Journal of Clinical Densitometry: Assessment & Management of Musculoskeletal Health, vol. 29, no. 1, 101666, 2026

1094-6950/29:101666/$36.00
https://doi.org/10.1016/j.jocd.2025.101666

Osteoporosis prediction from Frontal Lumbar Spine X-rays

Ryusei Inamori,[1,2] Tomoya Kobayashi,[1,2,**] Eichi Takaya,[1,2] Junya Iwazaki,[1] Carlos Makoto Miyauchi,[1] Saori Ikumi,[1,3] Yoshikazu Okamoto,[1,2] Cheng Wei Lin,[4] Sheng Che Hsiao,[4] Qingzong Tseng,[4] and Shinya Sonobe[1,5,6,7,*]

[1] AI Lab, Tohoku University Hospital, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan; [2] Department of Diagnostic Imaging, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan; [3] Department of Anesthesiology and Perioperative Medicine, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan; [4] Alpha Intelligence Manifolds, Inc., New Taipei City, Taiwan; [5] Department of Neurosurgery, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan; [6] Department of Translational Neuroscience, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan; and [7] Department of Neurosurgical Engineering, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan

Abstract

Background: This study aimed to evaluate the performance of DeepXray™ Spina, a software that estimates bone mineral density (BMD) and T-scores from frontal lumbar spine X-ray (FLS-X), in predicting osteoporosis.

Methodology: Patients from a Japanese cohort who underwent both FLS-X and dual-energy X-ray absorptiometry (DXA) using Hologic systems within 30 days at Tohoku University Hospital (May 2014-April 2024) were included. BMD was estimated from FLS-X using DeepXray™ Spina, which was developed using dataset from a Taiwanese Cohort. BMD assessed by DXA (observed BMD) and BMD estimated from FLS-X by DeepXray™ Spina (estimated BMD) were compared using Pearson's correlation coefficient (PCC) and normalized root mean square error (NRMSE). T-scores were converted to osteoporosis classifications as normal, osteopenia, or osteoporosis following the World Health Organization criteria. Classification performance was evaluated by accuracy, sensitivity, specificity, Cohen's kappa, and quadratic-weighted Cohen's kappa.

Results: The correlation between estimated and observed BMD was strong, with a PCC of 0.901 and an NRMSE of 0.070. For osteoporosis classification, the accuracy, sensitivity, specificity, and Cohen's kappa were as follows: 0.902, 1.000, 0.842, and 0.803 for normal; 0.854, 0.729, 0.924, and 0.673 for osteopenia; 0.951, 0.810, 1.000, and 0.863 for osteoporosis. The quadratic-weighted Cohen's kappa was 0.884.

Conclusion: This study evaluated the performance of Deep Xray™ Spina in predicting osteoporosis from FLS-X. The software is a practical and reliable tool for predicting osteoporosis, with high performance and robustness.

Keywords: Osteoporosis; Bone mineral density; Frontal lumbar spine X-ray; Dual-energy X-ray absorptiometry; Machine learning.

Introduction

Osteoporosis affects approximately 200 million people worldwide, and its prevalence is expected to increase further.[1-3] Failure to achieve timely diagnosis can lead to fractures, especially among the elderly.[4-9] Early detection is crucial for extending healthy life expectancy in an aging society.[10]

**Co-corresponding author.
Received 09/29/25; Revised 12/19/25; Accepted 12/25/25.
*Correspondence author at. Department of Translational Neuroscience, Tohoku University Graduate School of Medicine, 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan. E-mail: shinya.sonobe.d7@tohoku.ac.jp

Diagnosis is made through the assessment of bone mineral density (BMD) and T-scores. Dual-energy X-ray absorptiometry (DXA) is the gold standard, but it requires specialized equipment and is therefore available only in limited facilities and limited examination slots, which may restrict patient access. Many individuals with osteoporosis remain undiagnosed, and it would be useful if osteoporosis could be predicted from examinations that can be performed more widely.

Recent studies have reported artificial intelligence (AI) models predicting osteoporosis by BMD estimation based on lumbar spine X rays.[11-14] These models have generally been developed using data from geographically homogeneous populations, and their generalizability across geographically distinct populations remains unclear. DeepXray™ Spina (Alpha Intelligence Manifolds, Inc., Taiwan) is one such model, which estimates BMD and T scores from frontal lumbar spine X-ray images (FLS-X).

This study aimed to validate the performance of DeepXray™ Spina in predicting osteoporosis in a geographically distinct population.

