scientific reports

OPEN

Prediction of osteoporosis in patients with rheumatoid arthritis using machine learning

Chaewon Lee[1], Gihun Joo[2], Seunghun Shin[2], Hyeonseung Im[1,2,3,6✉] & Ki Won Moon[2,4,5,6✉]

Osteoporosis is a serious health concern in patients with rheumatoid arthritis (RA). Machine learning (ML) models have been increasingly incorporated into various clinical practices, including disease classification, risk prediction, and treatment response. However, only a few studies have focused on predicting osteoporosis using ML in patients with RA. We aimed to develop an ML model to predict osteoporosis using a representative Korean RA cohort database. The KORean Observational study Network for Arthritis (KORONA) database, established by the Clinical Research Center for RA in Korea, was used in this study. Among the 5077 patients registered in KORONA, 2374 patients were included in this study. Four representative ML algorithms were used for the prediction: logistic regression (LR), random forest, XGBoost (XGB), and LightGBM. The accuracy, F1 score, and area under the curve (AUC) of each model were measured. The LR model achieved the highest AUC value at 0.750, while the XGB model achieved the highest accuracy at 0.682. Body mass index, age, menopause, waist and hip circumferences, RA surgery, and monthly income were risk factors of osteoporosis. In conclusion, ML algorithms are a useful option for screening for osteoporosis in patients with RA.

Osteoporosis is a serious health problem in patients with rheumatoid arthritis (RA). The prevalence of osteoporosis in patients with RA is approximately two times higher than that in the general population[1]. A cohort study reported that osteoporosis was present in approximately 30% of patients with RA, whereas the proportion of age- and sex-matched controls with osteoporosis was 17%[2]. Bone fragility in RA results from complex mechanisms, such as inflammatory cytokine production, osteoclast activation, and use of glucocorticoids for RA treatment. Therefore, it is crucial to detect osteoporosis in patients with RA to prevent osteoporotic fractures. Bone mineral density (BMD) measurement using dual-energy X-ray absorptiometry (DXA) is a standard diagnostic tool for osteoporosis. However, a significant number of patients with RA do not undergo a BMD measurement test in clinical practice. According to the Consortium of Rheumatology Researchers of North America registry, only 11% of patients underwent DXA during the first year of follow-up[3]. It would be thus useful to predict osteoporosis in patients with RA who are at high risk. Recently, machine learning (ML) algorithms have been extensively applied to various clinical practices, including disease classification, risk prediction, and treatment response. Although studies on the prediction of osteoporosis or fractures using ML models in the general population have been conducted, those on patients with RA are lacking. Therefore, we aimed to develop an ML model to predict osteoporosis using a representative Korean RA cohort database. This study aimed to predict osteoporosis in patients with RA using clinical data. The contributions of our paper are summarized as follows:

- Our study is the first to investigate osteoporosis prediction models in patients with RA.
- We apply the four representative ML algorithms and show that their prediction performance is comparable to previous studies on the general population, confirming the effectiveness of ML for osteoporosis prediction in patients with RA.
- By considering 83 features and their importance, we found new predictive factors like socioeconomic status, including monthly income and education, which have not been considered well before.

[1]Department of Convergence Security, Kangwon National University, Chuncheon, South Korea. [2]Interdisciplinary Graduate Program in Medical Bigdata Convergence, Kangwon National University, Chuncheon, South Korea. [3]Department of Computer Science and Engineering, Kangwon National University, Chuncheon, South Korea. [4]Division of Rheumatology, Department of Internal Medicine, Kangwon National University Hospital, Chunchoen, South Korea. [5]Department of Internal Medicine, Kangwon National University School of Medicine, 1 Kangwondaehak-gil, Chuncheon 24341, South Korea. [6]These authors contributed equally: Hyeonseung Im and Ki Won Moon. ✉email: hsim@kangwon.ac.kr; kiwonmoon@kangwon.ac.kr

