Risk prediction of osteoporosis for postmenopausal women using machine learning methods

Sung Kean Kim

The Graduate School
Yonsei University
Graduate Program in Biomedical Engineering

Risk prediction of osteoporosis for postmenopausal women using machine learning methods

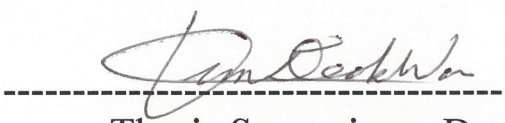
Directed by Professor Deok Won Kim

The Master's Thesis
submitted to the Graduate Program in Biomedical Engineering,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Master of Biomedical Engineering

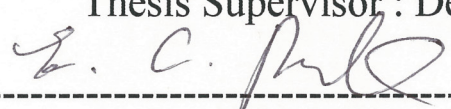
Sung Kean Kim

June 2014

This certifies that the Master's Thesis of
Sung Kean Kim is approved.



Thesis Supervisor : Deok Won Kim



Thesis Committee Member#1 : Eun-Cheol Park

Thesis Committee Member#2 : Sung-Bae Cho

The Graduate School
Yonsei University

June 2014

ACKNOWLEDGEMENTS

의학공학교실에 처음 발을 내딛은지도 벌써 2년이 넘어 이제 졸업을 앞두고 있습니다. 그 동안 도움을 주신 많은 분들께 감사의 마음을 전하고자 합니다.

먼저 그간 여러모로 많이 부족한 저를 항상 세심한 지도와 아낌없는 배려로 이끌어 주시고 학문의 방향을 제시해 주신 김덕원 교수님께 진심으로 감사의 말씀을 올립니다. 그리고 바쁘신 와중에도 본 논문을 심사해 주시고 좋은 조언을 해주신 박은철 교수님과 조성배 교수님께도 감사의 말씀을 드립니다. 또한, 수업을 통하여 많은 가르침을 주신 김남현 교수님, 유선국 교수님, 정혜정 박사님께도 감사의 마음을 전합니다. 의공학의 기초를 잡아 주신 학부 교수님들께도 감사 드립니다.

짧은 시간이었지만 항상 사람 좋은 미소로 좋은 이야기를 해주던 경환이 형과 재밌고 의리 있는 친구이자 연구실 선배 재림이에게도 감사의 마음을 전합니다. 그리고 통계를 잘 알려주고 즐거운 시간을 함께한 민경이, 성실하고 열심히 하는 모습으로 배울 점이 많았던 준열이, 입학 동기로서 의지할 수 있고 힘이 되었던 정모, 잘생기고 남자다운 완형이 형, 많은 것을 알려주며 연구에 흥미를 불어 넣어 줬던 태근이에게도 감사함을 전합니다. 또한, 종종 연구원들에게 맛있는 식량을 제공해주신 성우 형, 비상한 머리로 많은 도움을 주었던 수범이 형, 몸을 아끼지 않고 열심히 하는 재원이, 항상 신경 써주고 여러 방면에서 배울 점이 많았던 지수에게도 감사의 마음을 표합니다. 그리고 의학공학교실의 살림을 책임지시고 계신 정숙이 누나에게도 감사 드립니다.

오랜 친구이자 힘들 때 큰 힘이 되어주고 평생 우정을 함께 할 승열이, 동생이지만 친구 같은 이수, 듬직하고 믿음직스러운 웅이, 나의 정신적 지주 민재 형, 산전수전 다 겪고 인생을 말해 주시는 진형이 형, 마음이 따뜻한 성혜 누나, 언제 봐도 즐겁고 기분 좋은 성준이, 대건이, 우상이, 한준이, 동규, 동현이, 일경이 등 오랜 친구들에게도 감사함을 전합니다. 그리고 고등학교 수학 선생님이시자 내가 무척이나 좋아하는 이경일 선생님께도 감사의 마음을 표합니다.

저를 지금까지 잘 키워주시고 항상 응원해주시는 아버지, 어머니께 진심으로 감사 드리며, 앞으로는 더 멋진 아들이 되도록 노력하겠습니다. 사랑합니다. 마지막으로 어머니보다 더 어머니 같으신 외할머니께 항상

너무나도 감사하고 사랑한다는 말씀 드립니다.

2014년 6월
김성권 올림

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

ANN: Artificial neural networks
AUC: Area under the curve
BMD: Bone mineral density
CI: Confidence interval
DEXA: Dual X-ray absorptionmetry
FN: False negative
FP: False positive
KNHANES: Korea National Health and Nutrition Examination Surveys
LR: Logistic regression
ORAI: Osteoporosis risk assessment instrument
OSIRIS: Osteoporosis index of risk
OST: Osteoporosis self-assessment tool
RF: Random forests
ROC: Receiver operating characteristic
SCORE: Simple calculated osteoporosis risk estimation
SVM: Support vector machines
TN: True negative
TP: True positive
WHO: World Health Organization
yr: Year

<ABSTRACT>

Risk prediction of osteoporosis for postmenopausal women using machine learning methods

Sung Kean Kim

Graduate Program in Biomedical Engineering
The Graduate School, Yonsei University

(Directed by Professor Deok Won Kim)

Osteoporosis is a major public health concern worldwide. In particular, osteoporosis is common in postmenopausal women but is asymptomatic until a fracture occurs. A number of clinical decision tools for osteoporosis risk assessment have been developed to select postmenopausal women for the measurement of bone mineral density. We developed and validated machine learning models with the aim of more accurately identifying the risk of osteoporosis in postmenopausal women compared to the ability of conventional clinical decision tools.

We collected medical records from Korean postmenopausal women based on the Korea National Health and Nutrition Examination Surveys (KNHANES V-1) conducted in 2010. The training data set was used to construct models based on popular machine learning algorithms such as support vector machines (SVM), random forests, artificial neural networks (ANN), and logistic regression (LR) based on simple surveys. The machine learning models were optimized using 10-fold cross validation and were compared to four conventional clinical decision tools: osteoporosis self-assessment tool (OST), osteoporosis risk assessment instrument (ORAI), simple calculated osteoporosis risk estimation (SCORE), and osteoporosis index of risk (OSIRIS). The internal validation set was used to assess the performance of the models for predicting osteoporosis risk using accuracy and area under the curve (AUC) of the receiver operating characteristic. External validation was also performed using the KNHANES V-2 conducted in 2011 to evaluate the models.

SVM had significantly better AUC than ANN, LR, OST, ORAI, SCORE, and OSIRIS

for the training set. The validation on the internal validation set showed that SVM predicted osteoporosis risk with an AUC of 0.827, accuracy of 76.7%, sensitivity of 77.8%, and specificity of 76.0% at total hip, femoral neck, or lumbar spine. In the external validation, SVM predicted osteoporosis risk with an AUC of 0.833, accuracy of 76.5%, sensitivity of 70.3%, and specificity of 80.3% at any site. The significant factors selected by SVM were age, height, weight, body mass index, duration of menopause, duration of breast feeding, estrogen therapy, hyperlipidemia, hypertension, osteoarthritis, and diabetes mellitus.

Considering various predictors associated with low bone density, the machine learning methods may be effective tools for identifying postmenopausal women at high risk for osteoporosis.

Key words: osteoporosis, postmenopausal women, machine learning, risk assessment, clinical decision tools, support vector machines

Risk prediction of osteoporosis for postmenopausal women using machine learning methods

Sung Kean Kim

Graduate Program in Biomedical Engineering
The Graduate School, Yonsei University

(Directed by Professor Deok Won Kim)

I. INTRODUCTION

As the world population continues to age, the prevalence of osteoporosis and the incidence of osteoporosis-related fractures will sharply increase in most developing and developed countries[1]. Osteoporosis is a multi-factorial systemic skeletal disorder, characterized by low bone mass and micro-architectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture[2]. Fracture due to osteoporosis is one of the major factors of disability and death in elderly persons[3]. Osteoporosis is common in postmenopausal women but is asymptomatic until a fracture occurs. In women, bone loss begins predominantly after menopause, while bone loss occurs gradually with age in men[4]. The World Health Organization (WHO) estimates that 30% of all postmenopausal women have osteoporosis, which is defined as bone mineral density (BMD) 2.5 standard deviations below the young healthy adult mean (T-score$\leq$-2.5)[5]. Dual X-ray absorptionmetry (DEXA) of total hip, femoral neck, and lumbar spine is the most widely used tool for diagnosing osteoporosis. However, mass screening using DEXA is not widely recommended as it is a high-cost method of evaluating BMD[6]. Current research shows that too few DEXA scans are obtained among high-risk patients[7], while too many DEXA scans are obtained among low-risk

postmenopausal women[8].

