



## ORIGINAL ARTICLE

# Comparison of multivariate linear regression and a machine learning algorithm developed for prediction of precision warfarin dosing in a Korean population

Van Lam Nguyen<sup>1,2</sup>  | Hoang Dat Nguyen<sup>1,2</sup> | Yong-Soon Cho<sup>1,2</sup> | Ho-Sook Kim<sup>1,2</sup> | Il-Yong Han<sup>3</sup> | Dae-Kyeong Kim<sup>4</sup> | Sangzin Ahn<sup>1,2</sup>  | Jae-Gook Shin<sup>1,2,5</sup>

<sup>1</sup>Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, Busan, Korea

<sup>2</sup>Center for Personalized Precision Medicine of Tuberculosis, Inje University College of Medicine, Busan, Korea

<sup>3</sup>Department of Thoracic and Cardiovascular Surgery, Inje University Busan Paik Hospital, Busan, Korea

<sup>4</sup>Division of Cardiology, Department of Internal Medicine, Inje University Busan Paik Hospital, Busan, Korea

<sup>5</sup>Department of Clinical Pharmacology, Inje University Busan Paik Hospital, Busan, Korea

## Correspondence

Jae-Gook Shin and Sangzin Ahn, Department of Pharmacology and Pharmacogenomics Research Center, Inje University College of Medicine, 633-165 Gaegum-Dong, Jin-Gu, Busan, Korea. Email: phshinjg@gmail.com (J.S.); sangzinahn@gmail.com (S.A.)

## Funding information

Ministry of Health & Welfare, Republic of Korea, Grant/Award Number: HI15C1537; National Research Foundation of Korea, Grant/Award Number: 2018R1A5A2021242

## Abstract

**Background:** Personalized warfarin dosing is influenced by various factors including genetic and non-genetic factors. Multiple linear regression (LR) is known as a conventional method to develop predictive models. Recently, machine learning approaches have been extensively implemented for warfarin dosing due to the hypothesis of non-linear association between covariates and stable warfarin dose.

**Objective:** To extend the multiple linear regression algorithm for personalized warfarin dosing in a Korean population and compare with a machine learning--based algorithm.

**Method:** From this cohort study, we collected information on 650 patients taking warfarin who achieved steady state including demographic information, indications, comorbidities, comedications, habits, and genetic factors. The dataset was randomly split into training set (90%) and test set (10%). The LR and machine learning (gradient boosting machine [GBM]) models were developed on the training set and were evaluated on the test set.

**Result:** LR and GBM models were comparable in terms of accuracy of ideal dose (75.38% and 73.85%), correlation (0.77 and 0.73), mean absolute error (0.58 mg/day and 0.64 mg/day), and root mean square error (0.82 mg/day and 0.9 mg/day), respectively. VKORC1 genotype, CYP2C9 genotype, age, and weight were the highest contributors and could obtain 80% of maximum performance in both models.

**Conclusion:** This study shows that our LR and GMB models are satisfactory to predict warfarin dose in our dataset. Both models showed similar performance and feature contribution characteristics. LR may be the appropriate model due to its simplicity and interpretability.

## KEYWORDS

anticoagulants, linear models, supervised machine learning, thrombosis, warfarin

### Essentials

- Linear regression (LR) and gradient boosting (GBM) models predicting warfarin dose were compared.
- The models were trained on 29 features from 650 Korean patients.
- VKORC1 genotype, CYP2C9 genotype, age, and weight were the highest contributors in both models.
- Both LR and GBM models showed similar characteristics and performance in our dataset.

## 1 | INTRODUCTION

Warfarin is an extensively prescribed anticoagulant drug that works by inhibiting the activity of the vitamin K epoxide reductase multiprotein complex. This drug has been used to treat and prevent harmful blood clots, such as venous thrombosis and pulmonary embolism.<sup>1,2</sup> Predicting optimal warfarin dose is influenced by many factors including genetic and non-genetic attributes such as vitamin K epoxide reductase complex subunit 1 (VKORC1) genotype, cytochrome P450 family 2 subfamily C member 9 (CYP2C9) genotype, age, body weight, concurrent medication, and vitamin K intake.<sup>3-7</sup> Due to its narrow therapeutic index and high interpatient variability, personalized warfarin dosing to maximize efficacy and safety continues to be challenging.<sup>8</sup>

Therefore, many predictive warfarin dosing algorithms have been developed at the global level<sup>1</sup> and in individual ethnic populations.<sup>9-14</sup> Most of them have been shown to be cost-effective<sup>15,16</sup> and thus are extensively used in the personalized dose prediction of warfarin in clinical practice. In Korea, our group has developed a warfarin dose prediction algorithm in patients taking warfarin after valvular heart replacement surgery.<sup>13</sup> One limitation of our study is that it covers only patients with mechanical heart valve replacement (MHVR) as an indication of warfarin. It has been reported that there are differences in optimizing warfarin dose in patients with MHVR and with other indications.<sup>17,18</sup> Thus, an algorithm that can cover more indications of warfarin for a higher number of patients must be continuously developed.

Multivariate linear regression is known as a conventional approach to developing a predictive model because of its simple development process and high interpretability. It has been reported that a complex non-linear relationship may exist between stable warfarin dose and genetic and clinical factors.<sup>19</sup> Therefore, a machine learning approach has been extensively implemented in developing predictive algorithms of personalized precision warfarin dosing.<sup>14,20-22</sup>

In this study, we first aimed to extend our previous multiple linear regression predictive model of warfarin dosing for the Korean population. Additionally, to explore whether machine learning models, which can solve non-linear relationships, may have better performance than conventional linear regression models, a machine learning model was developed on the same dataset; then, we compared the performance and structure of both models.

## 2 | METHODS

### 2.1 | Cohort dataset

This cohort study was approved by the Institutional Review Board (IRB) of the Inje University Busan Paik Hospital, Busan, Korea (IRB approval number: 16-0232). The patients treated with warfarin were enrolled between 2005 and 2019. The data up to 2015 were collected retrospectively and data after that time were collected prospectively. The inclusion criteria consisted of agreement of genotyping, at least 18 years old or older, administration of warfarin, accessible medical records, and routine international normalized ratio (INR) measurement (every 4–8 weeks for reliable and medically stable patients). Those recorded as being noncompliant with warfarin medication were excluded from the study. All recruited participants provided written informed consent. The patients were followed-up to getting at least steady-state. The follow-up time was from 3 months to 2 years.