Methods

Data source and participants

The data used in this study were obtained from the KORean Observational study Network for Arthritis (KORONA) database established by the Clinical Research Center for RA in Korea[4]. KORONA includes a cohort of patients with RA recruited between July 2009 and Mar 2012. Individuals with RA who fulfilled the 1987 American College of Rheumatology classification criteria for RA and who were older than 18 years were enrolled at 23 centers in South Korea. A total of 5077 patients (4327 women, 750 men) were registered at baseline and underwent annual follow-up evaluations. All participants were provided informed consent prior to enrollment in the study. Among them, 1758 patients who had never undergone DXA were excluded. The dataset consisted of clinical information at initial enrollment as well as annual follow-up data for 5 years. Missing values of continuous features for each patient were replaced with their mean values across all follow-up data, and missing values of categorical features for each patient were replaced with their most frequent values across all follow-up data. When there was no most frequent value, that patient's data was excluded. After this preprocessing, the total number of patients considered in the study was 2374 (Fig. 1), of which 2118 patient records were imputed.

BMD measurement

BMD measurements were conducted using the Hologic QDR (Waltham, MA, USA) and GE Lunar Prodigy (Madison, WI, USA) systems in accordance with the standard scan and positioning protocols of the manufacturers. Osteoporosis was defined as an L-spine total or hip total value of -2.5 or less, in accordance with the World Health Organization (WHO) classifications[5]. We classified patients with RA into osteoporosis and control groups based on their BMD values.

Feature selection

In the KORONA database, there are over 1000 features. Principal component analysis (PCA) was performed to prevent multicollinearity. PCA was performed by grouping continuous features with correlations of ≥ 0.7. The distribution of some representative variables obtained by PCA is shown in Supplementary Fig. 1. Continuous variables were normalized using min–max normalization. Finally, 83 features were selected for the development of the prediction models. A list of selected features is presented in Supplementary Table S1.

Model development

Prediction was performed using four representative ML algorithms: logistic regression (LR), random forest (RF)[6], XGBoost (XGB)[7], and LightGBM (LGBM)[8]. LR is a model that uses regression to predict the probability of data falling into a category and classifies it as belonging to a more likely category. We used LR as the baseline method. RF is a bagging-based ensemble model composed of multiple decision trees and uses the most common value among the predicted values made by those decision trees as the final prediction. RF is robust and interpretable, and generally avoids overfitting, but is computationally expensive. XGB is an ensemble algorithm based on gradient boosting that uses a combination of multiple decision trees. It is faster than existing boosting models because it enables parallel learning on a general boosting model that combines multiple models. It is also generally more efficient than RF, though it tends to be more prone to overfitting. LGBM uses a leaf-centered tree-splitting method instead of the general balanced tree-splitting method used in gradient boosting machine models. It is typically faster and more memory-efficient than XGB. However, it is more susceptible to overfitting and can be harder to interpret.

For hyperparameter tuning, we employed GridSearchCV for LR and RandomizedSearchCV for RF, XGB, and LGBM, using fivefold cross-validation[9]. Specifically, the dataset was randomly partitioned into five subsets for each hyperparameter combination and model. In each iteration, one subset was used for testing while the
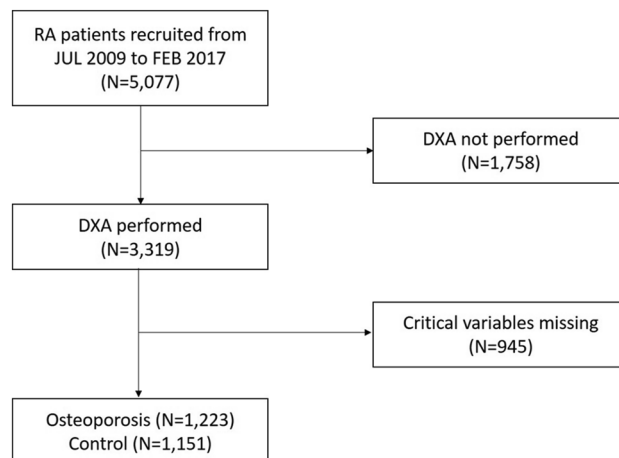

Figure 1. Flow diagram of study population.

remaining four as the training set. Each model underwent five evaluations, each time utilizing a different subset as the test set. The average result from the five iterations was then used as the final performance for that particular hyperparameter combination. Finally, we selected the optimal hyperparameter combination for each model and re-evaluated their performance using another fivefold cross-validation for comparison.