Although the WHO provides FRAX® on their website, which was developed for fracture risk assessment, recent studies showed that FRAX® does not have a better sensitivity for fracture prediction than low BMD (T-score≤-2.5)[9]. Asian women were reported to have lower bone mass than Caucasians even after the differences between the body sizes of Caucasian and Asian women were taken into account[10]. Moreover, compared to Western countries, less is known about the characteristics of osteoporosis in Asian women[11]. Some reports and guidelines have proposed that women over the age of 65 years should be screened by DEXA[8]. However, the diagnosis rate has been reported to be lower than one-third among postmenopausal women in Korea[12]. The prevalence of osteoporosis is high in Korea compared to Western countries[13]. Moreover, Koreans are increasingly at high risk of osteoporosis due to a deficiency of vitamin D, nutritional imbalance, and lifestyle factors[14]. Therefore, an effective prescreening tool is necessary for Korean postmenopausal women to increase the possibility of early treatment.

The risk factors of osteoporosis are well-known and include older age, gender, history of fracture, low body weight, estrogen deficiency at an early age, low calcium intake, vitamin D deficiency, and lifestyle factors such as smoking or drinking to excess[15,16]. There has been a great deal of research assessing the combination of risk factors that would be of most help to physicians. A number of epidemiological studies have developed clinical decision tools for osteoporosis risk assessment to select postmenopausal women for the measurement of BMD. The purpose of these clinical decision tools is to help estimate the risk for osteoporosis, not to diagnose osteoporosis.

The osteoporosis self-assessment tool (OST) is a simple formula based on age and body weight[17]. Although OST uses only two factors to predict osteoporosis risk, it has been shown to have good sensitivity with an appropriate cutoff value[18]. The osteoporosis risk assessment instrument (ORAI), simple calculated osteoporosis risk estimation (SCORE), and osteoporosis index of risk (OSIRIS) are more complex decision tools using other risk factors[19]. They include not only age and body weight, but also estrogen therapy, history of fracture, and rheumatoid arthritis. Nevertheless, ORAI, SCORE, and

OSIRIS have not shown significantly better performance than OST in predicting osteoporosis risk[19]. These four risk assessment tools have been proposed for increasing awareness of osteoporosis and for encouraging more efficient use of BMD measurements in patients who are likely to have low bone mass, especially in asymptomatic postmenopausal women[18]. However, all of these decision tools have the limitation of low accuracy for clinical use[8]. In recent years, new additional risk factors of osteoporosis have been investigated based on individual conditions and risk profile for osteoporosis to enhance sensitivity and specificity[20].

Machine learning is an area of artificial intelligence research which uses statistical methods for data classification. Several machine learning techniques have been applied in clinical settings to predict disease and have shown higher accuracy for diagnosis than classical methods[21]. These mathematical algorithms have the ability to classify large amounts of data into a useful format[22]. The classifiers take the medical data of each patient and predict the presence of diseases based on underlying patterns. Support vector machines (SVM), random forests (RF), and artificial neural networks (ANN) have been widely used approaches in machine learning[21]. They are the most frequently used supervised learning methods for analyzing complex medical data. Logistic regression (LR) is another machine learning technique and is a widely used standard regression model for binary data[21]. SVM, RF, ANN and LR are the models of choice in many tasks in medicine and bioinformatics for selecting informative variables or genes and predicting diseases more accurately.

Several studies have shown that SVM, RF, and ANN could help predict low BMD using diet and lifestyle habit data[23-26]. Although these studies considered risk factors, they did not select informative variables that could contribute to osteoporosis. Moreover, previous studies had no objective comparisons of the performance of osteoporosis prediction developed by epidemiological data among the machine learning methods and clinical decision tools. Therefore, a structural design is needed for constructing the models along with a comparative study of various analytical methods for predicting osteoporosis risk.

In this study, we developed and validated machine learning models with the aim of

identifying the risk of osteoporosis in postmenopausal women. The objective of this study was to select patients who were candidates for DEXA in order to increase the effectiveness of screening for osteoporosis. We developed the prediction models for osteoporosis using various machine learning methods including SVM, RF, ANN, and LR. The performance of machine learning methods and conventional clinical decision making tools including OST, ORAI, SCORE, and OSIRIS was compared in respect to accuracy and area under the curve (AUC) of the receiver operating characteristic (ROC).

II. MATERIALS AND METHODS

1. Data source and subjects

We collected data from Korean postmenopausal women, based on the Korea National Health and Nutrition Examination Surveys (KNHANES V-1) conducted in 2010[27]. The KNHANES V-1 was a cross-sectional survey conducted by the Division of Chronic Disease Surveillance, Korea Centers for Disease Control and Prevention. The survey is divided into a health interview survey, a nutrition survey, and a health examination survey. Each data set contains BMD measurements at total hip, femoral neck, and lumbar spine as well as medical characteristics. BMD was measured by DEXA using Hologic Discovery (Hologic Inc., Bedford, MA, USA). Among 4843 women who participated in the KHANES V-1, BMD was measured in 3891 participants. We excluded women who were not in postmenopausal status and who were in postmenopausal status after bilateral oophorectomy. Finally, 1674 women who were determined to have physiologically postmenopausal status were included in this study (Figure 1). We categorized the postmenopausal women into a control group and an osteoporotic group with low BMD (T-score $\leq$-2.5) at any site among total hip, femoral neck, or lumbar spine measurements. There were several modifications for data analysis. If an answer for a question in the KNHANES V-1 was 'don't know,' we regarded it as missing data and estimated the answer using a nearest neighbor algorithm[28]. This algorithm found the most similar samples to the real values present to estimate the missing values.

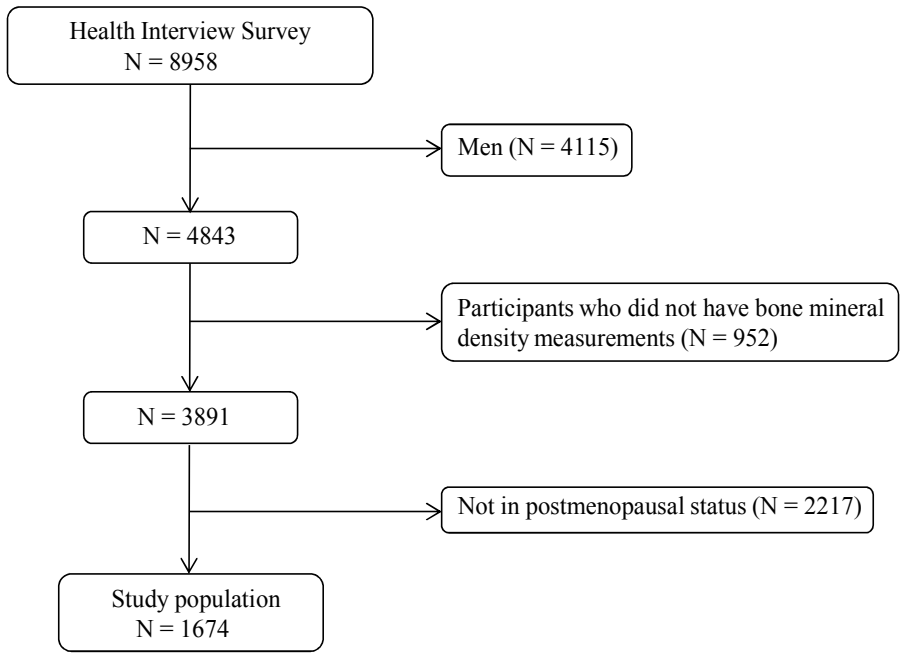


Figure 1. Flow diagram of inclusion or exclusion of study participants in the KNHANES V-1

2. Data analysis

The data were separated randomly into two independent data sets: training and internal validation sets. The training set, comprised of 60% (1000 participants) of the entire dataset, was used to construct models based on SVM, RF, ANN, and LR. The scores of the clinical decision tools for screening osteoporosis including OST, ORAI, SCORE, and OSIRIS were calculated according to each formula (Table 1). These four conventional clinical decision tools are the most widely used indices for predicting osteoporosis risk[18]. Because the KNHANES V-1 did not have specific information concerning fracture type but did indicate simple fracture histories at various sites, the fracture histories were used for the scoring of non-traumatic fracture in SCORE and the history of low impact fracture in OSIRIS. The prediction models were internally validated using 10-fold cross validation[29]. We designed the 10-fold cross validation not only to assess performance, but also to optimize prediction models using machine learning techniques. We used 10-fold cross validation on the training set, and the performance was measured on the internal validation set. The training set was partitioned into 10 equal-size subsets. One subset was used as a testing set for a model trained on the 9 remaining subsets. The cross validation process was repeated 10 times, allowing each subset to serve once as the testing set. The internal validation set, comprised of 40% (674 participants) of the entire dataset, was used to assess ability to predict osteoporosis in postmenopausal women. Validation using the internal validation set avoided potential bias of the performance due to over-fitting of the models to training set. Figure 2 shows the overview of trained machine learning models to predict osteoporosis.