From the medical record of patients and direct questionnaires, clinical and dietary information were collected for gender, age, weight, height, warfarin indication, therapeutic range of INR, resulting INR values and warfarin doses, comorbidities, comedication, dietary supplements, and smoking status. The self-monitoring tool was developed by us based on a Korean food composition table by the Korea Rural Development Administration<sup>23</sup> and was used to evaluate vitamin K intake. If more than 85 mg/day of vitamin K was administered, they were assumed to have an INR-decreasing dietary supplement.<sup>13,24</sup> Food and drugs potentially interacting with warfarin were characterized by referring to the MicroMedex Healthcare Series online database.<sup>25</sup> Nine hundred and twenty patients were enrolled, and 270 patients were dropped because of withdrawal of consent, inaccessible medical records, or inability to reach to stable warfarin dose. Details of data description and processing are shown in Figure S1 and Table S1 in supporting information. Finally, a dataset of 650 subjects with 29 independent features and 1 primary endpoint feature was used for modeling. Steady-state warfarin dose (mg/day) was the primary endpoint of this study. A steady-state warfarin dose was defined when three consecutively measured INR from the same mean daily dose for more than 4 weeks were within a therapeutic range, which was described from our previous report.<sup>13</sup>

## 2.2 | Splitting dataset

In this study, the whole dataset was randomly split into a training set (90%) and a test set (10%). The training set was used for model development and the test set was used for evaluating the performance of models. The equal stratification of variables between the training set and test set was examined by performing Student's *t*-test or Wilcoxon's rank-sum test for continuous variables after the assessment of normal distribution using a Shapiro-Wilk test while the categorical variables was compared with  $\chi^2$  test or Fisher's exact test in case of less than 5 in any cell frequency and contingency  $2 \times 2$ . The detail of the model development strategy is illustrated in Figure 1.

## 2.3 | Model development

### 2.3.1 | Multiple linear regression model development

Because the observed warfarin dose was not normally distributed, log transformation was applied to achieve normal distribution. To develop the multiple linear regression (LR) model of the warfarin dose, stepwise regression with bidirectional elimination was used to select the covariates and calculate the regression coefficients of

the model. The *P*-values for including and excluding covariates in the model were set at 0.1 and 0.05, respectively.

### 2.3.2 | Machine learning model development

Seven different predictive models (least absolute shrinkage and selection operator: LASSO, elastic net: EN, k-nearest neighbors: KNN, classification and regression tree: CART, support vector regression: SVR, random forest: RF, and gradient boosting machine: GBM) were trained on the training dataset with five-fold cross-validation using scikit-learn default settings. Based on the root mean square error (RMSE) and mean absolute error (MAE), the model that had the smallest error was chosen. As a result, GBM was the best-performing model and was selected for further development.

In the GBM model development, first, recursive feature elimination with five-fold cross-validation (RFECV) was used to eliminate unimportant features. Second, the remaining features were used for hyperparameter tuning. Boosting (*n\_estimators*) and tree-specific (*max\_depth*, *max\_features*, *subsample*) hyperparameters were tuned using grid search with five-fold cross-validation. Finally, the performance of the final model was evaluated on the test set.

### 2.3.3 | Software

In this study, model development and data analysis were done by using R (version 3.6.2) and Python (version 3.7.0). Machine learning algorithms were built based on the scikit-learn package version 0.22.2.

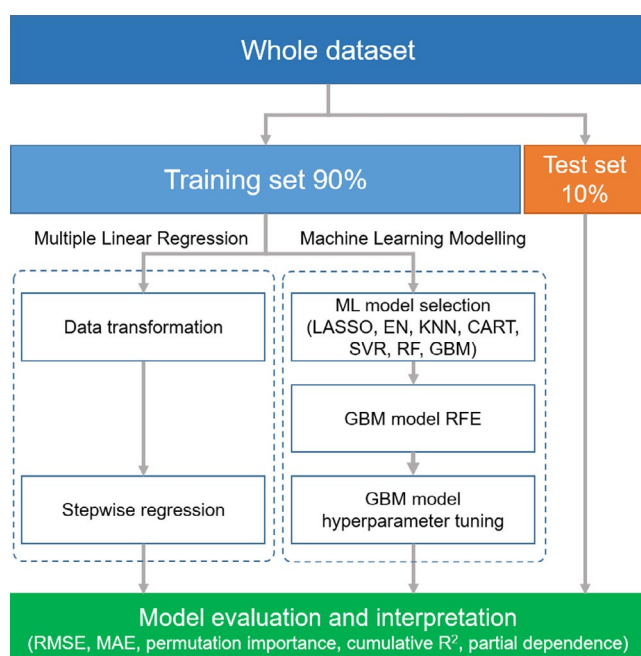
## 2.4 | Comparison of the performance of linear regression and gradient boosting machine model

MAE, RMSE, correlation (*R*), and accuracy of ideal warfarin dose (within 20% of the actual dose) were computed to compare the performance of the two proposed models. The acceptable error range of 20% of the actual dose was set similarly with other warfarin dosing model studies.<sup>1,12,20,21</sup>

A permutation importance score was measured to assess how much each feature contributes to the models following equation 1.

$$\text{Permutation importance score} = \frac{\sum_1^{100} \text{baseline } R^2 - \text{permuted } R^2}{100} \quad (1)$$

wherein baseline  $R^2$  was defined as the original  $R^2$ , which was obtained from the original test set; permuted  $R^2$  was defined as  $R^2$ , which was computed after shuffling a single feature of the test set. A single feature was randomly shuffled 100 times, then the mean and standard deviation of permutation importance score were calculated.



**FIGURE 1** The model development strategy. The whole dataset was divided into training set (90%) and test set (10%). Model development and model structure analysis were done using the training set. The performance of models was evaluated on the test set. CART, classification and regression tree; EN, elastic net; GBM, gradient boosting machine; KNN, k-nearest neighbors; LASSO, least absolute shrinkage and selection operator; ML, machine learning; RF, random forest; SVR, support vector regression

Additionally, a cumulative  $R^2$  of each model was evaluated by adding features ordered by the permutation importance score. First, all features in the test set were imputed using mean value (continuous features) or most abundant value (categorical features). After that, the imputed values were replaced with actual observed values for each feature in descending order of importance score, and on each step the predictive  $R^2$  value was re-calculated to illustrate the effect of adding each feature on the performance of the model. Partial dependence plots were used to show the marginal effect of the common four most important features of both models. Partial dependence was computed on the training dataset using a partial\_dependence module from the scikit-learn package with default settings.

