BMJ Open

Predictive value of machine learning on fracture risk in osteoporosis: a systematic review and meta-analysis

Yanqian Wu [ORCID] , Jianqian Chao, Min Bao, Na Zhang

Key Laboratory of Environmental Medicine Engineering of Ministry of Education/Health Management Research Center, School of Public Health, Southeast University, Nanjing, Jiangsu, China

Correspondence to
Dr Jianqian Chao;
chaoseu@163.com

ABSTRACT

Objectives Early identification of fracture risk in patients with osteoporosis is essential. Machine learning (ML) has emerged as a promising technique to predict the risk, whereas its predictive performance remains controversial. Therefore, we conducted this systematic review and meta-analysis to explore the predictive efficiency of ML for the risk of fracture in patients with osteoporosis.

Methods Relevant studies were retrieved from four databases (PubMed, Embase, Cochrane Library and Web of Science) until 31 May 2023. A meta-analysis of the C-index was performed using a random-effects model, while a bivariate mixed-effects model was used for the meta-analysis of sensitivity and specificity. In addition, subgroup analysis was performed according to the types of ML models and fracture sites.

Results Fifty-three studies were included in our meta-analysis, involving 15 209 268 patients, 86 prediction models specifically developed for the osteoporosis population and 41 validation sets. The most commonly used predictors in these models encompassed age, BMI, past fracture history, bone mineral density T-score, history of falls, BMD, radiomics data, weight, height, gender and other chronic diseases. Overall, the pooled C-index of ML was 0.75 (95% CI: 0.72, 0.78) and 0.75 (95% CI: 0.71, 0.78) in the training set and validation set, respectively; the pooled sensitivity was 0.79 (95% CI: 0.72, 0.84) and 0.76 (95% CI: 0.80, 0.81) in the training set and validation set, respectively; and the pooled specificity was 0.81 (95% CI: 0.75, 0.86) and 0.83 (95% CI: 0.72, 0.90) in the training set and validation set, respectively.

Conclusions ML has a favourable predictive performance for fracture risk in patients with osteoporosis. However, most current studies lack external validation. Thus, external validation is required to verify the reliability of ML models.

PROSPERO registration number CRD42022346896.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The latest systematic review and meta-analysis conducted to assess machine learning (ML) models for fracture risk.
⇒ We performed a quantitative synthesis to enhance the comparability of ML models.
⇒ C-index, sensitivity and specificity were performed to evaluate the performance of ML models.
⇒ Several studies were included in the systematic review but excluded from subsequent meta-analyses.
⇒ Most of the included studies lack external validation.

serious health issue after cardiovascular diseases.[3] This condition may cause fragility fractures that commonly occur in the wrist, spine and hip. Spine and hip fractures may lead to disability, which not only affects the quality of life and longevity of patients but also causes enormous medical expenses and a heavy burden of care.[4 5]

Machine learning (ML), a subfield of artificial intelligence, enables computers to 'learn' through programmes. Compared with traditional statistical methods, ML emphasises more on the accuracy of prediction and can detect regularities in multi-dimensional data sets. ML algorithms can be basically divided into supervised learning and unsupervised learning.[6] ML has been applied in the field of osteoporosis, providing a novel method for the prediction of fracture risk. A review by Ferizi *et al*[7] (2019) summarised relevant studies on the application of artificial intelligence to the prediction of osteoporosis. It drew a conclusion that ML methods for automatic image segmentation and fracture risk prediction showed a promising clinical value. A systematic review by Smets *et al*[8] (2021) reviewed the state-of-the-art ML methods and their application in osteoporosis diagnosis and fracture prediction. Another review by Anam *et al*[9] (2021) explored the prediction performance of MRI for osteoporosis in trabecular bone from a methodology-driven and application perspective. Most studies

INTRODUCTION

Osteoporosis is a systemic metabolic bone disease characterised by decreased bone mass and degraded bone microarchitecture, leading to an increased risk of bone fragility fracture.[1] Due to high disability and morbidity rates, high treatment costs and low quality of life of patients, it has emerged as a global health concern.[2] According to the WHO, osteoporosis is the second most

focused on the role of ML in the prediction of osteoporosis indicators, such as bone mineral density (BMD), or in the automatic segmentation of the images of patients at risk of osteoporosis. However, the efficiency of ML in predicting osteoporotic fractures is understudied.

The present study evaluated the predictive performance of ML for fracture risk in osteoporosis patients, providing an evidence-based medical basis for the application of ML in clinical practice.

MATERIALS AND METHODS

This study was conducted in accordance with the Preferred Items of Systematic Review and Meta-Analysis (PRISMA) statement (online supplemental table S1).[10] The protocol was registered on the international prospective register of systematic reviews (PROSPERO) (Registration No. CRD42022346896). Relevant studies were retrieved from Pubmed, Embase, Cochrane Library and Web of Science, and the retrieval was as of 31 May 2023. Two researchers independently searched the literature. The search strategy is shown in online supplemental table S2.

Inclusion criteria were as follows: (1) patients were diagnosed with osteoporosis; (2) ML was applied to predict fracture risk; (3) at least one measure of model performance (discrimination or calibration) was reported; (4) study population included adult patients older than 18 years, mainly including adults, older people and postmenopausal women. Exclusion criteria were as follows: (1) studies that only analysed risk factors without building complete ML models; (2) studies that only included osteoporosis but did not mention fracture risk; (3) studies without available full text (or only abstract available) or data; (4) meta-analyses, reviews, case reports, editorial materials, letters, protocols, errata, and notes.