Table 1. Calculation of clinical decision tools for screening osteoporosis

Factor	Score
OST	
Body weight (kg)	+0.2 × body weight
Age (years)	-0.2 × age
ORAI	
Age ≥75 years	+15
Age 65-74 years	+9
Age 55-64 years	+5
Body weight <60 kg	+9
Body weight 60-70 kg	+3
Estrogen therapy	+2 if not currently using estrogen
SCORE	
Race other than black	+5
Rheumatoid arthritis	+4
Non-traumatic fracture	+4 per fracture
Age	+3 for each decade
Estrogen therapy	+1 if never
Weight	-1 for each 10 lb (1 kg = 2.204623 lb)
OSIRIS	
Body weight (kg)	+0.2 × body weight
Age (years)	-0.2 × age
History of low impact fracture	-2
Estrogen therapy	+2

OST: osteoporosis self-assessment tool, ORAI: osteoporosis risk assessment instrument, SCORE: simple calculated osteoporosis risk estimation, OSIRIS: osteoporosis index of risk

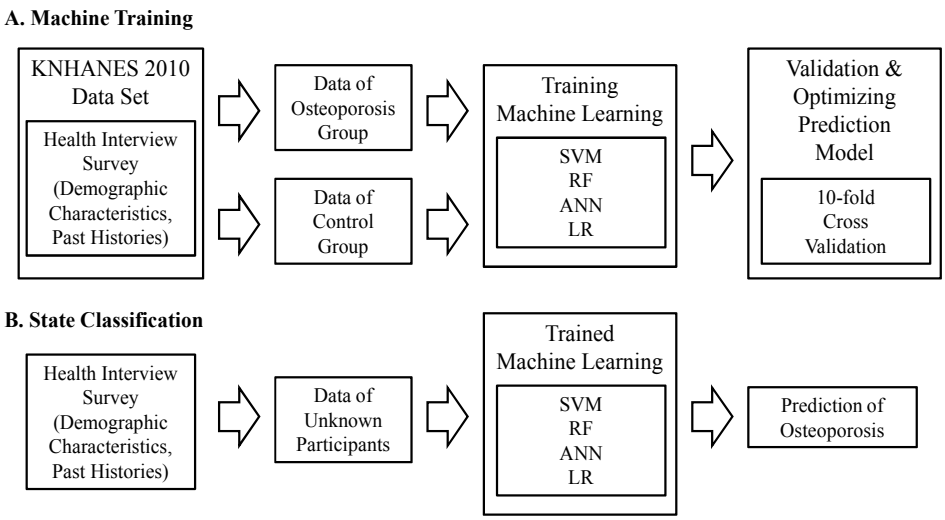


Figure 2. Overview of models to predict osteoporosis. The flow of training machine learning models (A) and the flow of state classification with unknown data (B).

3. Machine learning techniques

SVM is arguably the single most important development in supervised classification of recent years. SVM is fairly insensitive to the curse of dimensionality and is efficient enough to handle very large-scale classification in both samples and variables[30]. The main idea of SVM is to map data to a higher dimensional space via a kernel function, and then choose the maximum-margin hyper-plane that separates training data. The hyper-plane is based on a set of boundary training data, called support vectors[31]. New instances are classified according to the side of the hyper-plane they fall into (Figure 3). Thus, the goal of the SVM is to improve accuracy by the optimization of space separation using a kernel function.

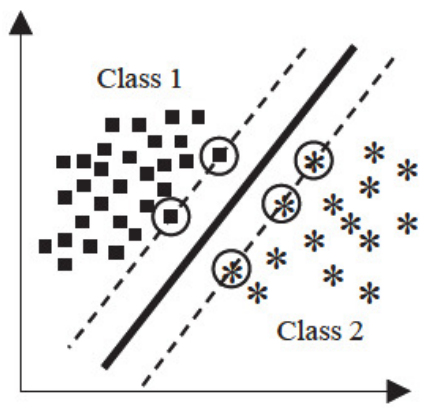


Figure 3. Algorithm of SVM model. SVM selects a hyper-plane (bold line) that maximizes the width of the gap (margin) between the two classes. The hyper-plane is specified by boundary training instances, called support vectors shown with circles.

RF is a classification algorithm that uses an ensemble of unpruned decision trees (Figure 4), each of which is built on a bootstrap sample of the training data using a randomly selected subset of predictors[32]. The final result is based on a majority vote of all trees[21]. There is little need to fine-tune parameters to achieve excellent performance for RF, the number of trees for each forest and the number of predictors at each node[33]. RF

can deal with high dimensional data in training faster than other methods.

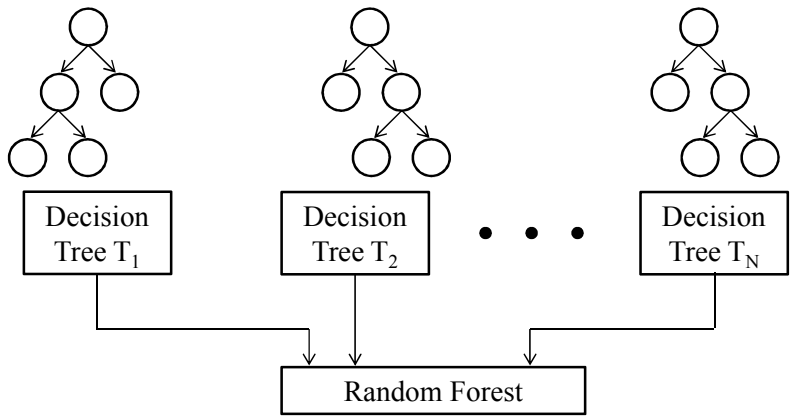


Figure 4. Algorithm of RF model

ANN, comprised of several layers and connections, is a computational model that emulates biological neural networks[34]. ANN is considered powerful in modeling non-linear relationship and has been applied increasingly in medical diagnosis and prediction[35]. The most common type of neural network is a multi-layer perceptron with feed forward, which consists an input layer, several hidden layers, and one output layer as shown in Figure 5. Each layer consists of a number of elementary processing units called neurons. Neurons between layers are heavily interconnected with different weights. The weights were adjusted by training algorithms. After repetitive cycle of training, similar to human's learning form experience, the weights were adjusted to achieve the best prediction ability[21]. ANN with too few hidden neurons would be incapable of differentiating between complex patterns, leading only to a linear estimate of the actual trend. In contrast, if an ANN model had too many hidden neurons, it would follow the noise in the data due to over-fitting, leading to poor generalization for untrained data[36]. Although the optimal number of hidden neurons was determined by empirical method, it has been recently determined depending on the number of input and output variables[37].

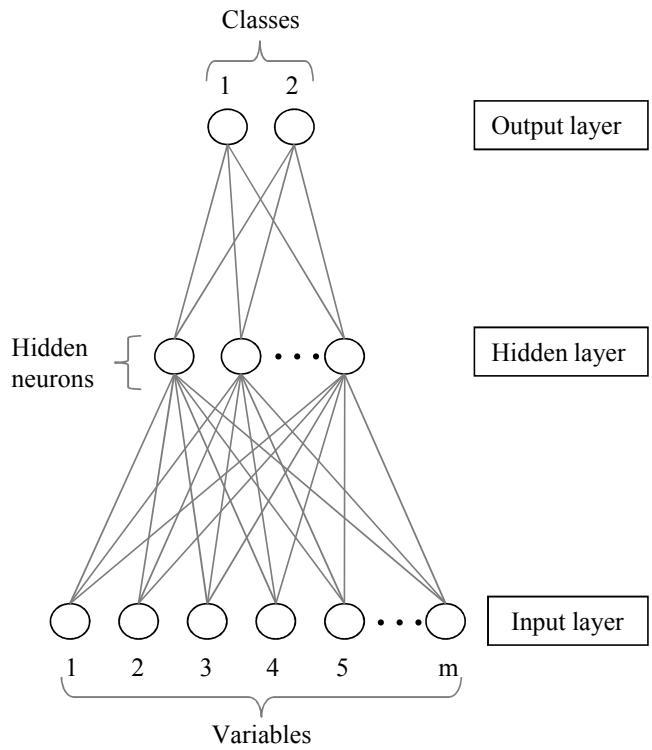


Figure 5. Algorithm of ANN model

LR is the gold standard method for analyzing binary medical data because it provides not only a predictive result, but also yields additional information such as a diagnostic odds ratio by fitting data to a logistic function (1), which always takes on values between zero and one[38].

$$P(x) = \frac{1}{1+e^{-(\alpha+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\ldots+\beta_d x_d)}} \qquad (1)$$

In equation (1), $\beta_1 \sim \beta_d$ are called the regression coefficients of input variables, respectively. Each of the regression coefficients describes the size of the contribution of that risk factor. LR is usually used as a base for comparison in machine learning studies[34].

4. Model selection and statistical analysis

We used the 10-fold cross validation scheme to construct machine learning models. The purpose of the machine learning models was to predict osteoporosis risk using the health interview survey concerning demographic characteristics and past histories listed in Table 2. Due to high dimensionality, variable selection was a necessary technique to make an effective prediction model and to improve prediction performance[39]. We also obtained an insight into factors related to osteoporosis through the variables that were entered into the classifiers. Eighty-one variables in the data of the characteristics including alcohol, smoking, stress status, and physical activity were initially selected to design the model to predict osteoporosis risk. We adopted a feature selection method of wrapper-based feature subset evaluation for SVM, RF, and ANN[21,40]. We determined the order of the variables with the embedded methods of each machine learning method and decreased the number of variables to determine the best subset using backward elimination[39]. The remaining features that indicated the highest accuracy in 10-fold cross validation were the selected subset for prediction. For LR, we used the backward stepwise method for variable selection.

