JBM
J Bone Metab 2023;30(3):245-252
https://doi.org/10.11005/jbm.2023.30.3.245
pISSN 2287-6375 eISSN 2287-7029

Original Article

Effect of Artificial Intelligence or Machine Learning on Prediction of Hip Fracture Risk: Systematic Review

Yonghan Cha[1], Jung-Taek Kim[2], Jin-Woo Kim[3], Sung Hyo Seo[4], Sang-Yeob Lee[4], Jun-Il Yoo[5]

[1]Department of Orthopedic Surgery, Daejeon Eulji Medical Center, Eulji University School of Medicine, Daejeon;
[2]Department of Orthopedic Surgery, Ajou Medical Center, Ajou University School of Medicine, Suwon;
[3]Department of Orthopedic Surgery, Nowon Eulji Medical Center, Eulji University, Seoul;
[4]Department of Biomedical Research Institute, Gyeongsang National University Hospital, Jinju;
[5]Department of Orthopaedic Surgery, Inha University Hospital, Inha University School of Medicine, Incheon, Korea

Corresponding author
Jun-Il Yoo
Department of Orthopaedic Surgery, Inha University Hospital, Inha University School of Medicine, 27 Inhang-ro, Jung-gu, Incheon 22332, Korea
Tel: +82-32-890-3663
Fax: +82-55-754-0477
E-mail: furim@hanmail.net

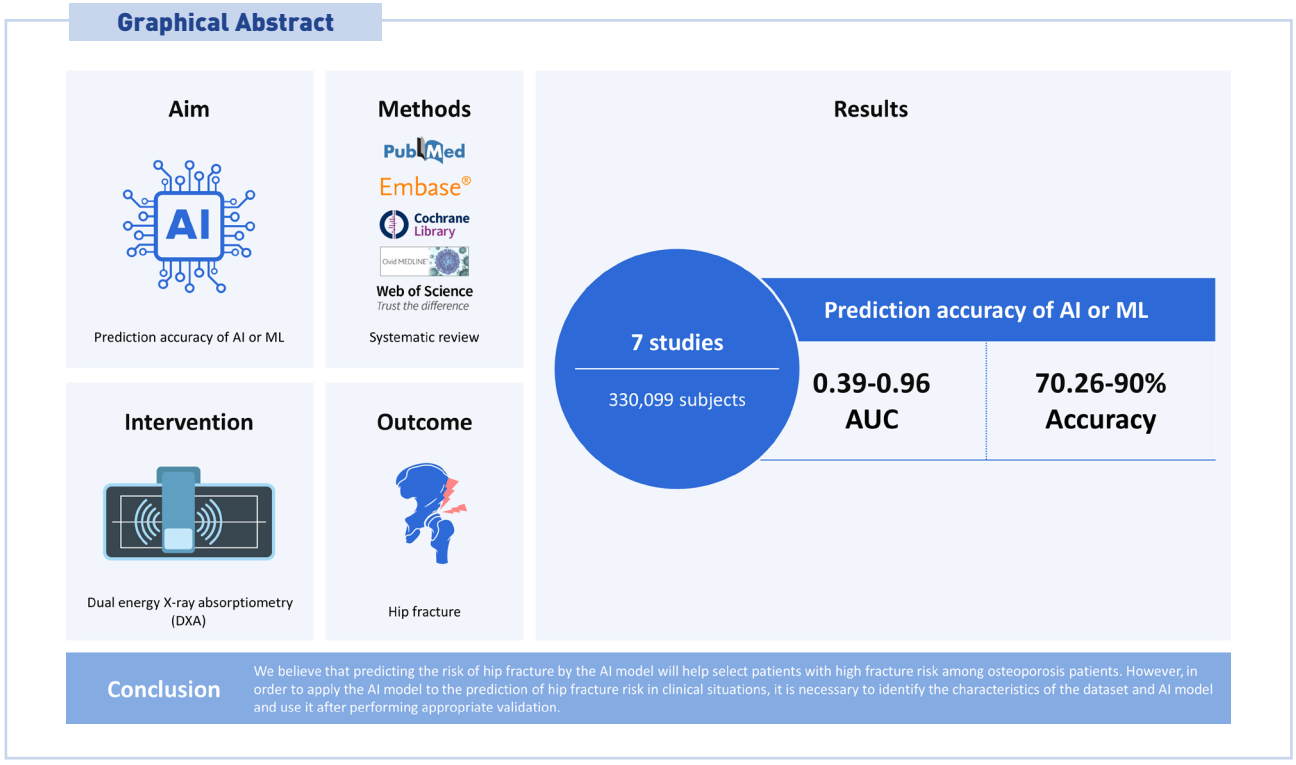
Received: April 19, 2023
Revised: May 12, 2023
Accepted: May 29, 2023