We tuned the following hyperparameters. For the LR model, the regularization intensity (C) was adjusted. For the RF model, max_depth (maximum tree depth), n_estimators (number of decision trees), min_samples_split (minimum number of samples required to split a node), and min_samples_leaf (minimum number of samples required to form a leaf note) were adjusted. For the XGB model, eta (learning rate), gamma (minimum loss reduction necessary to create additional partitions at the leaf nodes of the tree), max_depth, subsample (sub-sample ratio of training instances), and colsample_bytree (subsample ratio of columns for building each tree) were adjusted. For the LGBM model, learning rate, max_depth, subsample, and colsample_bytree were adjusted.

The accuracy, F1 score, and area under the receiver operating characteristic (ROC) curve (AUC) of each model were measured using fivefold cross-validation as mentioned above. The AUC was primarily used to compensate for the limitations of the accuracy when class distributions were different. The F1 score is the harmonic mean of precision and recall and is often used when the class distribution is imbalanced.

Statistical analysis

Continuous and categorical variables are presented as mean ± standard deviation and number (%), respectively. Baseline characteristics between the two groups were compared using Student's t-test for continuous variables and the chi-square test for categorical variables. Statistical significance was defined as a p-value of < 0.05. The default value of 0.5 was used for the threshold of the F1 score. Statistical analyses were performed using Python 3.8.8.

Ethical consideration

This study was conducted in compliance with the World Medical Association Declaration of Helsinki and approved by the Ethical Review Board of the Kangwon National University Hospital (IRB approval No. KNUH-2022-03-021).

Results

Demographic and clinical features

Baseline clinical characteristics of both groups are shown in Table 1. Compared with the control group, the proportion of women and age were higher, BMI was lower, and previous fracture history was higher in the osteoporosis group. The proportion of current smokers and drinkers was higher in the control group. There was no statistically significant difference in oral glucocorticoid use between the two groups.

Model evaluation

The AUC, accuracy, and F1 score of each model were measured using fivefold cross-validation after selecting an optimal hyperparameter combination for each model using either GridSearchCV or RandomizedSearchCV. The performance of each model in predicting osteoporosis using 83 clinical features is shown in Table 2, with 95%

Variables	Control (N = 1151)	Osteoporosis (N = 1223)	P value
Female sex	1024 (89)	1146 (94)	< 0.05
Age (years)	54.81 ± 1.05	62.45 ± 0.9	< 0.05
Body mass index (kg/m^2)	23.09 ± 0.29	22.09 ± 0.2	< 0.05
Previous fracture	117 (10)	243 (20)	< 0.05
Family history of RA	98 (9)	108 (9)	0.84
Current smoking	85 (7)	51 (4)	< 0.05
Current drinking	221 (19)	193 (12)	< 0.05
Oral glucocorticoid use	837 (73)	921 (75)	0.16

Table 1. Baseline characteristics for the control and osteoporosis groups. Values denote number (%) or mean ± standard deviation unless stated otherwise.

Model	AUC	Accuracy	F1 score
Logistic regression	0.750 (0.717–0.783)	0.681	0.700
Random forest	0.747 (0.715–0.781)	0.681	0.705
XGBoost	0.749 (0.72–0.779)	0.682	0.700
LightGBM	0.744 (0.715–0.772)	0.678	0.693

Table 2. Performance of each prediction model using fivefold cross-validation, with 95% confidence intervals for the AUC.

confidence intervals for the AUC. The AUC was highest in the LR model at 0.750, the accuracy was highest in the XGB model at 0.682, and the F1 score was highest in the RF model at 0.70, although the difference between the models was not significant. Figure 2 shows the ROC curve for each ML model.

Intraclass correlation coefficient (ICC)

We also calculated the ICC[10] to see the prediction consistency between the ML models for each patient. We considered the ML models as a fixed set of raters, their prediction probabilities as ratings, and the patients as targets. The ICC value ranges from 0 to 1, with values between 0.75 and 0.9 indicating good reliability and values greater than 0.90 indicating excellent reliability[10]. Figure 3 shows the prediction probabilities of each ML model for a sample of 30 out of 2374 patients, together with the ICC (3,1) value of 0.909 (95% CI 0.904–0.915), which indicates that the ML models are highly consistent and reliable.