Data sets in this study were class-imbalanced because the control group contained significantly more samples than the osteoporotic group. Applying a classifier to the imbalanced data could produce undesirable lower performance[41]. Therefore, it was important to improve prediction models for the imbalanced data. To obtain the optimal result, we adopted a grid search in which a range of parameter values were tested using the 10-fold cross validation strategy. For SVM, we chose a Gaussian kernel function among the available kernel functions based on the recommendation of a practical guide[42]. Two key parameters for the Gaussian kernel function, a penalty parameter C and scaling factor σ, need to be pre-selected to generate an optimal SVM model. The parameter C controls over-fitting of the model by specifying tolerance for misclassification and the σ controls the degree of nonlinearity of the model[43]. The parameter values included the number of trees and predictors for RF, and the number of hidden nodes and learning rate for ANN. We found the best classification model and employed its parameters for

prediction. Due to the imbalanced data problem in this study, prediction accuracy might not be a good criterion for assessing performance since the minor class has less influence on accuracy than the major class[44]. Therefore, we evaluated diagnostic abilities including not only accuracy, sensitivity, and specificity, but also AUC.

$$Sensitivity = TP / (TP + FN) \quad (1)$$

$$Specificity = TN / (TN + FP) \quad (2)$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

True positive (TP): Number of osteoporotic women correctly identified as osteoporosis
True negative (TN): Number of healthy women correctly identified as normal
False positive (FP): Number of healthy women incorrectly identified as osteoporosis
False negative (FN): Number of osteoporotic women incorrectly identified as normal

The AUC is known as a strong predictor of performance, especially with regard to imbalanced problems because it is the average sensitivity over all possible specificities[41]. The AUC may range from 0 to 1, with an area of 1.0 representing perfect discrimination and an area of 0.5 representing what is expected by chance alone[45]. To compare the performance of models, we generated the ROC curves which provide a graphical representation of the overall accuracy of a test by plotting sensitivity against (1-specificity) for all thresholds. Then we selected cut-off points as the points on the ROC curve closest to the upper left corner. This method maximized the Youden's index, giving equal weight to sensitivity and specificity[46]. ROC curve analysis is the most commonly used method in clinical analysis for establishing the optimal cut-off point. The cut-offs of the OST, ORAI, SCORE, and OSIRIS were calculated using ROC curve analysis. To discriminate osteoporosis, the following cut-offs were used: <-1 for OST, >16 for ORAI, >15 for SCORE, and <-1 for OSIRIS, respectively. We used MATLAB 2010a (Mathworks Inc., Natick, MA, USA) for the analysis of machine learning, SPSS 18.0 (SPSS Inc., Chicago, IL, USA) for LR and statistical analysis, and MedCalc 12.3 (MedCalc Inc., Mariakerke, Belgium) for ROC analysis.

5. External validation

Performance of the prediction models was evaluated in independent data, the KNHANES V-2 conducted in 2011[47]. For selecting eligible participants, the process of inclusion or exclusion in the KNHANES V-1 was applied in the data. A total of 485 of 8518 participants having postmenopausal status were included for external validation set.

III. RESULTS

1. Osteoporotic and control groups

Of 1674 postmenopausal women from the KNHANES V-1, 583 (34.8%) had combined osteoporosis at any site including total hip, femoral neck, or lumbar spine. Among 583 women with osteoporosis, 95 had osteoporosis at the total hip, 331 at the femoral neck, and 473 at the lumbar spine. Table 2 shows the weighted characteristics of postmenopausal women in the KNHANES V-1 categorized by the presence of osteoporosis to represent the entire postmenopausal women in Korea. By comparison with women in the control group, women with osteoporosis were of higher age and lower height, weight, body mass index, and waist circumference. Women with osteoporosis were also less likely to take estrogen therapy. The number of pregnancies, duration of menopause, and duration of breast feeding were higher in the osteoporotic group. Women in this group were more likely to have history of fracture, hypertension, and rheumatoid arthritis and less likely to have hyperlipidemia. In the KNHANES V-2, 185 (38.1%) of 485 postmenopausal women had osteoporosis at any site among total hip, femoral neck, or lumbar spine. Table 3 shows the characteristics of postmenopausal women between the control and osteoporotic groups in the KNHANES V-2. Women with osteoporosis were of higher age and lower height, weight, body mass index, and waist circumference. Women with osteoporosis were also less likely to take estrogen therapy. Duration of menopause, duration of breast feeding, and the number of pregnancies were higher in the osteoporotic group. Women in this group were more likely to have history of fracture, osteoarthritis, and hypertension.

Table 2. Characteristics of postmenopausal women in the KNHANES V-1

Variable*	**Total (n = 1674)**	**Without osteoporosis (n = 1091)**	**With osteoporosis at any site (n = 583)**	***p value*†**
Age (yrs)	62.2±0.4	59.0±0.4	69.1±0.6	< 0.001
Height (cm)	153.6±0.2	155.0±0.2	150.7±0.3	< 0.001
Weight (kg)	57.4±0.3	59.1±0.3	53.5±0.4	< 0.001
Body mass index (kg/m^2)	24.3±0.1	24.6±0.1	23.5±0.2	< 0.001
Waist circumference (cm)	82.5±0.4	83.0±0.4	81.4±0.5	< 0.001
Pregnancy	5.0±0.1	4.8±0.1	5.5±0.1	< 0.001
Duration of menopause (yrs)	13.4±0.4	10.1±0.4	20.7±0.7	< 0.001
Duration of breast feeding (months)	53.5±1.9	43.3±2.0	75.8±3.2	< 0.001
Estrogen therapy	14.9 (1.1)	18.3 (1.4)	7.3 (1.4)	< 0.001
History of fracture	14.6 (1.2)	10.6 (1.3)	23.4 (2.1)	< 0.001
Hyperlipidemia	18.7 (1.2)	21.0 (1.5)	13.6 (1.9)	0.006
Hypertension	41.5 (1.7)	39.2 (2.2)	46.7 (2.7)	0.037
Rheumatoid arthritis	4.4 (0.7)	3.6 (0.7)	6.2 (1.3)	0.047
Osteoarthritis	32.3 (1.6)	30.5 (1.7)	36.2 (2.8)	0.063
Diabetes mellitus	11.7 (1.1)	11.7 (1.3)	11.5 (1.9)	0.909

*Table values are given as mean ± standard error or % (standard error) unless otherwise indicated.

†*p*-values were obtained by t-test and chi-square test.

Table 3. Characteristics of postmenopausal women in the KNHANES V-2

Variable*	**Total (n = 485)**	**Without osteoporosis (n = 300)**	**With osteoporosis at any site (n = 185)**	***p value*†**
Age (yrs)	63.1±9.2	59.4±7.8	69.0±8.2	< 0.001
Height (cm)	154.6±5.7	156.2±5.3	151.9±5.3	< 0.001
Weight (kg)	57.8±8.5	60.2±8.1	54.1±7.8	< 0.001
Body mass index (kg/m^2)	24.2±3.2	24.7±3.2	23.4±3.0	< 0.001
Duration of menopause (yrs)	14.3±10.5	10.5±8.6	20.4±10.4	< 0.001
Duration of breast feeding (months)	46.6±38.9	37.9±30.0	60.6±46.9	< 0.001
History of fracture	68 (14.0)	30 (10.0)	38 (20.5)	0.001
Estrogen therapy	76 (15.7)	58 (19.3)	18 (9.7)	0.005
Osteoarthritis	151 (31.1)	80 (26.7)	71 (38.4)	0.007
Hypertension	194 (40.0)	107 (35.7)	87 (47.0)	0.013
Waist circumference (cm)	82.6±9.3	83.4±9.4	81.3±9.0	0.015
Pregnancy	4.9±2.4	4.7±2.3	5.2±2.4	0.027
Hyperlipidemia	110 (22.7)	75 (25.0)	35 (18.9)	0.120
Rheumatoid arthritis	20 (4.1)	16 (4.7)	6 (3.2)	0.444
Diabetes mellitus	62 (12.8)	36 (12.0)	26 (14.1)	0.511

*Table values are given as mean ± standard deviation or number (%) unless otherwise indicated.

†*p*-values were obtained by t-test and chi-square test.

2. Selected variables and parameters

Table 4 describes the final multivariate LR derived from the training set using backward selection. Variables selected by LR were age, weight, duration of menopause, diabetes mellitus, hyperlipidemia, and osteoarthritis.

Table 4. Odds ratios for predicting osteoporosis risk using the multivariate logistic regression with backward selection models in the training set[48]

Variables	β-Coefficient	Odds ratio [95% CI]	*p value*
Age (per 1 year increase)	0.09	1.09 [1.06-1.12]	<0.001
Weight (per 1 kg increase)	-0.07	0.92 [0.91-0.94]	<0.001
Duration of menopause (per 1 year increase)	0.03	1.03 [1.01-1.05]	0.002
Diabetes mellitus	-0.53	0.58 [0.39-0.86]	0.007
Hyperlipidemia	-0.47	0.62 [0.43-0.89]	0.011
Osteoarthritis	0.46	1.59 [0.92-2.73]	0.094

CI: confidence interval

Table 5 summarizes the results of variable selection for the various machine learning and conventional methods in the training set. While the conventional methods selected two to five variables to obtain simplicity, the machine learning methods except LR selected more than 10 variables for better performance. The predictors of osteoporosis selected by SVM included age, height, weight, body mass index, duration of menopause, duration of breast feeding, estrogen therapy, hyerlipidemia, hypertension, osteoarthritis, and diabetes mellitus. RF and ANN showed similar results.

