

Sensor-based Explainable Deep Learning Framework for Osteoporosis Screening via Hip X-ray Interpretation

Shang-Wen Feng,¹ Szu-Yin Lin,^{2*} Yueh-Te Shen,³
Bo-Yu Chen,³ Yen-Chiu Chen,⁴ and Yi-Hung Chiang¹

¹Department of Orthopedics, National Yang Ming Chiao Tung University Hospital,
No. 169, Xiaoshe Rd., Yilan City, Yilan County 260, Taiwan

²Department of Management Science, National Yang Ming Chiao Tung University,
No. 1001, Daxue Rd., East Dist., Hsinchu City 300, Taiwan

³Department of Computer Science and Information Engineering, National Ilan University,
No. 1, Sec. 1, Shennong Rd., Yilan City, Yilan County 260, Taiwan

⁴Department of Information Management, Chung Hua University,
No. 707, Sec.2, WuFu Rd., Hsinchu, Taiwan

(Received May 1, 2025; accepted August 14, 2025)

Keywords: X-ray sensing, medical imaging sensors, osteoporosis, deep learning, explainable AI

Sensor-based X-ray imaging, made possible by high-sensitivity medical imaging sensors equipped with advanced detector materials, enables noninvasive and low-dose assessment of bone structures. Leveraging such sensor outputs, we applied deep learning and explainable AI methods to develop a sensor-driven module for osteoporosis detection. Three hundred fifty-nine hip X-ray images (hereafter referred to as sensor-based X-ray images) were collected to train and evaluate four convolutional neural networks. Among them, the unsegmented whole hip VGG16 model achieved the highest classification accuracy of 76%. Gradient-weighted class activation mapping (Grad-CAM), shapley additive explanations (SHAP), and local interpretable model-agnostic explanations (LIME) were applied to generate saliency heat maps, followed by pixel-level consensus analysis to enhance interpretability. Retraining using the consensus-cropped images increased accuracy to 79% with corresponding 2–3% improvements in sensitivity, specificity, and *F1-score*. A web-based survey with orthopedic surgeons showed that the visual explanations improved clinical trust in 70% of cases. These results demonstrate that advanced sensor-based X-ray imaging technologies, combined with explainable deep learning, can serve as a practical decision-support module within intelligent medical sensing platforms, especially where dual-energy X-ray absorptiometry (DXA) is unavailable.

1. Introduction

With global population aging, many countries and regions face increasing healthcare challenges. According to the United Nations' World Population Ageing report,⁽¹⁾ South Korea, Singapore, and Taiwan are among the fastest-aging societies. Aging significantly pressures healthcare systems, particularly due to osteoporosis, a common skeletal disease caused by age-

*Corresponding author: e-mail: stan@nycu.edu.tw
<https://doi.org/10.18494/SAM5723>

related bone density loss.⁽²⁾ Osteoporosis is often asymptomatic until fractures occur, with hip fractures posing the greatest risks of disability or mortality.⁽³⁾ AI, especially deep learning, has recently shown great potential in medical imaging by demonstrating the ability to handle complex diagnostic tasks with high accuracy.^(4–6) During the COVID-19 pandemic, deep learning models achieved promising results in classifying lung infections and predicting hip fractures from sensor-based X-ray images.^(7,8) These advances suggest that intelligent healthcare and sensor-driven medical analysis will become key areas for AI development. The hip X-ray images used in this study were generated by advanced flat-panel detector sensors employing high-efficiency scintillator materials, which convert X-ray photons into high-resolution digital signals. These sensor technologies provide the fine structural details necessary for AI-based osteoporosis screening, and their integration with AI can extend diagnostic capabilities to point-of-care and resource-limited settings.

Against this backdrop, early detection of osteoporosis becomes critical yet remains challenging owing to limitations in conventional diagnostic methods. Currently, dual-energy X-ray absorptiometry (DXA) is the gold standard for osteoporosis diagnosis owing to its precision and stability. However, its high cost limits widespread access, and diagnosis often occurs only after fractures. Comparative studies indicate that quantitative computed tomography (QCT) is less accurate and poses higher radiation risks, making it a suboptimal alternative.⁽⁹⁾ To address these limitations, deep learning models have been applied for osteoporosis screening with encouraging accuracy.⁽¹⁰⁾ While existing models have achieved reasonable levels of accuracy, the continuous development of novel architectures raises the potential for further improvement through updated transfer learning strategies and advanced data preprocessing techniques. However, the integration of deep learning also introduces challenges related to the opacity of decision-making processes, including concerns over decision authority and ethical implications.⁽¹¹⁾ As a result, the incorporation of explainable artificial intelligence (XAI) has become essential, and it remains to be investigated which XAI methods are most effective for osteoporosis diagnosis.

