Original Article Clinics in Orthopedic Surgery 2023;15:678-689 • https://doi.org/10.4055/cios22240

Development of Prediction Model Using Machine-Learning Algorithms for Nonsteroidal Anti-inflammatory Drug-Induced Gastric Ulcer in Osteoarthritis Patients: Retrospective Cohort Study of a Nationwide South Korean Cohort

Jaehan Jeong, MS*, Hyein Han, MS*, Du Hyun Ro, MD[†,‡], Hyuk-Soo Han, MD[†], Sungho Won, PhD*[,§,‖]

*Department of Public Health Sciences, Seoul National University, Seoul,
[†]Department of Orthopedic Surgery, Seoul National University Hospital, Seoul National University College of Medicine, Seoul,
[‡]CONNECTEVE, Co. LTD., Seoul,
[§]Institute of Health and Environment, Seoul National University, Seoul,
[‖]RexSoft Inc., Seoul, Korea

Background: Nonsteroidal anti-inflammatory drugs (NSAID) are currently among the most prescribed medications worldwide to relieve pain and reduce inflammation, especially in patients suffering osteoarthritis (OA). However, NSAIDs are known to have adverse effects on the gastrointestinal system. If a gastric ulcer occurs, planned OA treatment needs to be changed, incurring additional treatment costs and causing discomfort for both patients and clinicians. Therefore, it is necessary to create a gastric ulcer prediction model that can reflect the detailed health status of each individual and to use it when making treatment plans.

Methods: Using sample cohort data from 2008 to 2013 from the National Health Insurance Service in South Korea, we developed a prediction model for NSAID-induced gastric ulcers using machine-learning algorithms and investigated new risk factors associated with medication and comorbidities.

Results: The population of the study consisted of 30,808 patients with OA who were treated with NSAIDs between 2008 and 2013. After a 2-year follow-up, these patients were divided into two groups: without gastric ulcer (n=29,579) and with gastric ulcer (n=1,229). Five machine-learning algorithms were used to develop the prediction model, and a gradient boosting machine (GBM) was selected as the model with the best performance (area under the curve, 0.896; 95% confidence interval, 0.883–0.909). The GBM identified 5 medications (loxoprofen, aceclofenac, talniflumate, meloxicam, and dexibuprofen) and 2 comorbidities (acute upper respiratory tract infection [AURI] and gastroesophageal reflux disease) as important features. AURI did not have a dose-response relationship, so it could not be interpreted as a significant risk factor even though it was initially detected as an important feature and improved the prediction performance.

Conclusions: We obtained a prediction model for NSAID-induced gastric ulcers using the GBM method. Since personal prescription period and the severity of comorbidities were considered numerically, individual patients' risk could be well reflected. The prediction model showed high performance and interpretability, so it is meaningful to both clinicians and NSAID users.

Keywords: *Non-steroidal anti-inflammatory drugs, Osteoarthritis, Stomach ulcer, Prediction model, Machine learning*

Received July 20, 2022; Revised December 14, 2022; Accepted February 1, 2023
Correspondence to: Sungho Won, PhD
Department of Public Health Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
Tel: +82-2-880-2714, Fax: +82-303-0942-2714, E-mail: sunghow@gmail.com
Co-Correspondence to: Hyuk-Soo Han, MD
Department of Orthopedic Surgery, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
Tel: +82-2-2072-3212, Fax: +82-2-764-2718, E-mail: oshawks7@snu.ac.kr
Jaehan Jeong and Hyein Han contributed equally to this work.

Nonsteroidal anti-inflammatory drugs (NSAIDs) have been widely used to relieve pain, reduce inflammation, and lower the temperature of patients with osteoarthritis (OA).[1] NSAIDs were proven to be very effective short-term painkillers; specifically, the treatment group showed 15.6% of pain relief compared to the placebo group within 12 weeks.[1]

NSAIDs block the production of prostaglandins by inhibiting the two cyclooxygenase enzymes, cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2).[2] Prostaglandins are key factors in many cellular processes, such as gastrointestinal cytoprotection, hemostasis and thrombosis, inflammation, renal hemodynamics, turnover of cartilage, and angiogenesis.[2] NSAIDs are classified according to their COX selectivity as non-selective NSAIDs and COX-2 selective NSAIDs.[2]

Lots of side effects have been reported for NSAIDs, many of which can be explained by these pharmacological mechanisms.[3] Gastrointestinal and cardiovascular side effects, which account for a large proportion of them, can be life-threatening.[4-9] In particular, gastrointestinal side effects are known to occur in 10% to 60% of patients taking NSAIDs, and gastric or duodenal ulcers are known to occur in 20% to 30% of patients.[10] Additionally, serious complications such as severe bleeding, perforation, and obstruction in patients with peptic ulcers occur in approximately 1% to 2% of patients.[11] Taking into consideration the characteristics of NSAIDs and patient risk factors, making appropriate prescriptions and ensuring good management are required to avoid high medical expenses and high mortality.[11,12] To sum up, NSAIDs are commonly used for OA patients and biologically proven to induce gastric ulcers. Therefore, we regarded NSAIDs as a main risk factor for gastric ulcers in OA patients.

Several studies have highlighted the risk factors for NSAID-induced peptic ulcers such as a history of peptic ulcer, high-dose NSAIDs, concomitant antiplatelet agents, anticoagulants, or corticosteroids.[3,6,11] Patients with these risk factors are recommended to use COX-2 selective NSAIDs or non-selective NSAIDs in combination with gastrointestinal protective agents (such as proton pump inhibitors, H2 receptor antagonists, or misoprostol).[3]

In South Korea, 87.7% of arthritis patients took NSAIDs 3 months or more, and 47.2% of them took high doses.[13] Considering the large proportion of the elderly in OA patients, long-term use of high-dose NSAIDs may greatly increase the possibility of gastric ulcers. If a gastric ulcer occurs, planned OA treatment needs to be changed, which incurs additional treatment costs and causes discomfort for both patients and clinicians. Therefore, it is necessary to create a gastric ulcer prediction model that can reflect the detailed health status of each individual and to use it when making treatment plans. In summary, the gastric ulcer prediction enables prevention of unnecessary drug use and reduction of medical costs.

