Osteoporosis and Sarcopenia 11 (2025) 98–106

Contents lists available at ScienceDirect

Osteoporosis and Sarcopenia

journal homepage: www.elsevier.com/locate/afos

Original article

Gradient-enhanced neural network and extreme gradient boosting modeling for the prediction of the 3D bone mineral density distribution from 2D-DXA scans

Jiin Seo [a,1], Luca Quagliato [a,1], Yoon-Sok Chung [b,c], Taeyong Lee [a,d,*]

[a] Department of Mechanical and Biomedical Engineering, Ewha Womans University, Seoul, South Korea
[b] Department of Endocrinology and Metabolism, Ajou University School of Medicine, Suwon, South Korea
[c] Ajou Institute on Aging, Ajou University Medical Center, Suwon, South Korea
[d] Graduate Program in System Health Science and Engineering, Division of Mechanical and Biomedical Engineering, Ewha Womans University, Seoul, South Korea

ARTICLE INFO

Keywords:
Dual-energy X-Ray absorptiometry (DXA)
2D-3D mapping
Extreme gradient boosting (XGB)
Gradient-enhanced neural network (GENN)
Bone mineral density (BMD) prediction

ABSTRACT

Objectives: This study aims to predict the volumetric bone mineral density (BMD) distribution from a dual-energy X-ray absorptiometry (DXA) scan. By employing machine learning, this study bridges the gap between DXA and computed tomography (CT) in terms of volumetric bone assessment, suggesting an approach for a cost-effective and low-radiation alternative for bone health evaluation in a three-dimensional (3D) fashion from a two-dimensional (2D) scan.

Methods: Data from 34 participants included aligned DXA and CT scans for the proximal femur. Intensity values were extracted in Hounsfield units, with 3D information mapped as target variables and 2D information as features. Two machine learning models, Extreme Gradient Boosting (XGB) and Gradient-Enhanced Neural Network (GENN), were trained using 5-fold cross-validation strategy and show an average registration accuracy of 0.89 ± 0.04, assessed thorough structural similarity index measure.

Results: Both models were built to predict the statistics of the 3D structure of the bone from a 2D image. The GENN model outperformed XGB, achieving mean absolute percentage errors (MAPE) of 12.98 ± 1.70%, 13.28 ± 2.01%, and 9.63 ± 1.66% for minimum, maximum, and the number of nonzero pixel intensities, respectively. In contrast, XGB's errors exceeded 16% across these metrics. The loss stabilized within 100 epochs, indicating model robustness and reliability across diverse test sets.

Conclusions: The proposed GENN framework offers a method for predicting 3D BMD distributions from a 2D-DXA scan, rivaling CT-based assessments. This approach reduces costs and radiation exposure, presenting a viable solution for personalized bone health evaluation and early osteoporosis diagnosis.

1. Introduction

With the increase in the average age in our society, the prevalence of age-related skeletal diseases, such as osteopenia and osteoporosis, has significantly increased in parallel with the expansion of elderly population. Osteoporosis, a condition characterized by decreased bone density and structural deterioration, is one of the leading causes of fractures in the elderly population. Osteoporotic fractures, commonly occurring in hip, spine, and wrist, often result in severe long-term disabilities [1,2]. Beyond physical impairments, the fractures affect the quality of life with chronic pain and limited mobility even after recovery [3]. Furthermore, osteoporotic fractures elevate mortality risk within the post-injury period due to complications such as pressure ulcer or prolonged immobility [4,5]. In particular, hip fractures frequently require surgical intervention and are often accompanied by complications or secondary fractures within a short period [6,7], often leading to hospital readmissions. Considering the substantial medical costs for patients and healthcare systems [8–10] alike there is a growing global attention on the early diagnosis and treatment of osteoporosis.

The bone mineral density (BMD) measurement is widely recognized as the most reliable method for diagnosing osteoporosis and assessing fracture risk. In clinical practice, dual-energy X-ray absorptiometry

* Corresponding author. Department of Mechanical and Biomedical Engineering, Ewha Womans University, Seoul, 03760, South Korea.
E-mail address: tlee@ewha.ac.kr (T. Lee).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.afos.2025.09.002
Received 21 July 2025; Received in revised form 4 September 2025; Accepted 9 September 2025
Available online 19 September 2025

(DXA) is the most preferred technique and is considered the gold standard for osteoporosis diagnosis by the World Health Organization (WHO) [11]. DXA is favored for its short examination time, affordability, and minimal radiation exposure. However, it has inherent limitations from two-dimensional (2D) imaging that cannot fully capture the complexities of three-dimensional (3D) bone structures [12,13]. In contrast, quantitative computed tomography (QCT) provides examination reflecting volumetric structure of the bone and assesses trabecular and cortical bone separately. Despite its precision, QCT is not recommended for routine clinical use due to high radiation exposure [12,13]. It also requires additional materials for its calibration blocks, known as phantoms, adding complexity to the diagnosis procedure. Thus, there has been increasing interest in research aimed at utilizing DXA for volumetric BMD assessment. These studies focus on the bone density evaluation of the spine and femur, which are the primary sites for osteoporosis diagnosis.

As part of these efforts to address the shortcomings of current diagnostic methods, several researchers have proposed reconstructing 3D medical images from 2D scans. Common approaches for 3D reconstruction from 2D images are the active shape model (ASM) and the statistical shape model (SSM) [14–17] and utilize QCT scans of the targeted bones, such as vertebrae or femur, with a set of landmarks to create an average 3D femur template model. The projection images of this template model are then compared with actual DXA images captured from multiple angles and this sequence represents the training dataset of the shaping models. Besides, when using convolutional neural network (CNN) for reconstruction, digitally reconstructed radiographs (DRRs) from computed tomography (CT) scans have been presented [18] and represent a replacement for DXA scans from a user-defined angle.

As summarized so far, though various approaches have been proposed, the current state-of-the-art in the reconstruction of 3D models from 2D scans still presents various limitations. First, in the literature studies, the scanned images, such as CT and DXA scans, were sourced from distinct medical institutes, with each type of scan provided by a different facility to form the whole dataset [15,19]. This approach was employed to ensure sufficient number of images from each scanning device, claimed to help the neural network model to be more robust on unseen data. However, this inconsistencies within the dataset could rather cause potential instability that affect the model performance since various sources are likely to introduce noise and inaccuracy at different levels in the dataset, subsequently propagated during the training process. Moreover, most of these volumetric reconstruction methodologies of planar scan, especially when it applies CNN, use graphic processing units (GPUs) for the processing operations. Even though GPUs provide considerable advantages in deep learning, such as parallel processing capabilities and shorter training durations [20,21], there are also drawbacks to this modelling approach. The most pointed ones are the power consumption and memory resource limits, which result in higher operational costs or the limitation of the training dataset size [22,23] which, in turn, restricts the feature extraction during training or future dataset expansion, potentially leading to the limitation in the model performance [24].

