subjects who took a health medical examination (HME) in 2003, we excluded those subjects diagnosed with AF (I48), CAD (I21-25), HF (I50), hemorrhagic stroke (HS) (I60-62), or ischemic stroke (IS) (I63-69) before their 2003 HME result. We also excluded those subjects whose HME data contained a null or an incorrect value. Consequently, the analysis cohort consisted of a total of 297,875 subjects. We conducted 2-year and 10-year follow-up cohort analyses. For each analysis, all subjects were divided into two groups: disease group and no disease group. The former consisted of those patients who were diagnosed with CVD during the follow-up period and the latter consisted of the rest of the subjects (Fig. 1).

B. FEATURE ENGINEERING

Feature extraction was conducted as follows. First, the following twelve features were extracted from HME results: age, sex, blood sugar, body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), gamma-glutamyltransferase (GGT), hemoglobin level (HMG), presence of gross proteinuria (OLIG_PROTE), serum aspartate aminotransferase (SGOT_AST), serum alanine aminotransferase (SGPT_ALT), and total cholesterol. Second, the following five features were derived from the results of the questionnaire survey included in the HME: the exercise and smoking status, estimated total amount of smoking (pack-years), average number of drinking days per week (DRNK_HABIT), and average alcohol consumption. Next, the information on eight diseases was extracted from the accompanied treatment DB, which contained the treatment information for each subject at the clinic. The following

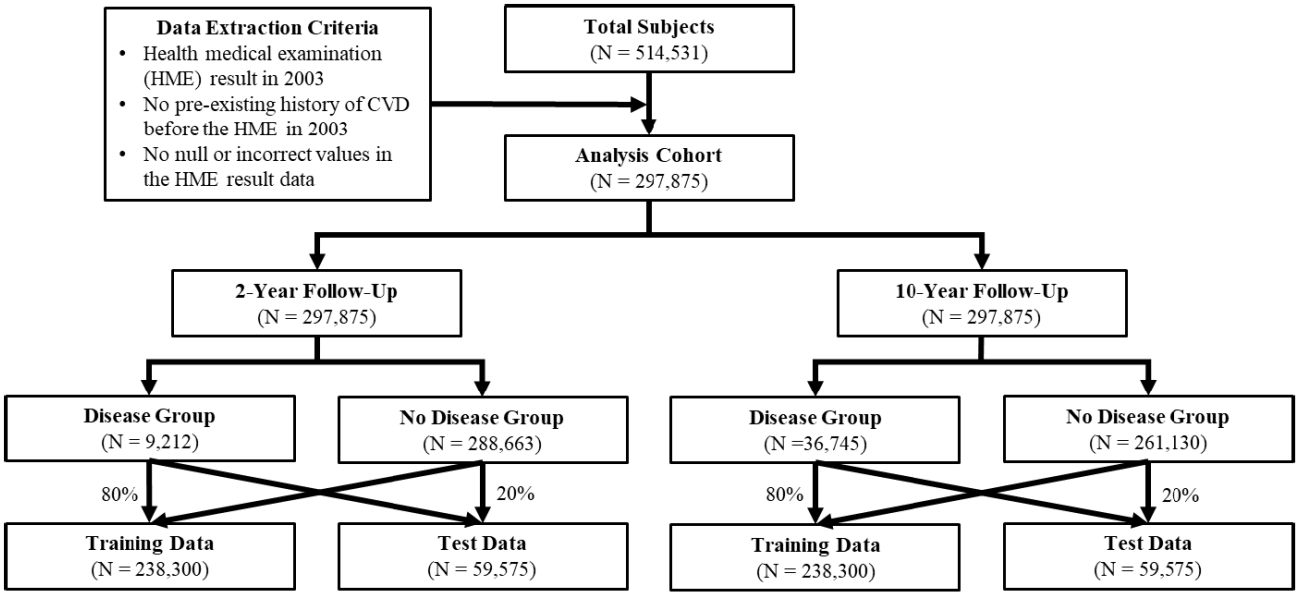


FIGURE 1. Study population and data extraction procedures. The analysis cohort was divided into disease and no disease groups, which were then divided into training and test sets. The training set was used to build a prediction model and the test set was used to validate the model.

diseases were considered: AF, CAD, cancer (C00-96), diabetes mellitus (DM) (E10-14), HF, hypertension (HTN) (I10-15), HS, and IS. When predicting the risk of CVD such as AF, CAD, HF, HS, and IS, the remaining three comorbidities, i.e., cancer, DM, and HTN, were exploited as binary input features of the prediction models. In particular, each comorbidity was considered only if the subject had been hospitalized more than once or had visited the clinic more than twice due to a comorbidity before the HME date. Lastly, we considered whether the subject took cardiovascular drugs such as antiarrhythmics, anticoagulants, antiplatelets, cardiotonics, and statins before the HME date. Note that all the subjects were free from CVD before their baseline HME. In total, 25 input variables and one target variable were considered.

C. DATASETS

From the analysis cohort, we constructed two datasets: one for the 2-year follow-up analysis and the other for the 10-year follow-up analysis. Both datasets consisted of the same subjects. The main difference is that for the former dataset, those who were diagnosed with CVD more than two years after their HME date were classified to the no disease group since we considered only 2 years of follow-up for this dataset. For each dataset, all subjects were exclusively divided into training and test sets. More specifically, for the training set, 80% of the subjects were randomly selected from both disease and no disease groups. The remaining 20% were used as the test set. The training set was used to build prediction models and the test set to evaluate the models. Fig. 1 illustrates the study population and data preprocessing steps and Table 1 summarizes the baseline characteristics of the two datasets, where we omit multi-valued categorical variables such as OLIG_PROTE, the smoking status, and the average number of drinking days per week.

D. LOGISTIC REGRESSION

Logistic regression (LR) is a linear regression model that estimates the value of a binary dependent (target) variable as a linear combination of independent variables (features). More specifically, given the independent variables $x_1, x_2, \ldots, x_n$, the dependent variable y is defined as follows:

$$y = \sigma(w_1x_1 + w_2x_2 + \ldots + w_nx_n + b)$$

where σ is a sigmoid function such that $\sigma(t) = 1/(1 + e^{-t})$. That is, LR models the risk of the target disease as a probability between 0 and 1. By using the training set, the best values for parameters $w_1, w_2, \ldots, w_n, b$ are computed such that the root-mean-square error is minimized. To develop an LR model, we used the LogisticRegression class provided in the Python Scikit-learn library (https://scikit-learn.org/stable/) with the gradient descent optimizer.

