Diagn Interv Radiol 2024; DOI: 10.4274/dir.2023.232116

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

Prediction of osteoporosis using MRI and CT scans with unimodal and multimodal deep-learning models

Yasemin Küçükçiloğlu[1,2]
Boran Şekeroğlu[3]
Terin Adalı[2,4,5]
Niyazi Şentürk[2,4]

[1]Near East University Faculty of Medicine, Department of Radiology, Nicosia, Cyprus

[2]Near East University, Center of Excellence, Tissue Engineering and Biomaterials Research Center, Nicosia, Cyprus

[3]Near East University, Applied Artificial Intelligence Research Center, Nicosia, Cyprus

[4]Near East University Faculty of Engineering, Department of Biomedical Engineering, Nicosia, Cyprus

[5]Sabancı University, Nanotechnology Research and Application Center, İstanbul, Turkey

PURPOSE

Osteoporosis is the systematic degeneration of the human skeleton, with consequences ranging from a reduced quality of life to mortality. Therefore, the prediction of osteoporosis reduces risks and supports patients in taking precautions. Deep-learning and specific models achieve highly accurate results using different imaging modalities. The primary purpose of this research was to develop unimodal and multimodal deep-learning-based diagnostic models to predict bone mineral loss of the lumbar vertebrae using magnetic resonance (MR) and computed tomography (CT) imaging.

METHODS

Patients who received both lumbar dual-energy X-ray absorptiometry (DEXA) and MRI (n = 120) or CT (n = 100) examinations were included in this study. Unimodal and multimodal convolutional neural networks (CNNs) with dual blocks were proposed to predict osteoporosis using lumbar vertebrae MR and CT examinations in separate and combined datasets. Bone mineral density values obtained by DEXA were used as reference data. The proposed models were compared with a CNN model and six benchmark pre-trained deep-learning models.

RESULTS

The proposed unimodal model obtained 96.54%, 98.84%, and 96.76% balanced accuracy for MRI, CT, and combined datasets, respectively, while the multimodal model achieved 98.90% balanced accuracy in 5-fold cross-validation experiments. Furthermore, the models obtained 95.68%–97.91% accuracy with a hold-out validation dataset. In addition, comparative experiments demonstrated that the proposed models yielded superior results by providing more effective feature extraction in dual blocks to predict osteoporosis.

CONCLUSION

This study demonstrated that osteoporosis was accurately predicted by the proposed models using both MR and CT images, and a multimodal approach improved the prediction of osteoporosis. With further research involving prospective studies with a larger number of patients, there may be an opportunity to implement these technologies into clinical practice.

KEYWORDS

Osteoporosis, dual-energy X-ray absorptiometry, lumbar vertebrae, deep learning, multimodal CNN

Corresponding author: Boran Şekeroğlu

E-mail: boran.sekeroglu@neu.edu.tr

Received 19 January 2023; revision requested 12 February 2023; last revision received 20 April 2023; accepted 06 May 2023.

Epub: 13.06.2023

Publication date: 08.01.2023

DOI: 10.4274/dir.2023.232116

Osteoporosis is a systemic skeletal degenerative disease characterized by the deterioration of the microstructure of the bone tissue and low bone mineral density (BMD), with a consequential increase in bone fragility and susceptibility to fracture.[1] The major complication of osteoporosis is fragility fractures that lead to morbidity, mortality, and decreased quality of life. The prevalence of the disease is rising as the proportion of the elderly population increases.[2] It was expected that by 2020, in the United States, approximately 12.3 million individuals older than 50 would have osteoporosis.[3] Tuzun et al.[4] showed that

You may cite this article as: Küçükçiloğlu Y, Şekeroğlu B, Adalı T, Şentürk N. Prediction of osteoporosis using MRI and CT scans with unimodal and multimodal deep-learning models. *Diagn Interv Radiol.* 2024;30(1):9-20.

the prevalence of osteoporosis among Turkish citizens increases with age, with 3%-4% affected at the age of 50 and more than 30% affected by the age of 80. These numbers are predicted to have increased by 64% (870,000 men and 1,841,000 women) in 2035.[4]

Accordingly, screening for osteoporosis is clinically advisable for fracture prevention. There are several imaging techniques, including radiography, ultrasonography, low-dose computed tomography (CT), and dual-energy X-ray absorptiometry (DEXA), which, with its negligible dose of radiation, is considered the gold standard imaging technique for the diagnosis of osteoporosis.[5] Recent research concluded that lumbar spine magnetic resonance imaging (MRI) and CT used for lower-back pain could be used to predict osteoporosis.[6] MRI provide accurate information on tissue structure, and CT images allow researchers to observe the anatomical structure of the vertebrae.[7] However, the advantages of using artificial intelligence (AI) in diagnosing osteoporosis have been rarely studied.[8]

AI techniques have gained significant ground in the field of computer vision, particularly in medical applications.[9,10] As a result, the use of AI has become common in the public health sector and now significantly impacts every aspect of early prediction and primary care.[11,12] Since deep-learning models can detect, learn, and predict indistinct and fuzzy patterns, they provide fast, effective, and reliable outcomes for the considered problem domain.[13] Furthermore, the synthesis and analysis of different images and data types have enabled AI and deep learning to make remarkable improvements in complex data environments in which the human capacity to identify high-dimensional relationships is limited in terms of processing a higher number of data and computational time.[14] However, the modification of recent deep-learning models and the proposal of particular architectures by considering the basic characteristics of specific applications has led to the achievement of more accurate results.

Main points

- This study considered two primary datasets that included magnetic resonance image (MRI) and computed tomography (CT) images and proposed specifically designed convolutional neural network (CNN) models for osteoporosis prediction. The proposed unimodal and multimodal CNN models included two parallelized blocks to extract and combine the loss of individual blocks based on the characteristics of lumbar scan images.
- The proposed unimodal CNN model outperformed the other models in predicting osteoporosis using MRI and CT images separately and obtained 96.54% and 98.84% balanced accuracy, respectively. Superior results were obtained using the proposed multimodal CNN model, and 98.90% balanced accuracy was achieved. Furthermore, a hold-out test set was used to test the models, and the proposed models outperformed the other considered models. Similarly, a superior result was obtained by the multimodal model (97.91%).
- The obtained results showed that the developed deep-learning models could produce accurate results in osteoporosis prediction using different imaging techniques.

Osteoporosis prediction, or identifying the presence of osteoporosis, is one of the primary aims of diagnostic imaging. Several types of AI research have been performed for this purpose,[15] and different imaging modalities and metadata have been considered in these studies. Xu et al.[16] considered micro-CT images for osteoporosis diagnosis. The classification followed several image pre-processing steps by support vector machine and k-nearest-neighbor algorithms. Lim et al.[8] performed machine-learning analyses using DEXA features and abdominopelvic CT to predict osteoporosis prevalence. To classify osteoporosis, Yamamoto et al.[17] trained five pre-trained convolutional neural network (CNN) models using hip radiographs and patient clinical covariates. Similarly, Jang et al.[18] proposed a deep neural network to predict osteoporosis using hip radiography images. Liu et al.[19] proposed a three-layered hierarchical model to distinguish osteoporosis and normal BMD using patients' clinical data. In that study, a logistic regression model achieved superior results by achieving receiver operating characteristics (ROC) area under the curve (AUC) scores of 0.818–0.962 for three layers.

All the abovementioned studies achieved reasonable and promising results. However, the use of MR and CT images in predicting osteoporosis with deep learning requires more investigation. In addition, the impact of the use of multimodal deep-learning models in diagnosing osteoporosis has not been studied adequately.

Based on this information, the current study aimed to accurately distinguish osteoporosis and normal BMD using different imaging modalities, including CT and MRI, to support and assist radiologists in clinical diagnoses. For this purpose, we considered two primary datasets, including lumbar CT and MRIs of patients who received both lumbar DEXA and MRI examinations or CT scans. We proposed a dual-block CNN-based model with different filter sizes and pooling operations and performed several experiments on the considered datasets to achieve a high-accuracy diagnosis of osteoporosis. The efficacy of different modalities on osteoporosis prediction was analyzed by considering CT and MRI scans in separate, combined, and multimodal implementations in unimodal and multimodal experiments. The proposed unimodal and multimodal CNN models were compared with six pre-trained and traditional CNN models.