In recent years, XAI has become increasingly active in the medical field. In many studies, XAI techniques have been applied to various medical diagnostic tasks.^(12–14) Sheu and Pardeshi⁽¹⁵⁾ conducted a comprehensive review of different XAI methods by analyzing and evaluating their applicability, highlighting the importance of human–machine interaction and the potential directions for future development. In medical imaging, XAI has attracted particular attention. According to statistical analyses by Velden *et al.*,⁽¹⁶⁾ visual explanation methods have been rapidly adopted in medical image analysis. Among these, commonly used techniques include class saliency heatmapping (CAM), gradient-weighted class saliency heatmapping (Grad-CAM), layer-wise relevance propagation (LRP), local interpretable model-agnostic explanations (LIME), and Shapley additive explanations (SHAP).

To address the challenges of limited accessibility to DXA-based osteoporosis diagnosis and the lack of model interpretability, we collaborated with a regional hospital to collect 359 hip sensor-based X-ray images. A deep-learning-based classification framework was established using *T-scores* from DXA results as ground truth, and various convolutional neural network models were compared. To enhance transparency, multimethod XAI techniques, including Grad-CAM, SHAP, and LIME, were applied to analyze model

decision features. A pixel-level consensus analysis was further conducted to identify key skeletal regions, followed by retraining using feature-cropped images to validate their clinical relevance. Finally, an expert survey involving orthopedic surgeons was conducted to assess the trustworthiness of the explanations. This work contributes to the development of an intelligent, sensor-driven osteoporosis screening platform, by integrating sensor-based imaging data with interpretable deep learning techniques for medical decision support. In this study, we aim to address both diagnostic accessibility and interpretability through a sensor-based explainable AI approach, validated with clinical feedback.

2. Research Methods and Experimental Design

2.1 Research framework

Figure 1 shows the overall research framework of this study. This study was conducted in collaboration with a regional hospital in Yilan, Taiwan, and approved by the Institutional Review

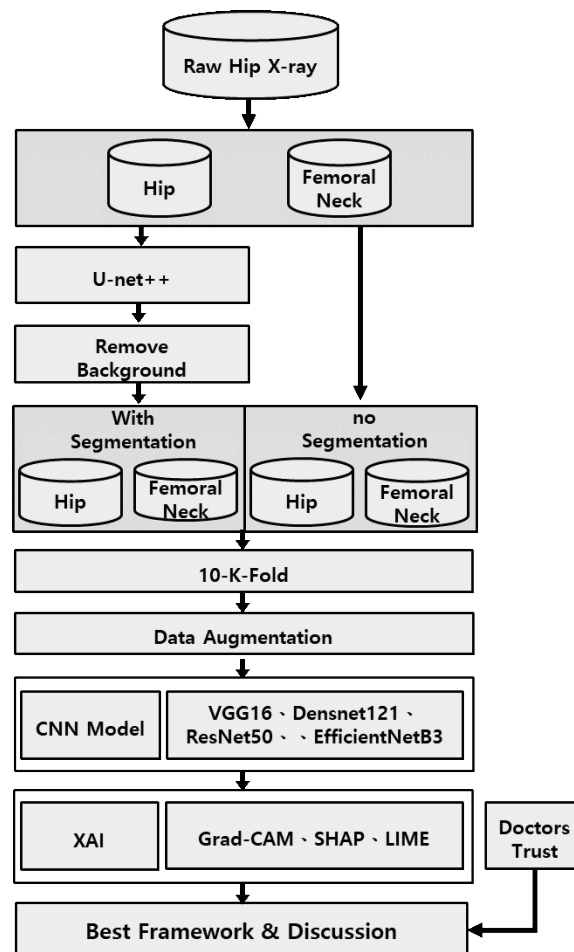


Fig. 1. Research framework.

Board of National Yang-Ming Chiao Tung University Hospital (IRB No. 2002A007). A retrospective dataset of 359 hip sensor-based X-ray images, regarded as outputs from a noninvasive medical sensor, was collected and categorized on the basis of DXA *T-scores*: osteoporotic ($T\text{-score} \leq -2.5$) and non-osteoporotic ($T\text{-score} > -2.5$). A binary classification pipeline was established to analyze the sensor-based images and predict osteoporosis status. The best-performing model identified in the initial experiments was selected for further feature analysis and validation using multimethod XAI techniques. Finally, a web-based survey of orthopedic surgeons was conducted to assess clinical trust in the model’s predictions.