Table 5. Variable selection in machine learning and conventional methods for osteoporosis risk of total hip, femoral neck, or lumbar spine in the training set[48]

Variable	Machine learning method				Conventional methods			
	SVM	RF	ANN	LR	OST	ORAI	SCORE	OSIRIS
Age	O	O	O	O	O	O	O	O
Height	O	O	O					
Weight	O	O	O	O	O	O	O	O
Body mass index	O	O	O					
Waist circumstance		O						
Pregnancy		O	O					
Duration of menopause	O	O	O	O				
Duration of breast feeding	O	O	O					
Estrogen therapy	O					O	O	O
History of fracture			O				O	O
Hyperlipidemia	O	O		O				
Hypertension	O	O						
Rheumatoid arthritis							O	
Osteoarthritis	O	O	O	O				
Diabetes mellitus	O	O	O	O				

SVM: support vector machines, RF: random forests, ANN: artificial neural networks, LR: logistic regression, OST: osteoporosis self-assessment tool, ORAI: osteoporosis risk assessment instrument, SCORE: simple calculated osteoporosis risk estimation, OSIRIS: osteoporosis index of risk

The optimal model of SVM was found using a Gaussian kernel function with a penalty parameter C of 100 and scaling factor σ of 30. In RF, the optimal number of trees was 100, and the number of predictors for each node was 3. The optimal ANN was set with 3 nodes of a hidden layer and learning rate of 0.1.

3. Performance of 10-fold cross validation

Figure 6 shows the prediction performance of 10-fold cross validation of the machine learning and conventional methods. The mean and standard deviation of AUCs were calculated from 10 validation results. We obtained the AUCs of SVM, RF, ANN and LR of 0.822, 0.808, 0.794, and 0.793, respectively. The predictors of the machine learning methods came from selected variables shown in Table 5. The AUCs of OST, ORAI, SCORE, and OSIRIS were 0.794, 0.791, 0.766, and 0.787, respectively. In Figure 6, we found that more complex discriminating functions such as SVM and RF showed better performance than simple linear functions such as LR, OST, ORAI, SCORE, and OSIRIS. For the AUCs, the SVM performed better than ANN ($p=0.028$), LR ($p=0.037$), OST ($p=0.037$), ORAI ($p=0.022$), SCORE ($p=0.005$), and OSIRIS ($p=0.009$) in 10-fold cross validation using a Wilcoxon signed rank test.

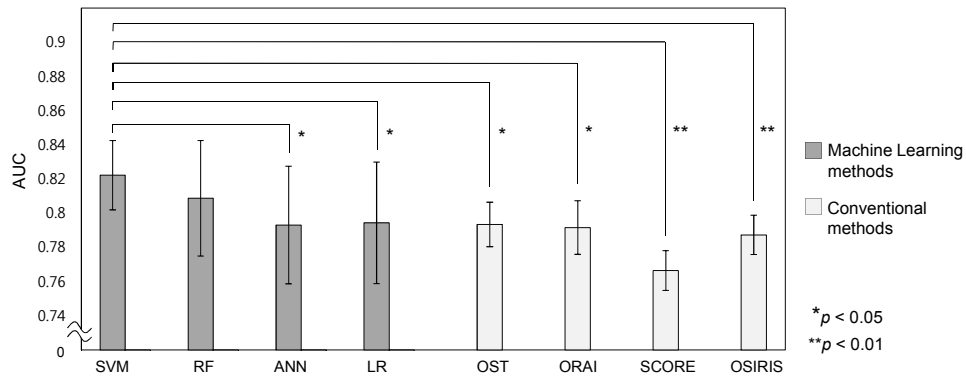


Figure 6. Performance results (AUC) of the machine learning and conventional methods using 10-fold cross validation.

Error bars indicate the standard deviation of the mean[48]. AUC: area under the curve, SVM: support vector machines, RF: random forests, ANN: artificial neural networks, LR: logistic regression, OST: osteoporosis self-assessment tool, ORAI: osteoporosis risk assessment instrument, SCORE: simple calculated osteoporosis risk estimation, OSIRIS: osteoporosis index of risk

4. Performance of internal validation set

To assess the ability of the models for predicting osteoporosis, we applied our methods to an internal validation set composed of the independent data. Table 6 shows the results of classifying the internal validation set for selecting women at risk of osteoporosis at various BMD measurement sites. As a result, the SVM model was the best discriminator between controls and women with osteoporosis. Considering osteoporosis at any site, SVM predicted osteoporosis risk with an AUC of 0.827, accuracy of 76.7%, sensitivity of 77.8%, and specificity of 76.0%. SVM also predicted osteoporosis at total hip, femoral neck and lumbar spine with the highest AUC and accuracy. SVM predicted osteoporosis at the total hip with an AUC of 0.921. On the other hand, osteoporosis at the lumbar spine was difficult to predict, with an AUC of 0.778. The ROC analysis demonstrated the SVM had the statistically significant better accuracy than OST at total hip, femoral neck and lumbar spine. Figure 7 shows the ROC curves of SVM, LR and OST in predicting osteoporosis at any site. Because SVM and OST had the highest AUC among the machine learning methods and conventional methods, respectively, we compared their ROC curves. LR was also included for comparison with SVM and OST. The AUCs of SVM, LR and OST were 0.827, 0.809 and 0.806, respectively (Table 6).

Table 6. Diagnostic performance of osteoporosis risk assessment methods for the internal validation set[48]

	Any site (total hip, femoral neck, or lumbar spine)					
	AUC [95% CI]	Accuracy (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	PPV (%)	NPV (%)
SVM	0.827 [0.788-0.854]	76.7 [74.5-78.6]	77.8 [72.6-83.0]	76.0 [72.0-80.1]	64.8	85.8
RF	0.824 [0.787-0.856]	76.5 [74.4-78.5]	76.6 [71.3-81.9]	76.5 [72.5-80.5]	64.9	85.2
ANN	0.807 [0.775-0.842]	75.2 [73.1-77.2]	76.6 [71.3-81.9]	74.4 [70.2-78.5]	62.9	84.8
LR	0.809 [0.776-0.846]	74.5 [72.3-76.6]	77.8 [72.6-83.0]	72.7 [68.5-76.9]	61.8	85.2
OST	0.806 [0.768-0.839]	74.0 [71.8-76.1]	75.4 [70.0-80.1]	73.2 [69.0-77.4]	61.5	84.0
ORAI	0.782 [0.741-0.815]*	74.7 [72.5-76.7]	65.1 [59.1-71.4]	79.7 [75.9-83.5]	64.6	80.1
SCORE	0.781 [0.742-0.811]*	68.2 [65.9-70.4]*	76.6 [71.3-81.9]	63.7 [59.1-68.2]	54.5	82.7
OSIRIS	0.798 [0.798-0.831]*	69.7 [67.4-71.8]*	82.7 [78.0-87.5]	62.7 [58.2-67.3]	55.8	86.5
	Total hip					
	AUC [95% CI]	Accuracy (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	PPV (%)	NPV (%)
SVM	0.921 [0.863-0.948]	90.5 [88.9-91.8]	84.6 [73.2-95.9]	90.8 [88.6-93.1]	36.2	98.9
RF	0.910 [0.851-0.941]	88.1 [86.4-89.6]*	79.5 [66.8-92.1]	88.6 [86.1-91.1]	30.0	98.5
ANN	0.902 [0.856-0.939]	87.1 [85.4-88.6]*	82.0 [70.0-94.0]	87.4 [84.8-89.9]	28.5	98.7
LR	0.911 [0.866-0.943]	88.2 [86.6-89.7]*	84.6 [73.2-95.9]	88.5 [86.0-90.9]	31.1	98.9
OST	0.903 [0.834-0.939]	81.7 [79.7-83.5]*	89.7 [80.2-99.2]	81.2 [78.2-84.2]	22.7	99.2
ORAI	0.875 [0.801-0.914]*	73.0 [70.7-75.0]*	89.7 [80.2-99.2]	71.9 [68.4-75.4]	16.4	99.1
SCORE	0.885 [0.815-0.932]*	82.2 [80.2-83.9]*	79.4 [66.8-92.1]	82.3 [79.3-85.3]	21.6	98.4
OSIRIS	0.891 [0.821-0.930]*	76.4 [74.2-78.3]*	89.7 [80.2-99.2]	75.5 [72.2-78.9]	18.4	99.1
	Femoral neck					
	AUC [95% CI]	Accuracy (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	PPV (%)	NPV (%)
SVM	0.874 [0.833-0.901]	80.4 [78.7-82.5]	81.3 [74.6-88.1]	80.5 [77.2-83.8]	49.7	94.8
RF	0.867 [0.829-0.899]	77.6 [75.4-79.5]*	81.3 [74.6-88.1]	76.6 [73.1-80.2]	45.2	94.5
ANN	0.845 [0.807-0.879]	77.1 [75.0-79.1]*	82.1 [75.5-88.7]	75.9 [72.3-79.5]	44.7	94.7
LR	0.855 [0.812-0.891]	76.6 [74.4-78.5]*	82.9 [76.4-89.4]	75.0 [71.4-78.6]	44.0	94.8
OST	0.855 [0.817-0.887]	77.6 [75.4-79.5]*	79.8 [72.9-86.7]	77.0 [73.5-80.5]	45.1	94.1
ORAI	0.828 [0.789-0.866]*	75.8 [73.7-77.9]*	82.1 [75.5-88.7]	74.3 [70.6-77.9]	43.0	94.6
SCORE	0.825 [0.782-0.860]*	72.8 [70.6-74.9]*	79.0 [72.0-86.0]	71.3 [67.5-75.1]	39.5	93.5
OSIRIS	0.845 [0.807-0.884]	79.0 [76.9-80.9]	72.0 [64.3-79.8]	80.7 [77.4-84.0]	46.9	92.4
	Lumbar spine					
	AUC [95% CI]	Accuracy (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	PPV (%)	NPV (%)
SVM	0.778 [0.727-0.805]	73.1 [70.9-75.2]	70.9 [64.5-77.3]	74.0 [70.0-77.9]	52.2	86.4
RF	0.770 [0.726-0.803]	72.7 [70.5-74.9]	66.8 [60.1-73.4]	75.1 [71.1-78.9]	51.8	84.9
ANN	0.761 [0.724-0.801]	70.4 [68.2-72.6]	72.5 [66.2-78.8]	69.6 [65.5-73.7]	48.9	86.3
LR	0.752 [0.705-0.788]	70.4 [68.2-72.6]	72.5 [66.2-78.8]	69.6 [65.5-73.7]	48.9	86.3
OST	0.758 [0.710-0.795]	67.9 [65.6-70.1]*	71.5 [65.1-77.8]	66.5 [62.3-70.7]	46.1	85.3
ORAI	0.726 [0.673-0.758]*	72.1 [69.9-74.2]	59.5 [52.6-66.0]	77.1 [73.3-80.8]	51.1	82.6
SCORE	0.729 [0.687-0.772]*	62.9 [60.4-65.1]*	74.0 [67.9-80.2]	58.4 [54.0-62.8]	41.6	84.8
OSIRIS	0.751 [0.703-0.787]	63.4 [60.8-65.5]*	79.7 [74.1-85.4]	56.7 [52.3-61.1]	42.5	87.5