### 3 | RESULTS

#### 3.1 | Cohort dataset

The clinical and genetic information of the studied population is shown in Table 1. The mean age of 650 patients was 58.4 years old, among them, 381 (58.6%) patients were male. The average weight of the population was 61.5 kg. The median target INR range in our dataset was from 1.75 to 2.50. Among those, most of patients' (99%) target INR was 2.25, while 1 patient (0.15%) was 2.50, 3 patients (0.45%) were 2.00, and 2 patients (0.30%) were 1.75. Mechanical heart valve replacement was the most common indication for warfarin administration, which was observed in 390 patients (60%), while other indications included arrhythmia (20.5%), stroke (14.3%), deep vein thrombosis (1.2%), pulmonary embolism (3.2%), heart valve disease (14.2%), and other thrombosis or embolism (0.9%). Congestive heart failure (CHF) and/or cardiomyopathy had the highest prevalence (23.8%) compared to other comorbidities, followed by hypertension (23.1%), diabetes mellitus (12.8%), cancer (2.3%), liver disease (1.4%), hypothyroidism (1.4%), hyperthyroidism (0.9%), and myocardial infarction (0.9%). Smoking and INR-decreasing supplementary intake were recorded in 18.0% and 8.2% of the population, respectively. Comedications that could affect the anticoagulant effect of warfarin including HMG-CoA reductase inhibitor, INR-decreasing drug, INR-increasing drug, and anti-platelet drug were administrated in 26.5%, 0.9%, 1.5%, and 13.5% of the population, respectively. Regarding genetic factors, CYP2C9\*1/\*1 was the most common genotype (89.2%) followed by CYP2C9\*1/\*3, CYP2C9\*1/\*13 (0.46%), CYP2C9\*1/\*14 (0.3%), and CYP2C9\*3/\*3 (0.3%). The frequency of GG, GA, and AA genotypes of VKORC1 (-1639G>A) were 0.3%, 16.5%, and 83.2%, respectively.

#### 3.2 | Multiple linear regression model development

The extended multiple linear regression model (Table 2) of warfarin dose includes 15 variables including VKORC1 1639 G>A

genotype, CYP2C9 genotype, PROS1, rs13062355 AA genotype, age, weight, CHF and/or cardiomyopathy, heart valve disease, pulmonary embolism, diabetes mellitus, INR-decreasing drug, INR-increasing drug, anti-platelet drug, HMG-CoA reductase inhibitor, current smoking, and INR-increasing diet. Among included variables, VKORC1-1639 G>A genotype and CYP2C9 genotype account for the highest variation of warfarin maintenance dose with 19.1% and 13%, respectively. The final model explained about 59% of the variation of inter-individual warfarin dose: maintenance dose =  $\exp(1.435 + 0.414 \times \text{VKORC1-1639 GA} + 0.608 \times \text{VKORC1-1639 GG} - 0.007 \times \text{Age [years]} - 0.373 \times \text{CYP2C9 1-LoF} - 1.502 \times \text{CYP2C9 2-LoF} - 0.114 \times \text{CHF and/or Cardiomyopathy} + 0.007 \times \text{Weight [kg]} + 0.443 \times \text{INR-decreasing drug} - 0.078 \times \text{Current smoking} - 0.095 \times \text{Antiplatelet drugs} - 0.116 \times \text{Heart Valve Disease} + 0.152 \times \text{Pulmonary Embolism} - 0.086 \times \text{HMG-CoA reductase inhibitor} + 0.078 \times \text{Diabetes Mellitus} + 0.098 \times \text{INR-decreasing dietary supplements} - 0.221 \times \text{INR-increasing drug} - 0.070 \times \text{PROS1 [rs13062355 GA]} - 0.075 \times \text{PROS1 [rs13062355 AA]})$ , where it was coded as 1 in the case of CHF and/or cardiomyopathy, heart valve disease, pulmonary embolism, diabetes mellitus, co-administration of INR-increasing drug, INR-decreasing drug, anti-platelet drugs, HMG-CoA reductase inhibitor, INR-increasing diet, and smoking. CYP 2C9 1-LoF variant include: CYP2C9\*1/\*3, CYP2C9\*1/\*13, CYP2C9\*1/\*14; CYP 2C9 2-LoF variants include CYP2C9 \*3/\*3 unless otherwise, it was coded as 0.

#### 3.3 | Machine learning model development

To identify the most suitable machine learning model for our dataset, we examined the performance of seven different algorithms (EN, LASSO, KNN, SVR, CART, RF, and GBM) using five-fold cross-validation on the training set. RMSE and MAE were used to evaluate the performance of each model. GBM was the best-performing model among these models in both RMSE ( $1.10 \pm 0.09$  mg/day) and MAE ( $0.79 \pm 0.03$  mg/day; Figure 2). Based on these results, the GBM algorithm was further used to develop a predictive warfarin dosing model of the machine learning approach. After RFECV and hyperparameter tuning processes, the final GBM model was developed on the training set data of 17 importance features and with hyperparameter number of estimators (n\_estimators): 90; maximum depth of tree (max\_depth): 4; maximum of features (max\_features): 2; and fraction of observations to be selected for each tree (subsampling): 1.0 (data not shown).

#### 3.4 | Comparison of the performance and structure of LR and GBM models

The performance of two developed models is shown in Table 3 and Figure S2 in supporting information. From the results, we did not find significant performance improvement of the GBM model compared

TABLE 1 Demographics and clinical characteristics of 650 patients from our dataset