This study was divided into two main parts. First, we developed an NSAID-induced gastric ulcer prediction model using machine-learning (ML)-based algorithms. Second, we tried to discover risk factors among individual comorbidities during patient hospital visits and medication adherence during the observation period, controlling for covariates such as established risk factors, socioeconomic variables, and Charlson comorbidity index (CCI).

METHODS

The protocol of the study was approved by the Institutional Review Board of Seoul National University and was conducted in accordance with research ethics (No. E2011/003-002). In addition, the data used in this study were officially provided by National Health Insurance Service (NHIS) after obtaining approval from NHIS Review Committee of Research Support (No. NHIS-2021-2-093).

Data Source

The NHIS is a health insurance program, and approximately 96% of the total population of South Korea are enrolled in the NHIS. Therefore, with the cohort provided

by NHIS public health researchers and policy makers with representatives, universally useful information regarding citizens' utilization of health insurance and health examinations could be generated.

This study used data from the NHIS National Sample Cohort (NHIS-NSC), a population-based cohort established by the NHIS. The NHIS-NSC was constructed based on data from the entire South Korean population collected in 2002. The NHIS-NSC contains approximately one million subscribers to the NHIS and the Medical Aid program extracted using stratified sampling methods in 2002, comprising approximately 2% of all Koreans. NHIS-NSC 1.0 comprises information from 2002 to 2013 (12-year cohort), while NHIS-NSC 2.0 is extended to 2015 (14-year cohort). A 14-year cohort (2002–2015) of the NHIS-NSC 2.0 was tracked in terms of socioeconomic variables (residence location, year and month of death, and income level) and medical treatment details (health examinations and medical care history). Diseases in the NHIS-NSC 2.0 are registered using the sixth edition of the Korean Classification of Disease, which was modified from the International Classification of Disease, Tenth Revision (ICD-10) for use by the NHIS and medical care institutions in South Korea.

Study Population

This was a population-based and retrospective cohort study in Korea. We focused on patients with OA that were treated with NSAIDs, and we developed models to predict NSAID-induced gastric ulcers. We chose 2008 to 2013 as the index date, with a 1-year washout period for NSAIDs and a 2-year follow-up period for each patient (Fig. 1).

First, OA patients (ICD-10: M15-19) treated with NSAIDs and with at least the Anatomical Therapeutic Chemical (ATC) code of NSAIDs between 2008 and 2013 were selected (n = 191,565). Next, previous NSAID users who had prescription records within 1 year of the index date were excluded (n = 152,012). We also excluded patients with any history of gastric cancer, ankylosing spondylitis, and psoriatic arthritis during the washout period (n = 1,703).[14] In addition, patients who died during the follow-up period (n = 3,298) and patients under 20 years of age (n = 3,735) at the index date were also excluded. Finally, 30,808 patients were used to train and validate the ML algorithms of the prediction model (Fig. 2).

Definition and Modeling Strategy

Patients with gastric ulcer diseases (GUD) were diagnosed with gastric ulcer (ICD-10: K25) using upper gastrointestinal endoscopy (operational code of upper gastrointestinal endoscopy: E7611). The input features for the prediction

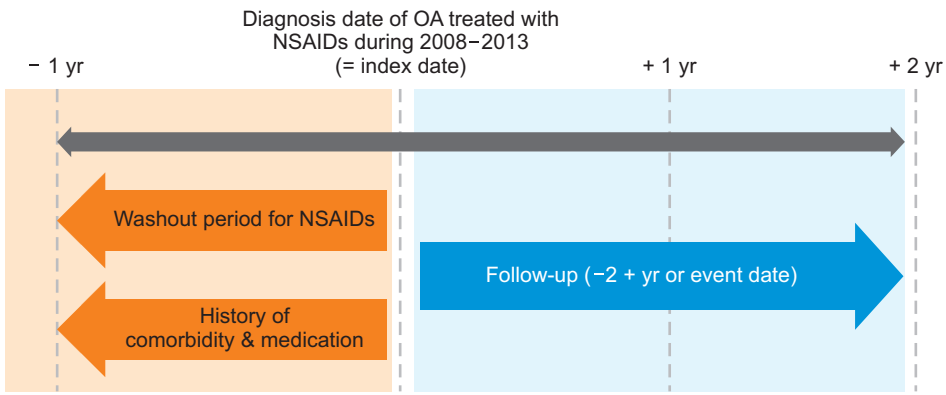

Fig. 1. Observation period (follow-up year) and washout period for this study. OA: osteoarthritis, NSAID: nonsteroidal anti-inflammatory drug.

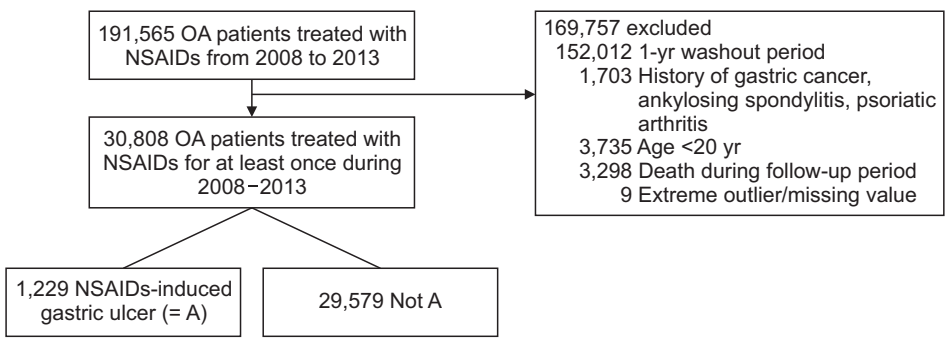

Fig. 2. Flowchart of the study design. OA: osteoarthritis, NSAID: nonsteroidal anti-inflammatory drug.