2.2 Dataset and data preprocessing

The collected sensor-based X-ray images represent structured outputs from a noninvasive biomedical sensing system and require appropriate preprocessing for downstream analysis. A total of 359 hip sensor-based X-ray images were retrospectively collected from 359 patients and categorized in accordance with DXA results. On the basis of a *T-score* threshold of -2.5 , images were labeled as osteoporotic ($T\text{-score} \leq -2.5$) or non-osteoporotic ($T\text{-score} > -2.5$). Two anatomical regions were evaluated: the whole hip and the femoral neck. For the whole hip region, the dataset initially included 124 osteoporotic and 235 non-osteoporotic cases. In the femoral neck region, there were 194 osteoporotic and 165 non-osteoporotic cases. To mitigate class imbalance, oversampling was applied during training to equalize the number of samples across categories, as summarized in Table 1.

In the preprocessing stage, all sensor-based X-ray images were processed using UNet++, a deep convolutional neural network architecture optimized for medical image segmentation. Prior to training the model, bone regions in the original images were manually annotated using the Labelme tool to generate ground truth masks. Once trained, the UNet++ model was used to predict segmentation masks for the remaining images, enabling background removal and isolation of the hip region, as illustrated in Fig. 2. To enhance model generalization, data augmentation was applied to the training set using the ImageDataGenerator() function, which performed four random transformations: rotation ($\pm 2^\circ$), width shift ($\pm 20\%$), height shift ($\pm 20\%$), and zoom (up to 30%).

2.3 Model evaluation indicators

Model performance was primarily evaluated using four standard metrics: accuracy, sensitivity, specificity, and *F1-score*. Accuracy is the measure of the proportion of correctly classified samples. Sensitivity (also known as recall or true positive rate) reflects the model’s

Table 1
Distributions of osteoporotic and non-osteoporotic cases before and after oversampling.

	Osteoporotic		Non-osteoporotic	
	Original	After oversampling	Original	After oversampling
Whole hip	124	235	235	235
Femoral neck	194	194	165	194

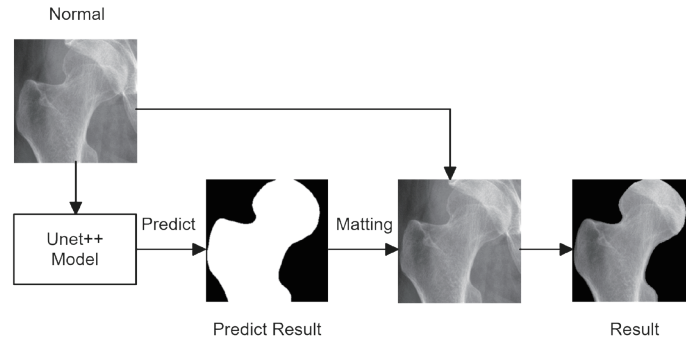


Fig. 2. UNet++-based background removal flowchart.

ability to identify positive cases, whereas specificity (true negative rate) assesses the model's ability to detect negative cases. The *F1-score*, which is the harmonic mean of precision and recall, accounts for both the accuracy and completeness of the classification. These metrics were computed on the basis of the confusion matrix structure shown in Table 2.

The metrics record the counts of true positives (*TP*), false positives (*FP*), false negatives (*FN*), and true negatives (*TN*). The formulas used are as follows.

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$$Sensitivity = TP / (TP + FN) \quad (2)$$

$$Specificity = TN / (TN + FP) \quad (3)$$

$$F1\text{-score} = 2(Precision * Recall) / (Precision + Recall) \quad (4)$$

To ensure robust evaluation and mitigate sampling bias, *k*-fold cross-validation was employed during model training and testing. The dataset was randomly divided into *k* equal-sized subsets. In each of the *k* iterations, one subset was used as the validation set while the remaining subsets were used for training. The process was repeated until every subset had served as the validation set once. The average performance across all *k* folds was then reported as the final evaluation. This method helps reduce overfitting and allows for the comprehensive utilization of limited data. However, it assumes that samples are independent and identically distributed; thus, its applicability is limited in datasets with inherent temporal or spatial dependences, where *k*-fold splitting may disrupt such structure.

2.4 Experimental design

This study comprises three experiments designed to evaluate model performance, interpretability, and clinical applicability using sensor-based X-ray data.

Table 2
Structure of the confusion matrix.