Methods

Dataset and study population

Study group

Lumbar DEXA examinations of 1,800 patients obtained between January 2018 and March 2021 from the Near East University Hospital's Radiology Department were evaluated retrospectively. A total of 1,554 patients with T-scores higher than –1 at levels L1–L4 and patients with severe scoliosis or lumbar deformity, spondyloarthrosis, inflammatory diseases (tuberculosis, brucella, ankylosing spondylosis, etc.), tumoral lesions (leukemia, lymphoma, multiple myeloma, vertebral metastasis, etc.), or a history of lumbar stabilization surgery were excluded from the study. Spondyloarthrosis can cause sclerosis at the vertebral plateau and osteophyte formations, which are bony spurs with high density. These lesions can cause higher BMD calculations and errors at DEXA examinations. More accurate data was aimed by excluding this patient group. A total of 246 patients with T-scores lower than –1 at levels L1–L4 were re-evaluated for the presence of recorded lumbar MRI or CT images obtained within six months.

The MRI study group consisted of 62 patients (2 males, 60 females), with ages ranging between 44 and 86 years [mean: 65, standard deviation (SD): +/–9.9]. A total of 535 T1-weighted sagittal MRI of these patients were included in the study. The study group for CT consisted of 50 patients (3 males, 47 females), with ages ranging between 46 and 83 years (mean: 68, SD: +/–8.7); 562 sagittal reformatted CT images of these patients were used in the study.

Control group

Lumbar MRI and lumbar CT images of patients, aged 18 to 44, obtained between January 2018 and March 2021 in Near East University Hospital's Radiology Department were evaluated retrospectively. Postmenopausal female patients and male patients over 50 years of age, patients with severe scoliosis or deformity, spondyloarthrosis, inflammatory diseases (tuberculosis, brucella, ankylosing spondylosis, etc.), tumoral lesions (leukemia, lymphoma, multiple myeloma, vertebral metastasis, etc.), or a history of lumbar stabilization surgery, glucocorticoid steroid use, or any disease that may cause secondary osteoporosis were excluded from the study. Furthermore, 526 sagittal T1-weighted MRI of 58 patients (26 males, 32 females, aged 20 to 44 years (mean: 32, SD: +/−8.3) and 534 sagittal reformatted CT images of 50 patients (30 males, 20 females, aged 18 to 44 years (mean: 28, SD: +/−7.6) were used in the study.

The BMDs of the patients were evaluated by DEXA (Lunar DPX, GE, Madison, USA), and MRI examinations were performed using a 1.5-T system (Magnetom Aera, Siemens Healthcare, Erlangen, Germany). The standard lumbar MRI protocol at Near East University Hospital's Radiology Department included sagittal T1- and T2-weighted sequences, sagittal short-tau inversion recovery sequences, and axial T2-weighted sequences. Sagittal T1-weighted images (repetation time: 400 ms, echo time: 7.7 ms, slice thickness: 3.5 mm, slice gap: 0.7 mm, matrix: 256 × 320, field of view: 30 cm) were used in this study. The CT examinations were performed using a 256-detector multislice CT scanner (Somatom Definition Flash, Siemens Healthcare, Erlangen, Germany). Figures 1 and 2 present the data selection procedure of this study for MRI and CT examinations, respectively.

Written informed consent was obtained from all individual participants included in the study.

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Near East University (30.09.2021/YDU/2021/95-1394).

Test set

It is not difficult to distinguish a young person's spine from that of an elderly person. To demonstrate that the dataset was not biased, a hold-out test set was extracted from

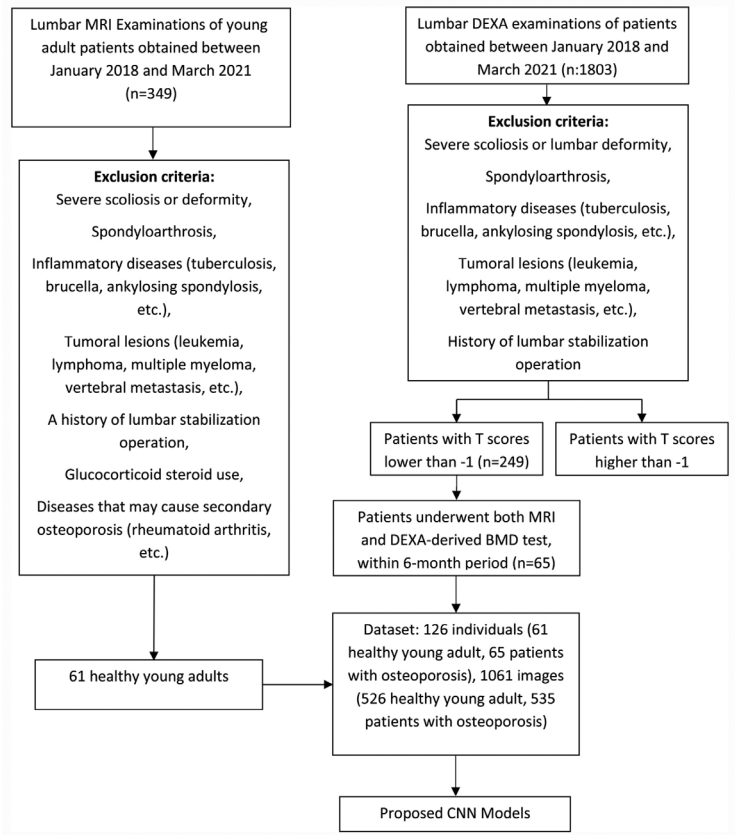

Figure 1. Data selection procedure of the study for magnetic resonance imaging examinations. MRI, magnetic resonance imaging; DEXA, dual-energy X-ray absorptiometry; CNN, convolutional neural network; BMD, bone mineral density.

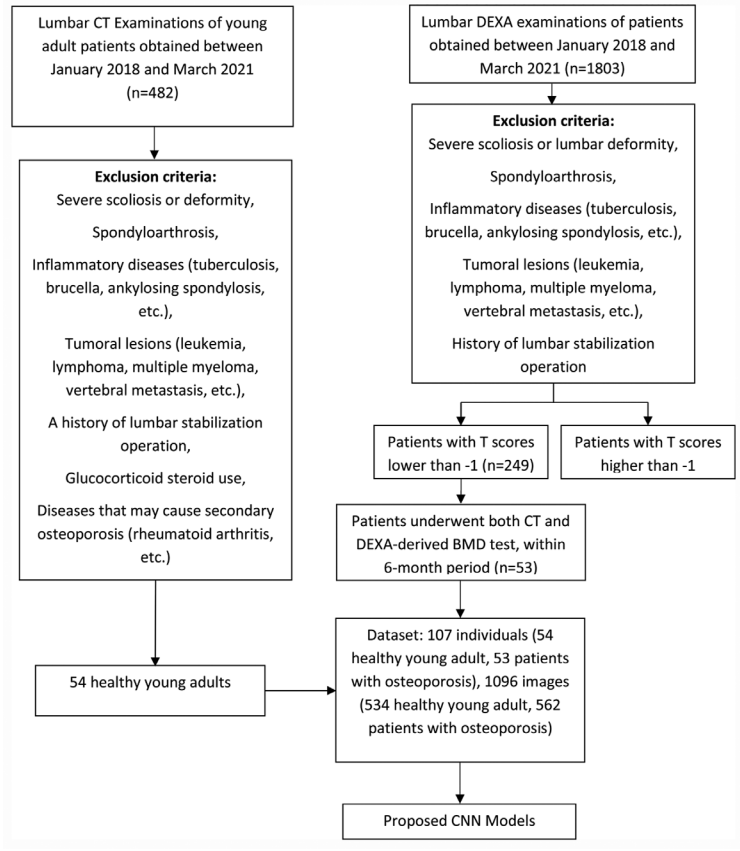

Figure 2. Data selection procedure of the study for computed tomography examinations. CT, computed tomography; DEXA, dual-energy X-ray absorptiometry; CNN, convolutional neural network; BMD, bone mineral density.

the dataset of the CT and MRI scans of the control and patient groups. Close age ranges were selected for both groups to observe the diagnostic abilities of all models and to check possible bias using the minimized age differences between the control and patient groups.