According to the background analysis provided so far, this research aims to reconstruct the 3D BMD distribution from a DXA scan images using an array-based approach. Derived from the assumption that the medical scans, either 3D or 2D, from the same patient shall have a quantifiable relationship, this study proposes a supervised machine learning methodology, based on two models: the Extreme Gradient Boosting (XGB) and Gradient-Enhanced Neural Network (GENN). The XGB and GENN models are widely used in various research fields for their ability to handle structured data and provide insights into complex relationships [25], making them particularly advantageous in tasks such as diagnostic modeling or outcome prediction [26], such as the one undevoured in this research. Both models have been trained considering data arrays in terms of spatial coordinates and Hounsfield unit (HU) data, the latter relevant to the BMD of each specific pixel in the 2D image and voxel in the 3D scan, respectively. To achieve parallel alignment between the 2D and 3D models of the femur bone, a registration algorithm was developed to minimize the error between DXA scans and the 2D projection of CT scans. The results show that the GENN model, thanks to the addition of gradients in the backpropagation, allows for a more robust and precise modeling for the 2D to 3D mapping scenario at hand, with average deviations equal to 12.43%, in contract to those of 16.96% of the XGB model. In this regard, the shallow width and extensive length of the database is seen to behave as a time series rather than a scalar database, justifying the better performance of the GENN algorithm to the XGB model. Collectively, the proposed approach was able to predict the statistics of the volumetric model of femur from the planar images, suggesting a novel way of evaluating the physical properties of the bones.

2. Methods

2.1. Subjects and data acquisition

The image database for this study included 34 patients (28 females and 6 males) with a mean age of 76.57 ± 8.38 years [range: 55–88 years], collected by Ajou University Hospital (Suwon, Korea) between December 2019 and May 2021 (IRB: AJIRB-MED-MDB-21-299). The study participants were excluded if they had a skeletal disease other than osteoporosis, as well as a femur implants and prosthesis, or if they had a history of hip fractures on at least one side of lower limbs. DXA scans were performed using Lunar Prodigy scanner (GE Healthcare) whereas CT scans were acquired using the SOMATOM Definition Edge (Siemens Healthineers, Forchheim, Germany). The region of interest (ROI) was set-up to include proximal femur of a single side for DXA scan, while that of CT scans comprised both femurs, placing the most proximal slice aligned with L4 of lumbar spine and the most distal slice below the distal femurs, with a matrix size of 512 × 512. All CT scans included in this study were non-contrast to avoid artificial enhancement of anatomical structures and the potential confounding of attenuation measurements. Due to the privacy concern, the dataset is not publicly available.

2.2. Segmentation and registration

The high-resolution in the images acquired by DXA and CT scans causes computationally intensive transformations and significantly long training times when deep learning models are used [27,28]. Additionally, the DXA scanning area was focused on the femoral joint. These being the case, the ROI was restricted to specific representative crop encompassing the proximal femur within the entire scanned domain. The cropping direction, whether left or right, was aligned with that of the DXA image. Femur segmentation from the DXA images was performed using the contour and auxiliary lines which divide major anatomical landmarks of femur, including femoral head, femoral neck, and intertrochanteric area. For the cropped CT images, bone segmentation was proceeded by applying the HU as a pixel intensity threshold. Given that HU are directly related to the density of matter, they were chosen as the representative metric for BMD in this research. The overall procedure of image segmentation is illustrated in Fig. 1.

After segmentation, point clouds were generated with nodes containing the coordinates and intensities of non-zero intensity pixels from both DXA and CT scans. Thereafter, the CT-based point cloud was refined using the open-source software MeshLab (Version 2013.12) for 3D mesh processing to isolate the femur region. The node coordinates and intensities of the point clouds were recorded in CSV files to enable matrix computations. The matrices computations were performed to align the positions of DXA- and CT-based point clouds. Since the dataset was obtained from two different devices, the coordinate systems of the image data were not inherently uniform. Therefore, it was necessary to

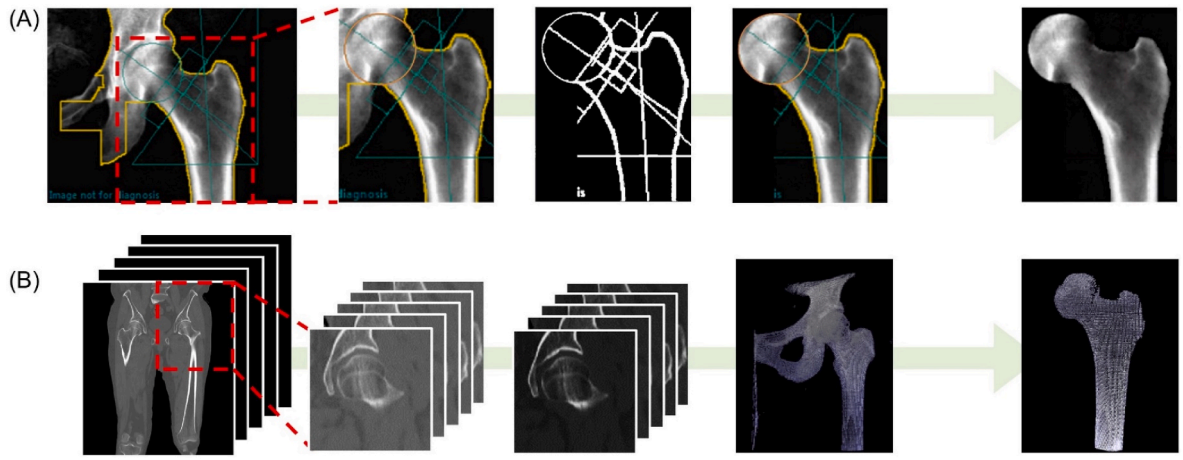

Fig. 1. Femur segmentation from DXA and CT scan images. (A) DXA scan segmentation (B) CT scans segmentation. DXA, dual-energy X-ray absorptiometry; CT, computed tomography.

align the two coordinate systems through image registration [29] to ensure that the same body parts of a given subject were positioned within the same pixel coordinates. To achieve the alignment, similarity transformation was applied to the point clouds, incorporating translation, rotation, and scaling [30,31]. The structural similarity index measure (SSIM) was used as the evaluation metric for the image registration [32]. Due to the inconsistency of dimension, SSIM was computed between DXA scan image and the projection image of CT scan-based volume. As a result of the registration, the SSIM between the femur DXA and CT image projections for the same patient was 0.89 ± 0.039. Figure 2A–E illustrate the difference between before and after the registration of point clouds based on the DXA and CT images. The top row depicts the location of point clouds in the pre-registration state and the post-registration state at the bottom row.

After the registration between the DXA- and CT-based point clouds, the coordinate and intensity of every node consisting the clouds were exported in CSV files for database composition.