Background: Dual energy X-ray absorptiometry (DXA) is a preferred modality for screening or diagnosis of osteoporosis and can predict the risk of hip fracture. However, the DXA test is difficult to implement easily in some developing countries, and fractures have been observed before patients underwent DXA. The purpose of this systematic review is to search for studies that predict the risk of hip fracture using artificial intelligence (AI) or machine learning, organize the results of each study, and analyze the usefulness of this technology. **Methods:** The PubMed, OVID Medline, Cochrane Collaboration Library, Web of Science, EMBASE, and AHRQ databases were searched including "hip fractures" AND "artificial intelligence". **Results:** A total of 7 studies are included in this study. The total number of subjects included in the 7 studies was 330,099. There were 3 studies that included only women, and 4 studies included both men and women. One study conducted AI training after 1:1 matching between fractured and non-fractured patients. The area under the curve of AI prediction model for hip fracture risk was 0.39 to 0.96. The accuracy of AI prediction model for hip fracture risk was 70.26% to 90%. **Conclusions:** We believe that predicting the risk of hip fracture by the AI model will help select patients with high fracture risk among osteoporosis patients. However, to apply the AI model to the prediction of hip fracture risk in clinical situations, it is necessary to identify the characteristics of the dataset and AI model and use it after performing appropriate validation.

Key Words: Artificial intelligence · Diagnosis · Hip fractures · Machine learning · Prognosis

INTRODUCTION

Worldwide, 158 million people over the age of 50 are estimated to have a high risk of osteoporotic fractures.[1] Thus, 1 in 5 men and 1 in 3 women over the age of 50 will experience an osteoporotic fracture.[2] If the high risk of osteoporotic fractures due to an aging population continues, it is predicted that the number of osteoporotic fractures will double by 2045.[1] After osteoporotic fractures, patients suffer reduced quality of life due to chronic pain, functional disability and depen-



dence, high morbidity and mortality.[3] Especially, elderly hip fractures cause a socioeconomic burden in developed countries.[4] Therefore, various programs to properly treat these patients and prevent re-fracture are being implemented in various countries.[5] However, it may be more important to prevent the occurrence of primary fractures by diagnosing and treating osteoporosis, the main cause of these fractures, at an early stage.

Dual energy X-ray absorptiometry (DXA) is one of the preferred modalities for screening or diagnosis of osteoporosis and can predict the risk of hip fracture to some extent.[6,7] However, a study by Hsieh et al. [8] found that 80% of patients between the ages of 40 and 90 who had visited their institution and had pelvis or spine radiographs did not have a DXA test. Also, the DXA test may be difficult to implement easily in some developing countries.[1] Although the fracture risk assessment tool (FRAX) can predict the risk of hip fracture, symptomatic lumbar or occult hip fractures and fractures have been observed before patients underwent DXA.[1] Therefore, if there is no hassle of performing additional tests and a method of predicting the risk of hip fracture only by taking radiographs is provided, it will be possible to reduce the radiation exposure of patients and reduce additional costs.

Artificial intelligence (AI) or machine learning (ML) is a computational modeling tool widely accepted for modeling complex real-world health problems.[9] AI has already been used in many fields of medicine, such as nephrology, microbiology, and radiology, and is being studied in various fields such as diagnosis of fractures and prediction of clinical courses in orthopedics.[10] Although it is possible to predict the risk of hip fracture only with bone mineral density (BMD) or clinical factors, the prediction model using AI can compensate for the shortcomings of existing examination methods and handle large numbers of input variables simultaneously. Also, if an automated system of AI is constructed, there is an advantage that the hassle of checking examinations can be solved.[11] However, it seems that little is known about the assessment of hip fracture risk by AI yet.