Materials and methods

Population

Patients who underwent both FLS-X examinations and DXA scans within 30 days at Tohoku University Hospital between May 2014 and April 2024, with the DXA scans performed using a Hologic DXA system, were included.

Data acquisition

BMD and T-scores assessed by DXA were acquired. Additionally, the FLS-X acquired on the date closest to the DXA scan was acquired. Kidney−ureter−bladder was also treated as FLS-X. FLS-X were obtained using two types of equipment manufactured by FUJIFILM Corporation (Tokyo, Japan) and KONICA MINOLTA Inc. (Tokyo, Japan). The exposure parameters of FLS-X were as follows (min, mean, max): KVP (66, 74.1, 98), Exposure Time (14, 106.6, 280), X-ray Tube Current (160, 307.1, 500), and source-to-image distance (984, 1095, 1806).

Estimation of BMD and T-scores from FLS-X

Separately from the acquired BMD and T-scores, estimated BMD and T-scores from FLS-X were output using DeepXray™ Spina. The software processes DICOM-format FLS-X and detects key anatomical landmarks from the T12 to L5 vertebrae, then outputs estimated BMD for each of L1 to L4 using a machine learning model for regression (for details, please refer to the Supplementary Material). These estimated BMD were calibrated for GE DXA systems; therefore, they were converted to Hologic-equivalent values using the following equation: Hologic-equivalent value = 0.856 × GE-equivalent value - 0.016.[16] The Hologic-equivalent values were converted to T-scores (estimated T-scores) using the Young Adult Mean values from the National Health and Nutrition Examination Survey database.[17] Images with invalid BMD estimation (for details, please refer to the Supplementary Material) for any of L2 to L4 were excluded.

Evaluation of estimation performance

A Scatter plot was created with BMD assessed by DXA (observed BMD) on the x-axis and estimated BMD on the y-axis. Pearson's correlation coefficient (PCC) and normalized root mean square error (NRMSE, normalized by the range [min−max]) between observed BMD and estimated BMD were calculated. T-scores assessed by DXA (observed T-scores) and estimated T-scores were converted to osteoporosis classifications following the World Health Organization criteria: normal (T-score ≥ -1.0), osteopenia ($-2.5 <$ T-score < -1.0), or osteoporosis (T-score ≤ -2.5).[18] A confusion matrix was created between osteoporosis classifications derived from observed T-scores and those based on estimated T-scores.

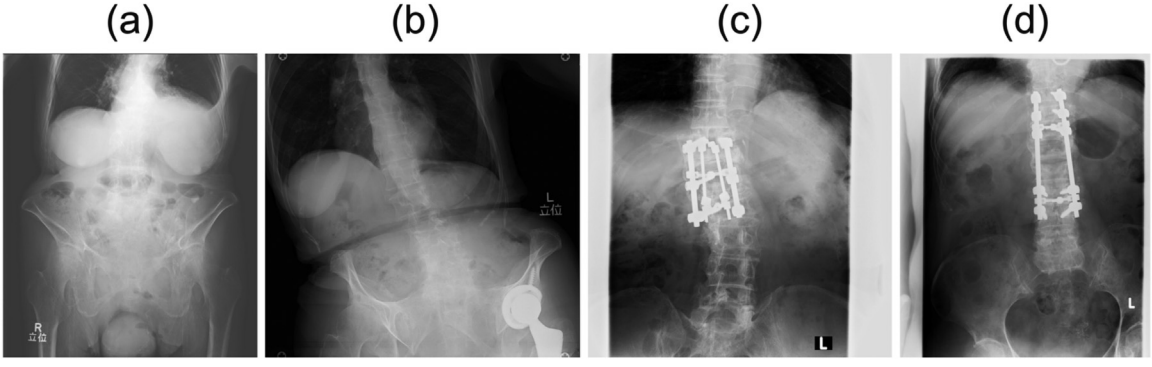

Fig. 1. Images Excluded Due to Invalid Estimated BMD.

The 4 images excluded due to invalid BMD estimation. The vertebrae were not properly detected due to nonvertebral structures (a, b) or metal implants (c, d).

Table 1
Patient Characteristics.