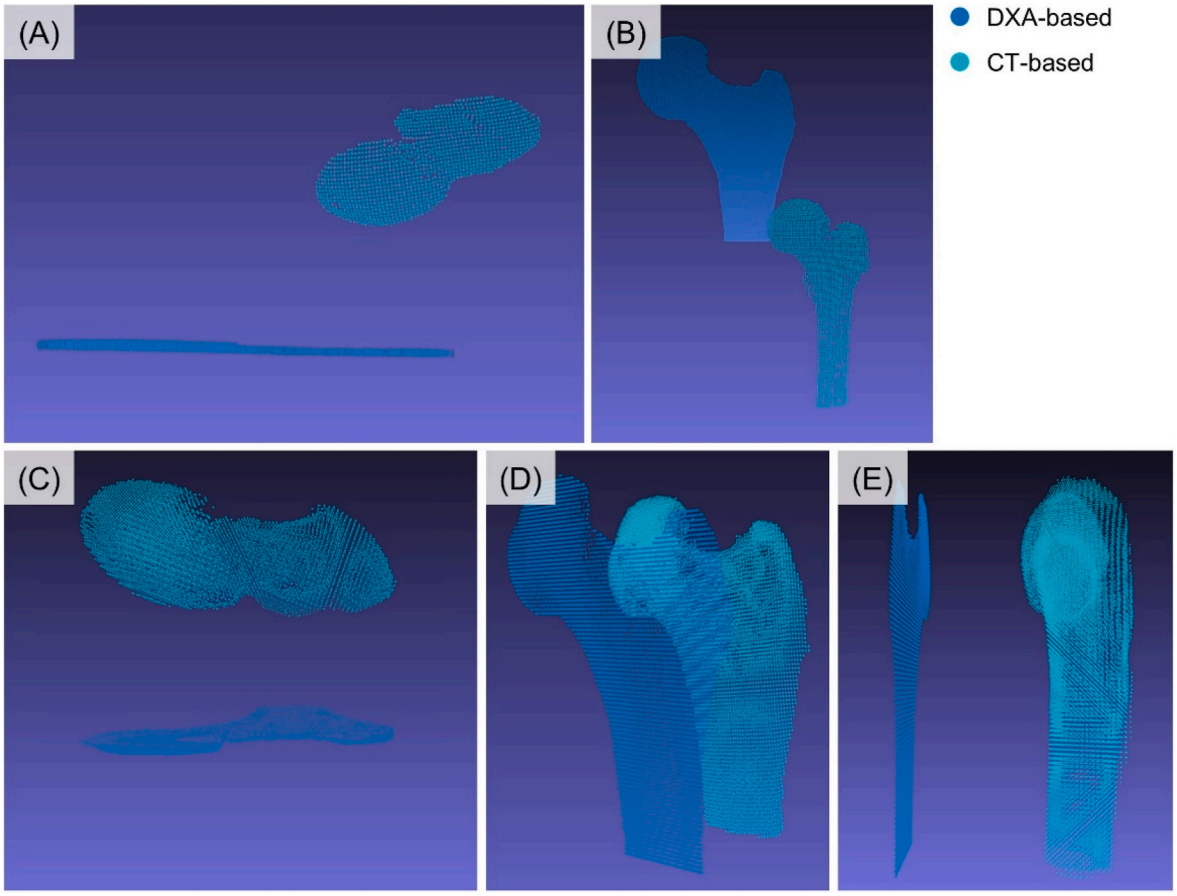

Fig. 2. The pre- and post-registration state of point clouds. The coordinate system inconsistency before the registration (A) top view (B) front view, and the point clouds alignment of post-registration. (C) Top view (D) front view (E) side view**.** DXA, dual-energy X-ray absorptiometry; CT, computed tomography.

2.3. Extreme gradient boosting (XGB) regression model

The extreme gradient boosting (XGB) is an advanced implementation of the gradient boosted decision tree and random forest algorithms. The XGB model is grounded on the gradient boosting (GB) algorithm that uses the multiple weak learning models combined sequentially. It is trained to improve its predictive performance by correcting the errors of the previous iteration by optimizing the weights of the model. The XGB is the parallelized model of GB, improving the computation time significantly and widely employed for various regression and classification problems due to its scalability and versatility [33]. It is developed based on multiple classification and regression tree repeatedly updated to reduce an objective function.

The specific equation used for ensemble formation is written as Eq. (1) utilizing K-trees, and the objective function (obj) in Eq. (2) is driven by both the loss function (l) defined as the mean square error (MSE) and regularization (ω) terms that prevent overfitting in the training process. The regularization function is formulated as in Eq. (3), where w is the vector of the scores of the leaves and T is the total number of trees, scaled by the parameter γ. The objective function for the t-step can be simplified as in Eq. (3), applying the definition of MSE in the loss function and the Taylor expansion up to the second order.

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \, \epsilon \, r \tag{1}$$

$$obj = \sum_{i=1}^{n} l(y_i, \widehat{y}_i^{(t)}) + \sum_{i=1}^{t} \omega(f_i) \; where \; \omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2}$$

$$obj^{(t)} = \left[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \left[\gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \right] \tag{3}$$

The parameter f in the equations can be found during the training process; however, some hyperparameters such as $T, \gamma,$ and w must be defined before beginning the training process. Several researchers found that hyperparameter values greatly impact the model performance [34, 35]. Therefore, optimal hyperparameters should be found through the hyperparameter tuning. The most common methods to optimize hyperparameters are grid search, random search, and Bayesian optimization. Considering its simplicity, the grid search method was applied in this study.

2.4. Gradient enhanced neural network (GENN) regression model

An artificial neural network (ANN) is a system that mimics how humans learn by identifying patterns in data. It receives inputs, multiplies them by weights, which determine the importance of each input, sums them up, and passes the result through the objective function to generate an output. It has the structure in which learning takes place by applying the transfer function to the sum of the weighted input values. The network is made up of layers: input layers receive the data, hidden layers process it, and the output layer produces the result.

A deep neural network (DNN) is a model that has multiple hidden layers between the input and output layers. A complex nonlinear mapping is implemented from the input to the output by the neurons of the network. DNNs are capable of interpreting complex data, such as recognizing objects in images or understanding natural language. The output of every neuron is sent to the following layer, in which it is multiplied by the weight of each neuron and added to the bias. This results in the execution of several nonlinear operations through the so-called activation functions. In this study, the neural network assumed rectified linear unit ($ReLu$) activation function for a generic input features' vector $\mathbf{x}$, defined as in Eq. (4).

$$\mathbf{h}_i = ReLu(\mathbf{W}_i\mathbf{x} + \mathbf{b}_i) \; where \; \mathbf{x} = [x_1, x_2, \ldots, x_n] \in R^n \tag{4}$$

A gradient enhanced neural network (GENN) is a type of neural network that incorporates gradient information into the training process to enhance the model performance. The gradient is a mathematical concept that shows how much the function output changes as the input changes. In a normal neural network, the goal of the model is to approximate a function $f_\theta(x)$, where x is the input and θ are the parameters, which are weights and biases of the network. The model predicts outputs $f_\theta(x)$, and training is proceeded in the way that minimizes a loss function that measures the difference between the prediction and true values.