Therefore, the purpose of this systematic review is to search for studies that predict the risk of hip fracture using AI or ML, organize the results of each study, and analyze the usefulness of this technology.

https://doi.org/10.11005/jbm.2023.30.3.245

METHODS

1. Study eligibility criteria

Studies were selected based on the following inclusion criteria: (1) studies using AI or ML techniques for prediction of hip fracture risk, such as femoral neck fracture, intertrochanteric fracture, or subtrochanteric fracture; and (2) studies reporting on a statistical analysis of area under the curve (AUC) or accuracy for prediction of hip fracture risk. Studies were excluded if they failed to meet the above criteria.

2. Search methods for identification of studies

The PubMed, OVID Medline, Cochrane Collaboration Library, Web of Science, EMBASE, and AHRQ databases were searched to identify relevant studies published up to June 2022 with English language restriction. The following search terms were used ("hip fractures"[MeSH Terms] OR ("hip"[All Fields] AND "fractures"[All Fields]) OR "hip fractures"[All Fields] OR ("hip"[All Fields] AND "fracture"[All Fields]) OR "hip fracture"[All Fields]) AND ("artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields]). A manual search was also conducted for possibly related references. Two authors, Yonghan Cha and Jung-Taek Kim, reviewed the titles, abstracts, and full texts of all potentially relevant studies independently, as recommended by the Cochrane Collaboration. Any disagreement was resolved by the third reviewer, Jun-Il Yoo. We assessed full-text articles of the remaining studies according to the previously defined inclusion and exclusion criteria, and then selected eligible articles. The reviewers were not blinded to authors, institutions, or the publication.

3. Data extraction

The following information was extracted from the included articles: authors, publication year, study period, number of patients, sex, age, AI algorithm variables for AI training, AUC, or accuracy for prediction of hip fracture risk.

RESULTS

The initial search identified 123 references from the selected databases. Eighty-two references were excluded by screening the abstracts and titles for duplicates, unrelated articles, case reports, and systematic reviews. The remaining 45 studies underwent full-text reviews and subsequently, 38 studies were excluded. Finally, 7 studies are included in this study.[8,12-17] The details of the identification of relevant studies are shown in the flow chart of the study selection process (Fig. 1).

The total number of subjects included in the 7 studies was 330,099 (Table 1). There were 3 studies that included only women,[12-14] and 4 studies included both men and women.[8,15-17] One study conducted AI training after 1:1

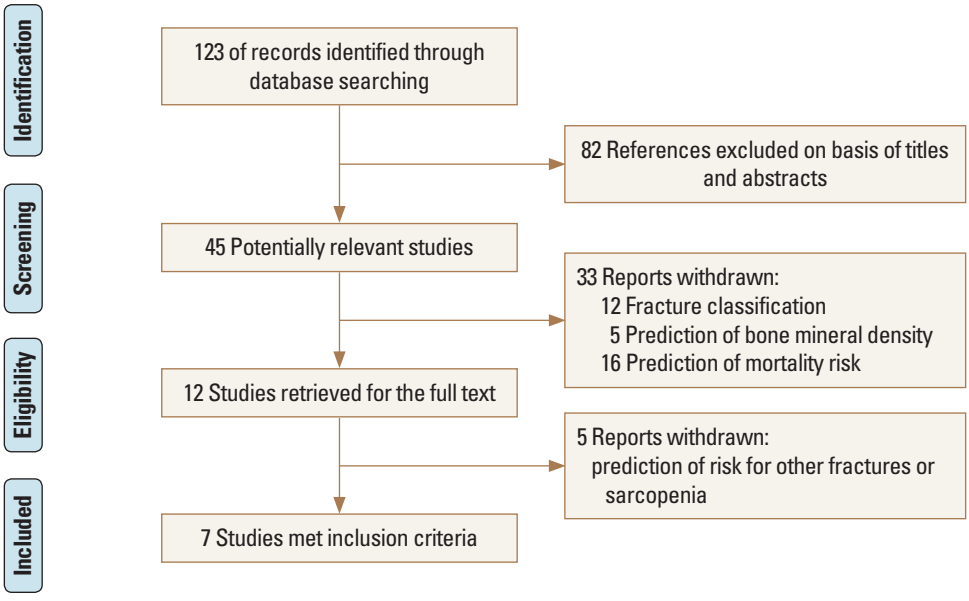


Fig. 1. The flow chart of the study selection process.