The control group of the set included 24 MRI (age: 39–44 years) and 34 CT scans (age: 40–44 years). The patient group of the set included 24 MRI (age: 44–46 years) and 34 CT scans (age: 46–48 years). Therefore, a total of 116 images were obtained (48 MRI/68 CT, 58 controls/58 patients).

The hold-out test set was not included in any part of the models' training, and additional experiments were performed. All models were trained using the rest of the images in the dataset, and the hold-out test set was used only for testing.

The distributional differences between the training and testing sets were analyzed in the patient and control groups. The skewness of distribution between the training and test groups is presented in Table 1. The skewness-of-distribution results showed that the training and test sets of the CT images were moderately skewed, while the MRI training and testing data had a fairly symmetrical distribution. These results suggest a minimal influence on the generalizability of our models.

Proposed model

A CNN is a deep-learning method that includes feature extraction and classification phases. Typical convolution layers consist of convolution operations using a number of predefined-sized kernels, an activation layer, and a pooling layer. In the classification phase, the extracted features are flattened and fed to the fully-connected layers for classification.

A 3 × 3 filter size effectively extracts low-level features of images with minimal noise; however, the connectivity of the features that provide significant distinguishable patterns according to the image characteristics might be lost by minimizing the filter size. On the other hand, bigger filters, such as 5 × 5, extract more details on superimposed regions using convolutions by considering more spatial pixels of the input images.

The pooling process aims to reduce the number of features by choosing the most informative one among the extracted features to decrease the computational cost of a model; however, relevant features might be eliminated.

In this study, a CNN-based architecture using parallelized dual blocks was designed to extract and combine different levels of features in accordance with the properties of the considered dataset. The first block included two convolutional layers with 3 × 3 filters followed by a 2 × 2 max-pooling operation. Similarly, the second block included two convolutional layers; however, the filter sizes were set to 5 × 5 to consider more spatial pixels in a wider region to detect the connected features. This enabled the extraction of both low- and high-level features and edges of the lumbar vertebrae. In addition, the max-pooling operation was not applied to the convolution layers in the second block, and the feature map size was reduced by shifting the filters by two spatial pixels (stride: 2). Therefore, the features obtained by the commonly used convolutional layers with 3 × 3 filters and a max-pooling operation were added to the features obtained by block 2. This provided new combined features using the variational properties of different blocks. Commonly, 32 filters and the rectified linear unit activation function are considered within block 1 and block 2. Since block 1 of the proposed model focuses on high-level features, such as general shape and intensity values, block 2 was used to extract more significant intensity values and provide more informative low-level features. Each convolutional layer of blocks was followed by batch normalization to avoid over-fitting.

A final convolutional layer was added to the proposed model to apply 5 × 5 filters to the added features and extract their most informative characteristics. Furthermore, the number of filters was increased, and the pooling operation was not considered to feed a fully connected layer with a maximum number of features to provide better convergence. The proposed model consisted of two fully connected layers with 32 and 16 neurons.

There are several approaches to creating multimodal models. One approach uses different planes of a single imaging technique (i.e., the axial, sagittal, and coronal planes of CT scans) as different modalities to create multimodal models. Another approach uses images acquired by different imaging devices in different modalities, such as using CT and MRI scans as separate modalities to implement a multimodal model. It is also possible to create multimodal models using images and text data as different modalities.[7,20] The use of modalities can include information from the same data or independent data for a common task. In addition, the multimodality of the models might include a single model for different modalities or independent models for the fed data.[21,22] As a result, the fusion of different modalities can be performed at the feature, classification, or decision level.[23] The fusion at the feature level includes the process of the different modality images, such as CT and MRI; it also unifies the extracted features and uses multimodal data representation to train a classifier. Conversely, the fusion of data at the classifier level uses the representation of independent features of different modalities in a concatenated feature set to train a multimodal model. Finally, fusing at the decision level trains an independent classifier for different modalities, and the outputs of each classifier are fused for the final decision.[23]

The proposed model was implemented as unimodal and multimodal approaches with common properties. In this study, the multimodality of the model was created by using two imaging techniques, MRI and CT, as separate modalities with two identical unimodal architectures. The loss functions (categorical cross entropy) of general unimodal architectures (L1 and L2) were used to determine the final loss (L3) of the multimodal model. Therefore, the multimodal model provided the common convergence of the CT and MRI scans and allowed us to test the model using either both modalities or a single modality simultaneously. The formula of binary cross-entropy is given in equation (1):

$$L_x = BCE = -(y_x \log(p_x) + (1 - y_x) \log(1 - p_x)) \quad (1)$$

where x represents the CT or MRI modality, and y and p denote the target and predicted classes. The final loss of the model is calculated as given in equation (2):

$$L_f = L_C + L_M, \quad (2)$$

Table 1. Skewness of distribution between the training and test groups

Group	Skewness of distribution
CT control group	0.519
CT patient group	–0.745
MRI control group	–0.037
MRI patient group	–0.129

CT, computed tomography; MRI, magnetic resonance imaging.

where L_f, L_c, and L_m denote the final, CT, and MRI modality losses. The general architecture of the proposed model with both unimodal and multimodal phases is shown in Figure 3.

Experimental design

This section presents the experimental design, validation strategy, and the considered evaluation metrics in detail. The experiments were performed in four stages to observe the accuracy of osteoporosis diagnosis using different imaging modalities.

The first two experimental stages involved training the proposed unimodal model using MRI and CT images, respectively. Considering separated MRI and CT modalities provided to analyze the effect of different medical imaging systems on osteoporosis diagnosis. The third stage was performed using the combined dataset created by shuffling both the CT and MRI datasets using the unimodal model. Finally, the fourth stage was implemented using the proposed model as a multimodal approach; CT and MRI scans were fed separately to the model and trained together.

A patient-based experiment was also performed to test the predictive ability of the models. A total of 27 patients (6 CT control + 6 CT osteoporosis + 8 MRI osteoporosis + 7 MRI control patients) and 228 corresponding images (57 for each of the control and osteoporosis groups) were extracted to represent different age samples from the dataset. The patient numbers varied because of the number of extracted slices of each patient. The unimodal and multimodal models were trained using the rest of the dataset independently to analyze the efficacy of the models for patient-based data.

The proposed model was compared in three stages to the traditional CNN model and six benchmark and recent pre-trained networks (EfficientNet B0,[24] InceptionV3,[25] ResNet50,[26] InceptionResNetV2,[27] EfficientNetV2S,[28] and ConvNeXt Tiny[29]) to demonstrate the efficiency of the proposed model.

The architecture of the traditional CNN model was determined after performing several experiments. The experiments were performed by adding and removing convolutional layers and pooling operations and by increasing and decreasing filter sizes and strides systematically and iteratively using MRI data. The architecture that produced the best results was considered in the comparisons. The final architecture of the traditional CNN included two convolutional layers (64 and 32 filters, respectively) and two fully connected layers (128 and 64 neurons, respectively). The abovementioned pre-trained networks were trained by adding a fully connected layer with 128 neurons for each and by using ImageNet weights. Therefore, the transfer-learning approach was used to transfer the acquired knowledge of the models to the diagnosis of osteoporosis.