	Actual positive	Actual negative
Predicted positive	True positive (<i>TP</i>)	False positive (<i>FP</i>)
Predicted negative	False negative (<i>FN</i>)	True negative (<i>TN</i>)

2.4.1 CNN-based classification modeling

In the first phase, we established a classification framework for analyzing sensor-based X-ray data acquired from a noninvasive imaging sensor. As shown in Fig. 3, original images were manually cropped into two anatomical regions: whole hip and femoral neck, each resized to 256×256 pixels. To reduce background noise in the sensor data, a UNet++ model was applied to remove nonbone regions. Four datasets were generated for model training: (a) whole hip, (b) segmented whole hip, (c) femoral neck, and (d) segmented femoral neck, as illustrated in Fig. 4. K -fold cross-validation ($K = 10$) with random sampling was used to evaluate model performance on independent test sets. All models were fine-tuned using the Adam optimizer with a learning rate of 1×10^{-5} and a batch size of 16. The loss was calculated using categorical cross-entropy. A dropout rate of 0.2 was applied after the convolutional layers to mitigate overfitting. Four deep learning models were evaluated using accuracy, sensitivity, specificity, and *F1-score*. The combination of input region and model that yielded the highest performance was selected for subsequent interpretability analysis.

2.4.2 Model interpretability analysis using XAI (Grad-CAM, SHAP, LIME)

The second experiment focused on interpreting the best-performing model using explainability techniques applied to sensor-based imaging data. Grad-CAM, SHAP, and LIME were applied simultaneously to individual test cases to generate heatmaps highlighting important regions. This multimethod approach mitigates bias from relying on a single explanation and verifies whether the model responds to valid features within the sensing input.

In the second phase, a pixel-level consensus analysis was performed using Grad-CAM outputs from correctly classified samples. Each heatmap was binarized into feature (1) and nonfeature (0) masks, then summed across all cases. Only pixels activated in more than half of the cases were retained as key regions, as illustrated in Fig. 5. This aggregation helps identify the most informative sensor-based regions contributing to model decisions.

In the final phase, the feature regions extracted from sensor outputs were validated to assess their effect on model performance. This verification process, illustrated in Figs. 6 and 7, allows for a robust assessment of feature importance by isolating specific regions and observing the impact on model performance. As shown in Fig. 6, cropped images containing only key, positive, or negative feature regions were used to retrain the model. If the accuracy remained stable despite reduced input size, this would confirm the importance of those regions. Conversely, significant performance degradations would suggest that the identified features were either incomplete or nonessential. In addition, nonsalient regions were tested separately by applying masking procedures, as illustrated in Fig. 7, to further evaluate the specificity of the model's focus.

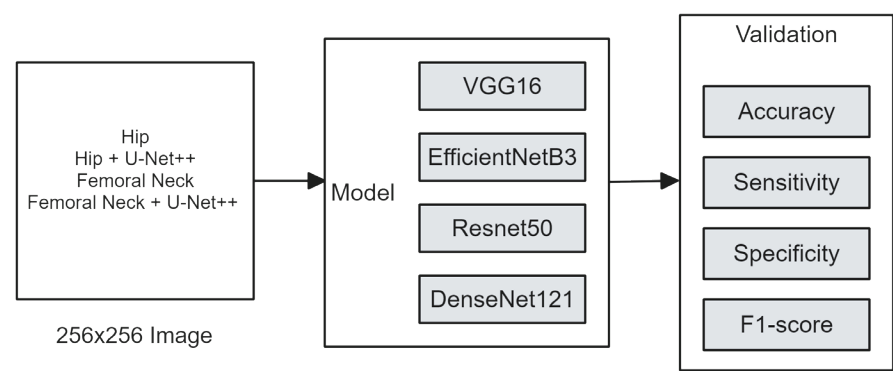


Fig. 3. Workflow of the first experiment: CNN-based classification modeling for hip X-ray images including anatomical region cropping, background removal, and deep learning model training.

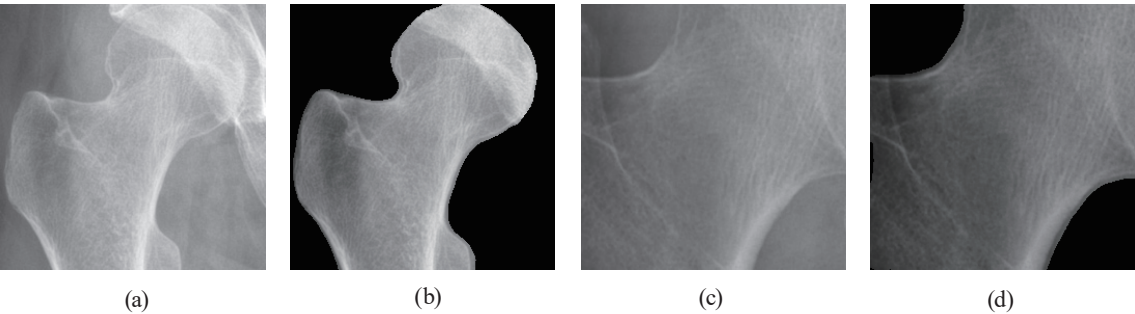


Fig. 4. Overview of the four training datasets derived from sensor-based X-ray images: (a) original whole hip, (b) segmented whole hip, (c) original femoral neck, and (d) segmented femoral neck.

