Mechanobiology in Medicine 3 (2025) 100130

Contents lists available at ScienceDirect

Mechanobiology in Medicine

journal homepage: www.sciencedirect.com/journal/mechanobiology-in-medicine

Full Length Article

Prognostic assessment of osteolytic lesions and mechanical properties of bones bearing breast cancer using neural network and finite element analysis☆

Shubo Wang[a], Tiankuo Chu[a], Murtaza Wasi[a], Rosa M. Guerra[b], Xu Yuan[c], Liyun Wang[a,*]

[a] *Department of Mechanical Engineering, University of Delaware, Newark, DE 19716, USA*
[b] *Department of Biomedical Engineering, University of Delaware, Newark, DE 19716, USA*
[c] *Department of Computer Science, University of Delaware, Newark, DE 19716, USA*

ARTICLE INFO

Keywords:
Metastatic osteolysis
Micro-computed tomography
Sequential scans
Fracture risk
Murine model

ABSTRACT

The management of skeletal-related events (SREs), particularly the prevention of pathological fractures, is crucial for cancer patients. Current clinical assessment of fracture risk is mostly based on medical images, but incorporating sequential images in the assessment remains challenging. This study addressed this issue by leveraging a comprehensive dataset consisting of 260 longitudinal micro-computed tomography (μCT) scans acquired in normal and breast cancer bearing mice. A machine learning (ML) model based on a spatial–temporal neural network was built to forecast bone structures from previous μCT scans, which were found to have an overall similarity coefficient (Dice) of 0.814 with ground truths. Despite the predicted lesion volumes (18.5 % ± 15.3 %) being underestimated by ~21 % than the ground truths' (22.1 % ± 14.8 %), the time course of the lesion growth was better represented in the predicted images than the preceding scans (10.8 % ± 6.5 %). Under virtual biomechanical testing using finite element analysis (FEA), the predicted bone structures recapitulated the loading carrying behaviors of the ground truth structures with a positive correlation ($y = 0.863x$) and a high coefficient of determination ($R^2 = 0.955$). Interestingly, the compliances of the predicted and ground truth structures demonstrated nearly identical linear relationships with the lesion volumes. In summary, we have demonstrated that bone deterioration could be proficiently predicted using machine learning in our preclinical dataset, suggesting the importance of large longitudinal clinical imaging datasets in fracture risk assessment for cancer bone metastasis.

1. Introduction

With great strides made in breast cancer screening, diagnosis, and treatment, the 5-year survival of breast cancer patients in the U.S. has increased from 60 % in the 1960s to 90 % in 2022.[1] Breast cancer patients and survivors are living longer, but also suffer many complications due to their breast cancer and/or cancer treatments.[2] Improving their quality of life has become increasingly important in patient care.[3] Bone is one common metastatic site in breast cancer, and bone metastasis results in skeletal-related events (SREs) like hypercalcemia, medullar compression, and pathological fractures that deprive patients of their autonomy and quality of life.[3–6] Pathological fractures are associated with elevated fatality in breast cancer patients.[5,7] However, it is challenging to predict and diagnose pathological fractures, which are sometime the first indication of bone metastases.[4] While prophylactic surgeries such as hemiarthroplasty or internal fixation with an intramedullary nail have been shown to improve survival rates better than osteosynthesis medications for some patients,[5,8] overtreatment with surgeries not only adds to the financial burden of patients but also causes complications and even mortality.[9–11] Thus, the management of bone fracture in cancer patients requires early detection, precise staging, and timely assessment of fracture risk to achieve optimal patient outcomes.

Imaging is essential in the management of breast cancer and bone metastases from diagnosis to treatment response assessment.[2,12,13]

This article is part of a special issue entitled: Bone Mechanobiology published in Mechanobiology in Medicine.

☆ Given her role as an Editor of this journal, Liyun Wang had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the peer-review process for this article was delegated to Lidan You.

* Corresponding author.
E-mail address: lywang@udel.edu (L. Wang).

https://doi.org/10.1016/j.mbm.2025.100130
Received 14 December 2024; Received in revised form 13 March 2025; Accepted 25 March 2025

Prevailing methods are morphological imaging techniques, including plain X-ray film, computed tomography (CT), magnetic resonance imaging (MRI), as well as nuclear medicine imaging modalities such as skeletal scintigraphy (also known as bone scan) and positron emission tomography (PET).[2] A key step in fracture risk assessment is the detection of bone metastases globally from skeletal scintigraphy, PET, and whole-body MRI.[14,15] The wide-spread spectrum of bone metastasis, from osteolytic, mixed, to osteoblastic or osteosclerotic, dampens the detection accuracy of a single imaging modality.[5] While bone scans show a higher sensitivity to detect osteosclerotic lesions, PET works well for osteolytic bone lesions.[14] Revealing local changes in bone structure, however, requires high spatial resolution CT scanning, which provides better morphological information of both cortical and trabecular bone.[5,14] Although Mirels and spinal instability neoplastic scores (SINS) are commonly used to assess the fracture risk qualitatively,[6,16] these methods are subjective with poor sensitivity and specificity that lead to overtreatment.[17,18] To understand the biomechanical behaviors of bone at risk of fracture, finite element (FE) methods have been used to calculate the strain and stress distribution on patient specific bone structures[19–22] and to quantify the reduction of the bone's load bearing capacity due to metastatic lesions.[23] Combining FE modeling and clinical imaging provides a quantitative tool in fracture risk assessment.