Table 1. Study, study period, demographic data of included studies

References	Year	Study period	Subjects	Total number of patients	Number of hip Fx patients (%)	Age (mean±SD)	Number of female (%)
Hsieh et al. [8]	2021	2006–2020	Aged 40–90 years who underwent hip and spine radiographs	23,339 (hip=5,164; spine=18,175)	No description	Hip=72.2±11.2; Spine=67.1±10.6	Hip=3,997 (77.4); Spine=14,469 (79.6)
Engels et al. [15]	2020	April 2008–March 2014	>65 years of age	288,086	7,644 (2.7)	75.67±6.2	140,709 (48.8)
Villamor et al. [12]	2020	No description	Postmenopausal women	137	89 (65.0)	81.4±6.95	137 (100.0)
Ho-Le et al. [13]	2017	No description	Women >60 years of age	1,167	90 (7.7)	No Fx=69.1±6.4; Hip Fx=76.8±7.5	1,167 (100.0)
Kruse et al. [16]	2017	1996–2006	Danish National Patient Registry	5,439 (men=717; women=4,722)	340 (6.3) (men=47; women=293)	• Men: no Fx=61.8; hip Fx=69.3 • Women: no Fx=59.7; hip Fx=74.5	4,722 (86.8)
Jiang et al. [14]	2015	No description	Postmenopausal women	11,497	186 (1.6)	• AI training group: no Fx=62±7; hip Fx=68.8±6.9 • AI validation group: no Fx=62±7.2; hip Fx=69.5±5.7	11,497 (100.0)
Tseng et al. [17]	2013	April 2004–January 2006	>60 years of age	434	217 (50.0) (men=68; women=149)	• Men: no Fx=78.4±7.9; hip Fx=70±7.4 • Women: no Fx=77.8±6.8; hip Fx=80.7±7.8	298 (68.7)

SD, standard deviation; Fx, fracture; AI, artificial intelligence.

matching between fractured and non-fractured patients. [17]

The AI algorithm used for prediction of hip fracture risk was very diverse, such as Artificial Neural Network, support vector machine, and K-nearest neighbors, etc. (Table 2). In addition, the variables used for AI training included not only demographic factors such as age, sex, body mass index, and past medical history, but also socioeconomic factors such as income level and education level. Also, Villamor et al. [12] used geometrical factors of femur from finite element analysis with patient's demographic factors for AI training. Hsieh et al. [8] used pelvis and lumbar spine radiographs and DXA as variables for hip fracture risk prediction. The AUC of AI prediction model for hip fracture risk was 0.39 to 0.96. The accuracy of AI prediction model for hip fracture risk was 70.26% to 90%.

DISCUSSION

The identification of high-risk individuals for hip fracture is clinically and socioeconomically important, because it could facilitate early intervention to reduce the burden of hip fracture in the general population.[13] However, osteoporosis is a silent disease that progresses before osteoporotic fractures.[18] After fracture has occurred, it increases mortality and morbidity in affected patients. Therefore, population-based screening is essential to identifying at-risk patients and implementing preventive services. But, prediction of hip fracture is very difficult, because it is influenced by multiple risk factors. Known risk factors for hip fracture are low BMD, previous history of hip fracture, female, advanced age, lower body weight and physical activity, sarcopenia, alcohol consumption, and smoking, etc.[19] Although the most clinically important risk factor is low BMD, considering BMD and other clinical factors together for predicting fracture risk can increase accuracy.[13] However, as in FRAX, the assessment of hip fracture risk using conventional methods may not include several important factors.[11] On the other hand, hip fracture prediction using ML can handle large numbers of input variables simultaneously and consider invisible relationships between variables.[11] Also, as Kruse et al. [16] showed in their study, there is an advantage of not having to go through input work by clinicians if the system is built to automatically analyze clinical data or image data in the AI model.