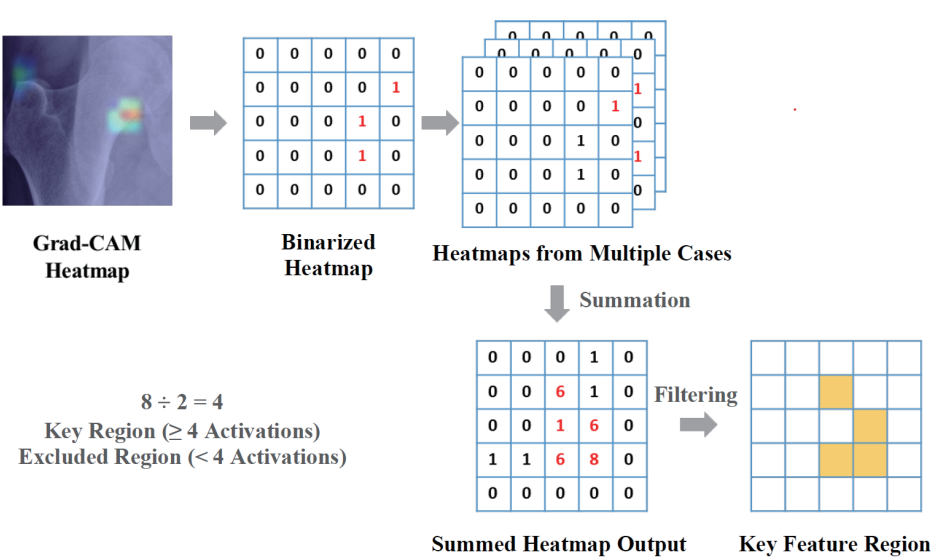


Fig. 5. (Color online) Pixel-wise consensus analysis workflow. This process aggregates binarized heatmaps from multiple correctly classified cases to statistically identify key feature regions where the model consistently focuses its attention.

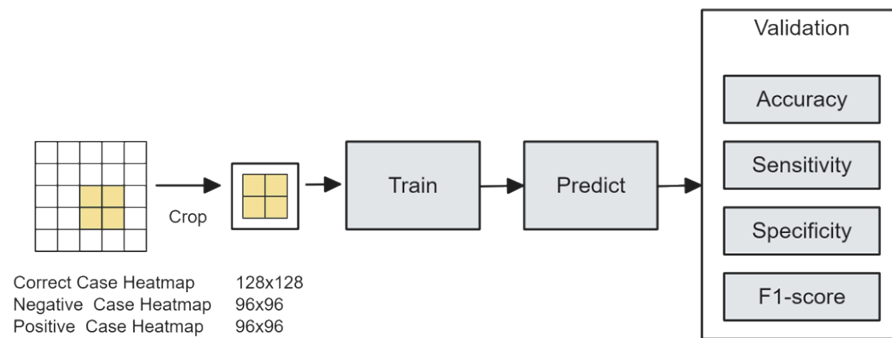


Fig. 6. (Color online) Verification workflow for key feature regions. This figure illustrates the process of feature isolation, where images are cropped to retain only the consensus regions before retraining to validate their impact.

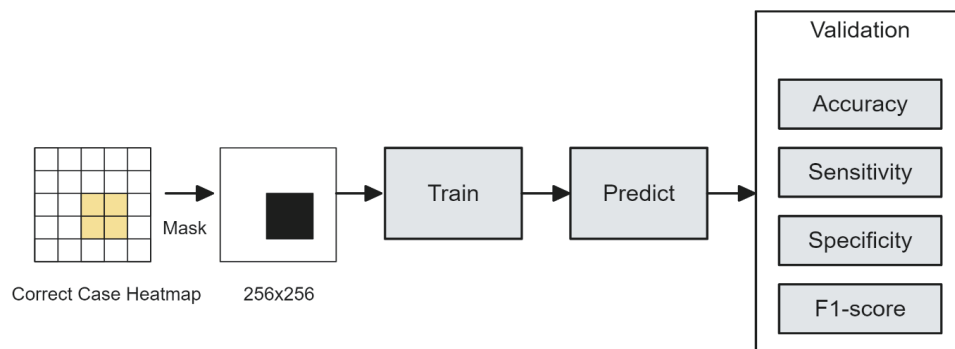


Fig. 7. Verification of nonsalient regions. This process involves masking the identified key feature regions to confirm that their removal negatively impacts model performance, thereby validating their importance.