Covariate/feature	All dataset (N = 650)	Training set (N = 585)	Test set (N = 65)	p value
Basic characteristics				
Age (years) <sup>a</sup>	58.4 (12.6)	58.3 (12.6)	59.9 (12.7)	.33
Male	375 (57.7)	341 (58.3)	34 (52.3)	.43
Height (cm) <sup>a</sup>	162.3 (8.4)	162.3 (8.4)	162.5 (8.4)	.83
Weight (kg) <sup>a</sup>	61.5 (11.0)	61.4 (10.8)	61.8 (12.3)	.81
Indication				
Mechanical heart valve replacement	390 (60.0)	352 (60.2)	38 (58.5)	.89
Arrhythmia	133 (20.5)	120 (20.5)	13 (20.0)	.95
Stroke	93 (14.3)	82 (14.0)	11 (16.9)	.65
Deep vein thrombosis	8 (1.2)	8 (1.4)	0 (0.0)	.72
Pulmonary embolism	21 (3.2)	19 (3.2)	2 (3.1)	.77
Heart valve disease	92 (14.2)	82 (14.0)	10 (15.4)	.91
Other thrombosis embolism	6 (0.9)	6 (1.0)	0 (0.0)	.89
Comorbidity				
Myocardial infarction	6 (0.9)	5 (0.9)	1 (1.5)	.89
CHF and/or cardiomyopathy	155 (23.8)	141 (24.1)	14 (21.5)	.76
Liver disease	9 (1.4)	9 (1.5)	0 (0.0)	.65
Hyperthyroidism	6 (0.9)	6 (1.0)	0 (0.0)	.89
Hypothyroidism	9 (1.4)	8 (1.4)	1 (1.5)	.65
Diabetes mellitus	83 (12.8)	70 (12.0)	13 (20.0)	.10
Hypertension	150 (23.1)	137 (23.4)	13 (20.0)	.64
Cancer	15 (2.3)	13 (2.2)	2 (3.1)	1.00
Habit				
Current smoking	117 (18.0)	101 (17.3)	16 (24.6)	.20
INR-decreasing dietary supplements <sup>b</sup>	53 (8.2)	50 (8.5)	3 (4.6)	.39
Genetics				
CYP2C9 genotype				
*1/*1	580 (89.2)	522 (89.2)	58 (89.2)	.95
1-LoF (*1/*3; *1/*13; *1/*14)	69 (10.6)	62 (10.6)	7 (10.8)	
2-LoF (*3/*3)	1 (0.2)	1 (0.2)	0 (0.0)	
VKORC1 (-1639G>A)				
GG	2 (0.3)	2 (0.3)	0 (0.0)	.87
GA	107 (16.5)	97 (16.6)	10 (15.4)	
AA	541 (83.2)	486 (83.1)	55 (84.6)	
PROS1 (rs13062355)				
GG	110 (16.9)	100 (17.1)	10 (15.4)	.20
GA	274 (42.2)	240 (41.0)	34 (52.3)	
AA	266 (40.9)	245 (41.9)	21 (32.3)	
CYP4F2 genotype				
*1/*1	339 (52.2)	304 (52.0)	35 (53.8)	.46
*1/*3	254 (39.1)	227 (38.8)	27 (41.5)	
*3/*3	57 (8.8)	54 (9.2)	3 (4.6)	
Comedication				
HMG-CoA reductase inhibitor	172 (26.5)	151 (25.8)	21 (32.3)	.34
INR-decreasing drug <sup>c</sup>	6 (0.9)	5 (0.9)	1 (1.5)	.89

(Continues)

TABLE 1 (Continued)

Covariate/feature	All dataset (N = 650)	Training set (N = 585)	Test set (N = 65)	p value
INR-increasing drug <sup>d</sup>	10 (1.5)	10 (1.7)	0 (0.0)	.60
Anti-platelet drug <sup>e</sup>	88 (13.5)	77 (13.2)	11 (16.9)	.52
Primary outcome				
Observed steady-state warfarin dose (mg/day) <sup>a</sup>	4.0 (1.5)	4.0 (1.6)	3.8 (1.3)	.32
Target INR range 1.50–2.00 (n = 2)	2.79 (0.29)	2.79 (0.29)	–	
Target INR range 1.70–2.80 (n = 644)	3.96 (1.53)	3.99 (1.56)	3.78 (1.29)	
Target INR range 1.50–2.50 (n = 3)	3.56 (1.87)	3.56 (1.87)	–	
Target INR range 2.00–3.00 (n = 1)	1.75 (–)	1.75 (–)	–	

Note: Statistical significance between training set and test set was conducted using Student's t-test, Wilcoxon rank-sum test,  $\chi^2$  test or Fisher's exact test appropriately.

Abbreviations: CHF, congestive heart failure; HMG-CoA reductase inhibitors, 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor; INR, international normalized ratio; LoF, loss-of-function.

<sup>a</sup>Data were shown as mean (standard deviation), others were shown as frequency (%).

<sup>b</sup>Dietary supplements with decreasing effect include broccoli, soybeans, nutrition pills containing vitamin K, and Korean ginseng.

<sup>c</sup>INR-decreasing drug includes rifampin or carbamazepine.

<sup>d</sup>INR-increasing drug includes amiodarone or fluconazole or doxifluridine, or cotrimoxazole.

<sup>e</sup>Anti-platelet drug includes aspirin or clopidogrel.

TABLE 2 Multiple linear regression model

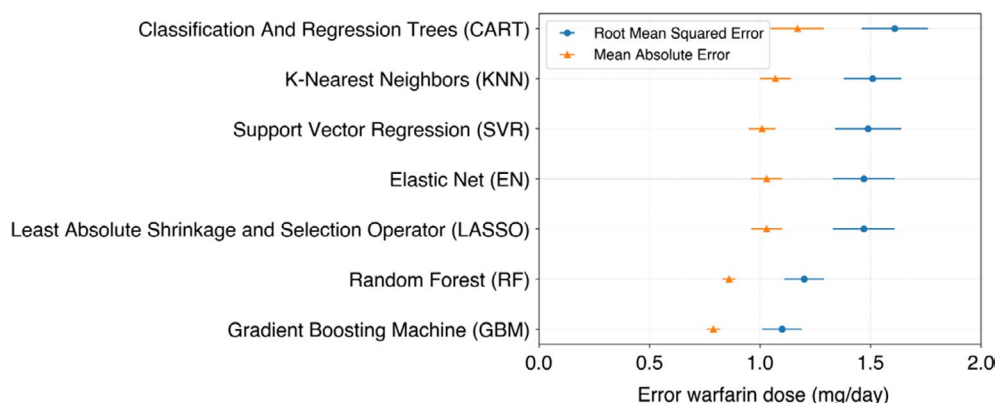
Covariates/features	Coefficients	p value	Cumulative R <sup>2</sup>	Univariate R <sup>2</sup>
Intercept	1.435	<.001		
VKORC1 (–1639 GA)	0.414	<.001	19.15	19.15
VKORC1 (–1639 GG)	0.608	.001		
Age	–0.007	<.001	32.26	12.43
CYP2C9 1-LoF (*1/*3, *1/*13, *1/*14)	–0.373	<.001	44.37	13.03
CYP2C9 2-LoF (*3/*3)	–1.502	<.001		
CHF and/or cardiomyopathy	–0.114 <sup>1</sup>	<.001	49.23	7.22
Weight (kg)	0.007	<.001	51.60	4.603
INR-decreasing drug	0.443	<.001	53.07	1.021
Current smoking	–0.078	.006	54.34	2.123
Anti-platelet drug	–0.095	.003	55.23	2.637
Valve heart disease	–0.116	.002	55.89	4.446
Pulmonary embolism	0.152	.01	56.52	2.159
HMG-CoA reductase inhibitor	–0.086	.001	57.13	2.684
Diabetes mellitus	0.078	.02	57.63	0.048
INR-decreasing dietary supplements	0.098	.009	58.02	1.651
INR-increasing drug	–0.211	.012	58.40	1.97
PROS1 (rs13062355 GA)	–0.070	.022	58.88	3.28
PROS1 (rs13062355 AA)	–0.075	.014		

Abbreviations: CHF, congestive heart failure; HMG-CoA reductase inhibitors, 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor; INR, international normalized ratio; LoF, loss-of-function.

to the LR model in terms of MAE (0.64 mg/day vs. 0.58 mg/day), RMSE (0.90 mg/day vs. 0.82 mg/day), correlation (0.73 vs. 0.77), and accuracy (73.85% vs. 75.38%).