All experiments were performed using 5-fold cross-validation to obtain consistent results. Cross-validation allowed the models to consider all images both in the training and testing phases. The models were trained five times, where the 4 of the folds were used for training, and the rest fold was used for testing. Therefore, the data dependency of the models was minimized in the evaluation. The data selection in the folds was performed randomly, and the final evaluation of the results was performed using the mean or sum of the correctly predicted samples ob-

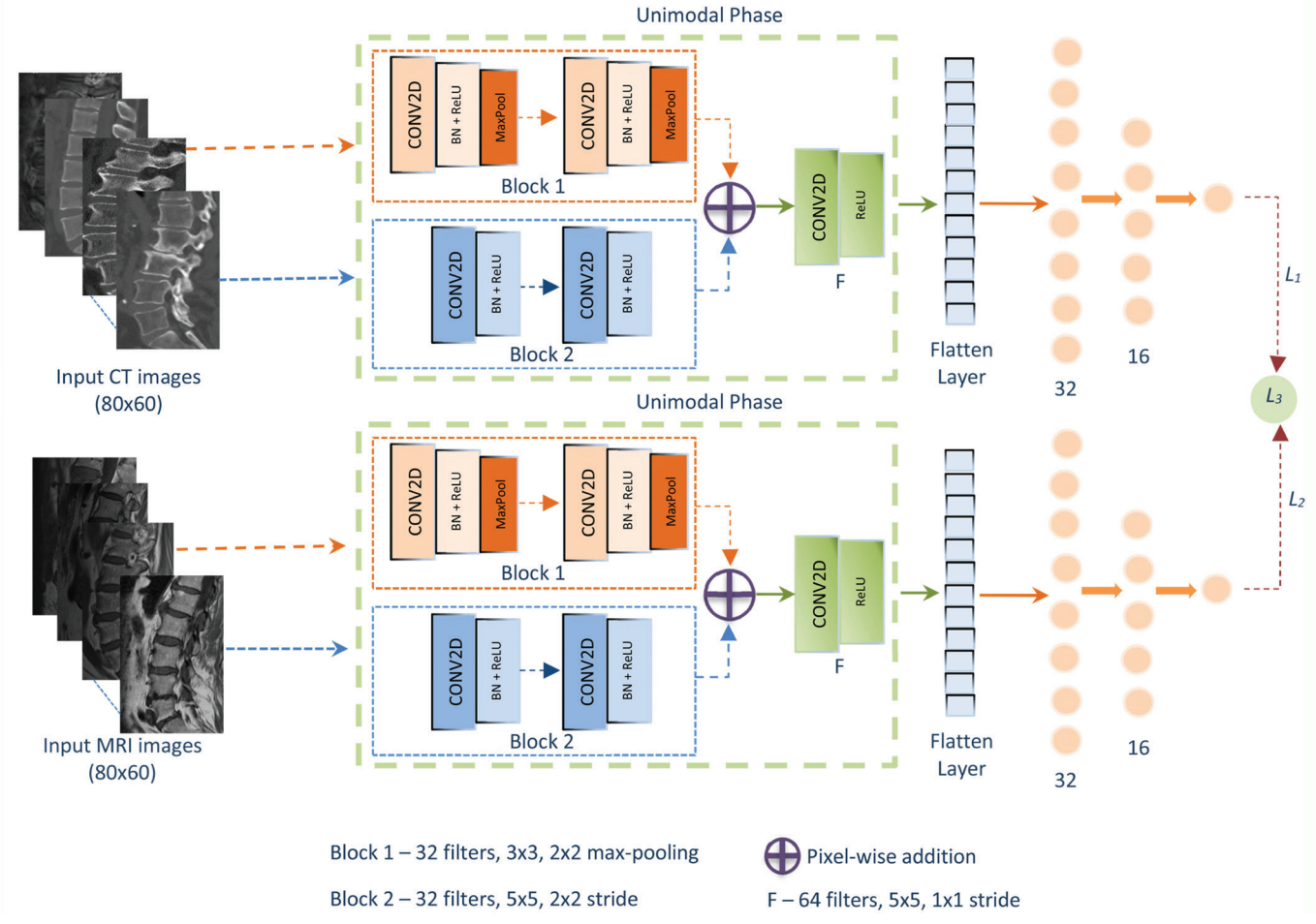

Figure 3. The multimodal architecture of the proposed convolutional neural network model for osteoporosis prediction.

tained in each fold. Data augmentation was not applied in the experiments.

Even though the datasets were balanced in this study, in which AI algorithms converged effectively,[30] we considered five different evaluation metrics–specificity, sensitivity, accuracy, ROC AUC score, and balanced accuracy for robust model evaluation.[31] Additionally, 95% confidence intervals (CIs) were also provided for accuracy, sensitivity, and specificity. Accuracy was the primary evaluation metric of the classification tasks and provided a reliable measurement for balanced datasets. Sensitivity and specificity were used to measure the models' abilities to predict a dataset's positive and negative samples. Additionally, balanced accuracy was considered to measure the minor variations caused by the minimal number of output differences in the datasets; it measures the general classification ability of the model by calculating the average of sensitivity and specificity metrics and eliminates the effect of different-sized datasets. The formula for balanced accuracy is shown in equation (3):

$$\text{Balanced } Accuracy = \frac{Sensitivity + Specificity}{2}, \quad (3)$$

where

$$Sensitivity = \frac{TP}{TP + FN}, \quad (4)$$

and

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

The terms *TN*, *TP*, *FP*, and *FN* denote the true negative, true positive, false positive, and false negative samples predicted by the models. The ROC AUC score determines how the model distinguishes the positive and negative classes efficiently. The higher the score, the better the performance.

The abovementioned test set was used to test all models after training using all the data from the dataset. This allowed us to observe the diagnostic ability of all models and to check possible bias using the minimized age differences between the control and patient groups. Table 2 shows the details of the experimental stages.

Statistical analysis

The adaptive moment estimation optimizer was used in all models, and the proposed models were trained for 20 epochs using 16 batch sizes. The models were implemented on a Windows 11 PC with an Intel® Core™ i7-9750H CPU, 16 GB memory, 1,660 GTX GPU, and Python 3.8.16 using TensorFlow Keras 2.11.0.

Results

Results of stage-1 experiments

In the stage-1 experiments, all models were trained using MRI. The InceptionV3 and EfficientNetV2S models obtained the lowest results for all evaluation metrics considered in this study. The InceptionResNetV2 model produced more stable results than InceptionV3. ResNet50 achieved the highest specificity rate of the study, with 97.45% (95% CI for specificity: 96.1%–98.8%). However, it was observed that difficulties occurred in learning both classes, and the sensitivity rate of the model was measured as 88.84% (95% CI for sensitivity: 86.1%–91.6%). This caused a decrease in the general ability of the model. The ResNet50 model achieved 93.18% accuracy (95% CI: 91.6%–94.7%), a ROC AUC score of 0.932, and 93.15% balanced accuracy. The CNN model achieved more reliable and consistent results than the abovementioned models, with 94.62% sensitivity, 95.10% specificity, 0.948 ROC AUC score, and 94.86% balanced accuracy. The overall prediction ability of the CNN was measured as 94.86% (95% CI: 93.5%–96.2%) accuracy. Even though the EfficientNet B0 model obtained the highest results compared with the other pre-trained models within the transfer-learning experiments, it did not outperform the unimodal implementation of the proposed model. It achieved 94.22%, 96.86%, 95.55% (95% CI: 94.3%–96.8%), 0.955, and 95.54% for sensitivity, specificity, accuracy, ROC AUC score, and balanced accuracy, respectively, while the proposed model achieved the highest sensitivity (96.01%), ROC AUC score (0.965), and balanced accuracy (96.54%) results of this study. The overall prediction ability of the proposed method was 96.54% (95% CI: 95.4%–97.7%). The ConvNeXt Tiny model achieved the same and highest sensitivity rate as the proposed method; however, it did not show the same performance for the other metrics. Table 3 presents the results of stage-1 in detail.

Results of stage-2 experiments

In this stage, CT images were used for training. All models increased the ability of osteoporosis prediction using CT images compared with the stage-1 experiments. However, the ResNet50 model obtained the lowest specificity (93.60% with 95% CI: 91.6%–94.7%), ROC AUC score (0.942), and balanced accuracy rates (94.24%). Even though the InceptionV3 model obtained higher specificity and balanced accuracy than ResNet50, it obtained a minimum sensitivity of 94.69% (95% CI for sensitivity: 92.8%–96.6%). The proposed model achieved superior results and outperformed all the models considered in this study (98.48%, 99.20%, 0.988, and 98.84% for sensitivity, specificity, ROC AUC score, and balanced accuracy, respectively). The overall accuracy of the proposed method was 98.84% (95% CI for sensitivity: 92.8%–96.6%).In contrast to its performance using MRI, the EfficientNetV2S model achieved 99.20% specificity (95% CI for specificity: 98.4%–100%); however, it did not obtain sufficient results to outperform the proposed model and the InceptionV3 model in other metrics. Similarly, the ConvNeXt Tiny model obtained the same specificity as both the proposed method and EfficientNetV2S. The EfficientNetV2S and InceptionResNetV2 models followed the proposed model for all metrics. Table 4 presents the results of stage-2 in detail.