	Overall	Female	Male	p-value
Patients, n (%)	323 (100)	253 (78.3)	70 (21.7)	
Images, n (%)	328 (100)	258 (78.7)	70 (21.3)	
Diagnosis, n (%)				
- Normal	126 (38.4)	86 (68.3)	40 (31.7)	
- Osteopenia	118 (36.0)	94 (79.7)	24 (20.3)	
- Osteoporosis	84 (25.6)	78 (92.9)	6 (7.1)	
Age, median (IQR), years	70.0 (60.0-77.0)	70.0 (60.0-77.0)	67.0 (60.0-73.0)	0.87
BMI, median (IQR), kg/m^2	23.68 (21.02-26.45)	24.25 (21.50-26.80)	23.55 (21.00-26.00)	0.40
BMD, median (IQR), g/cm^2	0.93 (0.79-1.07)	0.90 (0.80-1.00)	1.09 (1.00-1.15)	< 0.001
T-score, median (IQR)	-1.41 (-2.80 to -0.24)	-1.79 (-2.80 to -0.30)	-0.02 (-1.00 to 1.00)	< 0.001

Note: Median values were compared between groups using the Wilcoxon rank-sum test. Osteoporosis classifications were defined according to the World Health Organization criteria: normal (T-score ≥ -1.0), osteopenia ($-2.5 <$ T-score < -1.0), and osteoporosis (T-score ≤ -2.5). Abbreviations: IQR, interquartile range; BMI, body mass index; BMD, bone mineral density.

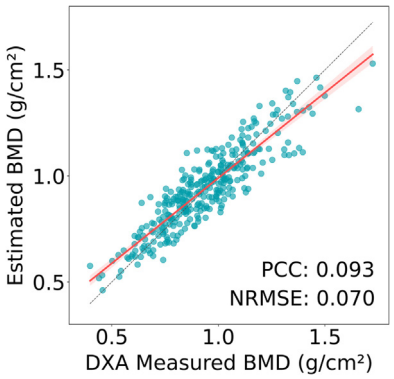

Fig. 2. Performance for BMD Estimation.
The scatter plot of bone mineral density (BMD) assessed by dual-energy X-ray absorptiometry (observed BMD) and BMD estimated from frontal lumbar spine X-ray (FLS-X) using DeepXray^{TM} Spina (estimated BMD). The red line with the light red shaded area indicates the regression line for observed BMD and estimated BMD, along with the 95% confidence interval. The black dashed line shows the line of identity.

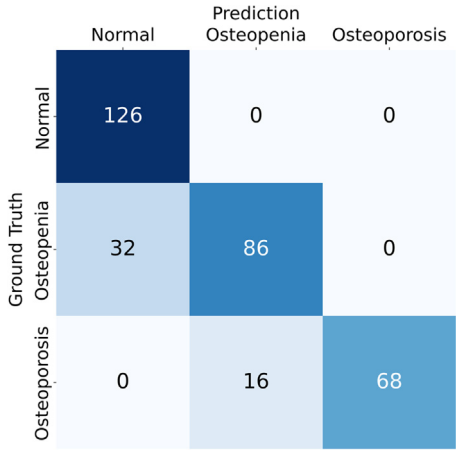

Fig. 3. Confusion Matrix for Osteoporosis Classification.
Confusion matrix comparing predictions and ground truth.

Based on the confusion matrix, accuracy, sensitivity, specificity, and Cohen's kappa were calculated according to the one-vs-rest strategy. In addition, the quadratic-weighted Cohen's kappa was calculated for multi-class agreement.

Results

Patient characteristics

A total of 332 images from 327 patients were included, as 5 patients had two images that met the inclusion criteria. Due to invalid BMD estimation, 4 images corresponding to 4 patients were excluded (Fig. 1). Consequently, 328 images from 323 patients were analyzed. The cohort consisted of 253 female and 70 male patients, with ages ranging from 13 to 91 years, and a mean age of 67.1 years. Among them, 126 were classified as normal, 118 as osteopenia, and 84 as osteoporosis (Table 1). There were 15 patients aged 40 years or younger, who are generally considered to have a low risk of primary osteoporosis; 7 patients were taking corticosteroids, and 6 patients had diseases that can lead to secondary osteoporosis.

Evaluation of estimation performance

The PCC and NRMSE were 0.901 and 0.070 between the observed BMD and estimated BMD (Fig. 2). The predictive performance of osteoporosis classification was calculated based on the confusion matrix (Fig. 3); the

Table 2
Predictive Performance of Osteoporosis Classification.

Classification	Accuracy	Sensitivity	Specificity	Cohen's kappa	Quadratic-weighted Cohen's kappa
Normal	0.902	1.000	0.842	0.803	0.884
Osteopenia	0.854	0.729	0.924	0.673	
Osteoporosis	0.951	0.810	1.000	0.863	

Note: T-scores estimated by DeepXray^{TM} Spina were converted to osteoporosis classifications following the World Health Organization criteria: normal (T-score ≥ -1.0), osteopenia ($-2.5 <$ T-score < -1.0), and osteoporosis (T-score ≤ -2.5). For each category, accuracy, sensitivity, specificity, and Cohen's kappa for prediction were calculated according to the one-vs-rest strategy.

accuracy, sensitivity, specificity, and Cohen's kappa were as follows: 0.902, 1.000, 0.842, and 0.803 for normal; 0.854, 0.729, 0.924, and 0.673 for osteopenia; 0.951, 0.810, 1.000, and 0.863 for osteoporosis. The quadratic-weighted Cohen's kappa was 0.884 (Table 2).