Resurgence of deep learning in computer vision since 2012 has spread to medical imaging and contributed to progresses in disease diagnosis, tumor/lesion segmentation, and anomaly detection.[24,25] Prediction of tumor growth from chronological image data has been performed and reported on brain tumors.[26–28] Conventionally, this problem was studied by using reaction diffusion-based models.[29–31] Zhang et al. recently applied deep learning to predict the growth of pancreatic tumor and brain tumor using LSTM (Long Short-Term Memory) based network and GANs (Generative Adversarial Networks).[32,33] Wang et al. trained an image-generating model based on multimodal imaging of MRI and CT data.[34] By far, only our previous study has attempted to predict volumetric images of osteolytic lesions.[35] Whereas the tumor is the main driving force of metastatic lesions in bone, structural integrity is crucial in determining bone strength. Under hormonotherapy, immune checkpoint therapies, and/or anti-bone resorption treatments, osseous tumor lesion type may shift over time between osteolytic and osteosclerotic.[36–38] Thus, leveraging the longitudinal image data is paramount to predict bone lesion development and progression for fracture risk assessment.

In this study, we present a proof of concept to assess the mechanical competence of bone at a future time point. Our objective was to test if deep learning could predict the progression of bone osteolysis, which could be further used to assess the mechanical behaviors of the bone. We retrospectively collected an extensive series of sequential μCT images from a murine model of bone metastases. A spatial–temporal neural network was developed and trained to generate future frames reflecting the disease's progression. These predicted μCT images were then utilized in Finite Element simulations of axial compression. The simulated mechanical behaviors demonstrated by the predicted images were found to be comparable to those of actual ground truth images. This study highlights (1) the importance of collecting sequential human imaging data to assess bone fracture risks in metastatic diseases, and (2) the potential of generative approaches in enhancing clinical datasets.

2. Methods and materials

2.1. Animal model of bone metastasis

A murine allograft model of breast cancer bone metastases was used in this study,[39] which was approved by the Institutional Animal Care and Use Committee (IACUC). Female C57BL/6J mice (Jackson Laboratory, Bar Harbor, ME, USA) were housed 4–5 mice per cage in an accredited animal facility at the University of Delaware. Mus musculus mesenchymal-like Py8119 cell line (ATCC, Manassas, VA, USA, CRL-3278™), originally derived from spontaneously arising tumors in MMTV-PyMY transgenic C57BL/6 female mice, were cultured in F–12K cell culture medium (ATCC, 30–2004™) supplemented with 5 % FBS at 37 °C degree and 5 % (v/v) CO2 in an incubator. Tumor cells were inoculated to mice by either intra-tibial injection into the proximal end of tibia (90 % of animals) or intracardiac injection (10 % of animals). Animals were subjected to different loading regimens, namely, sedentary, aerobic treadmill running, or localized tibial compression while their bones were scanned weekly.[39] Osteolytic lesions, the full thickness perforations of load-bearing cortices, occurred predominantly in the proximal metaphysis of tibiae and became visible in μCT images around week 2 to week 4.[39] Thus, the tibial proximal ends were the focus of this study.

2.2. μCT image acquisition and processing for machine learning and FE analysis

Detailed data acquisition and image processing can be referred to 39. In brief, μCT images of tumor-bearing and tumor-free mouse tibiae were acquired on a weekly basis using SkyScan® 1276 (Bruker, Kontich, Belgium). For each acquisition, 260 transmission frames were taken while the source-detector rotates over 180° with a step of 0.8° with the following parameters: 900 ms exposure time, 200 mA current, 50 kVp, and 0.5 mm Al filter. Volumetric CT images were reconstructed using NRecon® software (Bruker) with a voxel size of $7 \times 7 \times 7\ \mu m^3$. Proximal tibiae were automatically cropped using in-house python program into volumes of $3.50 \times 3.50 \times 1.68\ mm^3$ (500 × 500 × 240, pixel size). The first scan of each animal was registered to a common reference using the SimpleITK® package,[40] which was then used as references to register follow-up scans within the same animal. Cortical bone regions, the focus of this study, were automatically segmented as in our previous study.[39] Osteolytic lesions, often in cortical bone, were obtained by subtracting cortical region in registered image from that in the first scan when no lesion was developed. The percentage of lesion volume was calculated as lesion volume divided by the initial cortical bone volume.