Results of stage-3 experiments

In this stage, CT and MRI were combined as input patterns in training without distinguishing the differences. The ResNet50, InceptionV3, and InceptionResNetV2 models did not produce reasonable results compared with the other models. The CNN model and EfficientNet B0 obtained similar and relatively higher balanced accuracies of 95.74% and 95.73% (95% CI: 94.9%–96.6%), respectively. However, the conventional CNN mod-

Table 2. Details of the experiments and stages

Stage no	Exp. name	Image set	Validation	# of images	# of trained models	Type
1	Exp. 1 (MRI)	MRI	5-fold CV	1,013	6	Unimodal
2	Exp. 2 (CT)	CT	5-fold CV	1,028	6	Unimodal
3	Exp. 3 (Com)	Combined	5-fold CV	2,041	6	Unimodal
4	Exp. 4 (MM)	MRI + CT	5-fold CV	2,004	1	Multimodal
5	Exp. 5	Hold-out test set	-	116	9	Both
6	Patient-based	MRI + CT	Hold-out	2,004		Both

Exp., experiment; MRI, magnetic resonance imaging; CT, computed tomography; Com, combined; MM, multimodal; CV: cross-validation.

el produced a higher specificity rate, while the EfficientNet B0 model obtained a more accurate sensitivity rate. Similar to the previous experiments, the proposed model outperformed all models and achieved 96.69%, 96.83%, and 0.968 for sensitivity, specificity, and ROC AUC score. It realized a balanced accuracy of 96.76% (95% CI: 96.0%–97.5%). The EfficientNetV2S and ConvNeXt Tiny models obtained 96.25% and 95.98% accuracy, with 95% CIs of 95.4%–97.1% and 95.0%–96.7%, respectively, and it followed the proposed model. Table 5 presents the results of stage-3 in detail.

Results of stage-4 multimodal experiments

The proposed model was implemented as a multimodal approach in the multimodal experiments, and CT and MRI were fed to the model in different modalities. The models' training was performed using a total of 2,004 CT and MRI in separate unimodal blocks and fused at the feature level after the feature extraction process of independent unimodal blocks. Fusing the different modalities of osteoporosis images allowed us to test the proposed system on 1,002 images in a 5-fold cross-validation. Even though the model's training consisted of fewer samples than used in the combined datasets of stage-3, the multimodal approach achieved higher results in all metrics: 98.61%, 99.20%, 0.989, and 98.90% (95% CI: 98.4%–99.4%) for sensitivity, specificity, ROC AUC score, and balanced accuracy, respectively.

This experiment enabled us to observe the efficacy of feeding models with different image modalities instead of combining them into a single dataset. The obtained results showed that the multimodal image approach produced higher rates and was more effective in predicting osteoporosis. Table 6 presents the results obtained by the multimodal approach.

Figures 4 and 5 present the Grad-CAM++[32] and saliency maps (using SmoothGrad[33]) of the multimodal model for correctly predicted osteoporosis patients using MRI and CT scans, respectively. The figures show that the model focused on the lumbar vertebrae as expected to predict osteoporosis.

Results of stage-5 test-set experiments

All models, including the proposed multimodal and unimodal models (CT, MRI, and combined) and comparison models, were trained using the rest of the dataset and tested using the test set. In the CT experiments, the CT images of the test set were considered in the generalization phase, while in the MRI experiments, only the MRI were fed to the models. However, both modalities of the test sets were considered in the proposed combined unimodal and multimodal experiments. Table 7 presents the results obtained by all the models using the test set.

Additionally, the analysis of the prediction scores and DeLong statistical tests[34] were performed to evaluate the models' decision-making strengths and prediction capabilities and to compare the models' AUC scores statistically. As the same training and testing data were included for all models, a hold-out test was used for these analyses. The scores of correctly classified samples were considered, and the mean scores and SD were calculated for the unimodal and multimodal models. Even though all the models achieved reasonable scores, the results suggest that using a multimodal model increased the prediction scores and provided a more effective prediction of osteoporosis. Table 8 presents the mean and SD results obtained for each model using the hold-out test set.

Table 3. Results of stage-1 magnetic resonance imaging experiments

Model	Accuracy* (%)	Sensitivity* (%)	Specificity* (%)	Balanced accuracy (%)	ROC AUC score
Proposed model	**96.54 (95.4–97.7)**	**96.01 (94.3–97.7)**	97.06 (95.6–98.5)	**96.54**	**0.965**
CNN model	94.86 (93.5–96.2)	94.62 (81.5–87.8)	95.10 (93.2–97.0)	94.86	0.948
EfficientNet B0	95.55 (94.3–96.8)	94.22 (92.2–96.3)	96.86 (95.4–98.4)	95.54	0.955
ResNet50	93.18 (91.6–94.7)	88.84 (86.1–91.6)	**97.45 (96.1–98.8)**	93.15	0.932
InceptionV3	84.20 (82.0–86.5)	76.29 (72.6–80.0)	91.97 (89.6–94.3)	84.13	0.842
InceptionResNetV2	92.69 (91.1–94.3)	90.43 (87.9–93.0)	94.91 (93.0–96.8)	92.67	0.928
EfficientNetV2S	84.22 (82.1–86.6)	73.22 (69.5–76.93)	95.16 (93.3–97.1)	84.19	0.842
ConvNeXt Tiny	95.31 (94.1–96.6)	**96.01 (94.3–97.7)**	94.62 (92.7–96.5)	95.31	95.32

*Values in parentheses indicate a 95% confidence interval; CNN, convolutional neural network; ROC, receiver operating characteristics; AUC, area under the curve.

Table 4. Results of stage-2 computed tomography experiments

Model	Accuracy* (%)	Sensitivity* (%)	Specificity* (%)	Balanced accuracy (%)	ROC AUC score
Proposed model	**98.83 (98.2–99.5)**	**98.48 (97.4–99.5)**	**99.20 (98.4–100)**	**98.84**	**0.988**
CNN model	98.15 (97.3–99.0)	97.53 (96.2–98.9)	98.80 (97.8–99.8)	98.16	0.981
EfficientNet B0	97.85 (97.0–98.7)	97.15 (95.7–98.6)	98.61 (97.6–99.6)	97.87	0.978
ResNet50	94.26 (92.8–95.7)	94.88 (93.0–96.8)	93.65 (91.5–95.7)	94.24	0.942
InceptionV3	94.55 (93.2–95.9)	94.69 (92.8–96.6)	94.44 (92.4–96.4)	94.54	0.945
InceptionResNetV2	98.63 (97.9–99.3)	98.29 (97.2–99.4)	**99.00 (98.1–99.9)**	98.64	0.986
EfficientNetV2S	98.50 (97.8–99.4)	97.80 (96.4–99.2)	**99.20 (98.4–100)**	98.50	0.985
ConvNeXt Tiny	96.68 (95.6–97.6)	94.18 (92.4–96.2)	**99.20 (98.4–100)**	96.69	0.967

*Values in parentheses indicate a 95% confidence interval; CNN, convolutional neural network; ROC, receiver operating characteristics; AUC, area under the curve.

Table 5. Results of stage-3 combined experiments

Model	Accuracy* (%)	Sensitivity* (%)	Specificity* (%)	Balanced accuracy (%)	ROC AUC score
Proposed model	**96.76 (96.0–97.5)**	**96.69 (95.6–97.8)**	**96.83 (95.8–97.9)**	**96.76**	**0.968**
CNN model	95.73 (94.9–96.6)	95.33 (94.0–96.6)	96.14 (95.0–97.3)	95.74	0.957
EfficientNet B0	95.73 (94.9–96.6)	95.82 (94.6–97.0)	95.65 (94.4–96.9)	95.73	0.957
ResNet50	89.75 (88.4–91.1)	95.52 (94.3–96.8)	83.89 (81.6–86.2)	89.71	0.897
InceptionV3	89.75 (88.4–91.1)	88.04 (86.1–90.0)	91.50 (89.8–93.2)	89.77	0.899
InceptionResNetV2	92.65 (91.5–93.8)	88.62 (86.7–90.6)	96.73 (95.6–97.8)	92.68	0.926
EfficientNetV2S	96.25 (95.4–97.1)	96.20 (95.2–97.5)	96.28 (95.2–97.3)	96.24	0.962
ConvNeXt Tiny	95.98 (95.0–96.7)	95.94 (94.7–97.2)	96.03 (95.0–97.1)	95.99	0.960

*Values in parentheses indicate a 95% confidence interval; CNN, convolutional neural network; ROC, receiver operating characteristics; AUC, area under the curve.