Permutation importance score was computed to compare the contribution of each feature in the LR and GBM models. Figure 3A shows that VKORC1 (–1639 G>A) genotype, age, CYP2C9 genotype,





**FIGURE 2** The performance of seven different machine learning models based on RMSE and MAE on five-fold cross validation. The GBM model resulted in the smallest error in both RMSE and MAE, thus this model was used for downstream steps. The data were shown as mean and standard deviation. GBM, gradient boosting machine; MAE, mean absolute error; RMSE, root mean square error

Algorithms	Accuracy (%)	Correlation	MAE (mg/day)	RMSE (mg/day)	<i>p</i> -value <sup>a</sup>
Linear regression	75.38	0.77	0.58	0.82	.5398
Gradient boosting machine	73.85	0.73	0.64	0.90	

**TABLE 3** Comparison of the performance of multiple linear regression and gradient boosting machine algorithms

Abbreviations: MAE, mean absolute error; RMSE, root mean square error.

<sup>a</sup>*p*-value was computed using Paired Student's *t*-test on residual (predicted dose – actual dose) of two models.

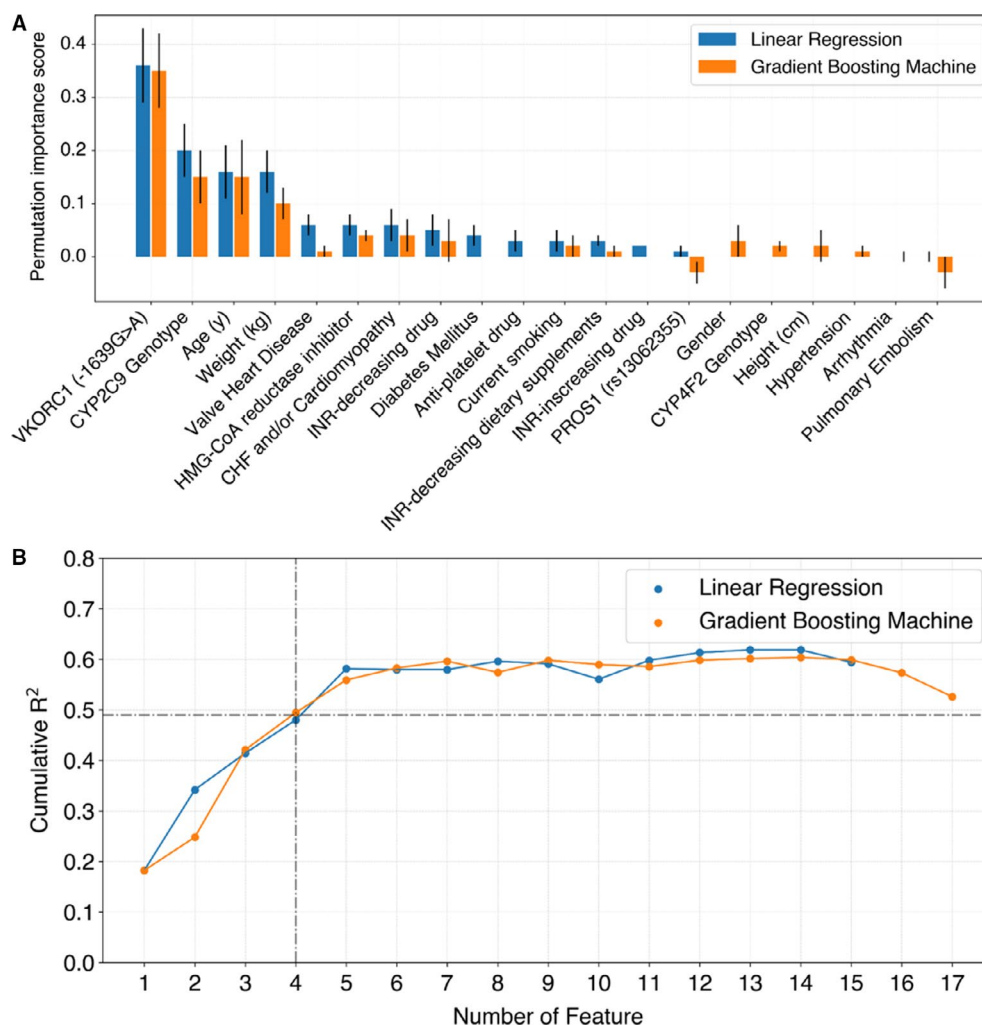
and weight are the most important features in warfarin dosing and the contribution of these features was similar between the two models. There were some slight differences in terms of the contribution of other features; however, their effects were very minor (Figure 3A). Generally, the LR model and GBM model showed similar performance when the same number of features was added to the model. In fact, the cumulative  $R^2$  on the test set of two models increased sharply at the first five most important features and leveled out around  $R^2 = 0.58$  in both models. In the case of adding more less-important features to the models, the performance of two models marginally fluctuated (Figure 3B). Moreover, the partial dependence plot of these four most important features demonstrated similar effects on predicting the stable warfarin dose in both models (Figure S3 in supporting information).

## 4 | DISCUSSION

Previously, our algorithm for predicting warfarin dose was built based on data of patients indicated with MHVR. However, with the limited sample size and indications, it was reasonable to extend a new model with a larger number of patients including internal medicine and surgery indications. The extended LR model explained 58.9% (adjusted  $R^2 = 57.6\%$ ; data not shown) variability of warfarin dose variation, which is higher than our previous model ( $R^2 = 55.8\%$ ). On the test set, the performance of our extended LR was better than our previous model in terms of all four criteria, namely MAE (0.58

and 0.67 mg/day), RMSE (0.82 and 0.86 mg/day), correlation (0.77 and 0.74), and accuracy of ideal dose (75.38% and 63.1%). This improvement can be explained by the effect of adding more indications (heart valve disease and pulmonary embolism), comorbidity (diabetes mellitus), comedications (HMG-CoA drugs, and CYP2C9 inducer) as well as genetic factors (PROS1 polymorphisms). As a vitamin K-dependent protein, PROS1 was reported to possibly affect warfarin variability.<sup>2,26</sup> In this study, however, we found that its contribution to warfarin dosing in our model was very small in comparison to other genetic factors, such as VKORC1 and CYP2C9 genotype. Therefore, adding PROS1 genotype information to our model may be useful in case this information is available preemptively. Otherwise, including VKORC1 and CYP2C9 genotype information may be sufficient.