E. DEEP NEURAL NETWORK

A deep neural network (DNN) is one of the most widely used deep learning methods, which can encode a nonlinear relationship between independent and dependent variables. A DNN normally consists of an input layer, several hidden layers, and an output layer (Fig. 2). Each layer consists of several nodes each of which is connected to every node in the previous and next layers, i.e., fully connected. The output of each node is basically defined to be a linear combination of the output values from the previous layer followed by an application of a nonlinear activation function. To develop a DNN model, we used Python 3 and the Keras 2.1.6 deep

TABLE 1. Baseline characteristics of the dataset.

Feature	2-Year Follow-Up			10-Year Follow-Up		
	All (N=297,875)	CVD (N=9,212)	No CVD (N=288,633)	All (N=297,875)	CVD (N=36,745)	No CVD (N=261,130)
Age (years)	52.44 ± 9.53	60.68 ± 9.82	52.18 ± 9.41	52.44 ± 9.53	59.52 ± 9.82	51.45 ± 9.06
Male	164,001 (0.55)	4,830 (0.52)	159,171 (0.55)	164,001 (0.55)	19,687 (0.54)	144,314 (0.55)
Blood sugar (mg/dL)	98.32 ± 32.72	104.04 ± 42.11	98.14 ± 32.35	98.32 ± 32.72	103.71 ± 40.89	97.57 ± 31.32
BMI (kg/m^2)	23.98 ± 2.99	24.3 ± 3.14	23.97 ± 2.99	23.98 ± 2.99	24.28 ± 3.14	23.94 ± 2.97
SBP (mmHg)	127.56 ± 18.16	134.66 ± 20.63	127.33 ± 18.03	127.56 ± 18.16	133.51 ± 19.51	126.72 ± 17.81
DBP (mmHg)	79.72 ± 11.65	82.42 ± 12.57	79.63 ± 11.61	79.72 ± 11.65	81.98 ± 12.11	79.4 ± 11.55
GGT (U/L)	37.76 ± 55.15	39.57 ± 59.64	37.7 ± 55	37.76 ± 55.15	40.16 ± 61.5	37.42 ± 54.19
HMG (g/dL)	13.93 ± 1.52	13.83 ± 1.51	13.93 ± 1.52	13.93 ± 1.52	13.88 ± 1.5	13.94 ± 1.52
SGOT_AST (U/L)	26.7 ± 17.44	27.47 ± 16.45	26.67 ± 17.47	26.7 ± 17.44	27.63 ± 18.28	26.57 ± 17.32
SGPT_ALT (U/L)	25.49 ± 20.3	25.14 ± 18.52	25.5 ± 20.35	25.49 ± 20.3	25.48 ± 19.76	25.49 ± 20.37
Total cholesterol (mg/dL)	200.18 ± 37.78	203.21 ± 40.45	200.08 ± 37.69	200.18 ± 37.78	203.65 ± 39.57	199.69 ± 37.5
Exercise	128,174 (0.43)	3,489 (0.38)	124,685 (0.43)	128,174 (0.43)	13,737 (0.37)	114,437 (0.44)
Drinking amount (50ml glass)	2.25 ± 3.41	1.58 ± 3	2.27 ± 3.42	2.25 ± 3.41	1.8 ± 3.16	2.31 ± 3.44
Smoking amount (pack-year)	88.86 ± 201.57	78.61 ± 198.06	89.19 ± 201.67	88.86 ± 201.57	89.25 ± 208.32	88.81 ± 200.6
Comorbidity						
Cancer	5,015 (0.02)	253 (0.03)	4,762 (0.02)	5,015 (0.02)	881 (0.02)	4,134 (0.02)
Diabetes mellitus	19,720 (0.07)	1,202 (0.13)	18,518 (0.06)	19,720 (0.07)	4,496 (0.12)	15,224 (0.06)
Hypertension	45,482 (0.15)	3,340 (0.36)	42,142 (0.15)	45,482 (0.15)	11,124 (0.3)	34,358 (0.13)
Medication						
Antiarrhythmic	266 (0.001)	77 (0.01)	189 (0.001)	266 (0.001)	157 (0.004)	109 (0.0004)
Anticoagulant	433 (0.001)	154 (0.02)	279 (0.001)	433 (0.001)	296 (0.01)	137 (0.001)
Antiplatelet	20,867 (0.07)	2,388 (0.26)	18,479 (0.06)	20,867 (0.07)	5,985 (0.16)	14,882 (0.06)
Cardiotonic	22,622 (0.08)	2,093 (0.23)	20,529 (0.07)	22,622 (0.08)	6,099 (0.17)	16,523 (0.06)
Statin	9,088 (0.03)	769 (0.08)	8,319 (0.03)	9,088 (0.03)	2,212 (0.06)	6,876 (0.03)
CVD						
Atrial fibrillation	880 (0.003)	880 (0.096)	-	4,722 (0.016)	4,722 (0.129)	-
Coronary artery disease	1,141 (0.004)	1,141 (0.124)	-	4,646 (0.016)	4,646 (0.126)	-
Heart failure	1,669 (0.006)	1,669 (0.181)	-	7,184 (0.024)	7,184 (0.196)	-
Hemorrhagic stroke	949 (0.003)	949 (0.103)	-	4,140 (0.014)	4,140 (0.113)	-
Ischemic stroke	5,179 (0.017)	5,179 (0.562)	-	25,052 (0.084)	25,052 (0.682)	-

Values denote means ± standard deviation (SD) or the n (%). BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; GGT, gamma-glutamyltransferase; HMG, hemoglobin level; SGOT_AST, serum aspartate aminotransferase; SGPT_ALT, serum alanine aminotransferase.

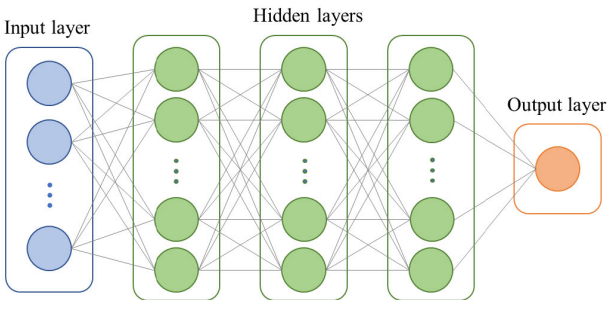


FIGURE 2. The DNN model used in the experiment consists of an input layer, three hidden layers, and an output layer.

learning library (https://keras.io/) with TensorFlow [18] as a backend. In particular, the hyperparameters of the DNN model were determined by using a grid search followed by a manual search. We varied the number of hidden layers (from 1 layer to 20 layers), number of nodes in each layer (from 10 nodes to 200 nodes), batch size, learning rate, activation function, and regularization methods. To this end, we used only 80% of the training data for learning the model and the remaining 20% as validation data. Then, the hyperparameters that led to the best performance on the validation data were used in the experiments. More specifically, we used three hidden layers each of which were composed of a fully connected layer with 30 nodes, followed by a batch normalization [19], ELU activation function [20], and dropout [21]. In the experimental results, the prediction performance of the DNN model was measured using the test data. Since the performance of the DNN model slightly varies with the initial parameter values, all measurements for the DNN model were averaged over 10 sample runs.