model were individual medical records, including the patient visit ratio for comorbidities and each NSAID medication adherence during the observation period. Twenty-six types of NSAIDs, such as aceclofenac, diclofenac, and meloxicam, were selected and prescription periods for each drug within the observation period were calculated. To keep track of the medication records, we used medication adherence, measured by the medication possession ratio (MPR). Originally, the MPR measures the percentage of days a patient was given the medication during the observation period.[15] Similarly, adherence of each NSAID was defined as follows:

$$\frac{\sum \textit{Prescribed days of medicine}}{\textit{Total observation time}}$$

In addition, the patient visit ratio, the severity of each disease, and time from the index date to the event date or the end of the follow-up period was calculated for 12 comorbidities. All diagnostic records (ICD-10) were grouped by the first three digits, which indicate the main disease category. Detailed codes for each disease are shown in Table 1. The patient visit ratio for comorbidities was defined as follows:

$$\frac{\sum \textit{Days of visiting hospital due to comorbidity}}{\textit{Total observation time}}$$

The numeric value of each dimension represents the total number of occurrences of a specific code between the index date and the event date divided by the follow-up days. Before developing the prediction model, covariates that could cause a serious bias to the outcome variable must be controlled. Therefore, 1:1 propensity score matching was performed on the training group. Covariates were divided into three parts: socioeconomic variables, well-known risk factors, and CCI. Age, sex, and socioeconomic status at the time of the index date were considered as variables indicating the basic demographic characteristics of the patients. Income level and NHIS insurance type were used as socioeconomic variables. Income level ranged from 0 to 10. The lower the level, the weaker the socioeconomic status. The NHIS insurance types were categorized as insured employees, insured self-employed individuals, and medical aid beneficiaries. All variables representing demographic characteristics are required to be in the controlled covariates because they may act as confounders in the analysis. Risk factors for NSAID-related peptic ulcers include a previous history of peptic ulcer and concomitant use of corticosteroids, anticoagulants, and antiplatelet agents. Patients with a previous history of peptic ulcer, including gastric and duodenal ulcers within the washout period, were defined as those with a diagnosis of the following codes (ICD-10: K25-28). The study population (n = 30,808) was randomly stratified into 75% training and 25% test groups, and propensity score matching of socio-

Table 1. Detailed ICD-10 Codes for Comorbidities

Disease	ICD code
Diabetes mellitus	E10-E14
Dyslipidemia	E78
Angina pectoris	I20
Myocardial infarction	I21-I23
Stroke	I60-I69
End stage renal disease	N18
Chronic obstructive pulmonary disease	J44
Cirrhosis	K74
Hypertension	I10
Gastroesophageal reflux disease	K21
Acute upper respiratory infection	J00-J06
Mental and behavioral disorders	F00-F99

ICD-10: International Classification of Disease, Tenth Revision.

Table 2. The Number of Prescriptions for 10 Most Frequently Prescribed Anti-ulcer Agents in the Whole Study Population

Medication	No. (%)
Ranitidine	12,221 (11.2)
Cimetidine	12,016 (11.1)
Rebamipide	11,462 (10.6)
Almagate	11,013 (10.2)
Artemisiae Argyi Folium 95% ethanol extract (20→1)	9,693 (9.0)
Alibendol	8,915 (8.0)
Levosulpiride	8,002 (7.4)
Mosapride	5,870 (5.4)
Itopride	2,946 (2.7)
Trimebutine	2,538 (2.3)

Only anti-ulcer drugs prescribed on the same day as nonsteroidal anti-inflammatory drugs were counted.

economic variables, existing risk factors, and CCI was performed in the training group with a ratio of 1:1 for GUD and non-GUD subjects.

The use of anti-ulcer agents was not considered a covariate. In our data, 27,619 people (90%) were prescribed NSAIDs and gastroprotective agents simultaneously, which indicates almost everyone used gastroprotective drugs prophylactically when taking NSAIDs. Statistically, it means the presence of multi-collinearity, and such imbalanced data can degrade the model performance.[16] Gastroprotective agents were defined as Korean Drug Classification codes 232, 233, 234, 235, 236, 237, 238, and 239.[17] Detailed prescription frequencies for anti-ulcer agents are shown in Table 2.

Machine Learning

For this study, 38 features (variables), including 12 comorbidities and 26 types of NSAIDs, were available from the NHIS-NSC for model development. To predict the occurrence of NSAID-induced gastric ulcers, we applied five machine learning algorithms: logistic regression (LR), support vector machine (SVM), random forest (RF), gradient boosting machine (GBM), and eXtreme Gradient Boosting (XGBoost).

LR is a statistical model fitting the logistic curve to the data.[18] LR is highly interpretable and computationally fast. However, it is vulnerable to multi-collinearity and requires large samples.[18] SVM conducts classification by drawing boundaries between classes that maximize the distance between the boundary and classes.[19] This can effectively perform a non-linear classification, but it is very complex and learning speed is slow.[19] RF is an ensemble-based method that operates based on several different decision trees.[18] This is robust against noise and overfitting, but is computationally intensive.[18] GBM is also an ensemble-based method that uses small (shallow) trees.[20] It is computationally less intensive than RF, but hyperparameter tuning is more difficult.[20] XGBoost is another version for GBM in which the regularization method is added.[21] However, still there are weaknesses related to tree-based algorithms, such as difficulty in interpretation. In summary, the advantages and disadvantages of each machine learning algorithm are very distinct, and those models were applied to find the best fit.