Two researchers independently extracted data using standardised tables. Any studies excluded after full-text review have been recorded with reasons for their exclusion. The list of extracted items was based on the CHARMS checklist,[11] and two data extraction sheets were prepared for developed and validated models, respectively. Finally, the extracted data included the first author, year of publication, country, study design, data source, population group, gender, age, fracture sites, types of predictive models, number of predictors and outcomes. The risk of bias was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST). The PROBAST contained a large number of questions in four distinct domains: participants, predictors, outcomes, and statistical analysis, reflecting the overall risk of bias and applicability.[12]

Meta-analysis of C-index, sensitivity and specificity was performed to evaluate the performance of ML models. If the C-index did not report 95% CIs and SEs, we estimated the SEs in reference to the study by Debray *et al.*[13] A C-index of 0.5 indicates low discrimination; 0.6 to 0.7 indicates modest discrimination; 0.71 to 0.8 indicates very good discrimination; and

greater than 0.8 indicates strong discrimination.[14] When original studies did not report the accuracy, we calculated it based on the sensitivity, specificity, the number of samples in each subgroup and the number of modelling samples.[13] Given the differences in variables, ML algorithms and parameters across the studies, the random-effects model was preferred for the meta-analysis of C-index, and the bivariate mixed-effects model was used for the meta-analysis of sensitivity and specificity. Heterogeneity was quantified using I^2 statistics. Sensitivity analysis was performed to further identify the source of heterogeneity by removing each study and re-calculating the pooled effect size of the remaining studies. The meta-analysis was performed using the software Stata V.15.1 (Stata Corporation) and R V.4.2.0 (R Development Core Team, Vienna, http://www.R-project.org). A p value less than 0.05 was considered statistically significant.

Patient and public involvement

No patients involved.

RESULTS

A total of 12 468 studies were searched from the databases, including 2409 from PubMed, 4387 from Embase, 170 from Cochrane Library and 5502 from Web of Science. After removing duplicates and screening titles and abstracts, 378 articles remained. According to a full-text review, 53 articles[13–67] were included. Fifty-three articles presented the development of one or more prediction models for osteoporotic fracture, while 26 articles described the validation of one or more models. The search process is shown in figure 1.

Fifty-three studies were ultimately included in our meta-analysis, involving 15 209 268 patients. Many studies originated from USA (n=11), European(n=11) and China (n=8). Most studies were cohort studies (n=46), and the rest were case–control studies (n=7). The median age of osteoporosis patients was 68.8 years (ranging from 48.5 to 84 years). The study population in most studies covered women (n=24). The fracture sites included multi-site (n=26), vertebra (n=14), hip (n=12) and femur (n=1). Most studies were based on clinical hospital data (n=19), while some used questionnaire collection data (n=10), osteoporosis registry data (n=9), electronic health records (n=7) and administrative data (n=6). Only 13 articles elucidated the cross-validation method. The baseline characteristics of the included studies are shown in online supplemental table S3.

There were 86 prediction models specifically developed for the osteoporosis population and 41 validation sets. Ninety-eight ML models reported the C-index or the area under the receiver operating characteristic curve (AUC), ranging from 0.50 to 0.98. Online supplemental table S4 shows all studies
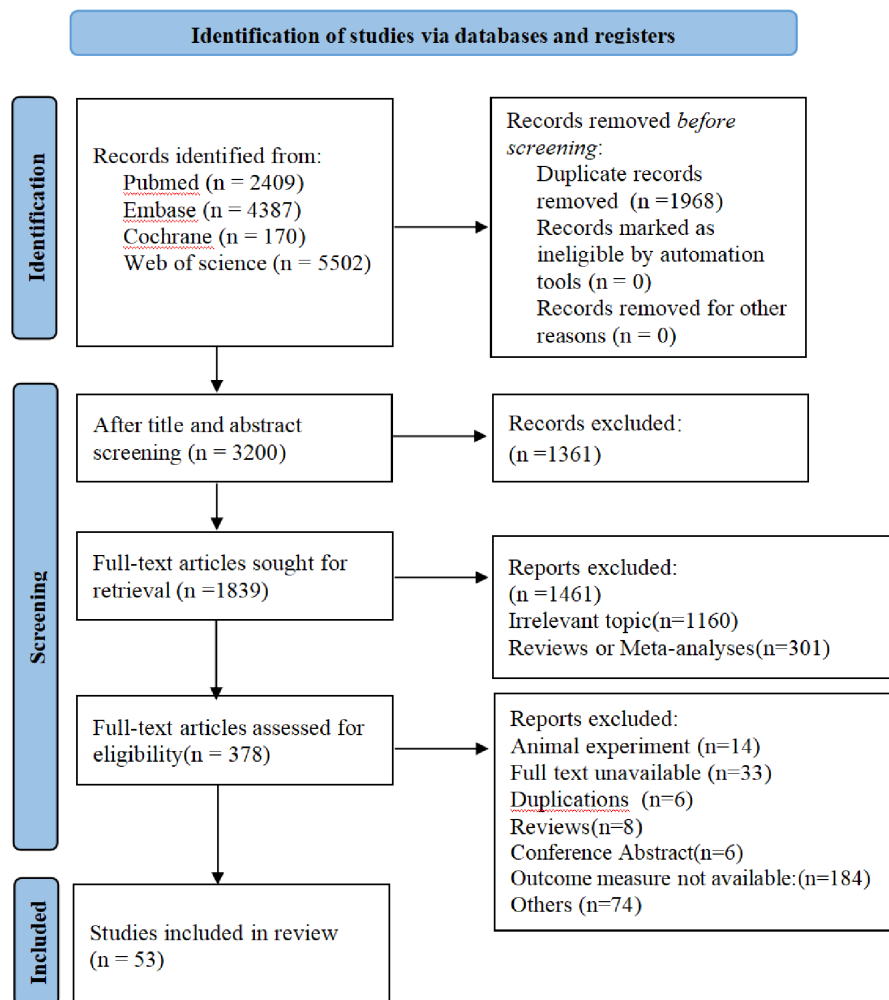
Figure 1 The flow chart of retrieval process.