F. RANDOM FORESTS

A random forest (RF) is a widely used ensemble learning method based on a bagging (bootstrap aggregating) approach [13]. It consists of multiple decision trees each of

which is built from a random sample drawn with replacement (called a bootstrap sample) from the training set, where the sample size is the same as the training set size. When each tree is built, only a random subset of features is used to split each node. Then, the prediction of the RF model is given as the averaged prediction of all decision trees to reduce the variance of the model and produce an overall better model. To develop an RF model, we used the RandomForestClassifier class provided in the Python Scikit-learn library. The hyperparameters of the RF model were determined by a grid search and 5-fold cross-validation on the training data. We varied the number of decision trees to build the RF model (from 100 to 1,000), maximum size of a random subset of features used to build each tree, maximum depth of each tree, and criterion function to measure the quality of a split.

G. LIGHTGBM

LightGBM is a fast and high-performance gradient boosting framework based on tree-based learning algorithms [14]. The gradient boosting decision tree (GBDT) [22] is also a widely used ensemble learning method like bagging, but it sequentially builds decision trees to reduce the bias of the model. Specifically, a new tree is built in such a way that it reduces the prediction error of the previously built trees. LightGBM is an efficient and effective implementation of GBDT with techniques called gradient-based one-side sampling and exclusive feature bundling. To develop a LightGBM model, we used the LGBMClassifier class provided in the Python LightGBM package (https://lightgbm.readthedocs.io). As for the RF model, the hyperparameters of the LightGBM model were determined by a grid search and 5-fold cross-validation on the training data. We varied the boosting type, number of boosted trees (from 50 to 200), maximum number of tree leaves, boosting learning rate, subsample ratio of the columns when constructing each tree, and regularization terms.

H. EXPERIMENTAL SETTINGS

All experiments were conducted on a workstation equipped with two octa-core Intel®Xeon®E5-2630 v3 2.40 GHz CPUs, 96 GB of main memory, and an NVIDIA GeForce GTX 1080Ti GPU with 11 GB memory. The host operating system was Ubuntu 16.04.3 LTS (64-bit) and all prediction models were implemented using Python 3, the Scikit-learn machine learning library, and the Keras 2.1.6 deep learning library. For the 2-year and 10-year follow-up analysis data, we implemented ML-based prediction models using LR, DNN, RF, and LightGBM with and without features for medications. To validate their effectiveness, we compared them with the baseline method derived from the ACC/AHA guidelines for estimation of 10-year risk for CVD [17]. More specifically, the baseline method was a simple logistic regression model built using the following seven features: age, sex, SBP, total cholesterol, smoking status, DM, and HTN. The performance of all prediction models was compared using the following metrics: the receiver operating characteristic (ROC) curves, precision-recall (PR) curves, sensitivity, specificity, and F1 score. Moreover, to find the most important features for estimating the risk of CVD, we analyzed the feature importance for each prediction model using the Shapley additive explanations (SHAP) method which is a game-theoretic, unified method to explain the output of any ML model [23].

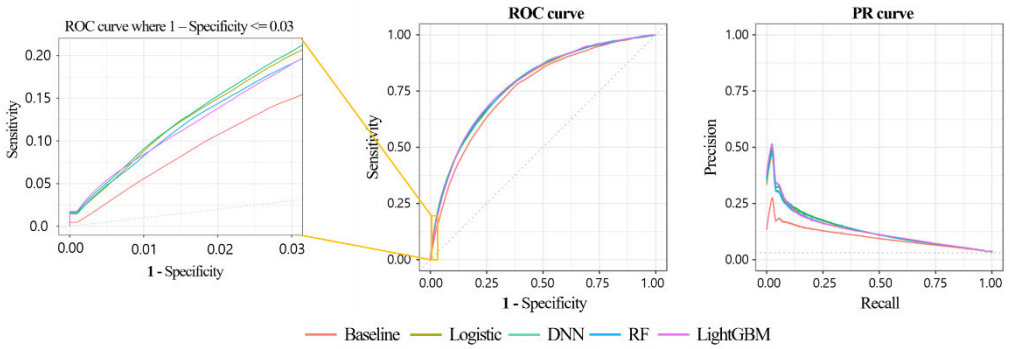
III. RESULTS

A. PREDICTION PERFORMANCE

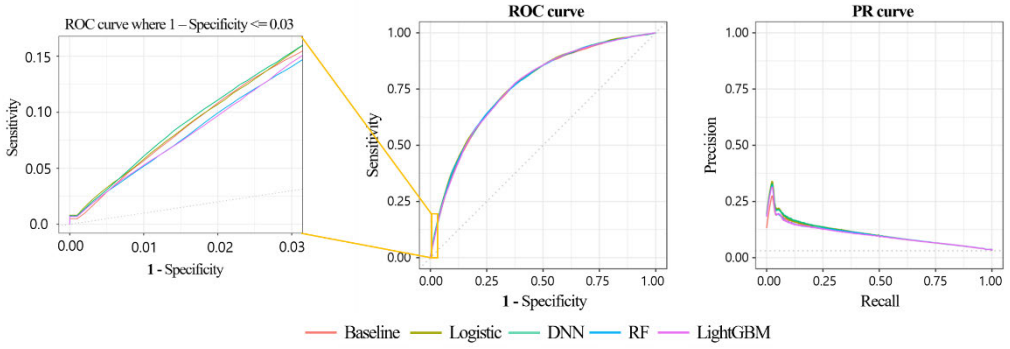
Figs. 3 and 4 show the performance of every prediction model, including the baseline model, for the test datasets in terms of the ROC and PR curves. Fig. 3 shows the results of the 2-year CVD risk prediction with and without medication features while Fig. 4 shows the results of the 10-year risk prediction. We also included the ROC curve regions of interest [24], that is, regions where the false positive rate, i.e., 1 – specificity, is less than or equal to the imbalance ratio of the dataset: 0.03 for the 2-year follow-up dataset and 0.125 for the 10-year follow-up dataset. Table 2 presents the area under an ROC curve (AUROC), area under a PR curve (AUPRC), sensitivity, specificity, and F1 score of each prediction model.

