



■ Original Article

# Application of Machine Learning Algorithms to Predict Osteoporotic Fractures in Women

Su Jeong Kang, Moon Jong Kim, Yang-Im Hur, Ji-Hee Haam, Young-Sang Kim\*

Department of Family Medicine, CHA Bundang Medical Center, CHA University, Seongnam, Korea

See editorial commentary page on 123

**Background:** Predicting the risk of osteoporotic fractures is vital for prevention. Traditional methods such as the Fracture Risk Assessment Tool (FRAX) model use clinical factors. This study examined the predictive power of the FRAX score and machine-learning algorithms trained on FRAX parameters.

**Methods:** We analyzed the data of 2,147 female participants from the Ansan cohort study. The FRAX parameters employed in this study included age, sex (female), height and weight, current smoking status, excessive alcohol consumption (>3 units/d of alcohol), and diagnosis of rheumatoid arthritis. Osteoporotic fracture was defined as one or more fractures of the hip, spine, or wrist during a 10-year observation period. Machine-learning algorithms, such as gradient boosting, random forest, decision tree, and logistic regression, were employed to predict osteoporotic fractures with a 70:30 training-to-test set ratio. We evaluated the area under the receiver operating characteristic curve (AUROC) scores to assess and compare the performance of these algorithms with the FRAX score.

**Results:** Of the 2,147 participants, 3.5% experienced osteoporotic fractures. Those with fractures were older, shorter in height, and had a higher prevalence of rheumatoid arthritis, as well as higher FRAX scores. The AUROC for the FRAX was 0.617. The machine-learning algorithms showed AUROC values of 0.662, 0.652, 0.648, and 0.637 for gradient boosting, logistic regression, decision tree, and random forest, respectively.

**Conclusion:** This study highlighted the immense potential of machine-learning algorithms to improve osteoporotic fracture risk prediction in women when complete FRAX parameter information is unavailable.

**Keywords:** Osteoporosis; Osteoporotic Fractures; Machine Learning; Risk Assessment

Received: September 4, 2023, Accepted: October 22, 2023

\*Corresponding Author: Young-Sang Kim <https://orcid.org/0000-0002-7397-5410>

Tel: +82-31-780-5360, Fax: +82-31-780-5944, E-mail: [zeroup@cha.ac.kr](mailto:zeroup@cha.ac.kr)

Copyright © 2024 The Korean Academy of Family Medicine

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

The principal objective of osteoporosis management is to reduce the fracture risk. Therefore, individuals at risk of fractures must be identified. According to the World Health Organization (WHO) diagnostic criteria for osteoporosis, treatment can be initiated for people when their bone mineral density (BMD) T score falls below -2.5; however, traditionally, treatment also can be started after screening people at risk through clinical risk factors. Attempts have been made to improve fracture risk assessment by combining BMD and clinical risk factors.<sup>1)</sup> The Fracture Risk Assessment Tool (FRAX) is one of the most widely used tools.

The FRAX, introduced by the WHO in 2008, is a computer-based algorithm that calculates the 10-year probability of hip and major osteoporotic fractures. It incorporates age, sex, height, weight, and a set of seven binary clinical risk factors including previous fractures, parent-fractured hip, current smoking, glucocorticoid use, excessive alcohol consumption, diagnosis of rheumatoid arthritis, and other causes of secondary osteoporosis. It can also include femoral neck BMD data.<sup>2)</sup> The FRAX fracture probability varies significantly from country to country. It was initially introduced in eight countries and has since been validated for use in more than 60 countries.<sup>3)</sup> Over the course of 2 decades, the FRAX has gained widespread utilization to the extent that it is included in clinical practice guidelines. However, it has certain limitations primarily stemming from its acceptance of only binary inputs. These limitations include the absence of information regarding the recency of previous fractures, the lack of dose-response data for glucocorticoid exposure, and the questionable accuracy of self-reported family history information.<sup>4)</sup>

Recently, artificial intelligence and machine-learning algorithms have gained significant attention in the field of osteoporosis. They are recognized for their potential in exploring new research fields, including the investigation of novel risk factors and the prediction of osteoporosis, falls, and fractures by leveraging biological testing, imaging, and clinical data.<sup>5)</sup> Currently, machine-learning algorithms, particularly supervised learning techniques, are the most extensively employed methods to predict outcomes. In this study, we aimed to construct a prediction model for osteoporotic fractures among female participants of a large-scale cohort study by employing machine-learning techniques and utilizing selected FRAX parameters, and to compare the performance of this model with conventional risk assessment tools.