Feature importance of each model

The feature importance of each model was calculated based on the fold with the highest AUC among the five folds used for cross-validation. The LR model was not considered because it was solely used as a benchmark to evaluate the efficacy of tree-based prediction models and did not fully resolve the issue of multicollinearity among the input features. The 20 most important features of the XGB model are shown in Fig. 4. The top 20 most important features of the other models are shown in Supplementary Figs. 2 and 3.

Discussion

We developed and evaluated ML models to predict osteoporosis risk in patients with RA. Osteoporotic fracture is a serious health issue in patients with RA. A British study showed that the relative risk (RR) of hip fracture of patients with RA was 2.0 (95% CI 1.8–2.3) and that of vertebral fracture was 2.4 (95% CI 2.0–2.8)[11]. These
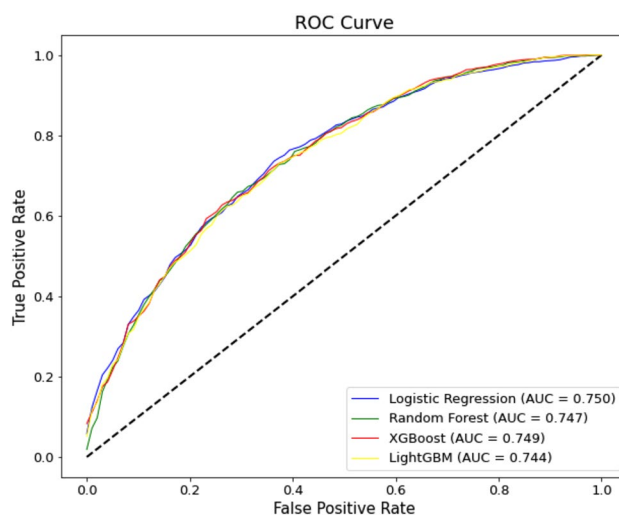

Figure 2. Receiver operating characteristic (ROC) curves for the prediction models.
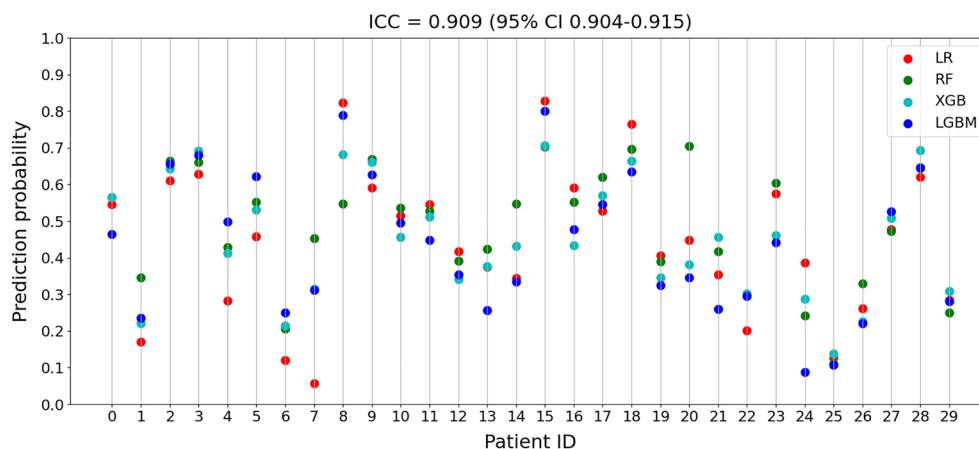