First, as shown in Figs. 3 and 4, every ML-based prediction model outperformed the baseline model in terms of ROC and PR curves, regardless of whether medication features were used or not. In particular, as shown in Table 2, with medication features, DNN achieved the best AUROC values, which are 2.36% and 1.31% higher than those of the baseline for 2-year and 10-year risk prediction, respectively. Meanwhile, with medication features, LightGBM achieved the best AUPRC values, which are 3.45% and 2.59% higher than those of the baseline, for 2-year and 10-year risk prediction, respectively. Moreover, the performance of the proposed ML models was mostly comparable to each other. In particular, there was no clear winner in terms of sensitivity, specificity, and F1 score as shown in Table 2.

Second, on the one hand, every ML-based prediction model achieved better performance for 2-year CVD risk prediction than 10-year risk prediction in terms of AUROC values and sensitivity. For example, with medication features, ML models achieved 3.04%-3.59% higher AUROC values for 2-year risk prediction than 10-year risk prediction. On the other hand, every ML-based prediction model achieved better performance for 10-year risk prediction than 2-year risk prediction in terms of AUPRC values and F1 scores. For example, without medication features, ML models achieved 18.37%-19.21% higher AUPRC values for 10-year risk prediction than 2-year risk prediction. The reason for this big difference in AUPRC values and F1 scores is that the 2-year follow-up dataset is more highly imbalanced than the 10-year follow-up dataset. Specifically, the subjects having CVD are only 3.09% in the former dataset whereas 12.34% in the latter dataset. Thus, the precision of 2-year risk prediction was much lower than that of 10-year risk prediction, and consequently so were AUPRC values and



(a) 2-year CVD risk prediction with medication features. Every ML-based prediction model outperforms the baseline model.



(b) 2-year CVD risk prediction without medication features. ML-based prediction models generally outperform the baseline model although all prediction models are comparable to each other under the ROC curve region of interest (leftmost figure).

FIGURE 3. ROC and PR curves of each prediction model for the 2-year CVD risk prediction.

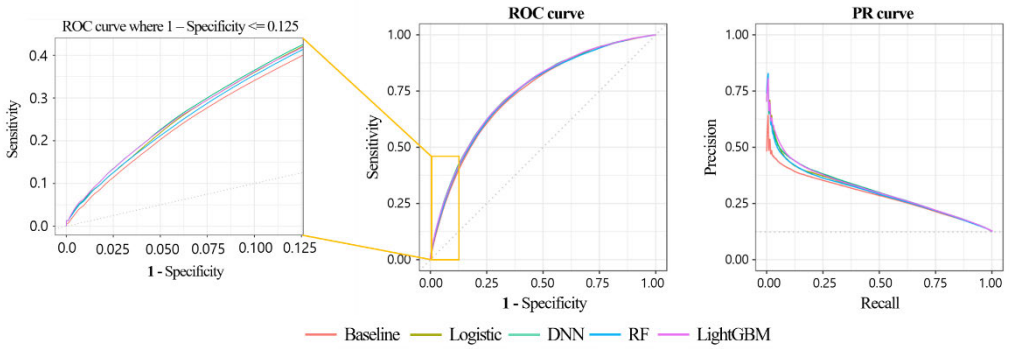
TABLE 2. Performance of all prediction models under various metrics.

Dataset	Model	AUROC	AUPRC	Sensitivity	Specificity	F1 Score
2-Year Follow-Up	LR	0.7800 ± (0.7781-0.7818)	0.1348 ± (0.1326-0.1370)	0.73	0.69	**0.13**
	DNN	**0.7834** ± (0.7816-0.7853)	0.1351 ± (0.1329-0.1373)	**0.76**	0.67	0.12
	RF	0.7813 ± (0.7794-0.7832)	0.1310 ± (0.1289-0.1331)	0.70	0.70	**0.13**
	LightGBM	0.7820 ± (0.7802-0.7839)	**0.1356** ± (0.1335-0.1376)	0.71	**0.72**	**0.13**
2-Year Follow-Up (without medication features)	LR	**0.7649** ± (0.7630-0.7668)	**0.1065** ± (0.1049-0.1081)	0.73	0.67	**0.12**
	DNN	0.7626 ± (0.7607-0.7645)	0.1063 ± (0.1047-0.1079)	0.74	0.66	**0.12**
	RF	0.7644 ± (0.7625-0.7663)	0.1040 ± (0.1024-0.1056)	**0.75**	0.68	**0.12**
	LightGBM	0.7608 ± (0.7589-0.7627)	0.1007 ± (0.0992-0.1021)	0.73	0.66	**0.12**
	Baseline	0.7598 ± (0.7579-0.7618)	0.1011 ± (0.0995-0.1026)	0.70	**0.70**	**0.12**
10-Year Follow-Up	LR	0.7496 ± (0.7486-0.7506)	0.3078 ± (0.3061-0.3095)	**0.70**	0.68	0.35
	DNN	**0.7512** ± (0.7502-0.7522)	0.3112 ± (0.3095-0.3129)	**0.70**	0.68	0.35
	RF	0.7454 ± (0.7443-0.7464)	0.3033 ± (0.3016-0.3050)	0.67	0.70	0.35
	LightGBM	0.7508 ± (0.7498-0.7519)	**0.3114** ± (0.3096-0.3131)	0.67	**0.71**	**0.36**
10-Year Follow-Up (without medication features)	LR	0.7440 ± (0.7430-0.7451)	0.2902 ± (0.2886-0.2918)	**0.70**	0.67	**0.35**
	DNN	0.7446 ± (0.7436-0.7456)	0.2906 ± (0.2890-0.2923)	0.69	0.68	**0.35**
	RF	0.7414 ± (0.7403-0.7424)	0.2885 ± (0.2868-0.2901)	0.66	**0.70**	**0.35**
	LightGBM	**0.7455** ± (0.7445-0.7465)	**0.2928** ± (0.2911-0.2945)	0.68	**0.70**	**0.35**
	Baseline	0.7381 ± (0.7371-0.7392)	0.2855 ± (0.2839-0.2872)	0.69	0.67	**0.35**