*AUC or accuracy is significantly different from the SVM at the level of $p<0.05$.

AUC: area under the curve, CI: confidence interval, PPV: positive predictive value, NPV: negative predictive value, SVM: support vector machines, RF: random forests, ANN: artificial neural networks, LR: logistic regression, OST: osteoporosis self-assessment tool,

ORAI: osteoporosis risk assessment instrument, SCORE: simple calculated osteoporosis risk estimation, OSIRIS: osteoporosis index of risk

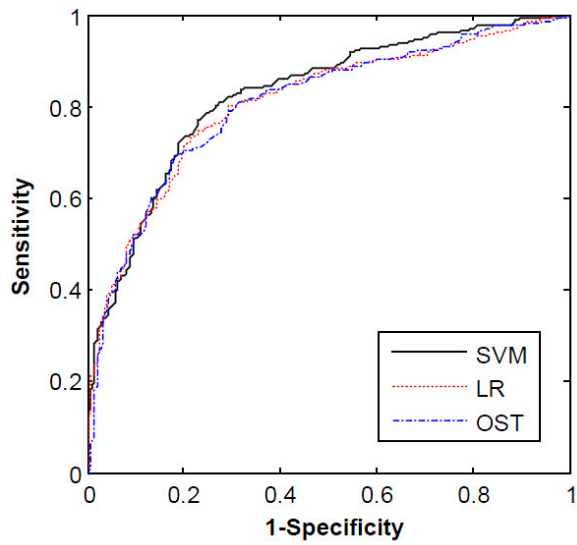


Figure 7. Receiver operating characteristic curves (ROC) of support vector machines (SVM), logistic regression (LR), and osteoporosis self-assessment tool (OST) in predicting osteoporosis risk at any site among total hip, femoral neck, or lumbar spine for the internal validation set[48]

5. Performance of external validation set

Additionally, we applied our methods to another independent data, the external validation set, for evaluation. Table 7 shows the results of classifying the external validation set for selecting women at risk of osteoporosis at any site. We evaluated the ability of SVM and OST since SVM and OST had the highest AUC among the machine learning methods and conventional methods, respectively, in 10-fold cross validation and internal validation. LR was also evaluated for comparison with SVM and OST. As a result, the SVM model showed the best performance in predicting osteoporosis risk. SVM predicted osteoporosis risk with an AUC of 0.833, accuracy of 76.5%, sensitivity of 70.3%, and specificity of 80.3%.

Table 7. Diagnostic performance of osteoporosis risk assessment methods for the external validation set

	Any site (total hip, femoral neck, or lumbar spine)					
	AUC [95% CI]	Accuracy (%) [95% CI]	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	PPV (%)	NPV (%)
SVM	0.833 [0.797-0.869]	76.5 [72.4-80.2]	70.3 [63.1-76.8]	80.3 [75.4-84.7]	68.8	81.4
LR	0.819 [0.781-0.857]	73.8 [69.6-77.6]	79.5 [72.9-85.0]	71.0 [64.8-75.5]	62.3	84.7
OST	0.822 [0.784-0.861]	74.6 [70.5-78.4]	79.5 [72.9-85.0]	71.7 [66.2-76.7]	63.4	85.0

AUC: area under the curve, CI: confidence interval, PPV: positive predictive value, NPV: negative predictive value, SVM: support vector machines, LR: logistic regression, OST: osteoporosis self-assessment tool

IV. DISCUSSION

Based on the various machine learning techniques, we investigated a new approach for predicting osteoporosis risk in postmenopausal women using data from the KNHANES V-1. To our best knowledge, this is the first report on application of conventional decision tools for BMD assessment in a Korean population. Among the machine learning and conventional methods, our SVM model discriminated more accurately between women with osteoporosis and control women. The 10-fold cross validation and ROC analysis indicated that the SVM had a statistically significant improvement in predicting osteoporosis. In other words, SVM was more effective in analyzing the epidemiological underlying patterns of osteoporosis compared with the other methods. This finding is consistent with a previous study on the comparison of machine learning methods in various complex discriminating problems for predicting disease[49]. RF has been known to have excellent performance in classification tasks, comparable to SVM[33]. In addition, RF has appealing theoretical and practical characteristics[50]. In this study, RF showed the second best performance among the prediction models, demonstrating it was comparable to SVM. Most experts have used conventional methods, including OST, ORAI, SCORE, and OSIRIS, due to their simplicity[51]. In this study, we applied complicated mathematical methods for more accurate prediction. Despite its complexity, our method is useful because computerized diagnostic decision supports have been increasingly easy to access due to the advancement of information systems for many medical problems[52].

If our prediction model retains good performance after validation in a larger population, it will be possible to use this technique as a cost-effective prescreening tool to determine candidates for evaluation with DEXA and also to prevent osteoporotic fracture in postmenopausal women at high risk. The patients in the high risk group categorized by this method should receive DEXA screening at the hospital. However, patients in the low risk group could postpone receiving a DEXA scan. Women experience menopause at 50 years old on average[53]. Accordingly, when we regard the Korean women who are over 50 years old as potential menopausal population, menopausal women account for 31.8% of all women in Korea. The 31.8% corresponds to approximately 8.5 million according to

the Korean National Statistical Office 2010. Although our SVM showed small improvement of 2.7% and 1.9% regarding accuracy compared to OST in the internal validation and external validation sets, respectively, the 2.7% and 1.9% correspond to approximately 230,000 and 162,000 women, which are not small population.

Our proposed SVM model included age, height, weight, body mass index, duration of menopause, duration of breast feeding, estrogen therapy, hyperlipidemia, hypertension, osteoarthritis, and diabetes mellitus as predictors (Table 5). Similar to earlier studies concerning prediction for osteoporosis, our results suggest that age is most closely associated with the development of osteoporosis, and weight is also an important factor. However, our findings also demonstrated different factors involved in osteoporosis such as height, duration of menopause, duration of breast feeding, and presence of chronic diseases such as hyperlipidemia, hypertension, osteoarthritis, and diabetes mellitus. Although many clinical studies have reported that height correlates with vertebral bone fractures, there has been no study using height as a predictor for osteoporosis at the total hip or femoral neck[54]. In our study, height showed a significant association with osteoporosis at any site and could enhance a decision tool for osteoporosis risk assessment. Estrogen therapy, duration of menopause, and duration of breast feeding were selected due to their association with exposure to estrogen in the endocrine system. Traditionally, estrogen, playing an important role in bone homeostasis, is thought to be the most important factor in bone growth and maturation in women[55,56]. Breast feeding might cause low BMD due to the removal of estrogen through breast milk and reduction of the cumulative exposure to estrogen[57].