## METHODS

### 1. Study Population

This study analyzed data from the Ansan cohort study, an ongoing prospective cohort study of the Korean Genome Epidemiology Study (KoGES). This cohort comprised a general population aged 40 years and above. The study participants were adults aged 40 years and above who were residents of Ansan region. The baseline study was conduct-

ed from June 25, 2001, to January 29, 2003, followed by biennial on-site interviews and comprehensive health examinations. The analysis included data from the seventh phase of KoGES. This study aligned with previously published research in terms of its cohort inclusion criteria and methodology.<sup>6,7)</sup>

From 3,975 individuals, we selected 2,238 female participants who had responded to the FRAX parameters for our analyses. We excluded 53 individuals due to missing weight and height information. In accordance with the research objectives, 2,147 participants who had answered “no” to questions regarding parental fractures, glucocorticoid use, and previous fractures were included in the final analysis.

This study was approved by the Institutional Review Board of CHA Bundang Medical Center (IRB protocol no., 2023-08-048), and informed consent was waived because of the retrospective nature of the study.

### 2. FRAX Parameters and Scores

The FRAX parameters employed in this study included age, sex (female), height and weight, current smoking, excessive alcohol consumption (>3 units/d of alcohol), and the diagnosis of rheumatoid arthritis. FRAX scores were calculated for the 10-year major osteoporotic fracture risk using the Korea-specific FRAX tool available online from the University of Sheffield (<https://www.sheffield.ac.uk/FRAX/tool.aspx?country=25>), without including BMD information.

### 3. Osteoporosis Self-Assessment Tool for Asians

The Osteoporosis Self-Assessment Tool for Asians (OSTA) score is used as a traditional and straightforward screening tool for predicting the risk of osteoporosis based on height and weight.<sup>8,9)</sup> The OSTA score is calculated by subtracting age (years) from weight (kg) and multiplying by 0.2.<sup>9)</sup> Individuals are stratified into being at low (OSTA >-1), medium ( $-1 \leq \text{OSTA} \leq -4$ ), and high risk (OSTA <-4) of sustaining osteoporosis.<sup>10)</sup>

### 4. Osteoporotic Fractures

An osteoporotic fracture was clinically defined as a fracture occurring spontaneously or following a minor trauma.<sup>11)</sup> We collected 10 years of cumulative data on fractures reported by patients through face-to-face and telephone interviews. Among these, we defined those occurring in the wrist, femur, spine, and upper arm as osteoporotic fractures.

### 5. Machine-Learning Algorithms and Features

The implemented machine-learning algorithms included gradient boosting (GB), random forest (RF), decision tree (DT), and logistic regression (LR). All models were performed with Scikit-learn in the Python 3.11 environment (Python Software Foundation, Wilmington, DE, USA). GB and RF are common tree-based ensemble methods known for their accuracy across various datasets. DT is a schematic representation of several decisions, each with a different probability of occurrence. LR models are widely used for multivariate analysis.<sup>12,13)</sup> During the training process, the hyperparameters were optimized us-

**Table 1.** Baseline characteristics

Characteristic	Overall	Without osteoporotic fracture	With osteoporotic fracture	P-value
No. of participants	2,147	2,071	76	
Age (y)	59.3±8.8	59.2±8.8	62.7±7.6	0.000
<50	412	405	7	
50–59	609	595	14	
60–69	808	768	40	
70–79	318	303	15	
Height (cm)	152.28±5.76	152.35±5.71	150.24±6.76	0.009
Weight (kg)	58.08±8.1	58.09±8.75	57.86±10.50	0.822
FRAX determinants (%)				
Current smoker	58 (2.9)	57 (2.8)	1 (1.3)	0.448
High alcohol consumption	525 (26.6)	512 (24.7)	13 (17.1)	0.129
Rheumatoid arthritis	212 (10.7)	199 (9.6)	16 (21.1)	0.001
FRAX score		5.0 (3.5 to 7.0)	6.0 (4.8 to 8.2)	0.001
OSTA score		-0.22 (-2.2 to 1.8)	-1.4 (-3.17 to 1.435)	0.020

Values are presented as number of participants, mean±standard deviation, number (%), or median (interquartile range). FRAX, Fracture Risk Assessment Tool; OSTA, Osteoporosis Self-Assessment Tool for Asians.

ing the grid-search approach.