BMD and FRAX are commonly used as key tools for as-

https://doi.org/10.11005/jbm.2023.30.3.245

Table 2. Prediction model of artificial intelligence and results of prediction for hip fracture risk in included studies

References	AI algorithm	Training variables for AI	AUC for prediction of hip Fx risk	Accuracy for prediction of hip Fx risk
Hsieh et al. [8]	No description	Pelvis and lumbar spine radiographs and DEXA	10-year risk=0.96	10-year risk=90
Engels et al. [15]	LR, Random Forest, SVM, RUSBoost, SuperLearner, XGBoost	Age, gender, prior fx history, medication use within administrative claims data	4-year risk: LR=0.695–0.704; Random Forest=0.685; SVM=0.650; RUSBoost=0.702; SuperLearner=0.698; XGBoost=0.703	No description
Villamor et al. [12]	SVM, LR, ANN, Random Forest	Age, height, weight, BMI, BMD, geometrical factors of femur from FEA	No description	SVM=78.35, LR=73.09, ANN=70.26, Random Forest=73.34
Ho-Le et al. [13]	LR, ANN, KNN, SVM	BMD, fx history, frequency of falls during the previous 12 months, calcium intakes, alcohol consumption, cigarette, metabolic equivalent index, height, weight	No description	10-year hip fx risk: ANN=87.3, LR=81.5, KNN=79.4, SVM=81.5
Kruse et al. [16]	Xtreme Gradient Boosting, Conditional Inference Random Forest, Generalized Additive Model, ADABoost, Random Forest, Generalized Linear Model, Bagged Multivariate Adaptive Regression Splines, Bayesian Generalized Linear Model, Bagged Flexible Discriminant, Bagged Tree, LR, Classification Tree, Stochastic Gradient Boosting, KNN	Medication use, total medication costs, ICD-10 codes, maximum length in years from occurrence to the scan date, CCI, yearly income, primary medical visit count and costs during both the prior and post periods, education level, job, ethnicity, age, sex, height, BMI, DEXA	• 5-year hip fx risk in women: Xtreme Gradient Boosting=0.92; Random Forest=0.91; Bagged Flexible Discriminant=0.91; Bagged Multivariate Adaptive Regression Splines=0.91; Generalized Additive Model=0.89; Conditional Inference Random Forest=0.86; Bagged Tree=0.87; LR=0.86; Generalized Linear Model=0.85; KNN=0.83 • 5-year hip fx risk in men: Xtreme Gradient Boosting=0.89; Conditional Inference Random Forest=0.88; Generalized Additive Model=0.89; ADABoost=0.84; Random Forest=0.81; Generalized Linear Model=0.83; Bagged Multivariate Adaptive Regression Splines=0.80; Bayesian Generalized Linear Model=0.77; Bagged Flexible Discriminant=0.74; Bagged Tree=0.68; LR=0.58; Classification Tree=0.57; Stochastic Gradient Boosting=0.39	No description
Jiang et al. [14]	SVM	Ethnicity, self-reported health, fx history, physical activity, smoking status, parent broke hip, corticosteroid use, diabetes treatment, age, height, weight, BMD, hip geometry	0.881	No description
Tseng et al. [17]	ANN	Monthly income, weight, height, leisure-time physical activity, MMSE score, peak expiratory flow rate, hand grip strength, BMD	0.868	No description

AI, artificial intelligence; LR, logistic regression; SVM, support vector machine; XGBoost, extreme gradient boosting; ANN, artificial neural network; KNN, K-nearest neighbor; ADABoost, adaptive boosting; DEXA, dual-energy X-ray absorptiometry; BMI, body mass index; BMD, bone mineral density; FEA, finite element analysis; ICD-10, International Classification of Diseases, tenth revision; CCI, Charlson Comorbidity Index; MMSE, Mini-Mental State Examination; AUC, area under the curve; Fx, fracture.

sessing hip fracture risk in clinical practice. However, several studies have reported using statistical regression analysis to predict hip fracture occurrence based on demographic factors of patients. Aldieri et al. [20] reported on the prediction of hip fracture using logistic regression analysis with demographic data (such as age, weight, and height), BMD, and quantitative computed tomography (QCT) data in a cohort of 100 Caucasian postmenopausal women aged 55 or older. Their study showed AUC values ranging from 0.64 to 0.92 depending on the combination of QCT data shape and intensity.[20] Baker-LePain et al. [21] conducted prediction analysis for hip fracture using a statistical technique