2.4.3. Clinical validation through expert survey

In the third experiment, we evaluated orthopedic surgeons' trust in AI predictions generated from sensor-based X-ray inputs, with and without explanation. A web-based clinical evaluation platform was developed to assess the trustworthiness of AI-assisted osteoporosis diagnosis with and without explainability support. The evaluation began with the presentation of the original X-ray image, followed by the model's prediction and three heatmaps.⁽¹⁷⁾ Surgeons were then asked to respond to six structured questions assessing their diagnostic confidence and trust in the model's predictions. In this experiment, whether visual explanations improve the clinical acceptance of sensing-based AI decision support was assessed.

3. Results and Discussion

In this study, we developed an explainable osteoporosis classification framework using a pipeline of data preprocessing, deep learning classification, XAI interpretation, and clinical evaluation using sensor-based X-ray images.

In the first experiment, four CNN models, namely, DenseNet121, EfficientNetB3, ResNet50, and VGG16, were trained to classify osteoporosis status using both whole hip and femoral neck regions with and without background removal. The VGG16 model trained on unsegmented whole hip images achieved the highest accuracy of 76%, as shown in Table 3, suggesting that preserving global anatomical context benefits model performance.

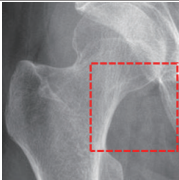
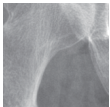
In the second experiment, we investigated model interpretability. Saliency heatmaps from Grad-CAM, SHAP, and LIME revealed that the model’s attention was concentrated on the femoral neck margins. To validate this, we extracted consensus feature regions by aggregating Grad-CAM outputs. Retraining the VGG16 model on images cropped to these consensus regions improved accuracy to 79%, with a corresponding 2–3% increase in sensitivity, specificity, and *F1-score*, as shown in Table 4. Conversely, masking these critical regions reduced accuracy to 70%, confirming their importance to the model’s decisions. This demonstrates that XAI-driven feature localization enhances both model focus and classification performance.

In the third experiment, we evaluated the clinical trustworthiness of the AI model through a web-based survey involving orthopedic surgeons. Each test case presented the original sensor-

Table 3
Classification performance of the VGG16 model on segmented vs unsegmented hip X-ray inputs.

Background removal	<i>Sensitivity</i>		<i>Specificity</i>		<i>F1-score</i>		<i>Accuracy</i>	
	Without	With	Without	With	Without	With	Without	With
Femoral neck	0.67	0.72	0.72	0.61	0.70	0.66	0.70	0.68
Whole hip	0.75	0.68	0.76	0.71	0.75	0.65	0.76	0.70

Table 4
(Color online) Model performances for original, cropped, and masked hip X-ray images. Training and testing times are reported in seconds.

	<i>Sensitivity</i>	<i>Specificity</i>	<i>F1-score</i>	<i>Accuracy</i>	Training time	Testing time
 Whole hip region	0.75	0.76	0.75	0.76	1201	0.23
 Cropped whole hip region	0.77	0.79	0.77	0.79	652	0.17
 Masked whole hip images (important regions removed)	0.436	0.82	0.63	0.70	NaN	NaN

* NaN indicates that retraining was not feasible owing to the removal of critical regions.

based X-ray image, the AI model's prediction, and three heatmaps generated by Grad-CAM, SHAP, and LIME. The survey comprised six structured questions assessing diagnostic confidence, perceived model reliability, and preference among different explainability techniques designed to evaluate the practical clinical utility of the system. The survey results indicated that visual explanations improved clinical trust in approximately 70% of the cases. Among the explainability methods, SHAP was identified as the most helpful, receiving 35% of the surgeon endorsements, followed closely by Grad-CAM at 33%.

To provide a comparative perspective, in this section, we summarize relevant studies on osteoporosis detection. Table 5 presents the best predictive results reported in related literature. In the study by Lee *et al.*,⁽¹⁸⁾ the highest reported accuracy was 71%, whereas our proposed framework achieved an improved accuracy of 79%. In terms of sensitivity, their method reached 0.81, whereas ours achieved 0.77; for specificity, their method obtained 0.60, whereas ours improved significantly to 0.79. Although there was a slight decrease of 4% in sensitivity, our framework improved overall performance, particularly with an 8% increase in accuracy and a 19% enhancement in specificity.