Discussion

This study evaluated the performance of DeepXray^{TM} Spina in predicting osteoporosis. In BMD estimation from FLS-X, the PCC and NRMSE were 0.901 and 0.070, respectively, between the observed BMD and estimated BMD. In osteoporosis classification, the quadratic-weighted Cohen's kappa was 0.884.

This study assessed the ability of DeepXray^{TM} Spina to estimate BMD from FLS-X. The PCC and the NRMSE demonstrated a very strong correlation between the observed BMD and the estimated BMD. The Cohen's kappa demonstrated substantial agreement in osteoporosis classification for normal and osteoporosis categories. These findings indicate that the software has the potential to be useful for osteoporosis screening. Indeed, the software demonstrated comparable or superior performance to previous reports (Table 3). Thus, the software is a reliable tool for predicting osteoporosis with high performance.

This study assessed the ability of DeepXray^{TM} Spina to generalize across different populations. External validation often results in significantly lower accuracy due to domain shifts caused by variations in imaging devices, imaging protocols, and patient characteristics.[19-21] The software was originally developed using approximately 15,000 DXA scans from four medical institutions in Taiwan, and its generalizability across different populations has not been evaluated. This study demonstrated that the software achieved high predictive performance in the Japanese cohort. Thus, the software is a reliable tool for predicting osteoporosis with high robustness.

The fact that DeepXray^{TM} Spina requires only X-rays is an essential advantage. X-rays are widely implemented and highly accessible. Therefore, the software may provide screening opportunities to patients who visit hospitals for unrelated purposes to osteoporosis (opportunistic screening). Also, X-rays can be performed using portable devices. In the future, the software may provide screening opportunities to patients who may have difficulty accessing medical

Table 3
Osteoporosis Prediction from X-ray using AI.

Author (Year)	Site	PCC	NRMSE	Predictive Performance of Osteoporosis Classification			Reference
				Normal	Osteopenia	Osteoporosis	
Hsieh et al. (2021)	Lumber	0.90	NA	NA	NA	NA	[12]
Boonrod et al. (2025)	Lumber	0.63	NA	NA	NA	0.696, 0.797, 0.665	[13]
Nguyen et al. (2025)	Lumber	0.87	NA	NA	NA	NA	[14]
Moro et al. (2025)	Lumber	0.89	NA	NA	0.850, 0.864, 0.841	0.924, 0.717, 0.947	[15]
Current study	Lumber	0.901	0.070	0.902, 1.000, 0.842	0.854, 0.729, 0.924	0.951, 0.810, 1.000	-

Note: "Site" refers to the anatomical location of the X-ray. "Predictive performance of osteoporosis classification for each category" includes accuracy, sensitivity, and specificity. Abbreviations: PCC, Pearson's correlation coefficient; NRMSE, normalized root mean squared error; NA, not available.

facilities due to transportation issues or physical limitations.[22,23] Thus, the software is a practical tool for predicting osteoporosis and may provide screening opportunities for patients in various medical settings.

This study has two limitations. First, it had a retrospective design. The reliability of the software was evaluated only in patients who had already undergone DXA scans; therefore, its reliability for screening candidates was not directly demonstrated. A future study with a prospective design involving a diverse general population would further enhance the reliability of the software as a screening tool. Second, we used the NHANES database to calculate T-scores because DeepXray™ Spina was developed based on the same database, and we aimed to ensure consistency with it. The use of reference values or databases specific to Japanese individuals may provide more appropriate clinical evaluation.

Conclusion

This study evaluated the performance of DeepXray™ Spina in predicting osteoporosis from FLS-X. The software is a practical and reliable tool for predicting osteoporosis, with high performance and robustness. It may provide screening opportunities for patients in various medical settings and contribute to extending healthy life expectancy among the elderly through early detection.

Data availability

The data are not publicly available due to privacy or ethical restrictions, which are available on request from the corresponding author.