on the development and validation of ML models for outcome prediction in patients with osteoporosis. Among all the identified prediction models, the logistic regression (31.4%) was the most commonly used algorithm, followed by the survival model (18%).

The most commonly used predictors in ML models were age (n=72), body mass index (BMI) (n=40), past fracture history (n=35), BMD T-score (n=33), history of falls (n=29), BMD (n=28), radiomics data (n=25), weight (n=24), height (n=23), gender (n=20), and other chronic diseases (n=20) (table 1).

The risk of bias assessment of the included studies is summarised in figure 2. More than half of these studies had a high risk of bias (67%). The risk of bias in most studies was low in terms of participants, predictors and outcome. However, a high or unclear risk of bias in the statistical analysis was observed in all studies. More details are shown in online supplemental table S5.

Sixty-six training datasets and 32 validation datasets were included in the meta-analysis of the C-index. Since substantial heterogeneity was present, we performed subgroup analyses based on fracture site and model type. Table 2 shows the results

of the meta-analysis of C-index of ML models in predicting osteoporosis. Logistic regression is the most widely used method. The forest plot of C-index is presented in online supplemental figures S1 and S2. The pooled C-index was 0.75 (95% CI: 0.72, 0.78) (I^2=99.7%, p<0.001) in the training set and 0.75 (95% CI: 0.71, 0.78) (I^2=99.8%, p<0.001) in the validation set. In the training set, other deep learning method showed the highest predictive performance (pooled C-index=0.97), followed by convolutional neural network (CNN) (pooled C-index=0.94), decision trees (pooled C-index=0.78) and logistic regression (pooled C-index=0.75). Furthermore, models for vertebral fracture (pooled C-index=0.80) and hip fracture (pooled C-index=0.76) outperformed those for multi-site fracture (pooled C-index=0.70). However, in the validation set, CNN (pooled C-index=0.98) showed the best performance, closely followed by other deep learning method (pooled C-index=0.82), logistic regression (pooled C-index=0.80) and support vector machines (SVMs) (pooled C-index=0.78). Models for vertebral fracture (pooled C-index=0.87) outperformed those for hip fracture (pooled C-index=0.73) and multi-site fracture (pooled C-index=0.71). Across

Table 1 Main sorts of predictors included in developed models for osteoporosis patients

Predictors	Number of models
Demographics	
Age	72
History of falls	29
Sex	20
Women's menopause age	8
Family genetic history	6
Race	5
Physical examination	
Body mass index	40
BMD t-score	33
BMD	28
Weight	24
Height	23
Motor ability	10
Lifestyle	
Alcohol consumption	13
Smoking	11
Physical activity	10
Lack of physical exercise	7
Daily activities	5
Limited physical activity	4
Frequent sun exposure	3
Comorbidity	
Past fracture history	35
Other chronic diseases	20
Osteoporosis	8
Rheumatoid arthritis	7
Genetic risk score	5
Fracture type	4
Backache	2
Drug and nutrient intake	
Use of hormonal drugs	8
Calcium intake	8
Nutritional status	6
Intake of other drugs	4
Radiomics	
Radiomic data	25
Mental state	
Cognitive performance	3
Anxiety/depression	2

Note: BMD (g/cm^2).
BMD, bone mineral density.

these studies, we extracted 57 estimates of balanced accuracy (the average of the reported sensitivity and specificity), ranging from 0.46 to 1.00. As presented in table 3, the mean sensitivity and specificity of models were 0.79 (95% CI: 0.72, 0.84) (I^2=99.2%, p<0.001) and 0.81 (95% CI: 0.75, 0.86) in the training set (I^2=99.9%, p<0.001) and 0.76 (95% CI: 0.80, 0.81) (I^2=98.9%, p<0.001) and 0.83 (95% CI: 0.72, 0.90) in the validation set (I^2=99.9%, p<0.001), respectively. The results of sensitivity analysis show that ML models built for different fracture sites have stable performance in the training and validation sets (online supplemental figures S3–S8).