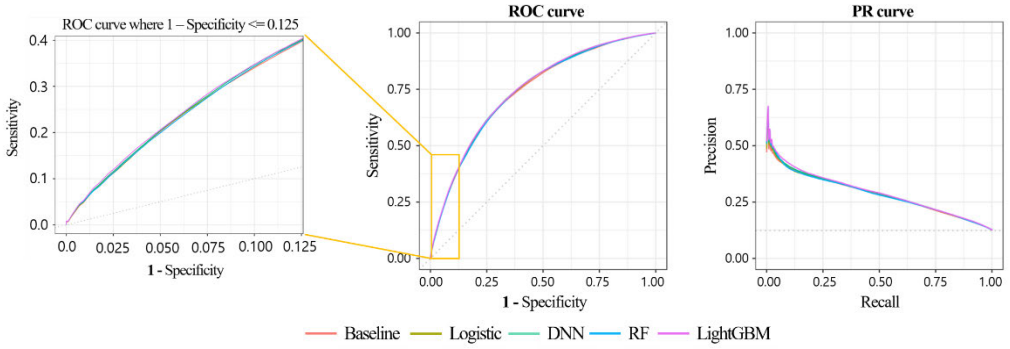
Values for AUROC and AUPRC denote the mean ± confidence interval (CI). AUROC, the area under a receiver operating characteristic curve; AUPRC, the area under a precision-recall curve; Precision = TP / (TP + FP), Sensitivity (Recall) = TP / (TP + FN), Specificity = TN / (TN + FP) where TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative; F1 score = 2(precision * recall) / (precision + recall). Thresholds to calculate sensitivity, specificity, and F1 score were chosen to maximize the geometric mean score, i.e., $\sqrt{\text{sensitivity} * \text{specificity}}$. Boldface numbers denote the best result for each metric for each dataset.

F1 scores. Still, the AUROC values and sensitivity of 2-year risk prediction were much higher than those of 10-year risk prediction.

Finally, in general, the accuracy of every prediction model was higher when considering medication features for both 2-year and 10-year follow-up datasets. For example,



(a) 10-year CVD risk prediction with medication features. Every ML-based prediction model outperforms the baseline model.



(b) 10-year CVD risk prediction without medication features. ML-based prediction models generally outperform the baseline model although all prediction models are comparable to each other under the ROC curve region of interest (leftmost figure).

FIGURE 4. ROC and PR curves of each prediction model for the 10-year CVD risk prediction.

LightGBM achieved a 2.12% higher AUROC value and a 3.49% higher AUPRC value with medication features than those obtained without medication features for the 2-year follow-up dataset. In particular, medication features were more effective for 2-year CVD risk prediction than 10-year risk prediction.

B. FEATURE IMPORTANCE

To estimate the contribution of each feature to the prediction, we analyzed the feature importance of each prediction model using the SHAP method [23]. Tables 3 and 4 show SHAP feature importance for each model for the 2-year and 10-year follow-up training datasets, respectively, where features with larger Shapley values are more important. Due to the high computational cost, the Shapley values for the DNN and RF models were computed using bootstrapping; thus, in Tables 3 and 4, the Shapley values should not be compared across different prediction models.

From the results, we make the following observations. First, each ML-based prediction model differently utilized the input features, due to their different characteristics. That is, an important feature in one prediction model was not necessarily important in another model. For example, in Table 3, exercise was the second most important feature for the LR model, but rarely utilized for the RF and LightGBM models. Second, among the seven features, namely, age, sex, SBP, total cholesterol, smoking status, DM, and HTN, suggested in the ACC/AHA guidelines, only age, SBP, and HTN were effectively utilized in every prediction model for both 2-year and 10-year follow-up datasets; they were all ranked in the top 11 important features. Third, medication features were more effectively utilized for 2-year CVD risk prediction than 10-year risk prediction. In particular, for 2-year risk prediction, antiplatelet and cardiotonic were included in the top 4 important features of DNN, RF, and LightGBM, and in the top 9 important features of LR. Finally, BMI, DRNK_HABIT, and exercise were more important for 10-year risk prediction than 2-year risk prediction; they were all ranked in the top 9 important features in Table 4 except that DRNK_HABIT was identified as the top 13 important feature of LightGBM.

IV. DISCUSSION

In this study, we developed various ML-based prediction models for estimating 2-year and 10-year risk of CVD by analyzing the Korean National Health Insurance Service–National Health Sample Cohort (KNHSC) data. Specifically, we developed prediction models based on LR, DNN, RF, and LightGBM and compared them with the baseline method derived from the ACC/AHA guidelines. We trained the ML-based prediction models with and without

TABLE 3. SHAP feature importance of each prediction model for 2-Year CVD risk prediction.

	LR	DNN	RF	LightGBM
1	Sex (1.0503)	Age (0.0157)	Age (0.0372)	Age (0.6444)
2	Exercise (1.0409)	Antiplatelet (0.0069)	Antiplatelet (0.018)	Antiplatelet (0.178)
3	Smoking status (0.5763)	Cardiotonic (0.0031)	HTN (0.006)	HTN (0.0932)
4	Age (0.5413)	DRNK_HABIT (0.0031)	Cardiotonic (0.0056)	Cardiotonic (0.0914)
5	DRNK_HABIT (0.3511)	HTN (0.0028)	SBP (0.004)	GGT (0.0803)
6	Antiplatelet (0.1172)	SBP (0.0028)	GGT (0.0031)	SBP (0.0778)
7	OLIG_PROTE (0.0677)	Smoking status (0.0018)	BMI (0.0021)	BMI (0.0611)
8	SBP (0.0636)	Sex (0.0014)	TOT_CHOLE (0.0018)	DRNK_HABIT (0.0604)
9	Cardiotonic (0.0603)	OLIG_PROTE (0.0013)	Blood sugar (0.0017)	Blood sugar (0.0394)
10	Smoking amount (0.049)	GGT (0.001)	HMG (0.0015)	TOT_CHOLE (0.0334)
11	HTN (0.0413)	Smoking amount (0.001)	DBP (0.0014)	DBP (0.032)
12	DBP (0.0403)	DM (0.001)	SGPT_ALT (0.001)	Smoking amount (0.0317)
13	GGT (0.0298)	BMI (0.0009)	Anticoagulant (0.001)	HMG (0.0261)
14	BMI (0.0269)	Statin (0.0009)	SGOT_AST (0.0009)	SGPT_ALT (0.0249)
15	Statin (0.0183)	Exercise (0.0007)	OLIG_PROTE (0.0009)	DM (0.0213)
16	DM (0.0179)	Drinking amount (0.0005)	Statin (0.0008)	Drinking amount (0.0211)
17	TOT_CHOLE (0.0111)	Anticoagulant (0.0005)	Smoking amount (0.0008)	SGOT_AST (0.0192)
18	Drinking amount (0.0109)	DBP (0.0004)	DRNK_HABIT (0.0008)	Statin (0.0175)
19	SGOT_AST (0.0105)	SGPT_ALT (0.0004)	DM (0.0006)	Exercise (0.0139)
20	Blood sugar (0.0105)	SGOT_AST (0.0004)	Smoking status (0.0005)	Smoking status (0.0135)
21	SGPT_ALT (0.0065)	HMG (0.0003)	Exercise (0.0004)	OLIG_PROTE (0.0126)
22	Anticoagulant (0.0058)	Blood sugar (0.0003)	Drinking amount (0.0003)	Anticoagulant (0.0066)
23	Cancer (0.0048)	Antiarrhythmic (0.0003)	Antiarrhythmic (0.0003)	Antiarrhythmic (0.0029)
24	HMG (0.0034)	TOT_CHOLE (0.0002)	Sex (0.0001)	Sex (0.0028)
25	Antiarrhythmic (0.0025)	Cancer (0.0001)	Cancer (0.0001)	Cancer (0.0018)