Table 6. Results of the proposed model in the multimodal experiment

Model	Accuracy* (%)	Sensitivity* (%)	Specificity* (%)	Balanced accuracy (%)	ROC AUC score
Proposed model	98.90 (98.4–99.4)	98.61 (97.9–99.3)	99.20 (98.6–99.8)	98.90	0.989

*Values in parentheses indicate a 95% confidence interval; ROC, receiver operating characteristics; AUC, area under the curve.

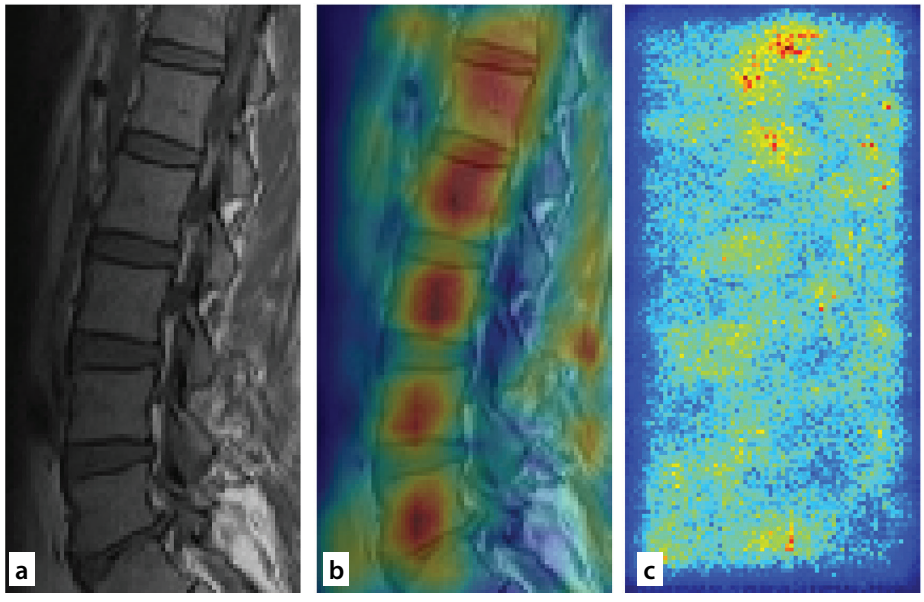

Figure 4. Grad-Cam++ and saliency visualization of the multimodal model for the magnetic resonance image (MRI) of correctly predicted osteoporosis: **(a)** original MRI, **(b)** Grad-CAM++, and **(c)** saliency map using SmoothGrad.

The *P* values obtained by performing the DeLong statistical test showed no significant differences; however, the multimodal model was slightly superior to the unimodal models. Table 9 presents the DeLong statistical test results.

Results of patient-based experiments

The proposed models were tested using representative patients. The unimodal model for CT scans achieved 100% specificity by predicting the control group correctly and 97.73% (95% CI: 88.9%–100%) sensitivity. The predictions of five patients were obtained accurately (100%); however, 70% (95% CI: 41.6%–98.4%) accuracy was obtained for a single patient.

The results of the unimodal model for patient-based MRI scans achieved slightly lower accuracy than CT scans, with 85.96% (95% CI: 76.9%–95.00%) sensitivity and 91.22% (95% CI: 83.9%–98.6%) specificity. Three of seven patients were accurately (100%) predicted for osteoporosis, and the remaining patients were predicted at between 66.67% (95% CI: 28.9%–100%) and 85.71% (95% CI: 59.8%–100%) accuracy.

The multimodal model obtained higher scores for MRI data and the same for CT data. The results for MRI data were 89.47% sensitivity (95% CI: 81.5%–97.4%) and 96.49% (95% CI: 91.7%–100%) specificity. The results obtained by the multimodal model for MRI and CT images were 92.98% (95% CI: 88.3%–97.7%) sensitivity and 98.24% (95% CI: 95.8%–100%) specificity. The accuracy was 95.61% (95% CI: 93.0%–98.3%). Table 10 shows the obtained patient-based results in detail.

Sample images, demo codes, and notebook implementations are available at: https://github.com/BoranSekeroglu/OSTEO_MODELS

Discussion

The obtained results are discussed in different sections to analyze the effect of different imaging modalities and multimodal approaches in predicting osteoporosis.

All the deep-learning models predicted osteoporosis at different rates. However, the proposed 5 × 5 convolutions in block 2 were used to predict more connected features at the vertebral bone boundaries. Adding the obtained features from both blocks and the consideration as new features enabled the proposed unimodal model to produce superior results compared with the other considered models. In addition, the multimodal approach of the proposed model resulted in superior prediction rates.

Although the pre-trained CNN models obtained fluctuating results in different experiments, it was observed that the ConvNeXt, EfficientNetV2S, EfficientNet B0, and InceptionResNetV2 models produced consistent and stable results, even though they did not outperform the proposed model. The fluctuations were considered to determine the consistency and stability of the models.

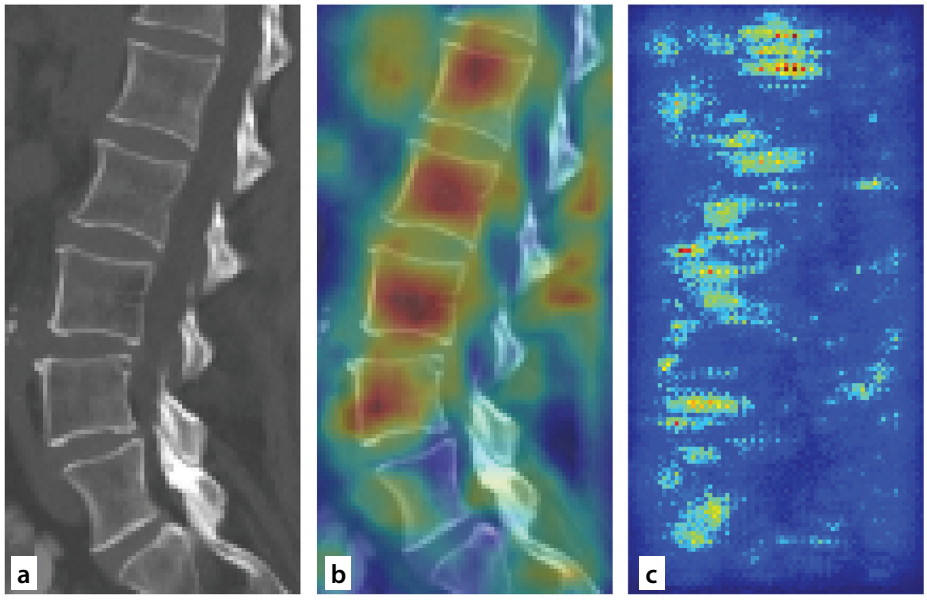

Figure 5. Grad-Cam++ and saliency visualization of the multimodal model for the computed tomography image of correctly predicted osteoporosis: **(a)** original computed tomography image, **(b)** Grad-CAM++, and **(c)** saliency map using SmoothGrad.