In a GENN, the network is trained not only on the error in the predicted function values, but also on the error in the gradients of the function. The gradient refers to how the output changes as the input changes. For a given input x, the output is $f_\theta(x)$, and the gradient of this function with respect to the input x is written as Eq. (5).

$$\nabla_x f_\theta = \frac{\partial f_\theta(x)}{\partial x_1}, \frac{\partial f_\theta(x)}{\partial x_2}, \ldots, \frac{\partial f_\theta(x)}{\partial x_n} \tag{5}$$

The gradient information is included in the loss function of a GENN. The extended loss function $L(\theta)$ could be expressed as Eq. (6). In Eq. (6), $f_\theta(x)$ is the predicted output, y as the true value, $g(x)$ the true gradient at point x, and λ is a weight balancing the importance of the gradient term.

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left((f_\theta(x_i) - y_i)^2 + \lambda \| \nabla_x f_\theta(x_i) - g(x_i) \|^2 \right) \tag{6}$$

The forward DNN employed in this study is defined between the input layers and output layers as depicted in Fig. 3, including the pre-backpropagation parts of the network based on the defined PDEs and relevant loss function.

Using Eqs. (4) and (5), the weights and biases for each layer are allocated at random for the first epoch, which enabled the estimation of the initial set of model predictions. The first and second-order partial derivatives of the model predictions to the input features are then computed, which could be formalized as Eqs. (7) and (8).

$$\frac{\partial \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{X}_{2D}} = \nabla_{X_{2D}} \boldsymbol{F}(\mathbf{x}) \; ; \; \frac{\partial \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{Y}_{2D}} = \nabla_{Y_{2D}} \boldsymbol{F}(\mathbf{x}) \; ; \; \frac{\partial \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{I}_{2D}} = \nabla_{I_{2D}} \boldsymbol{F}(\mathbf{x}) \tag{7}$$

$$\frac{\partial^2 \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{X}_{2D}^2} = \nabla_{X_{2D}} \left(\frac{\partial \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{X}_{2D}} \right) \; ; \; \frac{\partial^2 \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{Y}_{2D}^2} = \nabla_{Y_{2D}} \left(\frac{\partial \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{Y}_{2D}} \right) \; ; \; \frac{\partial^2 \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{I}_{2D}^2} = \nabla_{I_{2D}} \left(\frac{\partial \boldsymbol{F}(\mathbf{x})}{\partial \boldsymbol{I}_{2D}} \right) \tag{8}$$

2.5. Models' training and validation

In this study, XGB and GENN models were used to build the machine learning model that predict the statistics of CT-based 3D model from DXA image. The model inputs were pixels' 2D coordinates and intensities from the DXA image-based point cloud, and the output variables were average, nonzero minimum, maximum, and the number of nonzero-pixel values of the intensity of pixels from the CT image-based point cloud. The four output variables were computed among the pixels aligned along the same 2D pixel coordinates. The evaluation metric employed for both neural networks was the mean absolute percentage error (MAPE), which measures the prediction accuracy of the model by quantifying the relative error between the actual and predicted values. It provides an intuitive interpretation of the model performance in terms of percentage error. This makes it suitable especially for comparing model performance across various target variables. While metrics like root mean squared error or coefficient of determination are also useful, they do not offer the same convenience of interpretation for relative error, being not robust to outliers compared to MAPE. Both models included 5-fold cross-validation to ensure sufficient data for both training and validation. A randomly selected 80% of the database was employed for the training and the remaining 20% for the validation.

As the optimization of hyperparameters can affect the performance

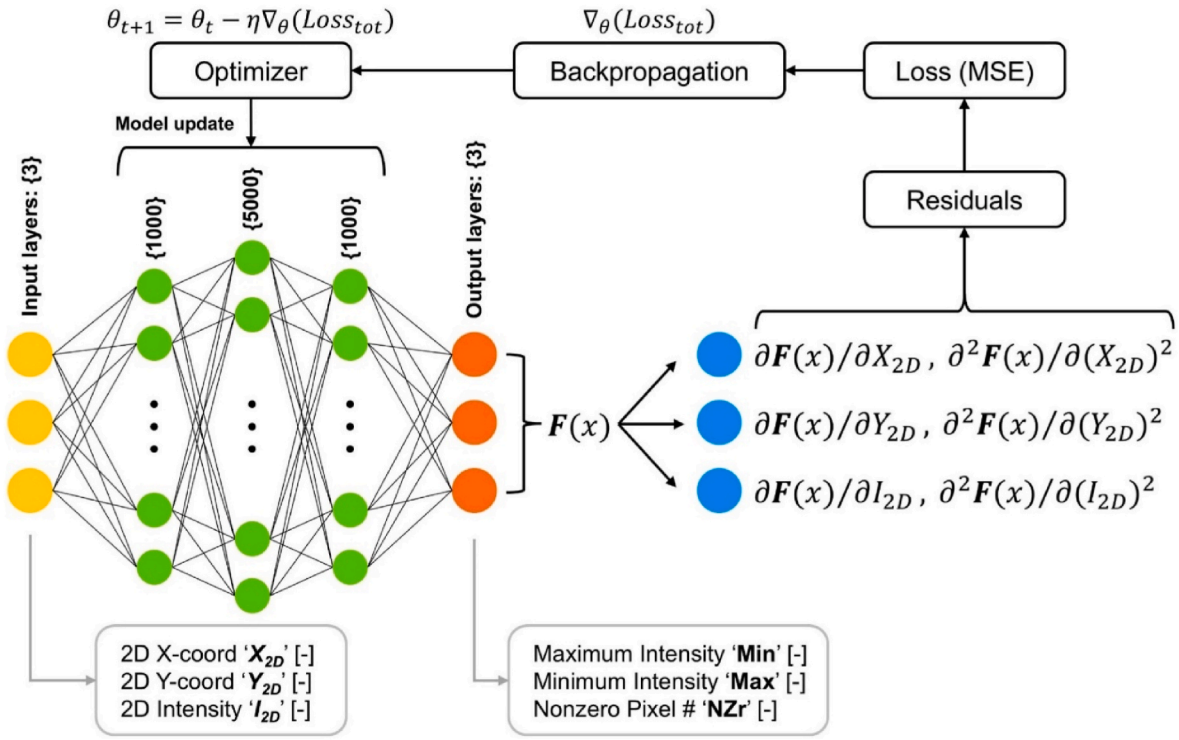

Fig. 3. The structure of gradient-enhanced neural network (GENN) used in the research. MSE, mean squared error.

of GB model, the modification of learning rate is also one of the methods that assist finding the optimal solution of neural networks. The learning rate works as a coefficient for the gradient to change the weights in smaller steps. The high learning rate could make a big change in the weights in relatively less steps, but meanwhile it could overshoot to the local minima. On the other hand, too small learning rate might prolong the training process. To find the most advantageous value, it is recommended to start with high learning rate and gradually reduce until the weight reaches the minimum. For the neural network employed in this study, the percentage error showed the tendency of convergence when the learning rate was 0.0001, as it was depicted in Fig. 4. Correspondingly, the learning rate was set as 0.0001 for the GENN model.