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; GGT, gamma-glutamyltransferase; HMG, hemoglobin level; OLIG_PROTE, presence of gross proteinuria; SGOT_AST, serum aspartate aminotransferase; SGPT_ALT, serum alanine aminotransferase; TOT_CHOLE, total cholesterol; DRNK_HABIT, the average number of drinking days per week; DM, diabetes mellitus; HTN, hypertension.

past cardiovascular medication information and compared their performance under various metrics. Every ML model achieved higher prediction accuracy with the medication features than without them, and significantly outperformed the baseline method when trained with the medication features. The SHAP feature importance analysis in Tables 3 and 4 also confirmed that the pre-existing use of cardiovascular drugs such as antiplatelets and cardiotonics was an important feature variable. With the medication features, for both 2-year and 10-year CVD risk prediction, the DNN model achieved the highest AUROC values, whereas the LightGBM model achieved the highest AUPRC values. Still, all the ML-based prediction models were comparable to each other in general in terms of AUROC, AUPRC, sensitivity, specificity, and F1 score.

Several previous studies also analyzed the KNHSC data to assess the risk of various cardiovascular events, but they mainly used statistical analyses [25]–[29]. Recently, many researchers have extensively been studying the use of ML and deep learning methods to build more accurate prediction models [5], [6], [8], [10]–[12], [30]–[35], mostly by analyzing medical images and wave signal data such as those obtained from MRI, CT, and electrocardiography. However, most of them only considered a much smaller number of subjects, from hundreds to tens of thousands. In contrast, in this study, 297,875 subjects were analyzed and thus the results can be applied more generally. Our work is similar in spirit to [5], [6] in that they also analyzed clinical big data and compared various ML-based prediction models for 5-year or 10-year risk of CVD. The main difference is that [5], [6] analyzed European populations, whereas we analyzed Korean populations, and thus our work is complementary to [5], [6] from the perspective of race and ethnicity. Moreover, in this work, to analyze the effect of short-term and long-term prediction, we compared the results of 2-year and 10-year risk prediction. In addition, we further analyzed the effect of past medication information on the prediction accuracy using SHAP feature importance, which has not been considered in [5], [6].

The prediction accuracy and performance of ML models vary depending on the data used. For example, a DNN-based model significantly outperformed LR- and RF-based models in [32], whereas an RF-based model outperformed other ML-based models including LR and neural network models in [36], [37]. In this study, the performance of the proposed ML models was mostly comparable to each other under various metrics. In many applications, DNN often achieves a higher prediction accuracy than LR and tree-based models such as RF and LightGBM. However, it is not the case in this study partly because the KNHSC dataset is a simple tabular dataset and thus it seems that there exists no particular complex nonlinear relationship between the features considered here. If we also considered spatial or sequential data such as images, regular laboratory data, and electrocardiograms,

TABLE 4. SHAP feature importance of each prediction model for 10-Year CVD risk prediction.

	LR	DNN	RF	LightGBM
1	Sex (0.8675)	Age (0.0627)	Age (0.1176)	Age (0.6245)
2	Exercise (0.8594)	HTN (0.0092)	HTN (0.0287)	HTN (0.0965)
3	Age (0.5976)	DRNK_HABIT (0.0092)	Antiplatelet (0.0219)	SBP (0.0888)
4	Smoking status (0.4593)	Exercise (0.0083)	SBP (0.0214)	BMI (0.0746)
5	DRNK_HABIT (0.3189)	Antiplatelet (0.0081)	Cardiotonic (0.0164)	Antiplatelet (0.0677)
6	SBP (0.0672)	SBP (0.0076)	DRNK_HABIT (0.0127)	GGT (0.0662)
7	HTN (0.0633)	Cardiotonic (0.0075)	Blood sugar (0.0126)	Cardiotonic (0.0504)
8	BMI (0.061)	Smoking status (0.0069)	BMI (0.0117)	Exercise (0.0485)
9	Antiplatelet (0.058)	BMI (0.0053)	Exercise (0.0116)	Smoking amount (0.0453)
10	OLIG_PROTE (0.0485)	DM (0.0041)	DBP (0.0109)	Blood sugar (0.0348)
11	Smoking amount (0.0426)	Smoking amount (0.0031)	DM (0.0095)	DBP (0.0326)
12	Cardiotonic (0.041)	GGT (0.003)	TOT_CHOLE (0.0095)	DM (0.0301)
13	DM (0.0353)	OLIG_PROTE (0.0029)	GGT (0.0094)	DRNK_HABIT (0.0292)
14	HMG (0.026)	HMG (0.0026)	HMG (0.0087)	TOT_CHOLE (0.0287)
15	TOT_CHOLE (0.025)	Sex (0.0026)	Smoking status (0.0084)	Smoking status (0.0282)
16	DBP (0.0246)	Blood sugar (0.0023)	Smoking amount (0.0075)	Drinking amount (0.026)
17	GGT (0.0237)	DBP (0.002)	Drinking amount (0.0067)	HMG (0.0249)
18	Blood sugar (0.0177)	TOT_CHOLE (0.002)	SGOT_AST (0.0067)	SGPT_ALT (0.0179)
19	Statin (0.0121)	Statin (0.0018)	SGPT_ALT (0.0061)	SGOT_AST (0.012)
20	Drinking amount (0.0075)	SGPT_ALT (0.0013)	OLIG_PROTE (0.0043)	Statin (0.0112)
21	SGOT_AST (0.0064)	Drinking amount (0.0013)	Statin (0.0041)	OLIG_PROTE (0.0082)
22	Anticoagulant (0.0064)	Anticoagulant (0.0009)	Sex (0.0037)	Sex (0.0067)
23	SGPT_ALT (0.0029)	Antiarrhythmic (0.0005)	Anticoagulant (0.0018)	Anticoagulant (0.0051)
24	Antiarrhythmic (0.0028)	SGOT_AST (0.0003)	Cancer (0.0009)	Antiarrhythmic (0.0019)
25	Cancer (0.0027)	Cancer (0.0003)	Antiarrhythmic (0.0007)	Cancer (0.001)

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; GGT, gamma-glutamyltransferase; HMG, hemoglobin level; OLIG_PROTE, presence of gross proteinuria; SGOT_AST, serum aspartate aminotransferase; SGPT_ALT, serum alanine aminotransferase; TOT_CHOLE, total cholesterol; DRNK_HABIT, the average number of drinking days per week; DM, diabetes mellitus; HTN, hypertension.

then DNN would be much more effective than other models. In such cases, convolutional neural networks or recurrent neural networks can even be more effective [12], [30], [31], [35].