called Active Shape Modeling in a cohort of 168 individuals aged 65 or older who had experienced hip fractures and 231 individuals who had not. The reported AUC values ranged from 0.631 to 0.835 in their studies.[21] Through these 2 studies, it was demonstrated that traditional statistical analysis using a well-combined set of variables can achieve accuracy similar to AI-based prediction. However, a limitation of such statistical analyses is that the selection of variables ultimately relies on human decision-making, making it challenging to establish a fully automatic system. In the studies included in our review, the reported AUC values ranged from 0.39 to 0.96, suggesting that AI models have the potential to achieve very high AUC values compared to previous studies utilizing traditional statistical analyses. However, direct comparisons between traditional statistical analysis and AI prediction in terms of accuracy for hip fracture prediction have not been made yet, indicating the need for further research.

Considering the results of the studies included in this review, AUC and accuracy for prediction of hip fracture risk by AI ranged from low to high. This seems to be because the variables and AI algorithms used are very diverse. These reports also make it difficult for clinicians to determine which model is the best. Therefore, when evaluating the results of studies on the prediction of hip fracture risk using the AI model, the following should be considered. The first is to check whether external validation is present. Kruse et al. [16] argued that it could be very dangerous to report the results after training using one AI model or dataset. They also said that some studies have reported the results of prediction models without external validation, and these problems are not well known. The second factor to pay attention to in hip fracture prediction analysis using AI model is overfitting. Ho-Le et al. [13] reported that overfitting could be a problem in any AI models, and that the number of hip fractures per risk factor was >10 to prevent overfitting. They argued that the consistency between training and test results of AI algorithm models is important to determine whether over-fitting is occurring. A third consideration is the problem of handling missing data in the dataset used for AI training. A lot of data is required for AI training, and especially big database facilitates this. However, not all patients in the database have all the data. Therefore, it is advisable to check whether techniques for missing data such as imputation are used. However, Jiang et al. [14] reported that these techniques were not necessary because the purpose of their study was primarily to demonstrate the potential increase in predictive ability by combining clinical and computational data.

There are several limitations to our study. First, we did not consider the degree of training of AI algorism. Second, because the occurrence of hip fractures in the elderly is multifactorial and influenced by various latent factors, we did not take into account the number and types of variables used in the prediction model when interpreting the study results. As a result of these influences, the AUC and accuracy values of the AI model are found to be diverse. Third, we have not been able to conclude which AI model is the most accurate for prediction of hip fracture yet. Fourth, the observed hip fracture rates in the 7 studies included in our systematic review exhibit significant variation. This suggests a high heterogeneity among the patients included in each study, which could potentially impact the accuracy of AI predictions for hip fractures.

We believe that predicting the risk of hip fracture by the AI model will help select patients with high fracture risk among osteoporosis patients. However, to apply the AI model to the prediction of hip fracture risk in clinical situations, it is necessary to identify the characteristics of the dataset and AI model and use it after performing appropriate validation.

DECLARATIONS

Funding

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI22C 0494). This work was supported by biomedical research instiute fund (GNUHBRIF-2019-0013) from the Gyeongsang National University Hospital.

Ethics approval and consent to participate

Not applicable.

Conflict of interest

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

https://doi.org/10.11005/jbm.2023.30.3.245

ORCID

Yonghan Cha *https://orcid.org/0000-0002-7616-6694*
Jung-Taek Kim *https://orcid.org/0000-0003-4243-5793*
Jin-Woo Kim *https://orcid.org/0000-0003-0186-5834*
Sung Hyo Seo *https://orcid.org/0000-0003-3327-5425*
Sang-Yeob Lee *https://orcid.org/0000-0001-8024-1135*
Jun-Il Yoo *https://orcid.org/0000-0002-3575-4123*