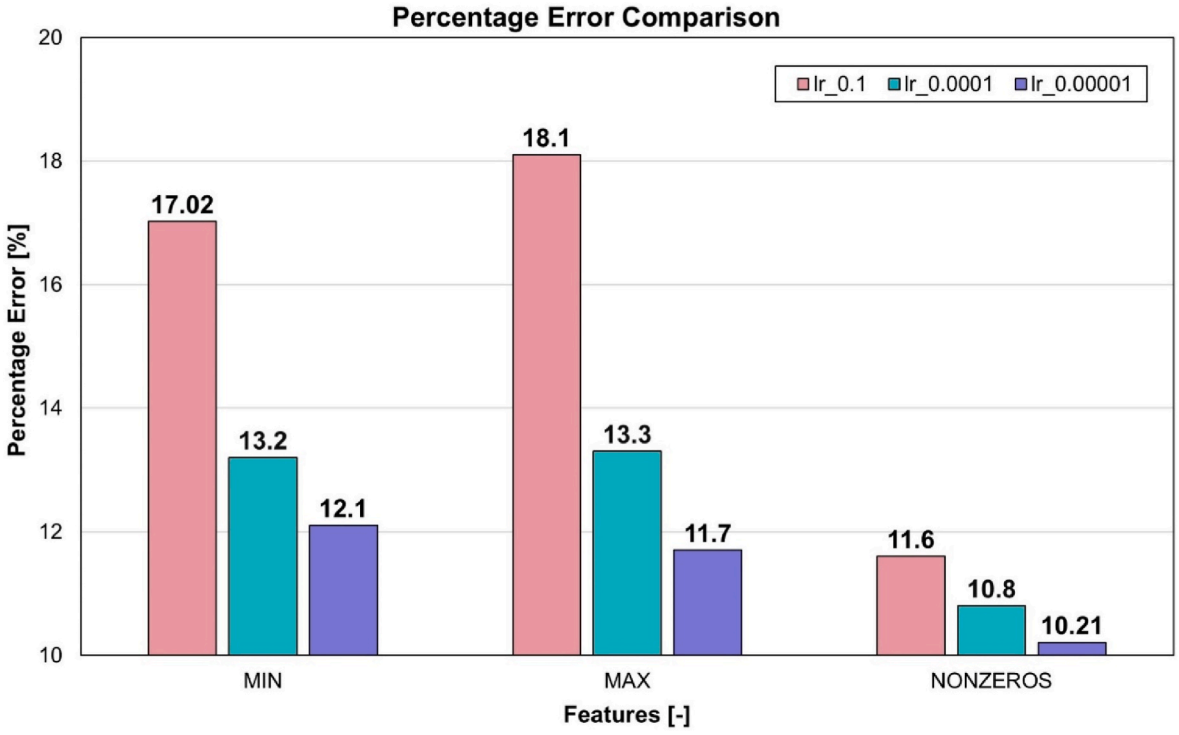

Fig. 4. Percentage error variance depending on learning rates when set as 0.1, 0.0001, and 0.00001.

3. Results

The proposed model architectures were validated by comparing the different subsets of the entire dataset against the CT scan images through k-fold cross validations, serving as the ground truth using the MAPE for each patient. With the XGB model, the MAPE of target variables was as follows: 16.28 ± 1.31%, 16.90 ± 1.17%, and 17.64 ± 2.55% for minimum, maximum, and the number of nonzero-pixel values in terms of pixel intensity, respectively. On the other hand, when the GENN model was employed, a significant change was observed with the predictive performance of the model. The MAPE of the neural network model was 12.98 ± 1.70%, 13.28 ± 2.01%, and 9.63 ± 1.66% for each target variable. The error deviation of both models was less than around 5% throughout the folds, showing the stability of the models and balance of the database. Figure 5A shows the comparison of the validation errors between the XGB and GENN models. As demonstrated, the MAPE of all the target variables were significantly lower with GENN model ($P < 0.05$) compared to the XGB model.

In addition, the total loss of the neural network model stabilized within 100 epochs. The inclination of each fold towards convergence is seen in Fig. 5B. The figure depicts that the five folds had a similar tendency of convergence with the subtle differences in the loss gradient value during cross-validation. The consistency across the folds implies that the model is robust enough to generalize well to various subsets of the dataset, ensuring reliable prediction performance. In addition, this also shows that the training dataset is uniform, without outliers.

As part of the model evaluation process, the prediction results using XGB and GENN at random five points of the femur point clouds were analyzed. The points were chosen from the key landmarks on the femur, including the femoral head, neck, trochanteric area, and the shaft. The specific position of the points at each point cloud are shown in Fig. 6, and the prediction results of the models were tabulated in Table 1. Although the range of MAPE in each point varies even when the same model was employed, the overall value of the metrics was significantly ($P < 0.05$) low within that of GENN.

4. Discussion

This study aimed to predict the 3D distribution of bone density from DXA scan images using XGB and GENN machine learning models. The pixel intensity from DXA and CT medical scans was used as the representing value of BMD, and the CT-based point cloud, considered the gold standard method, was applied as a reference. To maximize the utility of volumetric reconstruction of medical scan, the statistics that can assist understanding the BMD distributions were chosen as target variables: the minimum and maximum value of intensity, and the number of off-threshold pixels. The prediction model structure employed the XGB and GENN model individually, and the models were both built to work using numerical array for memory efficiency.

Both XGB and GENN models predicted the dependent variables with a certain level of high accuracy when it was validated against several patient cases. The evaluation metrics of all target variables were around or lower than 15%, which explains that both XGB and GENN models could predict the statistics of 3D pixel intensity from the single DXA image up to a certain level of accuracy. *Id est*, the MAPE of GENN model was fairly lower for the same target variables, indicating that the GENN model was more suitable for this task. Additionally, the total loss stabilized and converged within 100 epochs for all five folds in the cross-validation. This fact implies the learning process was consistent when the neural network-based model was employed.