Data were split into train and test sets to minimize overfitting problems. The train data were used for developing the prediction model, and the test data were used

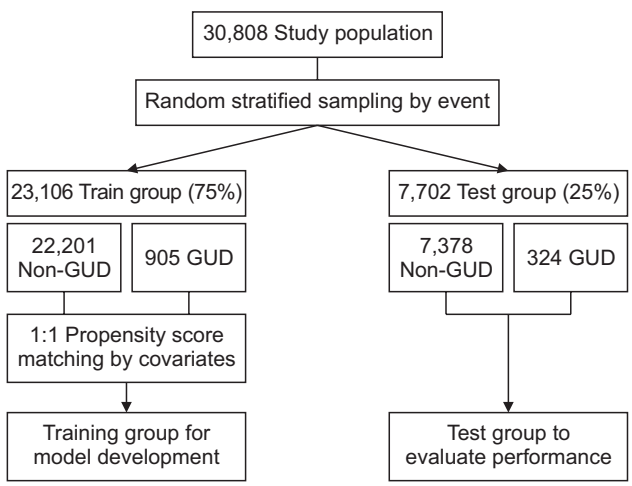

Fig. 3. Flowchart of the model evaluation strategy. GUD: gastric ulcer disease.

Table 3. Baseline Characteristics of GUD and Non-GUD Patients before Matching

Variable	Non-GUD (n = 29,579)	GUD (n = 1,229)	*p*-value
Age (yr)	55.44 ± 14.91	58.91 ± 13.59	< 0.001
Sex			0.428
Male	12,803 (0.43.3)	546 (44.4)	
Female	16,776 (0.56.7)	683 (55.6)	
Insurance type			0.153
Self-employed	10,405 (35.2)	414 (33.7)	
Insured	18,087 (61.1)	758 (61.7)	
Medical aid	1,087 (3.7)	57 (4.6)	
Income			0.644
Low (0–2)	5,535 (18.7)	239 (19.4)	
Middle (3–6)	9,416 (31.8)	377 (30.7)	
High (7–10)	14,628 (49.5)	613 (49.9)	
Risk factor			
Past PUD history	2,518 (8.5)	316 (25.7)	< 0.001
Corticosteroid	7 (0.02)	0	0.589
Anticoagulants	147 (0.5)	10 (0.8)	0.146
Antiplatelets	1,150 (3.9)	81 (6.6)	< 0.001
CCI	0.51 ± 0.94	0.96 ± 1.41	< 0.001

Values are presented as mean ± standard deviation or number (%). *p*-values were generated by two-sample *t*-tests for numerical variables, and chi-square or Fisher's exact tests for categorical variables.
GUD: gastric ulcer disease, PUD: peptic ulcer disease, CCI: Charlson comorbidity index.

for model evaluation.[19] A flowchart of the model development for our study is shown in Fig. 3.

Feature Selection, Cross-validation, and Visualization

Feature selection for the development of the gastric ulcer prediction model was performed by using the recursive feature elimination method. Feature selection tends to reduce computational costs and create a high-quality model. Model performance was evaluated by using 10-fold cross-validation (CV). Ten-fold CV randomly divides the dataset into 10 subsets with the same or similar number of events. Then, nine subsets were used as train sets, and the other was used as a test set. The performance was measured by combining the results from 10 test sets, and the model with the best performance was determined.

In addition, feature importance and the direction for the selected features were visualized. We considered a partial dependence plot because feature importance could not accurately explain the direction of each variable's effect on the outcome.[22] In summary, feature selection and CV were for obtaining high predictability, and visualizing was for high interpretability.

Statistical Analysis

In this study, the comparison of continuous variables between the two groups was performed using a two-sample *t*-test with mean and standard deviation. Comparisons of categorical variables between the two groups were assessed using the chi-square or Fisher's exact test and expressed as numbers and percentages. Drug adherence and medication frequency (the number of hospital visits for prescriptions) were also compared between GUD and non-GUD groups.

The performance of the learning model developed using the ML algorithm was evaluated using the area under the curve (AUC) of the receiver operating character-

Table 4. Baseline Characteristics of GUD and Non-GUD Patients for Training and Test Groups

	Training group			Test group		
	Non-GUD (n = 905)	GUD (n = 905)	*p*-value	Non-GUD (n = 7,378)	GUD (n = 324)	*p*-value
Age (yr)	59.81 ± 13.4	59.11 ± 13.35	0.296	55.44 ± 14.8	58.2 ± 14.44	0.001
Sex			1			0.492
Male	421 (46.5)	414 (45.7)		3,160 (42.8)	132 (40.7)	
Female	484 (53.5)	491 (54.3)		4,218 (57.2)	192 (59.3)	
Insurance type			0.594			0.307
Self-employed	290 (32.0)	309 (34.1)		2,560 (34.7)	105 (32.4)	
Insured	576 (63.6)	555 (61.3)		4,561 (61.8)	203 (62.7)	
Medical aid	290 (4.3)	41 (4.5)		257 (3.5)	16 (4.9)	
Income			0.955			0.198
Low (0–2)	163 (18.0)	168 (18.6)		1,354 (18.4)	71 (21.9)	
Middle (3–6)	286 (31.6)	284 (31.4)		2,367 (32.1)	93 (28.7)	
High (7–10)	456 (50.4)	453 (50.1)		3,657 (49.6)	160 (49.4)	
Risk factor						
Past PUD history	242 (26.7)	242 (26.7)	1	628 (8.5)	74 (22.8)	< 0.001
Corticosteroid	0	0	-	2 (0)	0	1
Anticoagulants	8 (0.9)	8 (0.9)	1	43 (0.6)	2 (0.6)	1
Antiplatelets	64 (7.1)	68 (7.5)	0.786	272 (3.7)	13 (4.0)	0.878
CCI	0.84 ± 1.2	0.90 ± 1.32	0.417	0.50 ± 0.91	0.95 ± 1.42	< 0.001

Values are presented as mean ± standard deviation or number (%). Baseline characteristics for the train group were derived after propensity score matching. *p*-values were generated by two-sample *t*-tests for numerical variables, and chi-square or Fisher's exact tests for categorical variables.
GUD: gastric ulcer disease, PUD: peptic ulcer disease, CCI: Charlson comorbidity index.

istic analysis. In addition, basic epidemiological indices, such as sensitivity, specificity, positive predicted value (PPV), and negative predicted value (NPV), were calculated. Statistical significance was set at $p < 0.05$. All development techniques were performed using the R software (ver. 3.3.3, R Foundation for Statistical Computing).