REFERENCES

1. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 2006;17:1726-33. https://doi.org/10.1007/s00198-006-0172-4.
2. Sànchez-Riera L, Carnahan E, Vos T, et al. The global burden attributable to low bone mineral density. Ann Rheum Dis 2014;73:1635-45. https://doi.org/10.1136/annrheumdis-2013-204320.
3. Saito T, Sterbenz JM, Malay S, et al. Effectiveness of anti-osteoporotic drugs to prevent secondary fragility fractures: systematic review and meta-analysis. Osteoporos Int 2017;28:3289-300. https://doi.org/10.1007/s00198-017-4175-0.
4. Cha YH, Ha YC, Lim JY, et al. Introduction of the cost-effectiveness studies of fracture liaison service in other countries. J Bone Metab 2020;27:79-83. https://doi.org/10.11005/jbm.2020.27.2.79.
5. Cha YH, Ha YC, Lim JY. Establishment of fracture liaison service in Korea: where is it stand and where is it going? J Bone Metab 2019;26:207-11. https://doi.org/10.11005/jbm.2019.26.4.207.
6. LeBoff MS, Greenspan SL, Insogna KL, et al. The clinician's guide to prevention and treatment of osteoporosis. Osteoporos Int 2022;33:2049-102. https://doi.org/10.1007/s00198-021-05900-y.
7. Cheng CT, Wang Y, Chen HW, et al. A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. Nat Commun 2021;12:1066. https://doi.org/10.1038/s41467-021-21311-3.
8. Hsieh CI, Zheng K, Lin C, et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. Nat Commun 2021;12:5472. https://doi.org/10.1038/s41467-021-25779-x.
9. Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods 2000;43:3-31. https://doi.org/10.1016/s0167-7012(00)00201-3.
10. Patel JL, Goyal RK. Applications of artificial neural networks in medical science. Curr Clin Pharmacol 2007;2:217-26. https://doi.org/10.2174/157488407781668811.
11. de Vries BCS, Hegeman JH, Nijmeijer W, et al. Comparing three machine learning approaches to design a risk assessment tool for future fractures: predicting a subsequent major osteoporotic fracture in fracture patients with osteopenia and osteoporosis. Osteoporos Int 2021;32:437-49. https://doi.org/10.1007/s00198-020-05735-z.
12. Villamor E, Monserrat C, Del Río L, et al. Prediction of osteoporotic hip fracture in postmenopausal women through patient-specific FE analyses and machine learning. Comput Methods Programs Biomed 2020;193:105484. https://doi.org/10.1016/j.cmpb.2020.105484.
13. Ho-Le TP, Center JR, Eisman JA, et al. Prediction of hip fracture in post-menopausal women using artificial neural network approach. Annu Int Conf IEEE Eng Med Biol Soc 2017;2017:4207-10. https://doi.org/10.1109/embc.2017.8037784.
14. Jiang P, Missoum S, Chen Z. Fusion of clinical and stochastic finite element data for hip fracture risk prediction. J Biomech 2015;48:4043-52. https://doi.org/10.1016/j.jbiomech.2015.09.044.
15. Engels A, Reber KC, Lindlbauer I, et al. Osteoporotic hip fracture prediction from risk factors available in administrative claims data - A machine learning approach. PLoS One 2020;15:e0232969. https://doi.org/10.1371/journal.pone.0232969.
16. Kruse C, Eiken P, Vestergaard P. Machine learning principles can improve hip fracture prediction. Calcif Tissue Int 2017;100:348-60. https://doi.org/10.1007/s00223-017-0238-7.
17. Tseng WJ, Hung LW, Shieh JS, et al. Hip fracture risk assessment: artificial neural network outperforms conditional logistic regression in an age- and sex-matched case control study. BMC Musculoskelet Disord 2013;14:207. https://doi.org/10.1186/1471-2474-14-207.
18. Nazrun AS, Tzar MN, Mokhtar SA, et al. A systematic review of the outcomes of osteoporotic fracture patients after hospital discharge: morbidity, subsequent fractures, and mortality. Ther Clin Risk Manag 2014;10:937-48. https://doi.org/10.2147/tcrm.S72456.
19. Marks R. Hip fracture epidemiological trends, outcomes, and risk factors, 1970-2009. Int J Gen Med 2010;3:1-17.
20. Aldieri A, Bhattacharya P, Paggiosi M, et al. Improving the

hip fracture risk prediction with a statistical shape-and-intensity model of the proximal femur. Ann Biomed Eng 2022;50:211-21. https://doi.org/10.1007/s10439-022-02918-z.

21. Baker-LePain JC, Luker KR, Lynch JA, et al. Active shape modeling of the hip in the prediction of incident hip fracture. J Bone Miner Res 2011;26:468-74. https://doi.org/10.1002/jbmr.254.

https://doi.org/10.11005/jbm.2023.30.3.245