It has been reported that genetics and clinical factors might have a non-linear association with warfarin dose,<sup>19</sup> which is a limitation of linear regression models.<sup>27</sup> Thus, many machine learning models that can solve the non-linear relationship have been developed.<sup>14,20–22</sup> In this study, we compared the performance of LR, which was developed using a conventional method and GBM, which is a typical machine learning model on the same dataset from the Korean population. No significant performance improvement of the GBM model was found in comparison to the LR model's performance in terms of MAE, RMSE, correlation, and accuracy of the ideal dose (Table 3). Interestingly, previous studies exploring the IWPC dataset reported that machine learning models had better performance when compared with multiple linear regression. Zhiyuan Ma et al. used stacked generalization frameworks, which combine



**FIGURE 3** Permutation feature importance score from both models (A); cumulative R<sup>2</sup> ordered by feature importance score of both models (B). The permutation importance score and cumulative R<sup>2</sup> were computed on the test set; the results showed the similarity in the structure and performance of both models as VKORC1, CYP2C9, age, and weight were the most important features and accounted for 80% of the maximum performance of models

multiple machine learning algorithms to estimate the optimal warfarin dose based on 5743 subjects from the International Warfarin Pharmacogenetics Consortium (IWPC) dataset. This study reported that the performance of stacked models was significantly higher than multiple linear models with MAE 8.31 mg/week and 8.53 mg/week and accuracy of ideal dose 47.8% and 46.3%, respectively.<sup>21</sup> Another study, which also used the IWPC dataset, combined two machine learning models to predict warfarin dose. The first model was designed to classify patients into two groups based on required maintenance dose and as consequence, the second model was used to predict warfarin dose based on the first model results. The result showed that the method had better performance in comparison to the IWPC model and Gage model in terms of MAE (8.4 mg/week, 9.1 mg/week, and 9.9 mg/week) and RMSE (11.6 mg/week, 13.8 mg/week, and 12.2 mg/week), respectively.<sup>22</sup> This can be explained that the larger sample size and homogeneity of the study population might improve the performance of the predictive machine learning models.

We also applied model-agnostic interpretation methods to further analyze the internal structure of LR and GBM models. The permutation importance score was calculated by measuring the performance reduction when shuffling each feature, and we found that VKORC1 (-1639G>A) genotype, CYP2C9 genotype, age, and weight were the most important factors. It is widely recognized that advanced age requires a lower dose of warfarin.<sup>28</sup> The potential mechanism is suspected to be due to the reduced blood flow to the liver, which leads to lower amount of warfarin transported to the liver or might be the poor capacity of vitamin K absorption in aged patients while weight was shown to have a positive correlation with warfarin dose.<sup>29,30</sup> Regarding the genetic factors, VKORC1 (-1639G>A) can explain approximately 25% variability of the warfarin dose requirement.<sup>31-33</sup> The substitution of amino acid in those having CYP2C9\*2 and CYP2C9\*3 resulted in a lower catalytic activity of CYP2C9; hence, lower warfarin dose was required in the patients who have this genotype.<sup>34,35</sup> These patterns were also confirmed by two proposed models when using the partial dependence plot to

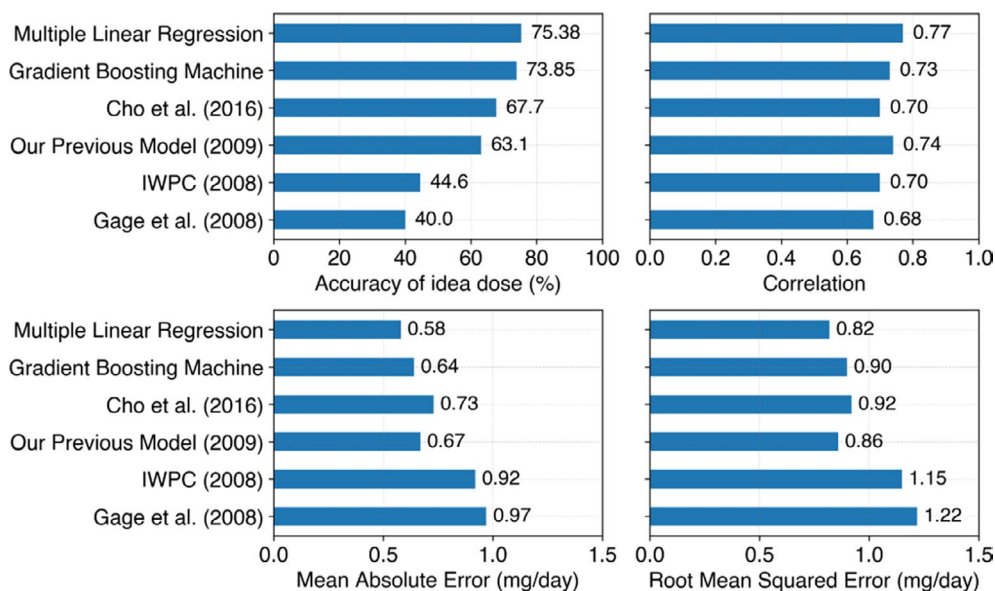


illustrate the marginal effect on the predicted warfarin dose (Figure S3). In clinical practice, it is difficult to collect information on all of these features (LR: 15 features; GBM: 17 features). Thus, we performed a cumulative  $R^2$  analysis to visualize the trade-off between the number of features and performance. As a result, the four most important features including genetic factors (VKORC1 and CYP2C9 genotype) and clinical factors (age and weight) accounted for more than 80% of the maximum performance of both LR and GBM models that were trained on our dataset (Figure S2). Also, partial dependence of the features showed similar patterns for each feature in our two models, depicting that the features have been similarly modeled in both LR and GBM (Figure S3).

In order to compare the performance with other known models, we evaluated these models' performance on the same testing dataset. The performance of our LR and GBM algorithms has higher accuracy of ideal dose (75.38% and 73.85%) in comparison to IWPC (40%); Gage et al.<sup>12</sup> (44%) and Cho et al.<sup>36</sup> (67%). Similarly, our models resulted in lower MAE (0.58 mg/day and 0.64 mg/day) than IPWC (0.92 mg/day), Gage et al.<sup>12</sup> (0.97 mg/day), and Cho et al.<sup>36</sup> (0.73 mg/day). Details of the performance of each model are shown in Figure 4. These results can be explained by the fact that the IWPC model was developed on a dataset including multiple populations.<sup>1</sup> The model was also adjusted by Asian and Black or African American races but not specified for the Korean population. Similarly, the model of Gage et al. was developed mainly on Caucasian information.<sup>12</sup> The performance of the Cho et al. model was higher than the IWPC and Gage's models presumably because it was built on Korean patient data.<sup>36</sup> Despite being better than two other models, the model by Cho et al. had lower performance in comparison to our models. The difference in sample size might be a reason for this result. While our models were developed on the training set with 585 subjects, Cho's model

was built on 101 subjects only. Another possible explanation might be that only stroke patients were considered in Cho's model,<sup>36</sup> while other indications and comorbidities were included in our models (Figure 3A). These data suggest that our models may be beneficial for optimizing warfarin dose in the Korean population.