Regarding model interpretability, Table 6 summarizes the results of studies incorporating explainable AI techniques. Jang *et al.*⁽¹⁹⁾ and Kumar *et al.*⁽²⁰⁾ both utilized Grad-CAM for visual explanations. However, relying solely on Grad-CAM may introduce bias in the interpretability results. To address this limitation, we additionally employed LIME and SHAP, enabling a multifaceted explanation analysis to mitigate potential unfairness. Moreover, beyond generating case-level visual explanations, our framework statistically aggregated feature regions to refine the system pipeline and improve classification performance. Finally, a web-based voting system was developed to collect orthopedic surgeons' feedback on case explanations, incorporating clinical insights to enhance the overall trustworthiness of the system.

Collectively, the results of these experiments suggest the feasibility of developing an explainable interpretation module within an intelligent medical sensing platform. By treating sensor-based X-ray imaging as a noninvasive sensing modality and coupling it with AI-driven analysis and interpretability techniques, the proposed system provides a practical decision-support tool for osteoporosis screening, particularly in settings where DXA is not readily available.

Building on these findings, there is potential to further integrate the proposed model into portable sensor-based X-ray diagnostic devices, enabling point-of-care osteoporosis screening in

Table 5
Comparative analysis results of prediction performance in related studies.

	Data	Class	Model	<i>Acc</i>	<i>Sen</i>	<i>Spe</i>	<i>F1-score</i>
Lee <i>et al.</i>	334 lumbar spine X-ray images	2	AlexNet = 0.71 VGG16 = 0.63 InceptionV3 = 0.6 ResNet50 = 0.68	0.71	0.81	0.60	0.73
Our proposed method	359 hip X-ray images	2	Densnet121 = 0.72 ResNet50 = 0.71 EfficientNetB3 = 0.7 VGG16 = 0.79	0.79	0.77	0.79	0.77

* *Acc* = Accuracy, *Sen* = Sensitivity, *Spe* = Specificity

Table 6
Comparative results analysis of XAI in related studies.

	Data	Class	Model	Acc	XAI	XAI analysis	Doctors' feedback
Jang <i>et al.</i>	1001 hip X-ray images of female patients over 55 years old	2	VGG16 (nonlocal neural network)	0.812	Grad-CAM	Confusion matrix	Physician discussion
Kumar <i>et al.</i>	457 knee X-ray images	2	Inception = 75.36% MobileNet = 78.26% ResNet50 = 72% VGG19 = 69% Osteonet = 84.06%	0.826	Grad-CAM	N/A	N/A
Our proposed method	359 hip X-ray images	2	DenseNet121 = 0.72 ResNet50 = 0.71 EfficientNetB3 = 0.70 VGG16 = 0.79	0.79	SHAP LIME Grad-CAM	Case discussion + Heatmap feature aggregation	Physician discussion + Physician average trust vote rate: 70%

* Acc = Accuracy

community or home settings. Moreover, the incorporation of Edge AI or real-time AI decision modules can facilitate on-device analysis, minimize latency, and enhance accessibility without relying on centralized computing resources. These advancements highlight the broader applicability of sensor-based AI modules in developing lightweight, explainable, and deployable diagnostic platforms for future medical sensing technologies.

4. Conclusions

In this study, we developed an explainable osteoporosis screening system based on hip X-ray images acquired from high-sensitivity medical imaging sensors. By integrating deep learning classification with explainable AI techniques, we achieved an initial accuracy of 76%. Feature localization using Grad-CAM, SHAP, and LIME identified the femoral neck margins as critical diagnostic regions, and retraining with these features improved accuracy to 79%. Furthermore, a clinical survey with orthopedic surgeons further demonstrated that visual explanations increased decision-making trust in approximately 70% of cases, highlighting the framework’s clinical relevance.

By directly leveraging outputs from advanced sensor-based X-ray imaging systems equipped with state-of-the-art detector materials, our framework can be integrated into next-generation smart medical sensing platforms. Such integration supports the development of portable, radiation-efficient, and real-time osteoporosis screening devices for hospitals, community health centers, and home-based care.

Future work will address current limitations by expanding the dataset to improve generalizability across diverse populations and refining methods to ensure stability and consistency of explanations across different XAI tools. These efforts will further enhance the framework’s readiness for clinical deployment and pave the way for sensor-based, explainable AI-enabled osteoporosis screening in community health programs and telemedicine services, potentially improving early detection rates and reducing fracture-related morbidity.