Some clinically interesting points were also found in our study. When we excluded the past cardiovascular medication information, the prediction accuracy of all prediction models significantly decreased. Note that the use of medication is an important part of a physician's judgment on a patient. If a doctor examines a patient and determines that CVD is likely to occur based on various clinical data, he/she may prescribe cardiovascular drugs as part of a preventive treatment. In other words, the use of such medications is the result of a doctor's analysis of various interrelated risk factors and the occurrence of diseases. Therefore, the prediction accuracy of all ML models was significantly improved when we included the medication features (see Figs. 3 and 4 and Table 2). Since we have found a paradoxical regularity that the incidence of CVD increases when cardiovascular drugs were used, we also developed prediction models using a dataset without the medication features to avoid the bias of a doctor's judgment and to check the performance of the ML models themselves (see Figs. 3 and 4 and Table 2). The SHAP feature importance analysis in Tables 3 and 4 also confirmed our finding. For example, the DNN, RF, and LightGBM models included the use of antiplatelets and cardiotonics in the top 4 important features among 25 input features for 2-year risk prediction. Moreover, the medication features were more effective for 2-year risk prediction than 10-year risk prediction. That is, the importance of past medication information was degraded for long-term prediction.

V. CONCLUSION

In this study, we analyzed the Korean National Health Sample Cohort big data and developed various ML-based prediction models to estimate the 2-year and 10-year risk of CVD. When we included the past medication information as input features, all proposed ML models significantly outperformed the baseline method derived from the ACC/AHA guidelines for estimating the 10-year risk for CVD, thus demonstrating the effectiveness of the ML methods in predicting CVD. However, whether the medication features were used or not, the performance of all ML models was mostly comparable to each other. Therefore, as future work, it will be interesting to investigate a more effective ML method for the CVD risk prediction. Meanwhile, since the use of medications by the physicians provided important information on the occurrence of diseases, when we included the past medication information as input features, all ML models achieved a higher prediction accuracy. In particular, the past medication information was more effective for a short-term prediction than a long-term prediction.

REFERENCES

[1] P. A. Wolf, R. D. Abbott, and W. B. Kannel, "Atrial fibrillation as an independent risk factor for stroke: The framingham study," *Stroke*, vol. 22, no. 8, pp. 983–988, Aug. 1991.