RESULTS

Baseline Characteristics of Study Population

This study included 30,808 patients after the exclusion process using specific criteria. The subjects were divided into GUD and non-GUD groups. Baseline characteristics of the 30,808 subjects before covariates matched by the propensity score matching method are presented in Table 3.

As all the covariates were categorical variables, except age and CCI, the frequencies were summarized with these variables, and the descriptive statistics (mean and standard deviation) were summarized for age and CCI variables. The proportion of women was higher than that of men; however, there were no significant differences between the two groups ($p = 0.428$). Similarly, other socioeconomic variables showed no significant difference between the two groups. In contrast, age, well-known risk factors, and CCI were significantly different between the two groups. The mean age of the GUD group (58.91 ± 13.59) was significantly higher than that of the non-GUD group (55.44 ± 14.91, $p < 0.001$). Among the risk factors, patients with a history of peptic ulcers and concomitant use of antiplatelet agents were significantly different between the two groups ($p < 0.001$). In the non-GUD group, the proportion of patients with a history of peptic ulcer was 8.5%, whereas it was 25.7% in the GUD group ($p < 0.001$). For antiplatelet drugs, the GUD group recorded a 2.7% higher rate than the non-GUD group, and the difference was statistically significant ($p < 0.001$).

Table 4 summarizes the baseline characteristics after propensity score matching of the training group and test group. The characteristics of the matched training group differed from those of the former training group. All covariates, including age, risk factors, and CCI, were evenly distributed between the two groups. In the test group, the prevalence of gastric ulcer was 4.2%, and significant differences in age, past peptic ulcer disease history, and CCI existed between the two groups.

Performance of the Prediction Models

The GBM algorithm-based prediction model was selected as the best predictive model among the five ML algorithms (AUC, 0.896; 95% confidence interval [CI], 0.883–0.909). Table 5 and Fig. 4 show the performance of the respective algorithms with seven significant features after recursive feature selection. LR was found to have the worst classification prediction performance at 0.636 (95% CI, 0.608–

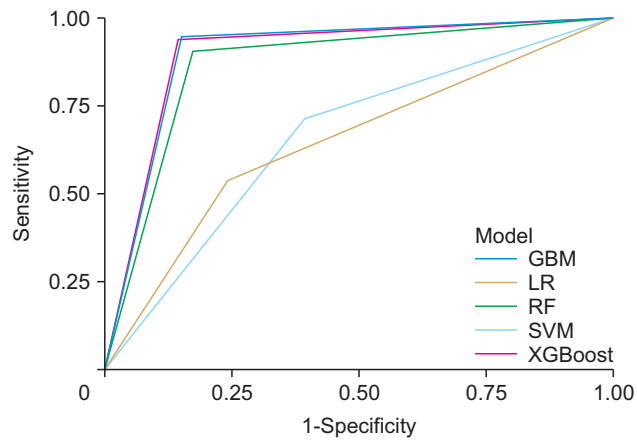

Fig. 4. Receiver operating characteristic curves comparing the area under the curve of each model. GBM: gradient boosting machine, LR: logistic regression, RF: random forest, SVM: support vector machine, XGBoost: eXtreme Gradient Boosting.

Table 5. Comparison of Model Performance to Predict NSAID-Induced Gastric Ulcer

Type	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV
LR	0.636 (0.608–0.663)	0.452	0.820	0.101	0.972
SVM	0.637 (0.609–0.664)	0.568	0.712	0.078	0.974
RF	0.862 (0.844–0.880)	0.886	0.839	0.195	0.994
GBM	0.896 (0.883–0.909)	0.944	0.847	0.214	0.997
XGBoost	0.893 (0.878–0.908)	0.923	0.863	0.201	0.996

After recursive feature selection, model performance was evaluated.
NSAID: nonsteroidal anti-inflammatory drug, AUC: area under the curve, CI: confidence interval, PPV: positive predicted value, NPV: negative predicted value, LR: logistic regression, SVM: support vector machine, RF: random forest, GBM: gradient boosting machine, XGBoost: eXtreme Gradient Boosting.

Table 6. The *p*-values for Testing the AUC in the Comparative Study of Two Diagnostic Tests

Model	LR	SVM	RF	GBM	XGBoost
LR	1	0.592	< 0.001	< 0.001	< 0.001
SVM	0.592	1	< 0.001	< 0.001	< 0.001
RF	< 0.001	< 0.001	1	< 0.001	< 0.001
GBM	< 0.001	< 0.001	< 0.001	1	0.103
XGBoost	< 0.001	< 0.001	< 0.001	0.103	1

p-values were generated by the Delong test.
AUC: area under the curve, LR: logistic regression, SVM: support vector machine, RF: random forest, GBM: gradient boosting machine, XGBoost: eXtreme Gradient Boosting.

0.663). Overall, tree-based classification prediction models such as RF, GBM, and XGBoost performed significantly better than other models like LR and SVM.

The sensitivity and specificity of the GBM model were 0.944 and 0.847, respectively, indicating the screening ability of the prediction model. The sensitivity is higher than specificity in tree-based prediction models; in contrast, the opposite is true for the other models. PPV and NPV were 0.214 and 0.997, respectively, for the developed GBM model. In addition, other prediction models also showed a high NPV of > 0.97.

Table 6 presents the results of verifying whether there is a significant difference between the derived prediction models. It can be confirmed that GBM is a significantly better model as compared to other models, except XGBoost, which has similar model performance ($p = 0.103$). In addition, the performances of the LR and SVM models were similar, with no statistically significant difference.