Acknowledgments

This research is supported by the National Science and Technology Council, Taiwan, under Grant numbers 112-2410-H-A49-087-MY2, 112-2410-H-197-002-MY2, 114-2410-H-A49-066-MY3, and NSTC 114-2221-E-216-002.

References

- 1 United Nations: World Population Ageing, 2019 Highlights: <https://digitallibrary.un.org/record/3846855> (accessed April 2023).
- 2 E.-L. Yong and S. Logan: Singapore Med. J. **62** (2021) 159. <https://doi.org/10.11622/smedj.2021036>
- 3 K. Kerschman-Schindl: Wien. Med. Wochenschr. **166** (2016) 22. <https://doi.org/10.1007/s10354-015-0417-y>
- 4 Y. Jiang, M. Yang, S. Wang, X. Li, and Y. Sun: Cancer Commun. **40** (2020) 154. <https://doi.org/10.1002/cac2.12012>
- 5 S. Wu, K. Roberts, S. Datta, J. Du, Z. Ji, Y. Si, S. Soni, Q. Wang, Q. Wei, Y. Xiang, B. Zhao, and H. Xu: J. Am. Med. Inf. Assoc. **27** (2020) 457. <https://doi.org/10.1093/jamia/ocz200>
- 6 A. Craik, Y. He, and J. L. Contreras-Vidal: J. Neural Eng. **16** (2019) 031001. <https://doi.org/10.1088/1741-2552/ab0ab5>
- 7 A. M. Ismael and A. Şengür: Expert Syst. Appl. **164** (2021) 114054. <https://doi.org/10.1016/j.eswa.2020.114054>
- 8 L. Tanzi, E. Vezzetti, R. Moreno, A. Aprato, A. Audisio, and A. Massè: Eur. J. Radiol. **133** (2020) 109373. <https://doi.org/10.1016/j.ejrad.2020.109373>
- 9 D. Dheeraj, U. Chauhan, M. Khapre, and R. Kant: Cureus (2022). <https://doi.org/10.7759/cureus.23131>
- 10 B. Zhang, K. Yu, Z. Ning, K. Wang, Y. Dong, X. Liu, S. Liu, J. Wang, C. Zhu, Q. Yu, Y. Duan, S. Lv, X. Zhang, Y. Chen, X. Wang, J. Shen, J. Peng, Q. Chen, Y. Zhang, X. Zhang, and S. Zhang: Bone **140** (2020) 115561. <https://doi.org/10.1016/j.bone.2020.115561>
- 11 J. M. Durán and K. R. Jongsma: J. Med. Ethics **47** (2021) 329. <https://doi.org/10.1136/medethics-2020-106820>
- 12 C.-T. Cheng, T.-Y. Ho, T.-Y. Lee, C.-C. Chang, C.-C. Chou, C.-C. Chen, I.-F. Chung, and C.-H. Liao: Eur. Radiol. **29** (2019) 5469. <https://doi.org/10.1007/s00330-019-06167-y>
- 13 V. Vila-Blanco, M. J. Carreira, P. Varas-Quintana, C. Balsa-Castro, and I. Tomás: IEEE Trans. Med. Imaging **39** (2020) 2374. <https://doi.org/10.1109/TMI.2020.2968765>
- 14 L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone: Comput. Methods Programs Biomed. **196** (2020) 105608. <https://doi.org/10.1016/j.cmpb.2020.105608>
- 15 R.-K. Sheu and M. S. Pardeshi: Sensors **22** (2022) 8068. <https://doi.org/10.3390/s22208068>
- 16 B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever: Med. Image Anal. **79** (2022) 102470. <https://doi.org/10.1016/j.media.2022.102470>
- 17 L. Zou, H. L. Goh, C. J. Y. Liew, J. L. Quah, G. T. Gu, J. J. Chew, M. P. Kumar, C. G. L. Ang, and A. W. A. Ta: IEEE Trans. Artif. Intell. **4** (2023) 242. <https://doi.org/10.1109/TAI.2022.3153754>
- 18 S. Lee, E. K. Choe, H. Y. Kang, J. W. Yoon, and H. S. Kim: Skeletal Radiol. **49** (2020) 613. <https://doi.org/10.1007/s00256-019-03342-6>
- 19 R. Jang, J. H. Choi, N. Kim, J. S. Chang, P. W. Yoon, and C.-H. Kim: Sci. Rep. **11** (2021) 19997. <https://doi.org/10.1038/s41598-021-99549-6>
- 20 A. Kumar, R. C. Joshi, M. K. Dutta, R. Burget, and V. Myska: Proc. 2022 45th Int. Conf. Telecommunications and Signal Processing (TSP) (2022) 91. <https://doi.org/10.1109/TSP55681.2022.9851342>