Figure 3. Intraclass correlation of part of the dataset using ICC (3, 1).
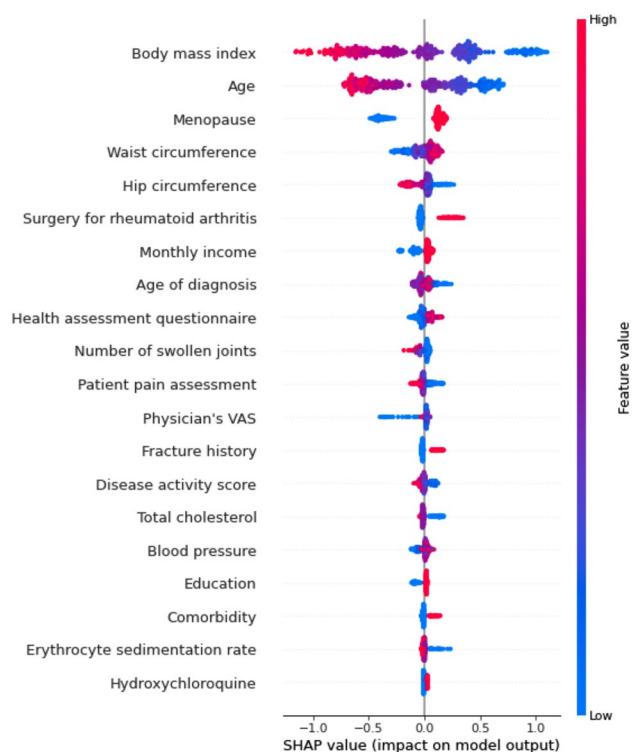
Figure 4. Top 20 features derived from the XGBoost model.

fragility fractures impair quality of life and increase healthcare costs and mortality rates. Korean national health claims data reported that the incidence of osteoporotic fractures is higher in patients with RA than in the general population over 3 years (17.4% vs. 11.6%), and the standardized mortality rate was 1.4 times higher in men and 1.3 times in women[12]. However, timely screening and appropriate management of osteoporosis remain inadequate in clinical practice. A retrospective study from the USA reported that only 30% of female patients with RA underwent DXA during 4.4 years[13]. A French study showed that DXA was performed in 21.6% of 1008 patients with RA, and approximately 12% of patients received osteoporosis medication[14]. A large observational study from the USA of 11,669 patients with RA followed from 2003 to 2014 revealed that approximately 50% of patients who required treatment did not receive medication for osteoporosis[15]. Identification of osteoporosis in high-risk patients with RA is an important issue in clinical practice.

Many researchers have focused on developing osteoporosis prediction tools using ML for the general population. Table 3 shows the characteristics and results of studies on osteoporosis prediction in the general population. The AUC scores in these studies ranged between 0.710 and 0.854. The AUC score in our study was 0.750, which is a relatively good performance compared with other studies. Other studies reported that the best performing models were artificial neural networks, XGB, RF, and multilayer perception models. Our results showed that the LR model had the best performance, whereas the other models showed similar performances. To the best of our knowledge, no studies have been conducted on the prediction of osteoporosis in patients with RA until now. Recently, one study investigated the prediction of fractures with ML in elderly patients with RA. Chen et al.[16] reported an ML model to predict the risk of fractures in patients with elderly-onset RA. They collected

Study	Number of patients	Number of selected features	Best performance model	Best performance (AUC)
Shim et al.[20]	1792	9	ANN	0.743
Erjiang et al.[21]	13,577	30	XGB	0.833
Yang et al.[22]	5982	16/19	RF	0.843/0.811
Wang et al.[23]	1419	18	ANN	0.762
Iliou et al.[24]	3426	2	MLP	0.710
Park et al.[19]	3309	20	XGB	0.730/0.790
Bui et al.[25]	1951	15	RF	0.854
Our study	2374	83	RL	0.750

Table 3. Key studies on the osteoporosis prediction using machine learning. *ANN* Artificial neural network; *LR* Logistic regression, *MLP* Multi-layer perceptron, *RF* Random forest, *XGB* XGBoost.

Ranking	Erjiang et al.[21]	Park et al.[19]	Bui et al.[25]	Our study
1	Age	Age	Age	Body mass index
2	Weight	Menopause	Weight	Age
3	Bisphosphonate use	Alkaline phosphatase	Height	Menopause
4	Body mass index	Weight	Uric acid	Waist circumference
5	Denosumab use	Forced vital capacity	Calcium	Hip circumference
6	Estrogen use	Sobriety	Cholesterol	Surgery for RA
7	Chronic respiratory disease	Aspartate aminotransferase	Creatinine	Monthly income
8	Osteopenia	Diastolic blood pressure	Free thyroxine level	Age of diagnosis
9	Height loss	Riboflavin intake	Glucose	Health assessment questionnaire
10	Hormonal therapy	Weight control by exercise	HbA1c	Number of swollen joints

Table 4. Comparisons with the top 10 ranked features from other studies on osteoporosis prediction using machine learning.

data from 487 elderly patients with RA and developed various ML models for fracture risk. The AUC of each model was 0.713–0.872.