Machine-learning models were trained to predict osteoporotic fractures using the selected FRAX parameters as input features. Among these parameters, age, height, and weight were continuous variables, whereas the others were categorical variables. All data were scaled using the StandardScaler method, which transforms continuous features to have a mean of 0 and a standard deviation of 1. The data were randomly divided into training and testing datasets with a 70:30 split, resulting in 1,502 in the training dataset and 645 in the test data. The prediction target of each model was a binary classifier, where “0” represented no fracture, where “1” represented one or more fractures accumulated over 10 years.

## 6. Performance Evaluation

The area under the receiver operating characteristic curve (AUROC) was calculated to compare the performance of each model.

## 7. Statistical Analysis

The baseline characteristics of individuals who experienced at least one fracture over a 10-year period were compared with those of individuals who did not experience any fracture during the same time-frame. Continuous variables were presented as means±standard deviations, while categorical variables were expressed as numbers (proportions). The FRAX and OSTA scores were presented as medians with interquartile ranges. To compare variables between the groups, we used the chi-square test or Fisher’s exact test for categorical variables. For continuous variables, the independent t-test and Mann-Whitney U test were used as appropriate. Statistical significance was defined as P-values <0.05. All statistical analyses were conducted using the IBM SPSS statistical package ver. 25.0 (IBM Corp., Armonk, NY, USA).

**Table 2.** Area under the receiver operating characteristic curve for the prediction of osteoporotic fracture

Variable	Osteoporotic fracture
Machine learning	
Gradient boosting classifier	0.662
Random forest classifier	0.637
Decision tree classifier	0.648
Logistic regression	0.652
FRAX score	0.617
OSTA score	0.579

FRAX, Fracture Risk Assessment Tool; OSTA, Osteoporosis Self-Assessment Tool for Asians.

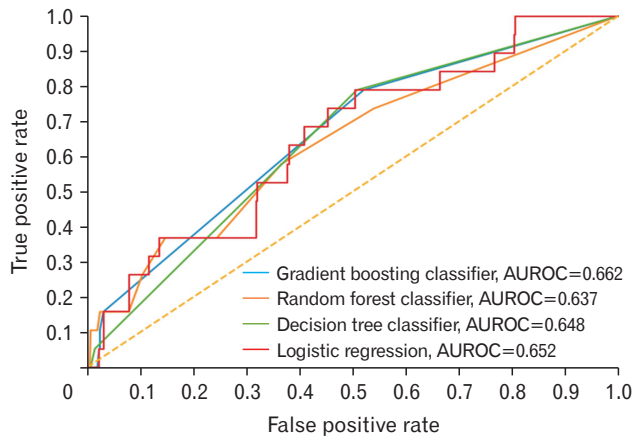
## RESULTS

### 1. Baseline Characteristics

We included 2,147 participants in the analysis (Table 1). The patients’ mean age was 59.3 years. During the follow-up, osteoporotic fractures occurred in 76 patients (3.5%). Patients who experienced osteoporotic fractures were older ( $P<0.001$ ) and had a shorter stature ( $P=0.009$ ). Furthermore, these patients were more frequently diagnosed with rheumatoid arthritis ( $P=0.001$ ). Additionally, their FRAX scores were higher ( $P=0.001$ ), and their OSTA scores were lower ( $P=0.020$ ).

### 2. Performance of Machine-Learning Models

The predictive results of each model are presented in Table 2. Among the four machine-learning models, GB showed the highest AUROC of 0.662. LR, DT, and RF yielded AUROC values of 0.652, 0.648, and 0.637, respectively (Figure 1). The AUROC values for the FRAX and OSTA were 0.617 and 0.579, respectively. When considering the AUROC for predicting osteoporotic fractures, the machine-learning models outperformed the FRAX and OSTA.



**Figure 1.** Receiver operating characteristic curve of each machine-learning model. AUC, area under the receiver operating characteristic curve.