DISCUSSION

ML is a popular research method that provides new tools for early detection of diseases. This study systematically explored the application of the latest ML methods in predicting fracture risk in osteoporosis. The most commonly used predictors in ML models are age, BMI, past fracture history, BMD T-score, history of falls, BMD, radiomics data, weight, height, gender and other chronic diseases. In general, most predictors included in model development studies are traditional risk factors. A recent study showed that the most common risk factors for fragility fractures encompassed decreased BMD, age, gender, low BMI, history of fragility fractures, family history of hip fractures, history of glucocorticoid therapy, smoking, excessive alcohol consumption, lack of vitamin D, early menopause and immobility.[68] This is consistent with some common fracture predictors identified in our study. Our study also finds that radiomics data are frequently used as a fracture predictor in ML models for osteoporosis. A retrospective, single-centre, preliminary investigation by Lim et al[69] reported that ML based on radiomics features and abdomen–pelvic CT for diagnosing osteoporosis showed high predictive performance, with accuracy, specificity, and negative predictive value exceeding 93%.

The present study found that ML methods commonly used in the field of osteoporosis included logistic regression, decision tree, random forest, survival model, SVM, ensemble learning, artificial neural network (ANN), CNN and the latest deep learning technology. ML has a good performance in the prediction and identification of osteoporosis and fracture. In terms of the models in the training sets, the prediction efficiency of other deep learning method is optimal, followed by CNN, decision trees and logistic regression. In the validation sets, CNN showed the best performance, closely followed by other deep learning method, logistic regression and SVM. Deep learning is more powerful than traditional ML algorithms, with a wide range of coverage. Its performance increases with the amount of data.[70] Deep learning has been successfully applied to assist in the diagnosis and prediction of osteoporotic fractures.[34 63] CNN, a core algorithm of deep learning, is widely used in the field of data analysis and disease prediction with high accuracy.[71] CNN techniques can effectively predict the risk of osteoporotic fractures, enabling clinicians to take timely treatment measures, thereby reducing the occurrence of fractures.[19 23 47] Additionally, logistic regression is an efficient, simple and easy-to-operate ML method that outputs calibrated predicted
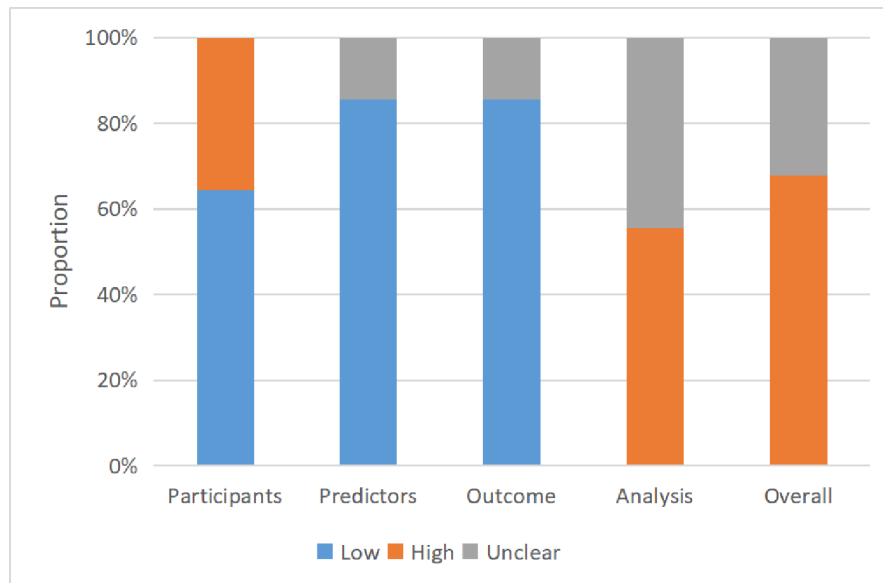
Figure 2 Risk of bias assessment (using Prediction Model Risk of Bias Assessment Tool) based on four domains across all machine learning models.

probabilities. An article on prediction models for the outcomes in patients with chronic obstructive pulmonary disease revealed that logistic regression was the most frequently used modelling method.[72] This is the same as recent findings reported by Silva *et al.*[73] Their ML models based on logistic regression outperformed those based on random forest and decision trees. Moreover, SVM adapts well to small samples and high-dimensional data with a

low misclassification rate, and therefore can be used for classification and regression analysis.[74]

Most included studies report multiple outcomes, such as sensitivity, specificity, AUC and ROC. The mean sensitivity of the models in the training set model is 0.79 (95% CI: 0.72, 0.84) greater than that of the models in the validation set. Most models are internally validated in the same population database and lack external validations

Table 2 Results of subgroup analysis of C-index by fracture site and machine learning type

Subgroup	Training dataset		Validation dataset	
	N	C-statistic (95% CI)	N	C-statistic (95% CI)
Fracture site				
Vertebral fracture	15	0.80 (0.74, 0.87)	6	0.87 (0.71, 1.00)
Hip fracture	20	0.76 (0.72, 0.81)	9	0.73 (0.65, 0.81)
Multi-site fracture	31	0.70 (0.67, 0.72)	17	0.71 (0.65, 0.76)
Model type				
LR	26	0.75 (0.72, 0.78)	7	0.80 (0.73, 0.87)
ANN	4	0.73 (0.64, 0.82)	3	0.66 (0.62, 0.70)
CNN	2	0.95 (0.94, 0.96)	1	0.98 (0.94, 1.00)
RF	3	0.70 (0.68, 0.72)	3	0.66 (0.59, 0.73)
SVM	5	0.72 (0.60, 0.85)	3	0.78 (0.59, 0.96)
DT	2	0.78 (0.56, 0.99)	1	0.69 (0.67, 0.70)
NB	2	0.74 (0.39, 1.00)	–	
kNN	1	0.51 (0.46, 0.55)	–	
Survival model	13	0.70 (0.69, 0.74)	9	0.68 (0.67, 0.69)
Boosted tree	5	0.71 (0.69, 0.74)	3	0.70 (0.69, 0.71)
Ensemble learning	1	0.72 (0.71, 0.73)		
Other DL	2	0.97 (0.96, 0.97)	1	0.82 (0.77, 0.87)
Overall	66	0.75 (0.72, 0.78)	32	0.75 (0.71, 0.78)

ANN, artificial neural network; CNN, convolutional neural network; DL, deep learnimg model; DT, decision tree; kNN, k-nearest neighbour; LR, logistic regression; NB, Naive Bayes; RF, random forests; SVM, support vector machine.