Table 7. Results of stage-5 test-set experiments

Model	Accuracy* (%)	Sensitivity* (%)	Specificity* (%)	Balanced accuracy (%)	ROC AUC score
CNN MRI	83.33 (72.8–93.9)	79.16 (62.9–95.4)	87.50 (74.3–100)	83.33	0.833
CNN CT	85.29 (76.9–93.7)	82.35 (69.5–95.2)	88.23 (77.4–99.1)	85.29	0.854
CNN combined	87.06 (81.0–93.2)	86.20 (77.3–95.1)	87.93 (79.5–96.3)	87.06	0.871
Efficient B0 MRI	93.75 (86.9–100)	91.66 (80.6–100)	95.83 (87.8–100)	93.75	0.978
Efficient B0 CT	92.64 (86.4–98.9)	91.17 (81.6–100)	94.11 (86.2–100)	92.64	0.927
Efficient B0 combined	92.24 (87.4–97.1)	91.37 (84.2–98.6)	93.10 (86.6–99.6)	92.24	0.924
ResNet50 MRI	77.08 (65.2–89.0)	79.16 (62.9–95.4)	75.00 (57.7–92.3)	77.08	0.771
ResNet50 CT	82.35 (73.3–91.4)	79.41 (65.8–93.0)	85.29 (73.4–97.2)	82.35	0.824
ResNet50 combined	83.05 (77.9–91.1)	81.03 (70.9–91.1)	85.00 (79.5–96.3)	83.01	0.840
InceptionV3 MRI	79.16 (67.7–90.7)	75.00 (57.7–92.3)	83.33 (68.4–98.2)	79.16	0.792
InceptionV3 CT	85.29 (76.9–93.7)	85.29 (73.4–97.2)	85.29 (73.4–97.2)	85.29	0.823
InceptionV3 combined	83.05 (77.9–91.1)	82.75 (73.0–92.5)	83.33 (77.3–95.1)	83.04	0.831
InceptionResNetV2 MRI	85.41 (75.4–95.4)	83.33 (68.4–98.2)	87.50 (74.3–100)	85.41	0.854
InceptionResNetV2 CT	94.11 (88.5–99.7)	94.11 (86.2–100)	94.11 (86.2–100)	94.11	0.942
InceptionResNetV2 combined	87.28 (83.1–94.5)	89.65 (81.8–97.5)	85.00 (79.5–96.3)	87.32	0.874
EfficientNetV2S MRI	93.75 (86.9–100)	88.26 (78.6–100)	96.91 (89.9–100)	93.75	0.978

The 5% change in specificity and sensitivity results assumed that the models were focused on a particular training data class during convergence, and their generalization ability was inconsistent on the test folds. The traditional CNN demonstrated that the correct determination of the architecture might cause higher classification rates than with the pre-trained models. However, determining the correct architecture is challenging and time-consuming, and it increases the computation cost of the studies.

The proposed model provided learning with few trainable parameters compared with other models and minimized the computational cost in terms of training time. Contrary to the high computational cost of the pre-trained models (i.e., 37 sec/epoch for EfficientNet B0 and 76 sec/epoch for ResNet50), the computation cost of the proposed unimodal, multimodal, and CNN models was an average of 1.2, 2.1, and 5.8 sec/epoch with the GPU, respectively.

Recent studies showed that obtaining above 74% for sensitivity, specificity, and accuracy in osteoporosis prediction is possible using AI and deep-learning models with different imaging modalities, such as CT, DEXA, and micro-CT.[8,15-18,35-37]

The results obtained using MRI of our study (stage-1) showed that MRI could be used effectively to predict osteoporosis with 96.54% accuracy (95% CI: 95.4%–97.7%) and eliminate the side effects of radiation-emitting devices.

On the other hand, the use of CT images resulted in more accurate results (98.84%), as they captured detailed anatomical structures. Combining CT and MRI without distinguishing the feature extraction process provided a limited contribution to osteoporosis prediction (96.76%). Even though a slight improvement occurred in the results compared with the results of the MRI experiments, there was a 2% decrease in the results obtained using CT images. However, the significant ability of deep-learning models in the feature extraction process provided superior results (98.90%) than those obtained in all experiments separately or combined using a multimodal approach. The consideration of both CT and MRI in individual unimodal blocks and the use of the loss of separate unimodal blocks to train the multimodal model provide further prediction contributions.

One of the most important outcomes of the experiments is that the trained model predicted osteoporosis using MRI or CT

Table 7. Continued

EfficientNetV2S CT	92.60 (86.3–98.8)	92.67 (83.1–100)	95.92 (88.0–100)	92.58	0.925
EfficientNetV2S combined	93.40 (87.2–99.6)	93.20 (86.1–100)	93.60 (87.1–100)	93.40	0.934
ConvNeXt Tiny MRI	93.75 (86.9–100)	91.66 (80.6–100)	95.83 (87.8–100)	93.75	0.978
ConvNeXt Tiny CT	90.56 (83.3–98.4)	85.29 (73.4–97.2)	95.83 (87.8–100)	90.56	0.906
ConvNeXt Tiny combined	90.55 (85.7–95.4)	93.20 (86.1–100)	87.93 (79.5–96.3)	90.56	0.905
Proposed unimodal MRI	95.83 (90.2–100)	91.66 (80.6–100)	100.00 (100–100)	95.83	0.959
Proposed unimodal CT	97.05 (93.0–100)	97.05 (91.4–100)	97.05 (91.4–100)	97.05	0.971
Proposed unimodal combined	95.68 (92.0–99.4)	94.82 (89.1–100)	96.55 (91.9–100)	95.68	0.957
Proposed multimodal	97.91 (95.1–100)	97.91 (95.1–100)	97.91 (95.1–100)	97.91	0.979

*Values in parentheses indicate a 95% confidence interval; MRI, magnetic resonance image, CT, computed tomography, ROC, receiver operating characteristics; AUC, area under the curve.

Table 8. Mean and standard deviation results of the prediction scores of the hold-out test set

Model	Mean ROC AUC score	Standard deviation
Proposed unimodal MRI	0.921	0.084
Proposed unimodal CT	0.952	0.056
Proposed unimodal combined	0.926	0.071
Proposed multimodal	0.965	0.039

MRI, magnetic resonance imaging; CT, computed tomography; ROC, receiver operating characteristics; AUC, area under the curve.

Table 9. Model comparison using the DeLong test for two correlated receiver operating characteristics area under the curve scores of the stage-5 test-set experiments

	Multimodal MRI vs. unimodal MRI	Multimodal CT vs. unimodal CT
Z value	0.574	0.444
P value	0.566	0.657
Confidence intervals	−0.021–0.038	−0.029–0.047

MRI, magnetic resonance imaging; CT, computed tomography.

Table 10. Results of patient-based experiments

Metric*	Unimodal CT	Unimodal MRI	Multimodal
Total accuracy (image-based)	97.38% (95% CI: 94.4%–100%)	88.59% (95% CI: 82.8%–94.4%)	95.61% (95% CI: 93.0%–98.3%)
Total sensitivity (image-based)	97.73% (95% CI: 88.9%–100%)	85.96% (95% CI: 76.9%–95.00%)	92.98% (95% CI: 88.3%–97.7%)
Total specificity (image-based)	100%	91.22% (95% CI: 83.9%–98.6%)	98.24% (95% CI: 95.8%–100%)
Maximum accuracy (patient group)	100%	100%	100%
Maximum accuracy (control group)	100%	100%	100%
Minimum accuracy (patient group)	70% (95% CI: 41.6%–98.4%)	66.67% (95% CI: 28.9%–100%)	87.5% (95% CI: 64.6%–100%)
Minimum accuracy (control group)	100%	77.78% (95% CI: 50.6%–100%)	88.88% (95% CI: 68.4%–100%)

*Total accuracy, total sensitivity, and total specificity indicate the image-based results of patients and the control group. The maximum and minimum scores indicate the highest and lowest scores from independent patient and control group analyses. MRI, magnetic resonance imaging; CT, computed tomography; CI, confidence interval.

images in a multimodal model, as the MRI and CT images used in this study were not obtained from the same patients. This will protect patients from being exposed to radiation from different imaging techniques. We believe that our study and its results demonstrate the efficiency of using deep-learning models, particularly the proposed unimodal and multimodal CNN models, in predicting osteoporosis more accurately.

Furthermore, using the hold-out test set with minimized age differences between the control and patient groups as well as patient-based experiments allowed us to observe the main prediction capabilities of the models and avoid any bias in the test data. As a result, all the proposed models outperformed the other models considered, and the unimodal models achieved 95.83% (95% CI: 90.2%–100%) and 97.05% (95% CI: 93.0%–100%) balanced accuracy on the hold-out test set for MRI and CT, respectively. Even though a slight decrease was observed in the combined unimodal model (95.68%), it also achieved higher scores than the other models. However, the multimodal model obtained superior results by obtaining 97.91% (95% CI: 95.1%–100%) balanced accuracy.