[2] L. A. Simons, J. McCallum, Y. Friedlander, and J. Simons, "Risk factors for ischemic stroke: Dubbo study of the elderly," *Stroke*, vol. 29, no. 7, pp. 1341–1346, Jul. 1998.
[3] T. J. Wang, M. G. Larson, D. Levy, R. S. Vasan, E. P. Leip, P. A. Wolf, R. B. D'Agostino, J. M. Murabito, W. B. Kannel, and E. J. Benjamin, "Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: The framingham heart study," *Circulation*, vol. 107, no. 23, pp. 2920–2925, Jun. 2003, doi: 10.1161/01.CIR.0000072767.89944.6E.
[4] M. F. Piepoli *et al.*, "2016 European guidelines on cardiovascular disease prevention in clinical practice: The sixth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts)developed with the special contribution of the European association for cardiovascular prevention & rehabilitation (EACPR)," *Eur. Heart J.*, vol. 37, no. 29, pp. 2315–2381, Aug. 2016, doi: 10.1093/eurheartj/ehw106.
[5] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?" *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0174944, doi: 10.1371/journal.pone.0174944.
[6] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," *PLoS ONE*, vol. 14, no. 5, p. 17, May 2019, Art no. e0213653, doi: 10.1371/journal.pone.0213653.
[7] S. S. Lee, K. A. Kong, D. Kim, Y.-M. Lim, P.-S. Yang, J.-E. Yi, M. Kim, K. Kwon, W. B. Pyun, B. Joung, and J. Park, "Clinical implication of an impaired fasting glucose and prehypertension related to new onset atrial fibrillation in a healthy asian population without underlying disease: A nationwide cohort study in korea," *Eur. Heart J.*, vol. 38, no. 34, pp. 2599–2607, Sep. 2017, doi: 10.1093/eurheartj/ehx316.
[8] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 211, Dec. 2019, doi: 10.1186/s12911-019-0918-5.
[9] J. Park, "Can artificial intelligence prediction algorithms exceed statistical predictions?" *Korean Circulat. J.*, vol. 49, no. 7, pp. 640–641, 2019, doi: 10.4070/kcj.2019.0110.
[10] Z. I. Attia, S. Kapa, F. Lopez-Jimenez, P. M. McKie, D. J. Ladewig, G. Satam, P. A. Pellikka, M. Enriquez-Sarano, P. A. Noseworthy, T. M. Munger, S. J. Asirvatham, C. G. Scott, R. E. Carter, and P. A. Friedman, "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram," *Nature Med.*, vol. 25, no. 1, pp. 70–74, Jan. 2019, doi: 10.1038/s41591-018-0240-2.
[11] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, "Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study," *Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018, doi: 10.1016/S0140-6736(18)31645-3.
[12] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.
[13] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 32–35, 2001, doi: 10.1023/a:1010933404324.
[14] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 3149–3157.
[15] National Health Insurance Service (NHIS), Wonju-si, South Korea. *Medical Check-Up Cohort DB*. Accessed: Jul. 9, 2020. [Online]. Available: https://nhiss.nhis.or.kr/bd/ab/bdaba022Heng.do
[16] J. Lee, J. S. Lee, S.-H. Park, S. A. Shin, and K. Kim, "Cohort profile: The national health insurance service–national sample cohort (NHIS-NSC), South Korea," *Int. J. Epidemiol.*, vol. 46, no. 2, Apr 2017, Art. no. e15, doi: 10.1093/ije/dyv319.
[17] D. C. Goff, D. M. Lloyd-Jones, G. Bennett, S. Coady, R. B. D'Agostino, R. Gibbons, P. Greenland, D. T. Lackland, D. Levy, C. J. O'Donnell, J. G. Robinson, J. S. Schwartz, S. T. Shero, S. C. Smith, P. Sorlie, N. J. Stone, and P. W. F. Wilson, "2013 ACC/AHA guideline on the assessment of cardiovascular risk," *Circulation*, vol. 129, no. 25, pp. S49–S73, Jun. 2014, doi: 10.1161/01.cir.0000437741.48606.98.
[18] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Oper. Syst. Design Implement.*, Savannah, GA, USA, 2016, pp. 1–21.
[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, Lille, France, vol. 37, 2015, pp. 1–11.
[20] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, May 2016, pp. 1–14.
[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[22] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 4768–4777.
[24] J. Brabec and L. Machlica, "Bad practices in evaluation methodology relevant to class-imbalanced problems," Tech. Rep., Dec. 2018, pp. 1–4. [Online]. Available: https://arxiv.org/abs/1812.01388
[25] M. K. Kim, K. Han, H.-S. Kim, Y.-M. Park, H.-S. Kwon, K.-H. Yoon, and S.-H. Lee, "Cholesterol variability and the risk of mortality, myocardial infarction, and stroke: A nationwide population-based study," *Eur. Heart J.*, vol. 38, no. 48, pp. 3560–3566, Dec. 2017, doi: 10.1093/eurheartj/ehx585.
[26] M. K. Kim, K. Han, Y.-M. Park, H.-S. Kwon, G. Kang, K.-H. Yoon, and S.-H. Lee, "Associations of variability in blood pressure, glucose and cholesterol concentrations, and body mass index with mortality and cardiovascular outcomes in the general population," *Circulation*, vol. 138, no. 23, pp. 2627–2637, Dec. 2018, doi: 10.1161/circulationaha.118.034978.
[27] E. Y. Lee, Y. Yang, H.-S. Kim, J.-H. Cho, K.-H. Yoon, W. S. Chung, S.-H. Lee, and K. Chang, "Effect of visit-to-visit LDL-, HDL-, and non-HDL-cholesterol variability on mortality and cardiovascular outcomes after percutaneous coronary intervention," *Atherosclerosis*, vol. 279, pp. 1–9, Dec. 2018, doi: 10.1016/j.atherosclerosis.2018.10.012.
[28] E. Roh, H. S. Chung, J. S. Lee, J. A. Kim, Y.-B. Lee, S.-H. Hong, N. H. Kim, H. J. Yoo, J. A. Seo, S. G. Kim, N. H. Kim, S. H. Baik, and K. M. Choi "Total cholesterol variability and risk of atrial fibrillation: A nationwide population-based cohort study," *PLoS ONE*, vol. 14, no. 4, p. 13, Apr. 2019, Art no. e0215687, doi: 10.1371/journal.pone.0215687.
[29] H. S. Chung, J. S. Lee, J. A. Kim, E. Roh, Y. B. Lee, S. H. Hong, H. J. Yoo, S. H. Baik, N. H. Kim, J. A. Seo, S. G. Kim, N. H. Kim, and K. M. Choi, "γ-glutamyltransferase variability and the risk of mortality, myocardial infarction, and stroke: A nationwide population-based cohort study," *J. Clin. Med.*, vol. 8, no. 6, Jun. 2019, Art. no. 832, doi: 10.3390/jcm8060832.
[30] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
[31] M.-Z. Poh, Y. C. Poh, P.-H. Chan, C.-K. Wong, L. Pun, W. W.-C. Leung, Y.-F. Wong, M. M.-Y. Wong, D. W.-S. Chu, and C.-W. Siu, "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart*, vol. 104, no. 23, pp. 1921–1928, Dec. 2018, doi: 10.1136/heartjnl-2018-313147.
[32] J.-M. Kwon, K.-H. Kim, K.-H. Jeon, H. M. Kim, M. J. Kim, S.-M. Lim, P. S. Song, J. Park, R. K. Choi, and B.-H. Oh "Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification," *Korean Circulat. J.*, vol. 49, no. 7, pp. 629–639, Jul. 2019, doi: 10.4070/kcj.2018.0446.
[33] M. Padmanabhan, P. Yuan, G. Chada, and H. V. Nguyen, "Physician-friendly machine learning: A case study with cardiovascular disease risk prediction," *J. Clin. Med.*, vol. 8, no. 7, p. 1050, Jul. 2019, doi: 10.3390/jcm8071050.
[34] S. J. Al'Aref *et al.*, "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging," *Eur. Heart J.*, vol. 40, no. 24, pp. 1975–1986, Jun. 2019, doi: 10.1093/eurheartj/ehy404.
[35] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Syst. Appl.*, vol. 115, pp. 465–473, Jan. 2019, doi: 10.1016/j.eswa.2018.08.011.

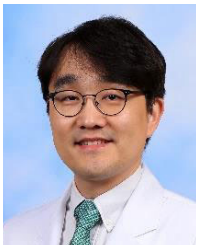
[36] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Comput. Methods Programs Biomed.*, vol. 130, pp. 54–64, Jul. 2016, doi: 10.1016/j.cmpb.2016.03.020.

[37] R. Shouval, A. Hadanny, N. Shlomo, Z. Iakobishvili, R. Unger, D. Zahger, R. Alcalai, S. Atar, S. Gottlieb, S. Matetzky, I. Goldenberg, and R. Beigel, "Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: An acute coronary syndrome israeli survey data mining study," *Int. J. Cardiol.*, vol. 246, pp. 7–13, Nov. 2017, doi: 10.1016/j.ijcard.2017.05.067.

GIHUN JOO received the B.S. and M.S. degrees in computer science from Kangwon National University, South Korea, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree in computer science. His research interests include machine learning, big data analysis and management, and precision medicine.

YEONGJIN SONG received the B.S. and M.S. degrees in computer science from Kangwon National University, South Korea, in 2018 and 2020, respectively. He is currently an Engineer with Classmethod, Japan. His research interests include programming languages, machine learning, and precision medicine.

HYEONSEUNG IM received the B.S. degree in computer science from Yonsei University, South Korea, in 2006, and the Ph.D. degree in computer science and engineering from Pohang University of Science and Technology (POSTECH), South Korea, in 2012. From 2012 to 2015, he was a Postdoctoral Researcher with the Laboratory for Computer Science, Université Paris-Sud, and Tyrex Team, Inria, France. He is currently an Associate Professor with the Department of Computer Science and Engineering, Kangwon National University, South Korea. His research interests include programming languages, logic in computer science, big data analysis and management, machine learning, precision medicine, and network security.



JUNBEOM PARK is currently an Associate Professor with the College of Medicine, Ewha Womans University. He is also the Director of the Cardiac Electrophysiology Laboratory, Ewha Womans University Medical Center. His research interests include mechanisms and predictors of atrial fibrillation, sinus node dysfunction, and clinical implication of AI in cardiovascular diseases.