Table 3 Results of subgroup analysis of sensitivity and specificity by fracture site and machine learning type

Subgroup	Training dataset			Validation dataset		
	N	Sensitivity (95% CI)	Specificity (95% CI)	N	Sensitivity (95% CI)	Specificity (95% CI)
Fracture site						
Vertebral fracture	10	0.73 (0.61, 0.82)	0.91 (0.86, 0.95)	3	0.87 (0.70, 0.95)	0.97 (0.94, 0.98)
Hip fracture	13	0.90 (0.82, 0.94)	0.82 (0.75, 0.88)	5	0.84 (0.77, 0.89)	0.85 (0.80, 0.89)
Multi-site fracture	18	0.71 (0.59, 0.81)	0.72 (0.60, 0.81)	8	0.66 (0.61, 0.70)	0.69 (0.53, 0.81)
Model type						
LR	17	0.70 (0.63, 0.77)	0.73 (0.67, 0.79)	4	0.66 (0.55, 0.75)	0.65 (0.50, 0.77)
ANN	4	0.91 (0.70, 0.98)	0.93 (0.75, 0.98)	3	0.78 (0.71, 0.83)	0.85 (0.71, 0.93)
CNN	3	0.83 (0.81, 0.84)	0.91 (0.79, 0.96)	1	0.98	0.95
RF	1	0.84	0.91	1	0.70	0.46
SVM	6	0.81 (0.63, 0.92)	0.63 (0.13, 0.95)	3	0.79 (0.72, 0.85)	0.89 (0.79, 0.94)
DT	2	0.97 (0.53, 1.00)	0.70 (0.67, 0.73)	–		
NB	2	0.63 (0.13, 0.95)	0.76 (0.70, 0.81)	–		
kNN	2	0.95 (0.39, 1.00)	0.80 (0.77, 0.83)	1	0.81	0.79
Survival model	1	0.81	0.52	–		
Boosted tree	1	0.59	0.67	1	0.70	0.95
Other DL	2	0.81 (0.72, 0.87)	0.96 (0.93, 0.98)	2	0.83 (0.74, 0.90)	0.95 (0.92, 0.97)
Overall	41	0.79 (0.72, 0.84)	0.81 (0.75, 0.86)	16	0.76 (0.80, 0.81)	0.83 (0.72, 0.90)

ANN, artificial neural network; CNN, convolutional neural network; DL, deep learnimg model; DT, decision tree; kNN, k-nearest neighbour; LR, logistic regression; NB, Naive Bayes; RF, random forests; SVM, support vector machine.

in other populations. Only ML models in six articles were externally validated.[32 34 36 48 55 67] Therefore, external validations of ML models for predicting fracture risk are needed. However, a single performance measure such as AUC or ROC is insufficient to recommend the application of ML models into clinical practice,[8] and multiple measures of performance should be combined.

This systematic review and meta-analysis summarised a large number of studies to comprehensively evaluate the predictive performance of ML for fracture risk in patients with osteoporosis. The characteristics of the established and validated models were described. We performed a quantitative synthesis that was never done in previous studies to compare these models. Furthermore, the meta-analysis of C-index was performed using the random-effects model, since the C-index was reported in most predictive models.[72 75] Meanwhile, the bivariate mixed-effects model was used for the meta-analysis of sensitivity and specificity. In the training dataset, the sensitivity of hip fracture was the highest, closely followed by multi-site fracture and vertebral fracture. For patients with hip fractures, radiographs may cause missed diagnosis and misdiagnosis, leading to poor prognosis.[76] ML models have been increasingly used to identify hip fracture risk with high accuracy.[31] ML has a stronger power to recognise images and can assist inexperienced clinicians in a highly accurate diagnosis.

Some limitations still need to be considered in the present study. Due to incomplete reporting of indicators, several studies were only included in the systematic review and were excluded from subsequent meta-analyses.[61 63]

Studies conducted in either Western or Asian populations lack external validation, and thus external validations in other populations are needed to widen the application of ML models. The risk of bias assessment demonstrated that most studies (67%) had a high risk of bias, regardless of whether they involved the development or external validation of a prediction model for the osteoporosis population. The main bias came from the statistical analysis because most studies did not properly handle continuous and categorical variables and reported no method for processing missing values. Only three articles reported the use of median imputation or multiple interpolation method to deal with missing values,[15 31 66] while others did not mention how to deal with missing values. These shortcomings in the methodology may be due to a lack of guidelines for the standard reporting of risk prediction studies at that time. In addition, some models were reported with little information, making it unable for other researchers to perform external validation, much less the application in clinical practice. For example, only 12 articles used the K-fold cross-validation method to improve the accuracy of their algorithms,[15 16 19 27 30 31 34 36 42 46 47 62] but most of the eligible articles did not. Models without stringent validation cannot be widely applied.[73] Many studies have limited applicability in clinical practice because of flawed methodologies or unrepresentative data sets. Future research should give priority to the development of practical algorithms. Furthermore, we observed large heterogeneity in the meta-analysis of C-statistics. Potential sources of heterogeneity may be the differences in patients' characteristics, data sources and analysis methods across

the studies. More than 30% of the research data came from clinical studies, and clinical data are heterogeneous and usually imbalanced. At last, most ML models did not report balanced accuracy and lacked calibration or external validation or decision curves. Thus, further research is required to address these issues, improving the generalisation of the models.