The patient-based analysis showed that the proposed models accurately predicted osteoporosis with superior specificity, particularly with CT images. However, the multimodal model provided better prediction ability with MRI. This proved once again that the combined use of different imaging modalities and the independent extraction of features during training improved the prediction capability of CNNs and might provide more accurate support for radiologists.

Even though the best prediction of osteoporosis was obtained using CT images in the unimodal experiments, considering MRI in unimodal and multimodal models could prevent patients from being exposed to radiation and assist radiologists in diagnosing osteoporosis.

This study has limitations. Patients with T scores higher than -1 and BMD levels within normal limits were excluded. This patient group could not be used as a control group because of the small number of patients. To protect non-indicated patients from DEXA-induced X-ray exposure, the DEXA data of the control group were not obtained. We are aware that the ideal scenario would have been to obtain the DEXA data of the control group, but we felt it would have been unethical. The control group was selected among pre-menopausal female patients and male patients under 50 years of age, and

patients with diseases and/or drug use that may cause secondary osteoporosis were excluded from the study. To demonstrate that the dataset was not biased, a hold-out test set was extracted from the dataset. However, the developed system was not tested with an external dataset, and the use of the proposed models in clinical practice requires further investigation.

Eliminating the abovementioned limitations might lead to more robust findings in further studies.

In conclusion, we considered two primary datasets that included MRI and CT images and proposed specifically designed CNN models for osteoporosis prediction. Several experiments were performed, and the obtained results were compared with those of the traditional CNN and six benchmark pre-trained models using the transfer-learning approach. The proposed unimodal CNN model outperformed the other considered models in predicting osteoporosis using MRI and CT images separately and obtained 96.54% and 98.84% balanced accuracy, respectively. Superior results were obtained using the proposed multimodal CNN model, and 98.90% balanced accuracy was achieved. Furthermore, a hold-out test and patient-based experiments were used to test the models, and the proposed models achieved superior results.

The obtained results showed that the developed deep-learning models could produce accurate results in osteoporosis prediction using different imaging techniques. However, considering MRI images, even in unimodal and multimodal models, could minimize DEXA and CT use and prevent patients from being exposed to radiation. Our future work will include risk assessment using MRI scans, and further studies might focus on increasing the accuracy obtained in this study using more patient data.

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Lungdahl BL. Overview of treatment approaches to osteoporosis. *Br J Pharmacol*. 2021;178(9):1891-1906. [CrossRef]
2. Burden AM, Tanaka Y, Ha XC,et al. Osteoporosis case ascertainment strategies in European and Asian Countries: a comparative review. *Osteoporos Int*. 2021;32(5):817-829. [CrossRef]
3. US Preventive Services Task Force; Curry SJ, Krist AH, et al. Screening for Osteoporosis to Prevent Fractures: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2018;319(24):2521-2531. [CrossRef]
4. Tuzun S, Eskiyurt N, Akarirmak U, et al. Incidence of hip fracture and prevalence of osteoporosis in Turkey: the FRACTURK study. *Osteoporos Int*. 2012;23(3):949-955. [CrossRef]
5. Dimai HP. Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-score, and reference databases. *Bone*. 2017;104:39-43. [CrossRef]
6. Balasubramanya R, Selvarajan SK. Lumbar spine imaging. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023. 2021. Accessed 26 Nov 2021. [CrossRef]
7. Li Y, Zhao J, Lv Z, Pan Z. Multimodal medical supervised image fusion method by CNN. *Front Neurosci*. 2021;15:638976. [CrossRef]
8. Lim HK, Ha HI, Park SY, Han J. Prediction of femoral osteoporosis using machine-learning analysis with radiomics features and abdomen-pelvic CT: A retrospective single center preliminary study. *PLoS One*. 2021;4:16(3):e0247330. [CrossRef]
9. Kocak B, Yardimci AH, Yuzkan S, et al. Transparency in artificial intelligence research: a systematic review of availability items related to open science in radiology and nuclear medicine. *Acad Radiol*. 2022. [CrossRef]
10. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2(2):e200029. [CrossRef]
11. Adali T, Sekeroglu B. Analysis of microRNAs by neural network for early detection of cancer. *Procedia Technology*. 2012;1:449-452. [CrossRef]
12. Hamamoto R. Application of artificial intelligence for medical research. *Biomolecules*. 2021;11(1):90. [CrossRef]
13. Kavur AE, Gezer NS, Baris M, et al. Comparison of semi-automatic and deep learning-based automatic methods for liver segmentation in living liver transplant donors. *Diagn Interv Radiol*. 2020;26(1):11-21. [CrossRef]
14. Smets J, Shevroja E, Hügle T, Leslie WD, Hans D. Machine learning solutions for osteoporosis-a review. *J Bone Miner Res*. 2021;36(5):833-851. [CrossRef]
15. Guglielmi G, Muscarella S, Bazzocchi A. Integrated imaging approach to osteoporosis: state-of-the-art review and update. *Radiographics*. 2011;31(5):1343-1364. [CrossRef]
16. Xu Y, Li D, Chen Q, Fan Y. Full supervised learning for osteoporosis diagnosis using micro-CT images. *Microsc Res Tech*. 2013;76(4):333-341. [CrossRef]
17. Yamamoto N, Sukegawa S, Kitamura A, et al. Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates. *Biomolecules*. 2020;10(11):1534. [CrossRef]
18. Jang R, Choi JH, Kim N, Chang JS, Yoon PW, Kim CH. Prediction of osteoporosis from simple hip radiography using deep learning algorithm. *Sci Rep*. 2021;11(1):19997. [CrossRef]
19. Liu L, Si M, Ma H, et al. A hierarchical opportunistic screening model for osteoporosis using machine learning applied to clinical data and CT images. *BMC Bioinformatics*. 2022;23(1):63. [CrossRef]
20. Miller SJ, Howard J, Adams P, Schwan M, Slater R. Multimodal classification using images and text. *SMU Data Science Review*. 2020;3:6. [CrossRef]
21. Wisely CE, Wang D, Henao R, et al. Convolutional neural network to identify symptomatic Alzheimer's disease using multimodal retinal imaging. *Br J Ophthalmol*. 2022;106:388-395. [CrossRef]
22. Li Y, Zhao J, Lv Z, Li J. Medical image fusion method by deep learning. *International Journal of Cognitive Computing in Engineering*. 2021;2:21-29. [CrossRef]
23. Guo Z, Li X, Huang H, Guo N, Li Q. Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes. *ISBI*. 2018:903-907. [CrossRef]
24. Tan M, Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks, ICML 2019 Conference: 36th International Conference on Machine Learning, Long Beach, California; 2019:6105-6114. [CrossRef]
25. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 CVPR Conference: 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016:2818-2826. [CrossRef]
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778. [CrossRef]
27. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, AAAI-2017 Conference: Thirty-First AAAI Conference on Artificial Intelligence. 2017:4278-4284. [CrossRef]
28. Tan M, Quoc V. Le. EfficientNetV2: smaller models and faster training. Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021 International Conference of Machine Learning. [CrossRef]

29. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. AConvNet for the 2020s, 2022 IEEE Conference on Computer Vision and Pattern Recognition (pp. 11976-11986). [CrossRef]
30. Kocak B. Key concepts, common pitfalls, and best practices in artificial intelligence and machine learning: focus on radiomics. *Diagn Interv Radiol*. 2022;28(5):450-462. [CrossRef]
31. Maier-Hein L, Reinke A, Christodoulou E, et al. Metrics reloaded: pitfalls and recommendations for image analysis validation. 2022. [CrossRef]
32. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for Deep Convolutional Networks, 2018 IEEE winter conference on applications of computer vision (WACV). [CrossRef]
33. Smilkov D, Thorat N, Kim B, Viégas F, Wattenberg M. Smoothgrad: removing noise by adding noise, workshop on visualization for deep learning, ICML 2017. [CrossRef]
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845. [CrossRef]
35. Zhang B, Yu K, Ning Z, et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: a multicenter retrospective cohort study. *Bone*. 2020;140:115561. [CrossRef]
36. Hsieh CI, Zheng K, Lin C, et al. Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun*. 2021;12(1):5472. [CrossRef]
37. Lee KS, Jung SK, Ryu JJ, Shin SW, Choi J. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med*. 2020;9(2):392. [CrossRef]