Important Features

Seven important features were selected in the GBM model, and the number of important features and the combinations which generate the best performance were chosen with a CV procedure. The patient visit ratios for acute upper respiratory tract infection (AURI) and gastroesophageal reflux disease (GERD) and medication adherence for loxoprofen, aceclofenac, talniflumate, meloxicam, and dexibuprofen were selected as key variables, and their effect sizes are shown in Fig. 5. This indicates that these seven variables were highly associated with the occurrence of gastric ulcers. Five NSAIDs, selected as important features, were also frequently used by the patients considered in this study. Meloxicam showed no significant difference in medication frequency, but represented a statistically significant difference in medication adherence between the two groups (Table 7).

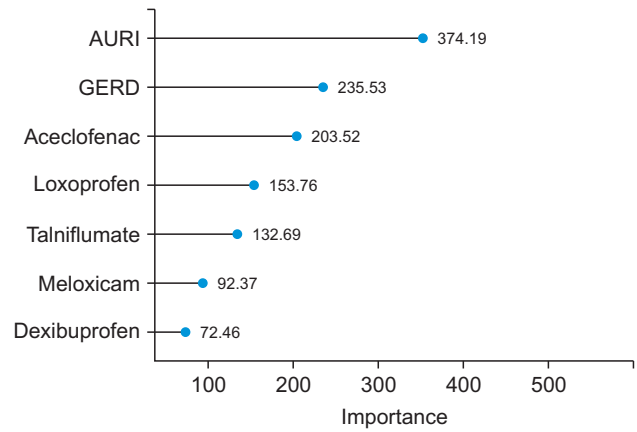

Fig. 5. Feature importance of seven variables in the best prediction model (gradient boosting machine). AURI: acute upper respiratory tract infection, GERD: gastroesophageal reflux disease.

Risk Factors

Fig. 6 shows the incidence of gastric ulcers tended to increase when the severity of the comorbidity (GERD) or the medication adherence (loxoprofen, aceclofenac, talniflumate, meloxicam, and dexibuprofen) increased despite non-smooth curves generated by the small sample sizes. AURI did not have a dose-response relationship, and it could not be interpreted as a significant risk factor even though it was initially detected as an important feature and improved the prediction performance. Finally, six factors, loxoprofen, aceclofenac, talniflumate, meloxicam, dexibuprofen, and GERD, were found to be definite risk factors in terms of causality.

DISCUSSION

This study was based on representative Korean population using data obtained from a well-established national cohort. Risk prediction models based on individual medi-

Table 7. Characteristics of NSAID Medication Frequency and Medication Adherence between GUD and non-GUD groups

Medication	Frequency of medication			Medication adherence		
	Non-GUD (n = 22,201)	GUD (n = 905)	*p*-value	Non-GUD (n = 22,201)	GUD (n = 905)	*p*-value
	Number (%)			Probability		
Aceclofenac	16,344 (55.3)	572 (46.5)	< 0.001	0.021	0.054	< 0.001
Loxoprofen	14,360 (48.6)	439 (35.7)	< 0.001	0.009	0.016	< 0.001
Talniflumate	11,286 (38.2)	371 (30.2)	< 0.001	0.007	0.012	0.000
Meloxicam	9,115 (30.8)	399 (32.5)	0.220	0.014	0.048	< 0.001
Dexibuprofen	8,465 (28.6)	253 (20.6)	< 0.001	0.003	0.007	0.002
Zaltoprofen	3,447 (11.7)	121 (9.9)	0.052	0.002	0.005	0.178
Ibuprofen	2,621 (8.9)	68 (5.5)	< 0.001	0.001	0.001	0.142
Mefenamic	2,379 (8.0)	71 (5.8)	0.004	0.001	0.001	0.322
Celecoxib	2,034 (6.9)	101 (8.2)	0.070	0.007	0.013	0.035
Nabumetone	1,809 (6.1)	70 (5.7)	0.547	0.003	0.013	0.103
Morniflumate	1,347 (4.6)	53 (4.3)	0.691	0.001	0.006	0.033
Etodolac	918 (3.1)	26 (2.1)	0.049	0.001	0.001	0.203
Naproxen	774 (2.6)	17 (1.4)	0.007	0.000	0.001	0.604
Diclofenac	614 (2.1)	19 (1.6)	0.200	0.001	0.002	0.165
Piroxicam	289 (1.0)	11 (0.9)	0.774	0.000	0.001	0.305

NSAID: nonsteroidal anti-inflammatory drug, GUD: gastric ulcer disease.

cal records can be easily implemented in medical practice, and we believe that our results are applicable to the general Korean population. NSAID-induced gastric ulcers are predictable, and individual patients could reduce their risks by modifying medication adherence and managing associated comorbidities. The prediction model proposed in this study was derived from a variety of candidate predictors, including the period of using NSAIDs and the severity of coexisting diseases during the observation period, while controlling for socioeconomic variables, established risk factors, and CCI. We developed a model for predicting the occurrence of NSAID-induced gastric ulcers and constructed a model that is most suitable for the patient by comparing the performance of all the models considered in this study using the AUC value. The results of this study confirmed that the ML-based prediction model could accurately predict disease occurrence based on individual patient's medical records. Other risk factors, including clinical measurement indices such as BMI and cholesterol, were not considered in this study because these indices could not represent the real-time status around the event date because of limitations of administrative data.

There are several distinctive features for the proposed method in this study. First, the previous investigation focused on the severe gastric ulcer with bleeding,[23,24] and we built a prediction model for an overall gastric ulcer. Second, NSAIDs are painkillers and their prescription periods differ by the degrees of symptoms. In order to reflect the dose-response relationship, the prescription day at the continuous scale was considered as a predictor. Third, hospital visit ratios for coexisting diseases were calculated and included as a measure of the severity of comorbidities. Such distinctiveness may allow us to obtain the prediction model with the best performance for NSAID-induced gastric ulcers using population-based big data.