Despite the limitations mentioned above, the present study can still provide meaningful recommendations for future research and practice. First, the major strength of our study is the rigorous literature search and methodology to provide reliable estimates. This is the latest systematic review and meta-analysis conducted to comprehensively assess ML models for fracture risk. Second, ML models can provide convincing evidence to assist clinicians in making more accurate judgments during highly complex decision-making processes, with certain clinical application values.[74] More rigorous, robust and comprehensive research is warranted to assess its clinical application and impact on clinicians and patients. Third, the advances in emerging technologies such as ML have opened a new era of clinical medical research, providing new directions for solving intricate problems with classical statistical methods. However, clinicians currently are not skillful in using such emerging technologies. Therefore, it is advisable for clinicians to improve their ability to use ML to make more accurate diagnoses.

In conclusion, ML has a favourable predictive performance for fracture risk in patients with osteoporosis and can be used as a potential tool for early identification of fracture risk in this population. However, most current studies lack external validation. Therefore, future research is needed to validate and improve the existing predictive models for osteoporosis risk rather than developing new models.

ORCID iD
Yanqian Wu http://orcid.org/0000-0002-0421-6111

REFERENCES

1 Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. WHO study group. *Osteoporos Int* 1994;4:368–81.

2 Veronese N, Kolk H, Maggi S. Epidemiology of fragility fractures and social impact. In: Falaschi P, Marsh D, eds. *Orthogeriatrics: The Management of Older Patients with Fragility Fracture*. Cham (CH): Springer, 2021: 19–34.

3 Piscitelli P, Feola M, Rao C, et al. Ten years of hip fractures in Italy: for the first time a decreasing trend in elderly women. *World J Orthop* 2014;5:386–91.

4 Borgström F, Karlsson L, Ortsäter G, et al. Fragility fractures in Europe: burden, management and opportunities. *Arch Osteoporos* 2020;15:59.

5 Zimmermann EA, Busse B, Ritchie RO. The fracture mechanics of human bone: influence of disease and treatment. *Bonekey Rep* 2015;4:743.

6 Gupta R, Srivastava D, Sahu M, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers* 2021;25:1315–60.

7 Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. *Curr Opin Rheumatol* 2019;31:368–75.

8 Smets J, Shevroja E, Hügle T, et al. Machine learning solutions for osteoporosis-a review. *J of Bone & Mineral Res* 2021;36:833–51. 10.1002/jbmr.4292 Available: https://asbmr.onlinelibrary.wiley.com/toc/15234681/36/5

9 Anam M, a/p Ponnusamy V, Hussain M, et al. Osteoporosis prediction for trabecular bone using machine learning: A review. *Computers, Materials & Continua* 2021;67:89–105.

10 Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6:e1000100.

11 Palazón-Bru A, Martín-Pérez F, Mares-García E, et al. A general presentation on how to carry out a CHARMS analysis for prognostic multivariate models. *Stat Med* 2020;39:3207–25.

12 Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020:m689.

13 Debray TP, Damen JA, Riley RD, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res* 2019;28:2768–86.

14 Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the C-Statistic and calibration measures? *Stat Methods Med Res* 2018;27:3505–22.

15 Wu Q, Nasoz F, Jung J, et al. Machine learning approaches for fracture risk assessment: a comparative analysis of genomic and phenotypic data in 5130 older men. *Calcif Tissue Int* 2020;107:353–61.

16 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.

17 van Geel TACM, Nguyen ND, Geusens PP, et al. Development of a simple prognostic nomogram for individualising 5-year and 10-year absolute risks of fracture: a population-based prospective study among postmenopausal women. *Annals of the Rheumatic Diseases* 2011;70:92–7.

18 Ulivieri FM, Rinaudo L, Piodi LP, et al. Bone strain index as a predictor of further vertebral fracture in osteoporotic women: an artificial intelligence-based analysis. *PLoS ONE* 2021;16:e0245967.

19 Yoda T, Maki S, Furuya T, et al. Automated differentiation between osteoporotic vertebral fracture and malignant vertebral fracture on MRI using a deep convolutional neural network. *Spine* 2022;47:E347–52.

20 Jiang X, Westermann LB, Galleo GV, et al. Age as a predictor of osteoporotic fracture compared with current risk-prediction models. *Obstet Gynecol* 2013;122:1040–6.

21 Schousboe JT, Rosen HR, Vokes TJ, *et al*. Prediction models of prevalent radiographic vertebral fractures among older women. *J Clin Densitom* 2014;17:378–85.

22 Sandhu SK, Nguyen ND, Center JR, *et al*. Prognosis of fracture: evaluation of predictive accuracy of the FRAX algorithm and garvan nomogram. *Osteoporos Int* 2010;21:863–71.