The different performances of the XGB and GENN models can be explained by their underlying learning paradigms and how they align with the target of this study. The XGB, or tree-based methods, undergo deterministic learning, focusing on minimizing prediction error through the optimization process without explicitly modeling uncertainty. This type of learning is proven to be helpful and applied to numerous classifications or regression problems because most issues in the real-world follow a structured decision-making process based on certain rules. On the other hand, neural network models adopt probabilistic learning through various elements, such as randomized data splits or network initialization [36,37]. These features allow the model to provide measures of uncertainty by predicting probabilities or distributions. The two models employed in the current study, XGB and GENN, were asked to make predictions based on medical data; thus, they are required to be flexible in handling uncertainty and ambiguity in order to provide the outcomes with an associated level of confidence. Taking these factors into account, the GENN model was specifically designed to combine both deterministic and probabilistic components. The integrative traits of the model likely contributed to its superior performance, as evidenced by its lower percentage error compared to the XGB model in the current study.

To the best of our knowledge, this is the first study that evaluated the ability of machine learning models built from an array-based method to predict the current health status of human proximal femur from statistics of volumetric models. Thus, in this study, the femoral volume was reconstructed from a single planar reference image (eg, DXA), consequently making them appropriate for subject-specific examination of volumetric BMD distribution, such as the one extract ed from a CT scan. In this regard, previous studies proposed the method of subject-specific local BMD evaluation [19,38] using mathematical equations and linear regression analysis to compute the volumetric BMD from the DXA scans. However, although regression algorithms have proven their usefulness in local BMD assessment, the statistical methods cannot fully reflect the geometrical information of bone microarchitecture. Since the structure

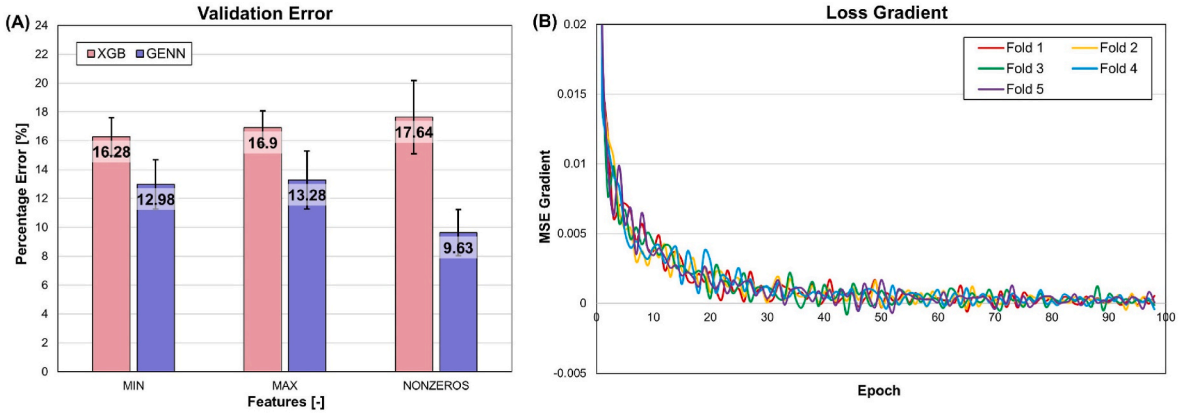

Fig. 5. (A) Validation error comparison between XGB and GENN model (B) loss gradient of GENN model. XGB, extreme gradient boosting; GENN, gradient-enhanced neural network.

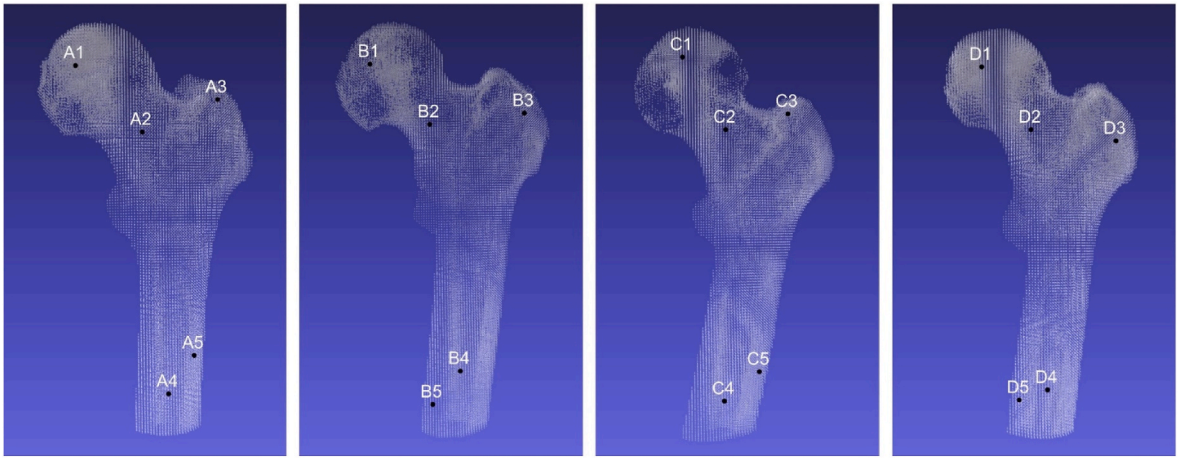

Fig. 6. Four examples of the CT-based point clouds. Five points for each model were chosen to validate the models' performance. The mean absolute percentage errors of each target variable for the points were shown in Table 1. CT, computed tomography.

Table 1
Mean absolute percentage error (MAPE) for each test point, as per Fig. 6.

Point #	XGB MAPE [%]			GENN MAPE [%]		
	MIN	MAX	NZr	MIN	MAX	NZr
A1	7.26	10.25	8.16	16.46	17.26	16.75
A2	13.46	11.05	14.71	17.59	11.23	10.01
A3	11.97	9.04	15.47	14.15	15.77	15.74
A4	15.27	16.55	10.06	13.55	12.18	17.87
A5	14.84	11.84	9.52	14.07	11.54	17.95
B1	10.69	11.66	9.75	16.66	17.01	10.06
B2	8.84	5.43	11.03	12.29	11.93	14.53
B3	15.32	18.22	10.68	16.69	19.19	14.36
B4	14.51	13.98	12.67	15.21	15.46	13.19
B5	11.59	12.83	11.61	13.64	16.86	14.80
C1	12.98	13.19	11.52	17.79	16.65	11.62
C2	5.43	7.69	6.67	14.55	14.26	14.20
C3	9.32	6.41	13.18	11.77	14.76	13.97
C4	11.28	14.35	4.79	18.11	19.29	14.21
C5	12.27	7.29	7.72	13.36	12.21	14.63
D1	11.22	14.44	10.67	16.26	17.01	17.22
D2	12.27	16.91	8.61	15.36	17.41	11.79
D3	11.77	15.41	9.01	12.39	17.56	10.13
D4	10.36	16.94	12.38	16.71	15.62	13.63
D5	12.92	14.07	10.24	17.92	13.18	17.08

XGB, extreme gradient boosting; GENN, gradient-enhanced neural network; MIN, minimum intensity; MAX, maximum intensity; NZr, number of nonzero-pixel intensity.

of the microarchitecture associated with the fracture risk is not in a regular form but rather uneven, the absence of geometrical data could cause the potential uncertainty in the computed result.