Our SVM model also used hyperlipidemia, hypertension, osteoarthritis, and diabetes mellitus for more accurate prediction. Hyperlipidemia and diabetes mellitus were also selected in LR with significant odds ratios in Table 4, although they would be counterintuitive to many clinicians. There have been several studies indicating the influence of chronic diseases on the BMD of patients[20,58,59]. However, these results also showed that the effect of chronic diseases on BMD was minor. Our prediction model was able to consider these chronic diseases in combination using a SVM characterized by nonlinearity and high dimension. Because the SVM model delicately handled a separating

space composed of these factors in high dimension, it was possible to consider all factors for the improvement of sensitivity and specificity in predicting osteoporosis.

We found in the present study that the optimal SVM adopted a penalty parameter C of 100 and scaling factor σ of 30. Our SVM had the similar parameters to the several previous studies[49,60]. However, most studies obtained the optimal parameters empirically, and did not show the specific optimal parameters[21,23,43]. Since there is no clear guideline on selecting the most effective parameters for certain classification problems[61], our results may be a guide to future epidemiologic researches for predicting osteoporosis.

There are several limitations to this study. First, the study was based on a cross-sectional survey which cannot establish a causal relationship and has several defects according to a medical view. In addition, several pieces of information were collected from the self-reported questionnaires, so reporting bias cannot be excluded. For example, the prevalence of disease was based on a health interview survey taken on one occasion. Weight, height, body mass index, and hormone therapy status, as well as BMD, could differ according to time of the measurement. Second, it was difficult to consider drug effects. For example, treatment with steroids for rheumatoid arthritis, asthma, dermatitis, or autoimmune diseases is known to cause glucocorticoid-induced osteoporosis[62]. Systematic approaches are warranted to consider the long-term effects of drugs in further studies. Third, our prediction models were developed for Korean women. Several studies have indicated the possibility that Korean people have physiologic differences in bone metabolism[63,64]. To validate our findings, further study is needed including large heterogeneous samples. Fourth, our study was characterized by imbalanced class distribution. Traditional classifiers have generally shown poor performance on imbalanced data sets because they are designed to find the best classification for the majority[44]. The imbalanced class was still critical even though we adopted a dense grid search to decide the optimal prediction models in order to overcome this problem. Therefore, if more patient data associated with osteoporosis were collected, the performance would be expected to improve. Furthermore, separating training and internal validation sets with balanced random selection considering the proportion of osteoporosis and control would help develop more reliable model.

V. CONCLUSION

In this study, we developed and validated prediction models using various machine learning methods including SVM, RF, ANN, and LR with the purpose of identifying the risk of osteoporosis in postmenopausal women and compared the performance of machine learning methods with that of conventional clinical decision tools including OST, ORAI, SCORE, and OSIRIS.

Previous studies indicated that SVM, RF, and ANN could help predict low BMD[23-26]. However, these studies did not apply the variable selection technique to select informative variables regarding osteoporosis. In addition, these studies had no objective comparisons of the performance of osteoporosis prediction. Therefore, we constructed a structural design for the prediction models along with a comparative study of various analytical methods. We used the 10-fold cross validation scheme to optimize the machine learning models. Internal and external validation was performed to assess the performance of the models for predicting osteoporosis risk. Consequently, the SVM model showed the best performance in comparison to other machine learning methods and the four conventional methods.

In conclusion, the most important finding of this study is the identification of postmenopausal women at high risk of osteoporosis to increase the possibility of appropriate treatment before fracture occurs. Machine learning methods might contribute to the advancement of clinical decision tools and understanding about the risk factors for osteoporosis. In particular, it is worth noticing that in addition to age and weight which are used in conventional methods, all our machine learning models included duration of menopause, osteoarthritis, and diabetes mellitus in common as informative factors associated with osteoporosis. It is needed to investigate the relationship between these factors and osteoporosis clinically. Further studies should be targeted at constructing an extended prediction model for progressive osteoporosis through the collection of prospective data, and the simultaneous prediction of osteopenia and osteoporosis using multi-category classification. If it is possible to simultaneously predict normal, osteopenia, and osteoporosis, it will be helpful for people with osteopenia who are required to prevent

in advance to delay or avoid the transition to osteoporosis. Furthermore, it is warranted to develop a prediction model based on clinical diagnosis by physicians not DEXA, which could improve performance of the model. We hope that this study enables women to reduce the risk of osteoporosis, which is the major cause of fracture.

REFERENCES

1. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 2006; 17:1726-33.
2. NIH consensus development conference. Diagnosis, prophylaxis and treatment of osteoporosis. Am J Med 1993; 94:646-50.
3. Marshall D, Johnell O, Wedel H. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. BMJ 1996; 312:1254-9.
4. Riggs BL, Melton Iii 3rd LJ, Robb RA, Camp JJ, Atkinson EJ, Peterson JM, Rouleau PA, McCollough CH, Bouxsein ML, Khosla S. Population-based study of age and sex differences in bone volumetric density, size, geometry, and structure at different skeletal sites. J Bone Miner Res 2004; 19:1945-54.
5. Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. Osteoporos Int 1994; 4:368–81.
6. Nayak S, Roberts MS, Greenspan SL. Cost-effectiveness of different screening strategies for osteoporosis in postmenopausal women. Ann Intern Med 2011; 155:751–61.
7. Elliot-Gibson V, Bogoch ER, Jamal SA, Beaton DE. Practice patterns in the diagnosis and treatment of osteoporosis after a fragility fracture: a systematic review. Osteoporos Int 2004; 15:767–78.
8. Raisz LG. Screening for osteoporosis. N Engl J Med 2005; 353:164–71.
9. Trémollieres FA, Pouillès JM, Drewniak N, Laparra J, Ribot CA, Dargent-Molina P. Fracture risk prediction using BMD and clinical risk factors in early postmenopausal women: sensitivity of the WHO FRAX tool. J Bone Miner Res 2010; 25:1002-9.
10. Barrett-Connor E, Siris ES, Wehren LE, Miller PD, Abbott TA, Berger ML, Santora AC, Sherwood LM. Osteoporosis and fracture risk in women of different ethnic groups. J Bone Miner Res 2005; 20:185-94.
11. Shin CS, Choi HJ, Kim MJ, Kim JT, Yu SH, Koo BK, Cho HY, Cho SW, Kim SW, Park YJ, Jang HC, Kim SY, Cho NH. Prevalence and risk factors of osteoporosis in

Korea: a community-based cohort study with lumbar spine and hip bone mineral density. Bone 2010; 47:378-87.

12. Choi YJ, Oh HJ, Kim DJ, Lee Y, Chung YS. The prevalence of osteoporosis in Korean adults aged 50 years or older and the higher diagnosis rates in women who were beneficiaries of a national screening program: The Korea National Health and Nutrition Examination Survey 2008-2009. J Bone Miner Res 2012; 27:1879-86.
13. Cui LH, Choi JS, Shin MH, Kweon SS, Park KS, Lee YH, et al. Prevalence of osteoporosis and reference data for lumbar spine and hip bone mineral density in a Korean population. J Bone Miner Metab 2008; 26:609–17.
14. Choi HS, Oh HJ, Choi H, Choi WH, Kim JG, Kim KM, et al. Vitamin D insufficiency in Korea--a greater threat to younger generation: the Korea National Health and Nutrition Examination Survey (KNHANES) 2008. J Clin Endocr Metab 2011; 96:643–51.
15. Khosla S, Melton LJ 3rd. Osteopenia. N Engl J Med 2007; 356:2293–300.
16. National Osteoporosis Foundation. Prevention: who's at risk? [accessed on 2014, March 24]. Available at: http://www.nof.org/.
17. Koh LK, Sedrine WB, Torralba TP, Kung A, Fujiwara S, Chan SP, et al. A simple tool to identify Asian women at increased risk of osteoporosis. Osteoporos Int 2001; 12:699–705.
18. Richy F, Gourlay M, Ross PD, Sen SS, Radican L, De Ceulaer F, et al. Validation and comparative evaluation of the osteoporosis self-assessment tool (OST) in a Caucasian population from Belgium. QJM 2004; 97:39–46.
19. Rud B, Hilden J, Hyldstrup L, Hróbjartsson A. The osteoporosis self-assessment tool versus alternative tests for selecting postmenopausal women for bone mineral density assessment: a comparative systematic review of accuracy. Osteoporos Int 2009; 20:599–607.
20. Hofbauer LC, Brueck CC, Singh SK, Dobnig H. Osteoporosis in patients with diabetes mellitus. J Bone Miner Res 2007; 22:1317–28.
21. Hsieh CH, Lu RH, Lee NH, Chiu WT, Hsu MH, Li YC. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines,