23 Rubin KH, Möller S, Holmberg T, *et al*. A new fracture risk assessment tool (FREM) based on public health registries. *J of Bone & Mineral Res* 2018;33:1967–79.

24 Pluskiewicz W, Adamczyk P, Franek E, *et al*. Ten-year probability of osteoporotic fracture in 2012 Polish women assessed by FRAX and nomogram by nguyen et al.-conformity between methods and their clinical utility. *Bone* 2010;46:1661–7.

25 Jang EJ, Lee Y-K, Choi HJ, *et al*. Osteoporotic fracture risk assessment using bone mineral density in Korean: a community-based cohort study. *J Bone Metab* 2016;23:34.

26 Monchka BA, Kimelman D, Lix LM, *et al*. Feasibility of a generalized convolutional neural network for automated identification of vertebral compression fractures: the manitoba bone mineral density registry. *Bone* 2021;150:116017.

27 Mehta SD, Sebro R. Computer-aided detection of incidental lumbar spine fractures from routine dual-energy X-ray absorptiometry (DEXA) studies using a support vector machine (SVM). *J Digit Imaging* 2020;33:204–10.

28 Langsetmo L, Nguyen TV, Nguyen ND, *et al*. Independent external validation of nomograms for predicting risk of low-trauma fracture and hip fracture. *CMAJ* 2011;183:E107–14.

29 Ioannidis G, Jantzi M, Bucek J, *et al*. Development and validation of the fracture risk scale (FRS) that predicts fracture over a 1-year time period in Institutionalised frail older people living in Canada: an electronic record-linked longitudinal cohort study. *BMJ Open* 2017;7:e016477.

30 Nishiyama KK, Macdonald HM, Hanley DA, *et al*. Women with previous fragility fractures can be classified based on bone microarchitecture and finite element analysis measured with HR-pQCT. *Osteoporos Int* 2013;24:1733–40.

31 Kruse C, Eiken P, Vestergaard P. Machine learning principles can improve hip fracture prediction. *Calcif Tissue Int* 2017;100:348–60.

32 Kolanu N, Brown AS, Beech A, *et al*. Natural language processing of Radiology reports for the identification of patients with fracture. *Arch Osteoporos* 2021;16.

33 Kim HY, Jang EJ, Park B, *et al*. Development of a Korean fracture risk score (KFRS) for predicting osteoporotic fracture risk: analysis of data from the Korean national health insurance service. *PLoS ONE* 2016;11:e0158918.

34 Hsieh C-I, 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.

35 Hong N, Park H, Kim CO, *et al*. Bone radiomics score derived from DXA hip images enhances hip fracture prediction in older women. *J Bone Miner Res* 2021;36:1708–16.

36 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.

37 Henry MJ, Pasco JA, Merriman EN, *et al*. Fracture risk score and absolute risk of fracture. *Radiology* 2011;259:495–501.

38 Galassi A, Martín-Guerrero JD, Villamor E, *et al*. Risk assessment of hip fracture based on machine learning. *Appl Bionics Biomech* 2020;2020:8880786.

39 FitzGerald G, Compston JE, Chapurlat RD, *et al*. Empirically based composite fracture prediction model from the global longitudinal study of osteoporosis in postmenopausal women (GLOW). *J Clin Endocrinol Metab* 2014;99:817–26.

40 Ferizi U, Besser H, Hysi P, *et al*. Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data. *Magnetic Resonance Imaging* 2019;49:1029–38. 10.1002/jmri.26280 Available: https://onlinelibrary.wiley.com/toc/15222586/49/4

41 Enns-Bray WS, Bahaloo H, Fleps I, *et al*. Biofidelic finite element models for accurately classifying hip fracture in a retrospective clinical study of elderly women from the AGES reykjavik cohort. *Bone* 2019;120:25–37.

42 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.

43 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.

44 Cheung EYN, Bow CH, Cheung CL, *et al*. Discriminative value of FRAX for fracture prediction in a cohort of Chinese postmenopausal women. *Osteoporos Int* 2012;23:871–8.

45 Chanplakorn P, Lertudomphonwanit T, Daraphongsataporn N, *et al*. Development of prediction model for osteoporotic vertebral compression fracture screening without using clinical risk factors, compared with FRAX and other previous models. *Arch Osteoporos* 2021;16.

46 Bredbenner TL, Mason RL, Havill LM, *et al*. Fracture risk predictions based on statistical shape and density modeling of the proximal Femur. *J Bone Miner Res* 2014;29:2090–100.

47 Beyaz S, Açıcı K, Sümer E. Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. *Jt Dis Relat Surg* 2020;31:175–83.

48 Berry SD, Zullo AR, Lee Y, *et al*. Fracture risk assessment in long-term care (frail): development and validation of a prediction model. *J Gerontol A Biol Sci Med Sci* 2018;73:763–9.

49 Beaudoin C, Jean S, Moore L, *et al*. Prediction of osteoporotic fractures in elderly individuals: a derivation and internal validation study using healthcare administrative data. *J Bone Miner Res* 2021;36:2329–42.

50 Baleanu F, Moreau M, Charles A, *et al*. Fragility fractures in postmenopausal women: development of 5-year prediction models using the FRISBEE study. *J Clin Endocrinol Metab* 2022;107:e2438–48.