## DISCUSSION

This study pioneered the use of machine-learning algorithms in predicting fractures in a community-based cohort of women using specific FRAX parameters. Although the FRAX is a widely adopted tool validated in over 60 countries, it has notable limitations. FRAX relies on a binary input for previous fractures, overlooking variations in fracture risk, such as the higher risk associated with two previous fractures compared to one. It does not account for the elevated recurrence risk of hip and vertebral fractures as opposed to distal fractures.<sup>14,15)</sup> Additionally, FRAX lacks information regarding the recency of fractures.<sup>16-18)</sup> Moreover, factors such as steroid exposure, smoking, and alcohol consumption are also likely to affect fractures in a dose-dependent manner.<sup>14,19-21)</sup>

Parental hip fracture is a well-established independent risk factor for osteoporotic fractures in offspring.<sup>4)</sup> However, self-reported family history information can be negatively influenced by recall bias or lack of medical knowledge.<sup>4)</sup>

This study aimed to predict osteoporosis by exclusively utilizing responses related to relatively objective and bias-free indicators, excluding items in the FRAX that have been identified for their limitations.

In the field of osteoporosis, machine-learning algorithms have been applied for various purposes, including predicting the presence of osteoporosis, forecasting fractures and falls, and discovering novel risk factors.<sup>5)</sup> Several studies have employed machine learning to predict fractures. In a study similar to ours that utilized Ansan cohort data, Kong et al.<sup>7)</sup> found that the CatBoost algorithm outperformed FRAX with BMD in predicting total fractures, with AUROC scores of 0.688 and 0.666, respectively. The AUROC for predicting major osteoporotic fractures using the FRAX score without BMD was 0.638, similar to our findings. Kong et al.<sup>7)</sup> trained their algorithm using additional clinical factors obtained from a cohort study and identified the top 20 risk factors for fractures. Their study also elucidated lesser-known novel factors such as arthralgia scores, homocysteine levels, and C-reactive protein levels, indicating their potential as new risk factors.<sup>7)</sup>

The reported AUROC for fracture prediction using FRAX varies from 0.6 to 0.79, and our study exhibited an AUROC of 0.61.<sup>22,23)</sup> The FRAX exhibits the best performance in predicting hip fractures in women when BMD information is available.<sup>23)</sup> Hence, when predicting hip fractures without BMD information and relying solely on clinical risk factors, as analyzed in this study, machine-learning algorithms can demonstrate better performance than that of the FRAX model. Therefore, the application of machine-learning algorithms in clinical practice is a meaningful endeavor.

This study had several strengths. To our knowledge, it was the first study to use machine learning to predict fractures by training on FRAX parameters from a large prospective cohort and comparing it with the FRAX score to explore the clinical applicability of machine learning. Particularly noteworthy is the validation of the potential for predicting fractures using the relatively objective FRAX parameters in situations where clinical information regarding these parameters is limited, potentially leading to the development of more effective tools.

This study had certain limitations. First, the evaluation of fractures was based on patient reports rather than radiographs, which may introduce discrepancies between the reported and actual conditions. The relatively low rate of osteoporotic fractures may be attributed to the fact that only self-reported fractures were considered. However, this cohort is known for well-structured research procedures, interview, and surveys. Another limitation was the absence of BMD data in our analysis. The predictive accuracy could potentially have been higher if BMD were included in the machine-learning training. This aspect should be investigated in future studies. Finally, this study aggregated all 10-year cumulative fractures into a single binary classification, which did not permit the prediction of fractures over time.

In conclusion, when obtaining complete FRAX information is challenging, machine-learning algorithms show promise for predicting osteoporotic fractures, particularly in women. This study highlights the advancements in the clinical application of machine learning.

## CONFLICT OF INTEREST

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

## FUNDING

This research was supported by the Korean Fund for Regenerative Medicine (KFRM) grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Health & Welfare) (code: KFRM 23C0102L1).