and artificial neural networks. Surgery 2011; 149:87–93.
22. Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, et al. Machine learning in bioinformatics. Brief Bioinform 2006; 7:86–112.
23. Ordóñez C, Matías JM, de Cos Juez JF, García PJ. Machine learning techniques applied to the determination of osteoporosis incidence in post-menopausal women. Math Comput Model 2009; 50:673–9.
24. De Cos Juez FJ, Suárez-Suárez MA, Sánchez Lasheras F, Murcia-Mazón A. Application of neural networks to the study of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. Math Comput Model 2011; 54:1665–70.
25. Lix LM, Yogendran MS, Leslie WD, Shaw SY, Baumgartner R, Bowman C, et al. Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. J Clin Epidemiol 2008; 61:1250–60.
26. Walid M, Ahmad S, Fadi C, Dima R. Intelligent predictive osteoporosis system. Int J Comput Appl 2011; 32:28–30.
27. Ministry for Health, Welfare and Family Affairs. The Fifth Korea National Health and Nutrition Examination Survey (KNHANES V), 2010-Health Examination. [accessed on 2014, March 24]. Available at: http://knhanes.cdc.go.kr/.
28. Batista GEAPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. Appl Artif Intell 2003; 17:519–33.
29. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI 1995; 14:1137–45.
30. Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics 2005; 21:631-43.
31. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995; 20:273–97.
32. Breiman L. Random forests. Mach Learn 2001; 45:5–32.
33. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006; 7:3.
34. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network

classification models: a methodology review. J Biomed Inform 2002; 35:352–9.
35. Drew PJ, Monson JR. Artificial neural networks. Surgery 2000; 127:3-11.
36. Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods 2000; 43:3-31.
37. Takasu A, Sakamoto T, Okada Y. Effect of induction rate for mild hypothermia on survival time during uncontrolled hemorrhagic shock in rats. J Trauma 2006; 61:1330-5.
38. Bishop CM. Pattern Recognition and Machine Learning. 4th ed. New York: Springer; 2006.
39. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007; 23:2507–17.
40. Dash M, Liu H. Consistency-based search in feature selection. Artif Intell 2003; 151:155–76.
41. Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review. GESTS Int T Comput Sci Eng 2006; 30:25–36.
42 Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. [accessed on 2014, March 24]. Available at: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf/.
43. Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Med Inform Decis Mak 2010; 10:16.
44. Sun Y, Kamel MS, Wong AKC, Wang Y. Cost-sensitive boosting for classification of imbalanced data. Pattern Recogn 2007; 40:3358–78.
45. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29-36.
46. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biometrical J 2005; 47:458–72.
47. Ministry for Health, Welfare and Family Affairs. The Fifth Korea National Health and Nutrition Examination Survey (KNHANES V), 2011-Health Examination. [accessed on 2014, June 4]. Available at: http://knhanes.cdc.go.kr/.

48. Yoo TK, Kim SK, Kim DW, Choi JY, Lee WH, Oh E, et al. Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. YMJ 2013; 54:1321-30.
49. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, et al. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. Lancet 2006; 368:1012–21.
50. Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 2008; 9:319.
51. Schwartz EN, Steinberg DM. Prescreening tools to determine who needs DXA. Curr Osteoporos Rep 2006; 4:148–52.
52. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009; 46:5–17.
53. Yeun EJ. A study on the health promoting lifestyle practices of middle-aged women in Korea. J Korean Soc Health Educ Prmot 2000; 17:41-59.
54. Ferrar L, Jiang G, Armbrecht G, Reid DM, Roux C, Glüer CC, et al. Is short vertebral height always an osteoporotic fracture? The Osteoporosis and Ultrasound Study (OPUS). Bone 2007; 41:5–12.
55. Khosla S, Melton LJ 3rd, Atkinson EJ, O'Fallon WM, Klee GG, Riggs BL. Relationship of serum sex steroid levels and bone turnover markers with bone mineral density in men and women: a key role for bioavailable estrogen. J Clin Endocrinol Metab 1998; 83:2266–74.
56. Riggs BL, Khosla S, Melton 3rd LJ. Sex steroids and the construction and conservation of the adult skeleton. Endocr Rev 2002; 23:279-302.
57. Lipworth L, Bailey LR, Trichopoulos D. History of breast-feeding in relation to breast cancer risk: a review of the epidemiologic literature. J Natl Cancer Inst 2000; 92:302–12.
58. Kim KC, Shin DH, Lee SY, Im JA, Lee DC. Relation between obesity and bone mineral density and vertebral fractures in Korean postmenopausal women. Yonsei Med J 2010; 51:857-63.

59. MacGregor GA, Cappuccio FP. The kidney and essential hypertension: a link to osteoporosis? J Hypertens 1993; 11:781–5.
60. Yang ZR, Chou KC. Bio-support vector machines for computational proteomics. Bioinformatics 2004; 20:735-41.
61. Liu HX, Zhang RS, Luan F, Yao XJ, Liu MC, Hu ZD, et al. Diagnosing breast cancer based on support vector machines. J Chem Inf Comput Sci 2003; 43:900-7.
62. Van Staa TP, Leufkens HG, Abenhaim L, Zhang B, Cooper C. Use of oral corticosteroids and risk of fractures. J Bone Miner Res 2000; 15:993–1000.
63. Lips P, Hosking D, Lippuner K, Norquist JM, Wehren L, Maalouf G, et al. The prevalence of vitamin D inadequacy amongst women with osteoporosis: an international epidemiological investigation. J Intern Med 2006; 260:245–54.
64. Choi JY, Shin A, Park SK, Chung HW, Cho SI, Shin CS, et al. Genetic polymorphisms of OPG, RANK, and ESR1 and bone mineral density in Korean postmenopausal women. Calcified Tissue Int 2005; 77:152–9.

< ABSTRACT(IN KOREAN)>

기계 학습 방법을 이용한 폐경기 후 여성의 골다공증 위험 예측

<지도교수 김 덕 원>

연세대학교 대학원 생체공학협동과정

김 성 권

골다공증은 전 세계적으로 중요한 공중 보건 관심사이다. 특히, 골다공증은 폐경기 후 여성에서 일반적이지만, 골절이 발생할 때까지 증상이 없다는 특징이 있다. 그 동안 폐경기 후 여성의 골다공증 위험도 평가를 위한 여러 임상 결정 도구들(clinical decision tools)이 개발되어 왔다. 본 연구는 기존의 임상 결정 도구들에 비해 더 정확하게 폐경기 후 여성의 골다공증 위험도를 식별할 수 있는 기계 학습 모델을 개발하고 검증하였다.

본 연구에서는 2010년 국민건강영양조사를 이용하여 폐경기 후 한국 여성의 자료를 수집하였다. Training data set은 널리 알려진 지원 벡터 기계, 랜덤 포리스트, 인공신경망, 로지스틱 회귀분석에 기반한 모델을 개발하는데 사용하였다. 기계 학습 모델은 10-fold cross validation을 통해 최적화하였고, 4가지 기존의 임상 결정 도구들(OST, ORAI, SCORE, OSIRIS)의 성능과 비교하였다. Internal validation set은 정확도와 AUC를 이용하여 모델의 성능을 평가하기 위해 사용하였다. 또한, 2011년 국민건강영양조사를 이용하여 external validation을 시행하였다.

Training set에서 지원 벡터 기계는 인공신경망, 로지스틱 회귀분석, OST, ORAI, SCORE, OSIRIS 보다 유의하게 더 높은 AUC를 보였다. Internal validation set에서 지원 벡터 기계의 AUC, 정확도, 민감도 및 특이도는 각각 0.827, 76.7%, 77.8%, 76.0%이었다. 또한, external validation set의 경우, 지원 벡터 기계의 AUC, 정확도, 민감도 및 특이도는 각각 0.833, 76.5%, 70.3%, 80.3%이었다. 지원 벡터 기계에 의해 선택된 유의한 요인으로는 나이, 신장, 체중, 체질량 지수, 폐경 기간, 모유 수유 기간, 여성호르몬제 복용여부, 고지혈증, 고혈압, 골관절염, 당뇨병이었다.

낮은 골밀도와 관련된 다양한 예측 변수들을 고려해봤을 때, 기계 학습 방법은 골다공증의 위험이 높은 폐경기 후 여성을 식별하는데 효과적인 도구가 될 것으로 사료된다.

핵심되는 말 : 골다공증, 폐경기 후 여성, 기계 학습, 위험도 평가, 임상 결정 도구, 지원 벡터 기계

PUBLICATION LIST

International Journal

1. Yoo TK, **Kim SK**, Kim DW, Choi JY, Lee WH, Oh E, et al. Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. YMJ 2013; 54:1321-30.

International Conference

2. **Kim SK**, Yoo TK, Park JS, Kim DW. Risk prediction of osteoporosis at any site among total hip, femoral neck, or lumbar spine using machine learning and conventional methods. 3th Asian Federation of Osteoporosis Societies, Seoul, Korea 2013 (Sep 6 - 8), (Traveling Award).

3. **Kim SK**, Yoo TK, Oh E, Kim DW. Osteoporosis risk prediction using machine learning and conventional methods. Conf Proc 35th IEEE Eng Med Biol Soc, Osaka, Japan 2013 (July 3 - 7).