51 Almog YA, Rai A, Zhang P, *et al*. Deep learning with electronic health records for short-term fracture risk identification: crystal bone algorithm development and validation. *J Med Internet Res* 2020;22:e22550.

52 Zagórski P, Tabor E, Martela-Tomaszek K, *et al*. Five-year fracture risk assessment in postmenopausal women, using both the POL-RISK calculator and the garvan nomogram: the silesia osteo active study. *Arch Osteoporos* 2021;16.

53 Díez-Pérez A, González-Macías J, Marín F, *et al*. Prediction of absolute risk of non-spinal fractures using clinical risk factors and heel quantitative ultrasound. *Osteoporos Int* 2007;18:629–39.

54 Lix LM, Leslie WD, Majumdar SR. Measuring improvement in fracture risk prediction for a new risk factor: a simulation. *BMC Res Notes* 2018;11.

55 Li Q, Long X, Wang Y, *et al*. Development and validation of a nomogram for predicting the probability of new vertebral compression fractures after vertebral augmentation of osteoporotic vertebral compression fractures. *BMC Musculoskelet Disord* 2021;22.

56 Lee S, Lee JW, Jeong J-W, *et al*. A preliminary study on discrimination of osteoporotic fractured group from nonfractured group using support vector machine. *Annu Int Conf IEEE Eng Med Biol Soc* 2008;2008:474–7.

57 Jacobs JWG, Da Silva JAP, Armbrecht G, *et al*. Prediction of vertebral fractures is specific for gender and site of bone mineral density measurement. *J Rheumatol* 2010;37:149–54.

58 Eller-Vainicher C, Chiodini I, Santi I, *et al*. Recognition of morphometric vertebral fractures by artificial neural networks: analysis from GISMO Lombardia database. *PLoS ONE* 2011;6:e2727711.

59 Zhong B-Y, He S-C, Zhu H-D, *et al*. Risk prediction of new adjacent vertebral fractures after PVP for patients with vertebral compression fractures: development of a prediction model. *Cardiovasc Intervent Radiol* 2017;40:277–84.

60 Xiao X, Wu Q. The utility of genetic risk score to improve performance of FRAX for fracture prediction in US postmenopausal women. *Calcif Tissue Int* 2021;108:746–56.

61 Du J, Wang J, Gai X, *et al*. Application of intelligent X-ray image analysis in risk assessment of osteoporotic fracture of femoral neck in the elderly. *Math Biosci Eng* 2023;20:879–93.

62 Wang M, Chen X, Cui W, *et al*. A computed tomography-based radiomics nomogram for predicting osteoporotic vertebral fractures: a longitudinal study. *J Clin Endocrinol Metab* 2023;108:e283–94.

63 Dong Q, Luo G, Lane NE, *et al*. Deep learning classification of spinal osteoporotic compression fractures on radiographs using an adaptation of the genant semiquantitative criteria. *Academic Radiology* 2022;29:1819–32.

64 Wen Z, Mo X, Zhao S, *et al*. Study on risk factors of primary non-traumatic OVCF in Chinese elderly and a novel prediction model. *Orthop Surg* 2022;14:2925–38.

65 Pluskiewicz W, Adamczyk P, Werner A, *et al*. POL-RISK: an algorithm for 10-year fracture risk prediction in postmenopausal women from the RAC-OST-POL study. *Pol Arch Intern Med* 2023;133:16395.

66 Kong X, Zhao Z, Zhang D, *et al*. Major osteoporosis fracture prediction in type 2 diabetes: a derivation and comparison study. *Osteoporos Int* 2022;33:1957–67.

67 Agarwal A, Baleanu F, Moreau M, *et al*. External validation of FRISBEE 5-year fracture prediction models: a registry-based cohort study. *Arch Osteoporos* 2022;18:13.

68 Pisani P, Renna MD, Conversano F, *et al*. Major osteoporotic fragility fractures: risk factor updates and societal impact. *WJO* 2016;7:171.

69 Lim HK, Ha HI, Park S-Y, *et al*. Prediction of femoral osteoporosis using machine-learning analysis with radiomics features and abdomen-pelvic CT: a retrospective single center preliminary study. *PLoS ONE* 2021;16:e0247330.

70 Avci O, Abdeljaber O, Kiranyaz S, *et al*. A review of vibration-based damage detection in civil structures: from traditional methods to machine learning and deep learning applications. *Mech Syst Signal Process* 2021;147:107077.

71 Shen S-Y, Peña Fernández M, Tozzi G, *et al*. Deep learning approach to assess damage mechanics of bone tissue. *J Mech Behav Biomed Mater* 2021;123:104761.

72 Bellou V, Belbasis L, Konstantinidis AK, *et al*. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ* 2019;367:l5358.

73 Silva KD, Lee WK, Forbes A, *et al*. Use and performance of machine learning models for type 2 diabetes prediction in community settings: a systematic review and meta-analysis. *Int J Med Inform* 2020;143:104268.

74 Jain R, Sontisirikit S, Iamsirithaworn S, *et al*. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infect Dis* 2019;19:272.

75 Fleuren LM, Klausch TLT, Zwager CL, *et al*. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020;46:383–400.

76 Hossain M, Akbar SA, Andrew G. Misdiagnosis of occult hip fracture is more likely in patients with poor mobility and cognitive impairment. *Acta Orthop Belg* 2010;76:341–6.