In RA, conventional risk factors for osteoporosis include female sex, smoking, old age, low BMI, menopause, diabetes, thyroid disorders, lack of physical activity, and glucocorticoid use[17]. Lee et al.[18] investigated risk factors for osteoporosis using a traditional statistical technique and the same database used in our study, the KORONA registry. In Lee et al.'s study, the number of patients with osteoporosis (n = 619) was smaller than that in our study (n = 1223) as enrolled patients were limited to postmenopausal women and men over 50 years old. The clinical characteristics were quite different between Lee et al.'s study and the present study. For example, in our study, the proportion of current smokers and drinkers was higher in the control group as this group included more male patients. We intended to develop a prediction model regardless of sex; therefore, we included both male and female patients. Lee et al. reported that older age, lower BMI, longer disease duration, higher cumulative glucocorticoid dose, and higher health assessment questionnaire scores were independent risk factors of osteoporosis[18]. However, our ML model suggested new features such as monthly income, education, surgical history, and marital status as predictive factors of osteoporosis. Interestingly, socioeconomic status, including monthly income and education, was selected as a predictive factor in addition to previously well-known conventional predictors. Another Korean study that developed an ML model using National Health and Nutrition Examination data showed that monthly income was an important predictive factor of osteoporosis[19]. Table 4 shows a comparison of the top 10 ranked features from other studies and our study. In most studies, age and BMI were identified as the highest-ranked predictive features. Menopause, bisphosphonates, and estrogen use were also ranked high. In addition to conventional risk factors, features including alkaline phosphatase, cholesterol, uric acid, and blood pressure are important for prediction. These findings provide new insights into the prediction of osteoporosis in clinical practice.

Our study had several strengths. First, we developed a prediction model for patients with RA who have not been covered in previous research. The performance of our model was comparable to that of previous studies on the general population. Second, our study used 83 clinical features for the development of the prediction models, which is much greater than the number of features used in other studies. We selected these features from more than 1000 features using PCA to avoid multicollinearity problems. However, our study has the following limitations. First, there was no separate dataset to validate this model. We could not find any RA patient cohort dataset that included BMD results. However, the KORONA data were collected from 23 separate hospitals in Korea, and heterogeneity existed in the dataset. Second, we could not develop a fracture-prediction model. Although clinical information on fractures was available, the number of fracture events was too small to create a prediction model.

In conclusion, we applied representative ML algorithms to predict osteoporosis using clinical data from patients with RA. By comparing with previous studies, we observed a comparable performance. ML methods have the potential to support practitioners in the detection of osteoporosis in patients with RA.

Data availability
Raw data were generated by the Clinical Research Center for RA in Korea. The data supporting the findings of this study are available from the corresponding authors upon request.

Code availability
The code for experiments is available at https://github.com/knu-plml/RA-osteoporosis-ml-prediction.

References
1. Haugeberg, G., Uhlig, T., Falch, J. A., Halse, J. I. & Kvien, T. K. Bone mineral density and frequency of osteoporosis in female patients with rheumatoid arthritis: Results from 394 patients in the Oslo county rheumatoid arthritis register. *Arthritis Rheum.* **43**, 522–530 (2000).
2. Hauser, B., Riches, P. L., Wilson, J. F., Horne, A. E. & Ralston, S. H. Prevalence and clinical prediction of osteoporosis in a contemporary cohort of patients with rheumatoid arthritis. *Rheumatology* **53**, 1759–1766 (2014).