The XGB and GENN models were provided the data within similar locations by relating the statistical information of BMD with corresponding coordinates. In addition, the prediction for the partial anatomy of the femur was evaluated using the proposed models, with both achieving successful results to a certain extent. The XGB model attained MAPE of 16.28 ± 1.31%, 16.90 ± 1.17%, and 17.64 ± 2.55% for minimum, maximum, and the number of nonzero-pixel intensity values, respectively, whereas the GENN model achieved MAPE values of 12.98 ± 1.70%, 13.28 ± 2.01%, and 9.63 ± 1.66% for the same target variables. The femoral head, neck, and shaft were chosen as the sites where the MAPE was computed individually and considering that the majority of osteoporotic fracture occurs in the femoral neck, the methodology could be accounted to be helpful in conducting early diagnosis of osteoporosis.

Aside from reconstructing a 3D model out of a planar medical scan, other studies have proposed using HU values of CT scans to predict BMD or T-score for assessment of bone health in terms of bone density [39,40]. The resultant values are employed as features for fracture risk estimation or classification of patients depending on the presence or severity of skeletal diseases. However, these studies focus on deriving the average value of the whole ROI, such as vertebrae or proximal femurs. It inevitably causes overlooking the partial condition of ROI or being ineffectual at narrowing down where the vulnerable section might be. The present study proposed the methodology of earning the map to BMD distribution using structural information, which enables understanding the local BMD specifically.

Although there are positive results in predicting the 3D distribution of four different variables from a DXA scan, especially for the case of the GENN model, this research is also limited by several aspects. First, the study population was constrained in terms of size and diversity. The final dataset consisted of a relatively small number of subjects, which resulted from the strict inclusion criteria requiring paired DXA and CT scans, absence of fractures or implants, and consistent image quality. In addition, due to the retrospective nature of the database, the age range of the subjects was mainly between 70 and 80 years old, with a few middle-aged participants. The age distribution reflects the population most at risk of osteoporosis and related fractures. It aligns with the goal of the study, but the model applicability may be limited due to the data absence from younger individual. Considering that scan image collection was conducted in a single institution and composed of Korean-only data, it can possibly reflect more information about such a group. These conditions limiting the data pool suggest that adding more diverse input data from different institutions to the dataset could enhance the model performance and robustness. Second, the reference 3D bone values were expressed in terms of CT-derived HU rather than absolute volumetric BMD (mg/cm^3, vBMD) or other alternative metrics, because QCT data acquisitions were not available in the institutional database. Although several studies [41,42] have proposed methods for converting HU-based measures into vBMD, the resulting values remain dependent on the specific CT scanner settings and acquisition protocols. Accordingly, the prediction in this research were maintained at the level of pixel intensity, which can be converted to BMD employing scanner-specific calibration setting, in terms of regression equation of other established conversion approaches. Last, the DXA scans included auxiliary lines dividing the anatomical areas, which covered the bone image partially. The lines were removed during the preprocessing steps, but there is a possibility that the point clouds built based on the processed images might have been affected during the procedure.

The present study focused exclusively on the proximal femur, and therefore, the trained model is not directly generalizable to other

skeletal sites. Applying the approach to alternative regions, such as wrist or spine, would require site-specific datasets and at least partial retraining of the model, due to the dependence on region-specific preprocessing and registration techniques. Nevertheless, the suggested framework is adaptable, and it has the potential to be extended to other regions following the suggested methodology with the appropriate retraining using location-specific data.

To summarize, the proposed methodology successfully demonstrated the feasibility of predicting the volumetric data of human proximal femur from a single DXA scan image using machine learning models. It is proved that leveraging array-based processing could achieve the results up to an acceptable level of prediction accuracy and robustness. Despite the strength of the proposed methodologies, the study is still limited by the homogeneity of the dataset and the computational simplifications, such as uniform spacing among the patients. Future work might be developed by expanding the database with more diverse subject populations and incorporating anatomical details or adaptive image processing of data to improve model generalization and prediction performance. These advancements open the way for refining the models to offer personalized and precise bone health assessments, ultimately aiding early diagnosis and treatment of osteoporosis.

5. Conclusions

This study successfully demonstrated the feasibility of predicting volumetric BMD distribution from DXA scans using machine learning models. By employing XGB and GENN, the proposed methodology achieved robust predictive accuracy, with GENN outperforming XGB across all metrics, which underscored its applicability for tasks that imply complex relationships and numerical precisions. Despite these advancements, limitations such as dataset homogeneity and potential error risk under preprocessing procedure were pointed out, suggesting the need for more future research with more diverse and anatomically detailed datasets. These findings bridge the gap between DXA and CT by offering cost-effective and low-radiation approaches to volumetric bone assessment, providing a foundation for advancing personalized bone health diagnostics. The research emphasizes the possibility of machine learning to improve as clinical instruments for early diagnosis and management of osteoporosis.

CRediT author statement

Jiin Seo: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing. Luca Quagliato: Conceptualization, Methodology, Validation, Formal analysis, Writing – Review & Editing. Yoon-Sok Chung: Investigation, Resources, Data curation, Supervision. Taeyong Lee: Supervision, Project Administration, Funding acquisition.

Conflicts of interest

The authors declare no competing interests.

Acknowledgments

This work was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education (MOE, South Korea), National Research Foundation of Korea (NRF-5199990614253). Professor Luca Quagliato was supported by the RP-Grant 2025 of Ewha Womans University. **ORCID** Jiin Seo: 0009-0008-8647-6427. Luca Quagliato: 0000-0002-5379-8306. Yoon-Sok Chung: 0000-0003-0179-4386. Taeyong lee: 0000-0002-0176-0525.