## ORCID

Su Jeong Kang: <https://orcid.org/0009-0006-1052-2832>  
 Moon Jong Kim: <https://orcid.org/0000-0002-7961-4616>  
 Yang-Im Hur: <https://orcid.org/0000-0002-2633-9980>  
 Ji-Hee Haam: <https://orcid.org/0009-0004-3131-0156>

Young-Sang Kim: <https://orcid.org/0000-0002-7397-5410>

## REFERENCES

1. Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int* 2007;18:1033-46.
2. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 2008;19:385-97.
3. Kanis JA, Harvey NC, Johansson H, Liu E, Vandendput L, Lorentzon M, et al. A decade of FRAX: how has it changed the management of osteoporosis? *Aging Clin Exp Res* 2020;32:187-96.
4. Yang S, Leslie WD, Yan L, Walld R, Roos LL, Morin SN, et al. Objectively verified parental hip fracture is an independent risk factor for fracture: a linkage analysis of 478,792 parents and 261,705 offspring. *J Bone Miner Res* 2016;31:1753-9.
5. Smets J, Shevroja E, Huggle T, Leslie WD, Hans D. Machine learning solutions for osteoporosis: a review. *J Bone Miner Res* 2021;36:833-51.
6. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* 2009;41:527-34.
7. Kong SH, Ahn D, Kim BR, Srinivasan K, Ram S, Kim H, et al. A novel fracture prediction model using machine learning in a community-based cohort. *JBMR Plus* 2020;4:e10337.
8. 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.
9. Muslim D, Mohd E, Sallehudin A, Tengku Muzaffar T, Ezane A. Performance of osteoporosis self-assessment tool for Asian (OSTA) for primary osteoporosis in post-menopausal Malay women. *Malays Orthop J* 2012;6:35-9.
10. Chen CC, Rau CS, Wu SC, Kuo PJ, Chen YC, Hsieh HY, et al. Association of osteoporosis self-assessment tool for Asians (OSTA) score with clinical presentation and expenditure in hospitalized trauma patients with femoral fractures. *Int J Environ Res Public Health* 2016;13:995.
11. Bessette L, Ste-Marie LG, Jean S, Davison KS, Beaulieu M, Baranci M, et al. The care gap in diagnosis and treatment of women with a fragility fracture. *Osteoporos Int* 2008;19:79-86.
12. Inui A, Nishimoto H, Mifune Y, Yoshikawa T, Shinohara I, Furukawa T, et al. Screening for osteoporosis from blood test data in elderly women using a machine learning approach. *Bioengineering (Basel)* 2023;10:277.
13. Matsuo K, Aihara H, Nakai T, Morishita A, Tohma Y, Kohmura E. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. *J Neurotrauma* 2020;37:202-10.
14. El Miedany Y. FRAX: re-adjust or re-think. *Arch Osteoporos* 2020;15:150.
15. Johansson H, Oden A, McCloskey EV, Kanis JA. Mild morphometric vertebral fractures predict vertebral fractures but not non-vertebral fractures. *Osteoporos Int* 2014;25:235-41.
16. Johansson H, Siggeirsdottir K, Harvey NC, Oden A, Gudnason V, McCloskey E, et al. Imminent risk of fracture after fracture. *Osteoporos Int* 2017;28:775-80.
17. Giangregorio LM, Leslie WD; Manitoba Bone Density Program. Time since prior fracture is a risk modifier for 10-year osteoporotic fractures. *J Bone Miner Res* 2010;25:1400-5.
18. Balasubramanian A, Zhang J, Chen L, Wenkert D, Daigle SG, Grauer A, et al. Risk of subsequent fracture after prior fracture among older women. *Osteoporos Int* 2019;30:79-92.
19. 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.
20. Kanis JA, Johnell O, Oden A, Johansson H, De Laet C, Eisman JA, et al. Smoking and fracture risk: a meta-analysis. *Osteoporos Int* 2005;16:155-62.
21. Kanis JA, Johansson H, Johnell O, Oden A, De Laet C, Eisman JA, et al. Alcohol intake as a risk factor for fracture. *Osteoporos Int* 2005;16:737-42.
22. Donaldson MG, Palermo L, Schousboe JT, Ensrud KE, Hochberg MC, Cummings SR. FRAX and risk of vertebral fractures: the fracture intervention trial. *J Bone Miner Res* 2009;24:1793-9.
23. Marques A, Ferreira RJ, Santos E, Loza E, Carmona L, da Silva JA. The accuracy of osteoporotic fracture risk prediction tools: a systematic review and meta-analysis. *Ann Rheum Dis* 2015;74:1958-67.