The results of this study revealed that the GBM algorithm had the best prediction performance (AUC, 0.896; 95% CI, 0.883–0.909). In addition, loxoprofen, aceclofenac, talniflumate, meloxicam, dexibuprofen, and GERD were found to be significant risk factors. In this

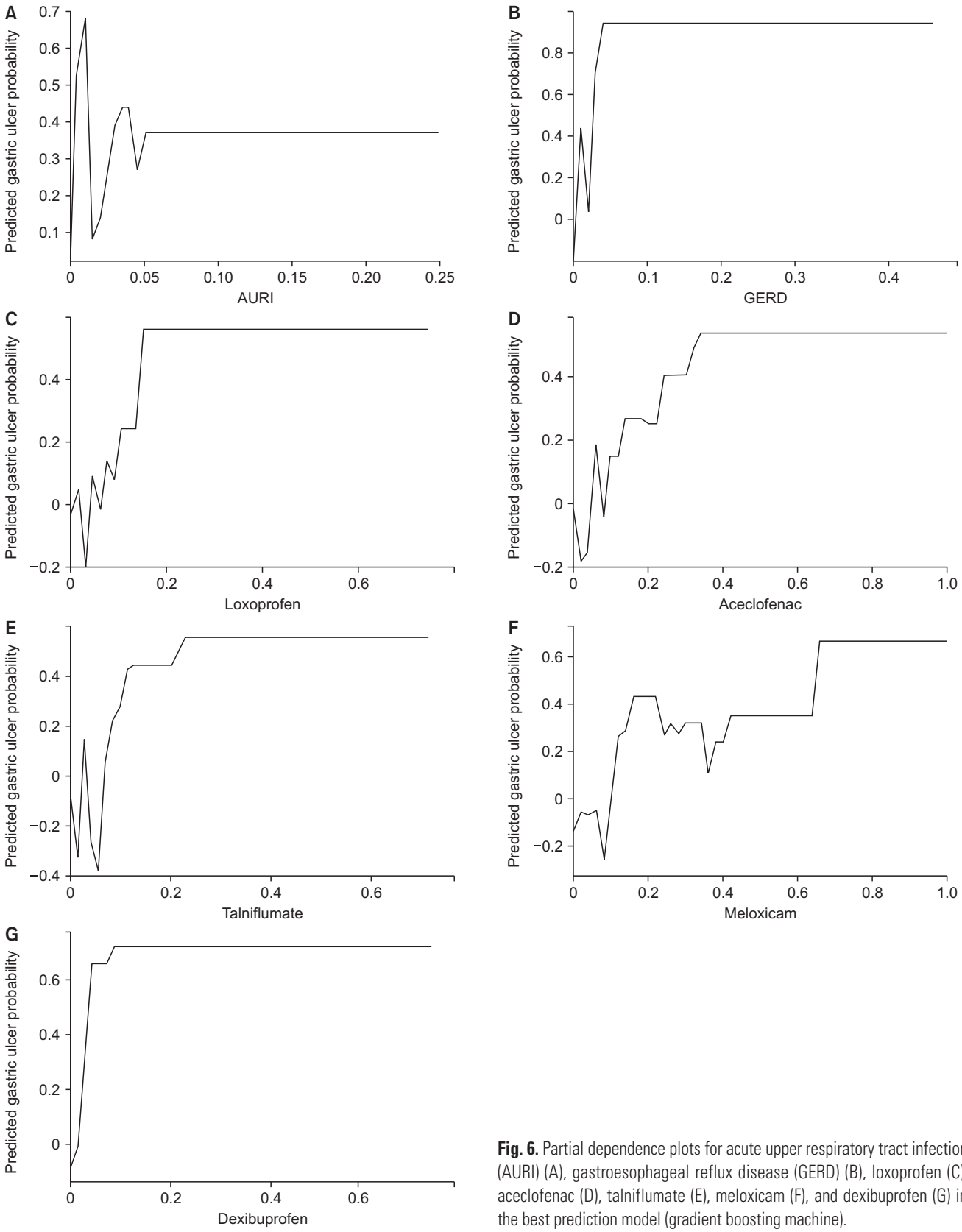

Fig. 6. Partial dependence plots for acute upper respiratory tract infection (AURI) (A), gastroesophageal reflux disease (GERD) (B), loxoprofen (C), aceclofenac (D), talniflumate (E), meloxicam (F), and dexibuprofen (G) in the best prediction model (gradient boosting machine).

study, aceclofenac and meloxicam were known to be risk factors for gastric ulcer, and these medications were found to possibly cause upper gastrointestinal side effects in a previous study.[25] Therefore, it seemed that the prediction model and risk factors developed in this study were reliable. However, in the PPVs of this study, the percentage of patients with a positive test for disease was lower than the sensitivity and specificity because of the imbalance of the outcome variable.

We used ML algorithms with excellent performance and investigated new risk factors. In our analysis, the proposed prediction model was developed based on individual medical records and validated using a test group, and the proposed model showed excellent performance. We expect this study can contribute to the prediction of NSAID-induced gastric ulcers using patient-level medical information.

However, this study has a few limitations. First, we were unable to obtain many concomitant cases of NSAIDs owing to the small sample size of the NHIS-NSC. Therefore, it was difficult to identify a significant correlation between the medications. Further studies using a full nationwide dataset are required to consider the correlation between medications. Second, specific and accurate information about the gastric ulcer status could only be obtained by gastric endoscopy and other clinical characteristics. To overcome this issue, we tried to utilize individual data as close to real-time as possible when performing predictive modeling. Although the NHIS-NSC data also include health screening data, it was insufficient to reliably utilize the data because the measurement period deviation from the onset of gastric ulcer or the end of the observation point varied depending on the patient. Therefore, it seems that it will be possible to construct a more accurate model for predicting gastric ulcer events using electronic medical records and lab data to fit each patient's characteristics.[26-28] In addition, gastrointestinal endoscopy images, helicobacter pylori test values, and numerical data based on electronic medical records could also be proficient in future studies.