References

[1] Alexiou KI, Roushias A, Varitimidis SE, Malizos KN. Quality of life and psychological consequences in elderly patients after a hip fracture: a review. Clin Interv Aging 2018:143–50.
[2] Katsoulis M, Benetou V, Karapetyan T, et al. Excess mortality after hip fracture in elderly persons from Europe and the USA: the CHANCES project. J Intern Med 2017;281:300–10.
[3] Lane MK, Nahm NJ, Vallier HA. Morbidity and mortality of bilateral femur fractures. Orthopedics 2015;38:e588–92.
[4] Hong S, Han K. The incidence of hip fracture and mortality rate after hip fracture in Korea: a nationwide population-based cohort study. Osteoporos Sarcopenia 2019; 5:38–43.
[5] Myers P, Laboe P, Johnson KJ, Fredericks PD, Crichlow RJ, Maar DC, et al. Patient mortality in geriatric distal femur fractures. J Orthop Trauma 2018;32:111–5.
[6] Ho AWH, Wong SH. Second hip fracture in Hong Kong–Incidence, demographics, and mortality. Osteoporos Sarcopenia 2020;6:71–4.
[7] Ruiz-Adame M, Correa M. A systematic review of the indirect and social costs studies in fragility fractures. Osteoporos Int 2020;31:1205–16.
[8] Cortez KA, Lai JGL, Tabu IA. Economic burden and the effects of early versus delayed hospitalization on the treatment cost of patients with acute fragility hip fractures under the UPM-PGH Orthogeriatric Multidisciplinary Fracture Management Model and Fracture Liaison Service. Osteoporos Sarcopenia 2021;7: 63–8.
[9] Carpintero P, Caeiro JR, Carpintero R, Morales A, Silva S, Mesa M. Complications of hip fractures: a review. World J Orthop 2014;5:402–11.
[10] Amphansap T, Sujarekul P. Quality of life and factors that affect osteoporotic hip fracture patients in Thailand. Osteoporos Sarcopenia 2018;4:140–4.
[11] 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.
[12] Glüer CC. 30 years of DXA technology innovations. Bone 2017;104:7–12.
[13] Luo Y. Empirical functions for conversion of femur areal and volumetric bone mineral density. J Med Biol Eng 2019;39:287–93.
[14] Borchardt G, Nickel B, Andersen L, et al. Femur and tibia BMD measurement in elective total knee arthroplasty candidates. J Clin Densitom 2022;25:319–27.
[15] Whitmarsh T, Fritscher KD, Humbert L, Del Río Barquero LM, Roth T, Kammerlander C, et al. A statistical model of shape and bone mineral density distribution of the proximal femur for fracture risk assessment. In: Fichtinger G, Martel A, Peters T, editors. Medical image computing and Computer-Assisted intervention – MICCAI 2011. Berlin (DE): Springer; 2011. p. 393–400.
[16] Kolta S, Le Bras A, Mitton D, Bousson V, de Guise JA, Fechtenbaum J, et al. Three-dimensional X-ray absorptiometry (3D-XA): a method for reconstruction of human bones using a dual X-ray absorptiometry device. Osteoporos Int 2005;16:969–76.
[17] Humbert L, Hazrati Marangalou J, del Río Barquero LM, van Lenthe GH, van Rietbergen B. Cortical thickness and density estimation from clinical CT using a prior thickness-density relationship. Med Phys 2016;43:1945–54.
[18] Almeida DF, Astudillo P, Vandermeulen D. Three-dimensional image volumes from two-dimensional digitally reconstructed radiographs: a deep learning approach in lower limb CT scans. Med Phys 2021;48:2448–57.
[19] Humbert L, Martelli Y, Fonolla R, Steghöfer M, Di Gregorio S, Malouf J, et al. 3D-DXA: assessing the femoral shape, the trabecular macrostructure and the cortex in 3D from DXA images. IEEE Trans Med Imaging 2016;36:27–39.
[20] Park H, Kim S. Hardware accelerator systems for artificial intelligence and machine learning. Adv Comput 2021;122:51–95.
[21] Inci A, Isgenc MM, Marculescu D. Efficient deep learning using non-volatile memory technology in GPU architectures. In: Pasricha S, Shafique M, editors. Embedded machine Learning for Cyber-Physical, IoT, and Edge Computing: Hardware Architectures. first ed. Cham: Springer; 2023. p. 225–52.
[22] Cano A. A survey on graphic processing unit computing for large-scale data mining, vol. 8. Wiley Interdiscip Rev Data Min Knowl Discov; 2018, e1232.
[23] Kurkure U, Sivaraman H, Vu L. Machine learning using virtualized GPUs in cloud environments. International Conference on High Performance Computing. Cham, Switzerland: Abstract; 2017 Jun. p. 591–604.
[24] Nickolls J, Dally WJ. The GPU computing era. IEEE Micro 2010;30:56–69.
[25] Lee S, Park J, Kim N, Lee T, Quagliato L. Extreme gradient boosting-inspired process optimization algorithm for manufacturing engineering applications. Mater Des 2023;226:111625.
[26] Lee S, Lim Y, Galdos L, Lee T, Quagliato L. Gaussian process regression-driven deep drawing blank design method. Int J Mech Sci 2024;265:108898.
[27] Wu D, Kim K, Li Q. Computationally efficient deep neural network for computed tomography image reconstruction. Med Phys 2019;46:4763–76.
[28] Tang S, Jing C, Jiang Y, Yang K, Huang Z, Wu H, et al. The effect of image resolution on convolutional neural networks in breast ultrasound. Heliyon 2023;9: e19253.
[29] Oliveira FP, Tavares JMR. Medical image registration: a review. Comput Methods Biomech Biomed Engin 2014;17:73–93.
[30] Thévenaz P, Unser M. Optimization of mutual information for multiresolution image registration. IEEE Trans Image Process 2000;9:2083–99.
[31] Zokai S, Wolberg G. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations. IEEE Trans Image Process 2005;14:1422–34.
[32] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13:600–12.

[33] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on knowledge Discovery and Data Mining; 2016 Aug. p. 785–94. San Francisco (CA). ACM; 2016.
[34] Zhang W, Wu C, Zhong H, Li Y, Wang L. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci Front 2021;12:469–77.
[35] Rathore PS, Dadich N, Jha A, Pradhan D. Effect of learning rate on neural network and convolutional neural network. Int J Eng Res Technol 2018;6:1–8.
[36] Dechter R. Reasoning with probabilistic and deterministic graphical models: Exact algorithms. Springer Nature; 2022.
[37] Ibrahim MH, Pal C, Pesant G. Improving probabilistic inference in graphical models with determinism and cycles. Mach Learn 2017;106:1–54.
[38] Dudle A, Gugler Y, Pretterklieber M, Ferrari S, Lippuner K, Zysset P. 2D-3D reconstruction of the proximal femur from DXA scans: Evaluation of the 3D-Shaper software. Front Bioeng Biotechnol 2023;11:1111020.
[39] Schreiber JJ, Anderson PA, Rosas HG, Buchholz AL, Au AG. Hounsfield units for assessing bone mineral density and strength: a tool for osteoporosis management. JBJS 2011;93:1057–63.
[40] Pu M, Zhang B, Zhu Y, Zhong W, Shen Y, Zhang P. Hounsfield unit for evaluating bone mineral density and strength: variations in measurement methods. World Neurosurg 2023;180:e56–68.
[41] Amin MFM, Zakaria WMW, Yahya N. Correlation between Hounsfield unit derived from head, thorax, abdomen, spine and pelvis CT and T-scores from DXA. Skelet Radiol 2021;50:2525–35.
[42] Lee S, Chung CK, Oh SH, Park SB. Correlation between bone mineral density measured by dual-energy X-ray absorptiometry and Hounsfield units measured by diagnostic CT in lumbar spine. J Korean Neurosurg Soc 2013;54:384–9.