Acknowledgement

This study was supported by the Japan Science and Technology Agency (JST) through its Core Research for Evolutional Science and Technology (CREST) program (Grant No. JPMJCR15D1) and its Support for Pioneering Research Initiated by the Next Generation (SPRING) program (Grant No. JPMJSP2114), as well as by the project “An advanced AI R&D and human resources development center aiming to solve Global x Local medical issues (Tohoku University Hospital Clinical AI)” funded by MEXT, Japan. We utilized ChatGPT for assistance with parts of the English proofing.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jocd.2025.101666.

References

1. Salari N, Ghasemi H, Mohammadi L, et al. 2021 The global prevalence of osteoporosis in the world: a comprehensive systematic review and meta-analysis. J Orthop Surg Res 16:609.
2. Sözen T, Özışık L, Başaran NÇ. 2017 An overview and management of osteoporosis. Eur J Rheumatol 4:46–56.
3. Wu J, Qu Y, Wang K, Chen Y. 2019 Healthcare resource utilization and direct medical costs for patients with osteoporotic fractures in China. Value Health Reg Issues 18:106–111.
4. Barrett-Connor E. 1995 The economic and human costs ofosteoporotic fracture. Am J Med 98:3S–8S.
5. Burge R, Dawson-Hughes B, Solomon DH, et al. 2007 Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. J Bone Miner Res 22:465–475.
6. Compston JE, McClung MR, Leslie WD. 2019 Osteoporosis. Lancet 393:364–376.
7. Chandran M, Ebeling PR, Mitchell PJ, et al. 2022 Harmonization of osteoporosis guidelines: paving the way for disrupting the status quo in osteoporosis management in the Asia Pacific. J Bone Miner Res 37:608–615.
8. Sànchez-Riera L, Carnahan E, Vos T, et al. 2014 The global burden attributable to low bone mineral density. Ann Rheum Dis 73:1635–1645.
9. Kanis JA, Johnell O. 2005 Requirements for DXA for the management of osteoporosis in Europe. Osteoporos Int 16:229–238.
10. Rachner TD, Khosla S, Hofbauer LC. 2011 Osteoporosis: now and the future. Lancet 377:1276–1287.
11. He Y, Lin J, Zhu S, et al. 2024 Deep learning in the radiologic diagnosis of osteoporosis: a literature review. J Int Med Res 52:3000605241244754.
12. Hsieh C-I, Zheng K, Lin C, et al. 2021 Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. Nat Commun 12:5472.
13. Boonrod Artit, Kittipongphat Nut, Piyaprapaphan Prarinthorn, Theerakulpisut Daris, Boonrod Arunnit. 2025 Assessing deep learning model performance in osteoporosis screening with lumbar spine radiographs. J Bone Miner Metab.
14. Nguyen Huy Gia, Nguyen Dinh-Tan, Tran Thach Son, Ling Sai Ho, Ho-Pham Lan Thuc, Nguyen Tuan Van. 2025 Artificial intelligence system for predicting areal bone mineral density from plain X-rays. Osteoporos Int: 2167–2176.
15. Moro Toru, Yoshimura Noriko, Saito Taku, et al. 2025 Development of artificial intelligence-assisted lumbar and femoral BMD estimation system using anteroposterior lumbar X-ray images. J Orthop Res: 1619–1631.
16. YAN Dong, Glen M Blake, SUN Xiao-lei, et al. (2022) Cross calibration of different DXA scanners and the standardization of normal reference for Chinese
17. 2019 ISCD Official Positions Adult. https://iscd.org/wp-content/uploads/2021/09/2019-Official-Positions-Adult-1.pdf
18. World Health Organization(WHO). 1994 Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: Report of a WHO study group. Genève, Switzerland: World Health Organization.
19. Subbaswamy A, Saria S. 2020 From development to deployment: dataset shift, causality, and shift-stable models in health AI. Biostatistics 21:345–352.
20. Valliani AA, Gulamali FF, Kwon YJ, et al. 2022 Deploying deep learning models on unseen medical imaging using adversarial domain adaptation. PLoS One 17:e0273262.
21. Yu AC, Mohajer B, Eng J. 2022 External validation of deep learning algorithms for radiologic diagnosis: A systematic review. Radiol Artif Intell 4:e210064.

22. Morishita F, Garfin AMCG, Lew W, et al. 2017 Bringing state-of-the-art diagnostics to vulnerable populations: the use of a mobile screening unit in active case finding for tuberculosis in Palawan, the Philippines. PLoS One 12:e0171310.

23. Dhoot R, Humphrey JM, O'Meara P, et al. 2018 Implementing a mobile diagnostic unit to increase access to imaging and laboratory services in western Kenya. BMJ Glob Health 3:e000947.