Our study also has limitations. First, only 650 subjects were used to develop and validate models. In fact, the sample size is known to have a high impact on the performance of machine learning models. This may be one of the reasons why the GBM model did not show any improvement compared to the LR model. Another limitation of this study is sparse data. For example, only 3 subjects with CYP2C9\*1/\*13, 2 subjects with CYP2C9\*1/\*14, and 1 subject with CYP2C9\*3/\*3 were recorded in our dataset, which may be difficult for pattern recognition and correctly modeling the effect of rare genotypes in the model. Previous studies also reported that CYP2C9\*13 and CYP2C9\*14 and CYP2C9\*3/\*3 are rare variants in Asian populations as well as the Korean population.<sup>13,37</sup> Regarding the difference in the genetic profiles of CYP2C9 and VKORC1 among ethnic populations,<sup>13,38-42</sup> when applying this model to populations other than East Asians or Koreans, careful interpretation is required. Another limitation worth mentioning is that it is a challenge when we want to apply a pharmacogenomics (PGx) dose prediction model for starting a treatment without genetic information. These models may be beneficial when combined with a preemptive PGx screening. Because the model in our study is focused on stable warfarin dose, application of our model needs caution when being applied to patients with other critical factors or when dosing unstable hospitalized patients. Finally, most patients in our dataset had a target INR of 2.25 (1.70–2.80). This means that our model's dose prediction would be more reliable when applied to patients in this range of



**FIGURE 4** Comparison of the performance of other published algorithms on the test set. Our proposed models had better performance compared to others in our dataset. These data suggest that our models would be beneficial for the Korean population

target INR. When applying a predictive model to a patient with an INR target that is different from the training data, careful interpretation is required.

In conclusion, we extended our previously reported multiple linear regression model for personalized warfarin dosing with a higher number of patients and covariates in a Korean population. The machine learning--based model (GBM) did not show significant improvement in comparison to multiple linear regression (LR) in our dataset. Although developing a multi-population machine learning--based prediction model may be the desired goal, a simple and interpretable linear regression model developed on a small ethnic group may be an appropriate strategy for predicting optimal warfarin dose. VKORC1 genotype, CYP2C9 genotype, age, and body weight were the most important contributors in both models and should be considered for preemptive warfarin dosing. Further studies need to be conducted to confirm its usability in clinical practice.

## ACKNOWLEDGMENTS

This work was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. HI15C1537) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; No. 2018R1A5A2021242). These funding sources had no role in the study design, in the data collection or analysis, in the preparation of the manuscript, or in the decision to submit the paper for publication.

## CONFLICTS OF INTEREST

The authors state that they have no conflicts of interest.

## AUTHOR CONTRIBUTIONS

Van Lam Nguyen and Hoang Dat Nguyen: conduct of the study, data analysis, and manuscript writing; Yong-Soon Cho and Ho-Sook Kim: study design and manuscript writing; Il-Yong Han and Dae-Kyeong Kim: conduct of the study and manuscript writing; Sangzin Ahn and Jae-Gook Shin: design of the study, conduct of the study, data analysis, and manuscript writing.

## ORCID

Van Lam Nguyen  <https://orcid.org/0000-0001-5643-3479>

Sangzin Ahn  <https://orcid.org/0000-0003-2749-0014>

## REFERENCES

1. Consortium IWP. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med*. 2009;360:753-764.
2. Wadelius M, Pirmohamed M. Pharmacogenetics of warfarin: current status and future challenges. *Pharmacogenomics J*. 2007;7:99-111. <https://doi.org/10.1038/sj.tpj.6500417>
3. Flockhart DA, O'Kane D, Williams MS, et al. Pharmacogenetic testing of CYP2C9 and VKORC1 alleles for warfarin. *Genet Med*. 2008;10:139-150. <https://doi.org/10.1097/GIM.0b013e318163c35f>
4. Takahashi H, Echizen H. Pharmacogenetics of CYP2C9 and interindividual variability in anticoagulant response to warfarin. *Pharmacogenomics J*. 2003;3:202-214.
5. Wynne H, Cope L, Kelly P, Whittingham T, Edwards C, Kamali F. The influence of age, liver size and enantiomer concentrations on warfarin requirements. *Br J Clin Pharmacol*. 1995;40:203-207.
6. Loebstein R, Yonath H, Peleg D, et al. Interindividual variability in sensitivity to warfarin-Nature or nurture? *Clin Pharmacol Ther*. 2001;70:159-164.
7. Nutescu EA, Shapiro NL, Ibrahim S, West P. Warfarin and its interactions with foods, herbs and other dietary supplements. *Expert Opin Drug Saf*. 2006;5:433-451. <https://doi.org/10.1517/14740338.5.3.433>
8. Li X, Li D, Wu J-C, Liu Z-Q, Zhou H-H, Yin J-Y. Precision dosing of warfarin: open questions and strategies. *Pharmacogenomics J*. 2019;19:219-229. <https://doi.org/10.1038/s41397-019-0083-3>
9. Yang M, Choi R, Kim JS, et al. Evaluation of 16 genotype-guided warfarin dosing algorithms in 310 Korean patients receiving warfarin treatment: poor prediction performance in VKORC1 1173C carriers. *Clin Ther*. 2016;38:2666.e1-2674.e1.
10. Choi JR, Kim JO, Kang DR, et al. Proposal of pharmacogenetics-based warfarin dosing algorithm in Korean patients. *J Hum Genet*. 2011;56:290-295. <https://doi.org/10.1038/jhg.2011.4>
11. Cini M, Legnani C, Cosmi B, et al. A new warfarin dosing algorithm including VKORC1 3730 G > A polymorphism: comparison with results obtained by other published algorithms. *Eur J Clin Pharmacol*. 2012;68:1167-1174. <https://doi.org/10.1007/s00228-012-1226-5>
12. Gage BF, Eby C, Johnson JA, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Ther*. 2008;84:326-331. <https://doi.org/10.1038/clpt.2008.10>
13. Kim HS, Lee SS, Oh M, et al. Effect of CYP2C9 and VKORC1 genotypes on early-phase and steady-state warfarin dosing in Korean patients with mechanical heart valve replacement. *Pharmacogenet Genomics*. 2009;19:103-112. <https://doi.org/10.1097/FPC.0b013e32831a9ae3>
14. Sharabiani A, Darabi H, Bress A, Cavallari L, Nutescu E, Drozda K. Machine learning based prediction of warfarin optimal dosing for African American patients. *IEEE International Conference on Automation Science and Engineering*. 2013;623-628. <https://doi.org/10.1109/CoASE.2013.6653999>
15. Kim DJ, Kim HS, Oh M, Kim EY, Shin JG. Cost effectiveness of genotype-guided warfarin dosing in patients with mechanical heart valve replacement under the fee-for-service system. *Appl Health Econ Health Policy*. 2017;15:657-667. <https://doi.org/10.1007/s40258-017-0317-y>
16. Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost-effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. *Ann Intern Med*. 2009;150:73-83.
17. Ageno W, Turpie AG. Exaggerated initial response to warfarin following heart valve replacement. *Am J Cardiol*. 1999;84:905-908.
18. Rose JP, Rihn TL, Long SF. Warfarin sensitivity after mechanical heart valve replacement. *Pharmacotherapy*. 1998;18:856-859.
19. Hamberg AK, Wadelius M, Friberg LE, Biss TT, Kamali F, Jonsson EN. Characterizing variability in warfarin dose requirements in children using modelling and simulation. *Br J Clin Pharmacol*. 2014;78:158-169.
20. Li Q, Tao H, Wang J, et al. Warfarin maintenance dose Prediction for Patients undergoing heart valve replacement— a hybrid model with genetic algorithm and Back-Propagation neural network. *Sci Rep*. 2018;8:9712. <https://doi.org/10.1038/s41598-018-27772-9>
21. Ma Z, Wang P, Gao Z, Wang R, Khalighi K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *PLoS ONE*. 2018;13:e0205872. <https://doi.org/10.1371/journal.pone.0205872>