In this study, we developed a high-quality prediction model for NSAID-induced gastric ulcers, taking into account drug adherence and severity of comorbidities as well as known risk factors. This result indicates the practical value of the proposed method for both patients and clinicians, and it helps us to decrease the medical burden on our NSAID-induced gastric ulcers.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

This research originated from Jaehan Jeong's master's thesis and was further improved. It was supported by a grant from the Patient-Centered Clinical Research Coordinating Center (PACEN) funded by the Ministry of Health & Welfare, Republic of Korea (grant no. HC19C0098).

ORCID

Jaehan Jeong https://orcid.org/0000-0002-9433-6052
Hyein Han https://orcid.org/0000-0001-9770-364X
Du Hyun Ro https://orcid.org/0000-0001-6199-908X
Hyuk-Soo Han https://orcid.org/0000-0003-1229-8863
Sungho Won https://orcid.org/0000-0001-5751-5089

REFERENCES

1. Crofford LJ. Use of NSAIDs in treating patients with arthritis. Arthritis Res Ther. 2013;15(Suppl 3):S2.
2. Gunaydin C, Bilge SS. Effects of nonsteroidal anti-inflammatory drugs at the molecular level. Eurasian J Med. 2018;50(2):116-21.
3. Lanza FL, Chan FK, Quigley EM; Practice Parameters Committee of the American College of Gastroenterology. Guidelines for prevention of NSAID-related ulcer complications. Am J Gastroenterol. 2009;104(3):728-38.
4. Fries JF, Williams CA, Bloch DA, Michel BA. Nonsteroidal anti-inflammatory drug-associated gastropathy: incidence and risk factor models. Am J Med. 1991;91(3):213-22.
5. Johnson RE, Hornbrook MC, Hooker RS, Woodson GT, Shneidman R. Analysis of the costs of NSAID-associated gastropathy: experience in a US health maintenance organisation. Pharmacoeconomics. 1997;12(1):76-88.
6. Schlansky B, Hwang JH. Prevention of nonsteroidal anti-inflammatory drug-induced gastropathy. J Gastroenterol. 2009;44 Suppl 19:44-52.
7. Singh G, Rosen Ramey D. NSAID induced gastrointestinal complications: the ARAMIS perspective: 1997: Arthritis, Rheumatism, and Aging Medical Information System. J Rheumatol Suppl. 1998;51:8-16.
8. Tramer MR, Moore RA, Reynolds DJ, McQuay HJ. Quanti-

tative estimation of rare adverse events which follow a biological progression: a new model applied to chronic NSAID use. Pain. 2000;85(1-2):169-82.

9. White TJ, Arakelian A, Rho JP. Counting the costs of drug-related adverse events. Pharmacoeconomics. 1999;15(5):445-58.
10. Lazzaroni M, Bianchi Porro G. Gastrointestinal side-effects of traditional non-steroidal anti-inflammatory drugs and new formulations. Aliment Pharmacol Ther. 2004;20 Suppl 2:48-58.
11. Park MG, Yoo JD, Lee KH. Current guidelines for nonsteroidal anti-inflammatory drugs. J Korean Orthop Assoc. 2020;55(1):9-28.
12. Lee SH, Han CD, Yang IH, Ha CW. Prescription pattern of NSAIDs and the prevalence of NSAID-induced gastrointestinal risk factors of orthopaedic patients in clinical practice in Korea. J Korean Med Sci. 2011;26(4):561-7.
13. Lee EY, Hong SJ, Park YB, et al. Gastrointestinal risk factors and non-steroidal anti-inflammatory drugs use in rheumatoid arthritis and osteoarthritis patients in Korea. J Rheum Dis. 2016;23(1):47-54.
14. Park HR, Cho SK, Im SG, et al. Treatment patterns of knee osteoarthritis patients in Korea. Korean J Intern Med. 2019;34(5):1145-53.
15. Andrade SE, Kahler KH, Frech F, Chan KA. Methods for evaluation of medication adherence and persistence using automated databases. Pharmacoepidemiol Drug Saf. 2006;15(8):565-77.
16. Cieslak, DA, Chawla NV. Learning decision trees for unbalanced data. In: Proceedings of Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD; 2008 Sep 15-19; Antwerp, Belgium. Philadelphia: Springer Berlin Heidelberg; 2008. 241-56.
17. Cho E, Kim SK. Prescribing superfluous gastroprotective agents: an indicator of polypharmacy. Korean J Clin Pharm. 2011;21(2):156-60.
18. Lorena AC, Jacintho LF, Siqueira MF, et al. Comparing machine learning classifiers in potential distribution modelling. Expert Syst Appl. 2011;38(5):5268-75.
19. Mahesh B. Machine learning algorithms-a review. Int J Sci Res. 2020;9(1):381-6.
20. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-232.
21. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining; 2016. 785-94.
22. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J Comput Graph Stat. 2015;24(1):44-65.
23. Guglielmi A, Ruzzenente A, Sandri M, et al. Risk assessment and prediction of rebleeding in bleeding gastroduodenal ulcer. Endoscopy. 2002;34(10):778-86.
24. Petersen J, Moller Hansen J, de Muckadell OB, Dall M, Hallas J. A model to predict the risk of aspirin/non-steroidal anti-inflammatory drugs-related upper gastrointestinal bleeding for the individual patient. Basic Clin Pharmacol Toxicol. 2020;126(5):437-43.
25. Castellsague J, Riera-Guardia N, Calingaert B, et al. Individual NSAIDs and upper gastrointestinal complications: a systematic review and meta-analysis of observational studies (the SOS project). Drug Saf. 2012;35(12):1127-46.
26. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc. 2017;24(2):361-70.
27. Sakr S, Elshawi R, Ahmed A, et al. Using machine learning on cardiorespiratory fitness data for predicting hypertension: the Henry Ford ExercIse Testing (FIT) Project. PLoS One. 2018;13(4):e0195344.
28. Ye C, Fu T, Hao S, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res. 2018;20(1):e22.