3. Aizer, J., Reed, G., Onofrei, A. & Harrison, M. J. Predictors of bone density testing in patients with rheumatoid arthritis. *Rheumatol. Int.* **29**, 897–905 (2009).
4. Sung, Y.-K. *et al.* Korean observational study network for arthritis (KORONA): Establishment of a prospective multicenter cohort for rheumatoid arthritis in South Korea. *Semin. Arthritis Rheum.* **41**, 745–751 (2012).
5. Organization, W. H. *Assessment of Fracture Risk and its Application to Screening for Postmenopausal Osteoporosis: Report of a WHO Study Group [meeting held in Rome from 22 to 25 June 1992].* (World Health Organization, 1994).
6. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
7. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016). https://doi.org/10.1145/2939672.2939785.
8. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).
9. Berrar, D. Cross-Validation. *Encycl. Bioinform. Comput. Biol.* **1**, 542–545 (2019).
10. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
11. Staa, T. P. V., Geusens, P., Bijlsma, J. W. J., Leufkens, H. G. M. & Cooper, C. Clinical assessment of the long-term risk of fracture in patients with rheumatoid arthritis. *Arthritis Rheum.* **54**, 3104–3112 (2006).
12. Kwon, H.-Y., Kim, H.-H., Sung, Y.-K. & Ha, Y.-C. Incidence and mortality of osteoporotic fracture in rheumatoid arthritis in South Korea using nationwide claims data. *J. Bone Metab.* **26**, 97–104 (2019).
13. Amarnath, A. L. D., Franks, P., Robbins, J. A., Xing, G. & Fenton, J. J. Underuse and overuse of osteoporosis screening in a regional health system: A retrospective cohort study. *J. Gen. Intern. Med.* **30**, 1733–1740 (2015).
14. Roux, C. *et al.* Management of bone fragility in patients with rheumatoid arthritis in France: An analysis of a national health insurance claims database. *Joint Bone Spine* **89**, 105340 (2022).
15. Ozen, G. *et al.* Trends and determinants of osteoporosis treatment and screening in patients with rheumatoid arthritis compared to osteoarthritis. *Arthritis Care Res.* **70**, 713–723 (2018).
16. Chen, R., Huang, Q. & Chen, L. Development and validation of machine learning models for prediction of fracture risk in patients with elderly-onset rheumatoid arthritis. *Int. J. Gen. Med.* **15**, 7817–7829 (2022).
17. Llorente, I., García-Castañeda, N., Valero, C., González-Álvaro, I. & Castañeda, S. Osteoporosis in rheumatoid arthritis: Dangerous liaisons. *Front. Med.* **7**, 601618 (2020).
18. Lee, J.-H. *et al.* The frequency of and risk factors for osteoporosis in Korean patients with rheumatoid arthritis. *BMC Musculoskelet. Disord.* **17**, 98 (2016).
19. Park, H. W. *et al.* Application of machine learning to identify clinically meaningful risk group for osteoporosis in individuals under the recommended age for dual-energy X-ray absorptiometry. *Calcif. Tissue Int.* **109**, 645–655 (2021).
20. Shim, J.-G. *et al.* Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. *Arch. Osteoporos.* **15**, 169 (2020).
21. Erjiang, E. *et al.* Machine learning can improve clinical detection of low BMD: The DXA-HIP study. *J. Clin. Densitom.* **24**(4), 527–537 (2021).
22. Ou Yang, W.-Y., Lai, C.-C., Tsou, M.-T. & Hwang, L.-C. Development of machine learning models for prediction of osteoporosis from clinical health examination data. *Int. J. Environ. Res. Public. Health* **18**, 7635 (2021).
23. Wang, Y. *et al.* Prediction model for the risk of osteoporosis incorporating factors of disease history and living habits in physical examination of population in Chongqing, Southwest China: Based on artificial neural network. *BMC Public Health* **21**, 991 (2021).
24. Iliou, T., Anagnostopoulos, C.-N. & Anastassopoulos, G. Osteoporosis detection using machine learning techniques and feature selection. *Int. J. Artif. Intell. Tools* **23**, 1450014 (2014).
25. Bui, H. M. *et al.* Predicting the risk of osteoporosis in older Vietnamese women using machine learning approaches. *Sci. Rep.* **12**, 20160 (2022).

Acknowledgements

Author contributions

K.W.M. designed the study; S.S. and G.J. collected and analyzed the data; C.L. and G.J. conducted experiments; C.L. drafted the manuscript; H.I. and K.W.M. supervised the study; G.J., H.I. and K.W.M. revised the manuscript; and all authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-48842-7.

Correspondence and requests for materials should be addressed to H.I. or K.W.M.

Reprints and permissions information is available at www.nature.com/reprints.