22. Sharabiani A, Bress A, Douzali E, Darabi H. Revisiting warfarin dosing using machine learning techniques. *Comput Math Methods Med.* 2015;2015:560108. <https://doi.org/10.1155/2015/560108>
23. Park S-H, Kim S-N, Lee SH, Choe J-S, Choi Y. Development of 9th revision Korean food composition table and its major changes. *Korean J Community Nutr.* 2018;23:352-365.
24. Booth SL, Charnley JM, Sadowski JA, Saltzman E, Bovill EG, Cushman M. Dietary vitamin K1 and stability of oral anticoagulation: proposal of a diet with constant vitamin K1 content. *Thromb Haemost.* 1997;77:504-509.
25. IBM Watson Health. Drug interactions: Warfarin. IBM Micromedex: Drug Interaction Checking. <https://www.micromedexsolutions.com/>. Accessed July 11, 2019.
26. Wadelius M, Chen LY, Eriksson N, et al. Association of warfarin dose with genes involved in its action and metabolism. *Hum Genet.* 2007;121:23-34.
27. Ugrinowitsch C, Fellingham GW, Ricard MD. Limitations of ordinary least squares models in analyzing repeated measures data. *Med Sci Sports Exerc.* 2004;36:2144-2148.
28. Gurwitz JH, Avorn J, Ross-Degnan D, Chodnovskiy I, Ansell J. Aging and the anticoagulant response to warfarin therapy. *Ann Intern Med.* 1992;116:901-904.
29. Woodhouse K, Wynne HA. Age-related changes in hepatic function. *Drugs Aging.* 1992;2:243-255.
30. Wynne HA, Cope LH, Mutch E, Rawlins MD, Woodhouse KW, James OF. The effect of age upon liver volume and apparent liver blood flow in healthy man. *Hepatology.* 1989;9:297-301.
31. Sconce EA, Khan TI, Wynne HA, et al. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood.* 2005;106:2329-2333.
32. Rieder MJ, Reiner AP, Gage BF, et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N Engl J Med.* 2005;352:2285-2293.
33. D'Andrea G, D'Ambrosio RL, Di Perna P, et al. A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin. *Blood.* 2005;105:645-649.
34. Gage BF, Eby C, Milligan PE, Banet GA, Duncan JR, McLeod HL. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. *Thromb Haemost.* 2004;91:87-94.
35. Crespi CL, Miller VP. The R144C change in the CYP2C9\* 2 allele alters interaction of the cytochrome P450 with NADPH: cytochrome P450 oxidoreductase. *Pharmacogenetics.* 1997;7:203-210.
36. Cho S-M, Lee K-Y, Choi JR, Lee K-A. Development and comparison of warfarin dosing algorithms in stroke patients. *Yonsei Med J.* 2016;57:635-640.
37. Bae J-W, Kim H-K, Kim J-H, et al. Allele and genotype frequencies of CYP2C9 in a Korean population. *Br J Clin Pharmacol.* 2005;60:418-422. <https://doi.org/10.1111/j.1365-2125.2005.02448.x>
38. Nakai K, Tsuboi J, Okabayashi H, et al. Ethnic differences in the VKORC1 gene polymorphism and an association with warfarin dosage requirements in cardiovascular surgery patients. *Pharmacogenomics.* 2007;8:713-719. <https://doi.org/10.2217/14622416.8.7.713>
39. Cho HJ, Sohn KH, Park HM, et al. Factors affecting the interindividual variability of warfarin dose requirement in adult Korean patients. *Pharmacogenomics.* 2007;8:329-337. <https://doi.org/10.2217/14622416.8.4.329>
40. Obayashi K, Nakamura K, Kawana J, et al. VKORC1 gene variations are the major contributors of variation in warfarin dose in Japanese patients. *Clin Pharmacol Ther.* 2006;80:169-178. <https://doi.org/10.1016/j.clpt.2006.04.010>
41. Lee SC, Ng SS, Oldenburg J, et al. Interethnic variability of warfarin maintenance requirement is explained by VKORC1 genotype in an Asian population. *Clin Pharmacol Ther.* 2006;79:197-205. <https://doi.org/10.1016/j.clpt.2005.11.006>
42. Veenstra DL, You JHS, Rieder MJ, et al. Association of Vitamin K epoxide reductase complex 1 (VKORC1) variants with warfarin dose in a Hong Kong Chinese patient population. *Pharmacogenet Genomics.* 2005;15:687-691. <https://doi.org/10.1097/01.fpc.0000174789.77614.68>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Nguyen VL, Nguyen HD, Cho Y-S, et al. Comparison of multivariate linear regression and a machine learning algorithm developed for prediction of precision warfarin dosing in a Korean population. *J Thromb Haemost.* 2021;19:1676–1686. <https://doi.org/10.1111/jth.15318>