To accommodate GPU memory limitations in deep learning and computational costs in FE modeling, the registered μCT images were further cropped and downsampled to volumes with pixel spacing of 35 μm (96 × 96 × 48, matrix size) using linear interpolation. Downsampled volumetric μCT data were smoothened and segmented into binary label maps. Details of the dataset (Table 1) include the number of the total sequences of μCT images with varying length equal to 2 time points and the number of those with osteolytic lesions (n = 34/63, osteolytic/total), 3 time points (n = 47/112), 4 time points (n = 30/52), and 5 time points (n = 16/33). The dataset was randomly split 70 %, 15 %, and 15 % (n = 182, 39, or 39 scans) for training, validation, and testing. The ratios of osteolytic samples with lesions were 49 % (89/182), 49 % (19/39), and 49 % (19/39) for training, validation, and testing, respectively (Table 1). For the FEA, 3D surface models of the bone specimens were exported as STL files using 3D Slicer® 4.0.[41] Disconnected bone pieces were semi-manually removed using the 3D Builder® (version 15.2, Microsoft Corp, Redmond, WA) and the wizard tool in Materialise® 3-matic (Materialise, Leuven, Belgium). Refined surface meshes were converted to 3D volumetric meshes and then exported in the Abaqus input file using the Materialise® 3-matic software.

Table 1
Summary of the dataset used in the study.

		# of Samples	Training	Validation	Testing
# of scans	2	63 (34)	36 (19)	10 (6)	17 (9)
	3	112 (47)	84 (34)	11 (5)	17 (8)
	4	52 (30)	39 (24)	8 (4)	5 (2)
	5	33 (16)	23 (12)	10 (4)	0
Total		260 (127)	182 (89)	39 (19)	39 (19)

Note: The dataset includes a total of 260 samples with 127 of them showing osteolytic lesions. The number of samples (with osteolytic lesions) is shown for sequences with 2–5 scans, as well as the corresponding number of samples used for training, validation, and testing.

2.3. Prediction of future frame by machine learning

2.3.1. Problem definition

Our task is to predict a volumetric μCT image $X(t_{n+1})$ taken at time point t_{n+1} given the scans $X(t_1 : t_n)$ taken at previous time steps in a self-supervised manner. Because the input and output belong to the same sequence, this task is similar to the video prediction process in computer vision, in which future frames are generated for a given short video sequence. However, there are several important differences between our application and the video prediction: a) The time interval is one week in our preclinical study and can be much longer (months) in clinical applications, thus the changes between neighboring frames can be more dramatic than the short video sequences; b) Each frame is a single-channel gray-scale 3D image in this study ($X(t) \in R^{48\times96\times96}$) that contains spatial information in three dimensions, in contrast with the 2D RGB images in videos; c) Our sequence lengths are much shorter and only one future frame is considered in our study rather than multiple frames in short video predictions; d) The underlying mechanism driving the chronologic changes is biological processes while video prediction usually deals with motions governed mainly by physics.

2.3.2. Deep learning model architecture

As the underlying issue of our problem is sequence prediction, LSTM (Long Short-Term Memory)[42] is chosen in this study to handle temporal changes within the sequences (Fig. 1A). A convolutional autoencoder is used to extract low dimensional features ($z \in R^{3\times4\times4}$) from each input 3D image. The encoder, composed of four residual blocks, downsamples the original input. Each residual block has two 3D convolutional (Conv3D) layers with skip connections similar to ResNet[43] (Fig. 1B). The decoders also have four residual blocks constructed with 3D transposed convolutional layers (Conv3DTranspose) to reconstruct the original shape of the input image.[44] Please note that the inputs and outputs in this model are greyscale 3D images.

2.3.3. Training and evaluation

The model was implemented in TensorFlow® 2.2 and Python® 3.7.9 on an Ubuntu 18 LTS machine with an AMD threadripper 1900x CPU (16 logic processors), 32 GB memory and NVIDIA 1080 TI GPUs (11 GB memory). We used 70 % of the data for training while the rest were split evenly for validation and evaluation. The model performance was evaluated in terms of PSNR (Peak signal to noise ratio, defined in Eqs. (1) and (2)),[45] Dice score (defined in Eq. (3)),[46] SSIM (Structural similarity, defined in Eq. (4)),[47] and LPIPS (Learned Perceptual Image Patch Similarity, defined previously[48]). PSNR and SSIM are commonly used for evaluation of image quality, while the Dice score is mostly employed in the tasks of segmentation to evaluate morphology. LPIPS is a deep learning-based evaluation method that compares the similarities between the activations of two images using pre-trained neural networks like VGG[48] or Alexnet,[49] which mimics human perceptions more closely than classic metrics like MSE, PSNR, and SSIM.[50] A version based on VGG[48] was used in this study.

$$MSE = \frac{1}{M \cdot N} \sum_{i,j}^{M,N} \left(X(i,j) - X'(i,j)\right)^2 \tag{1}$$

$$PSNR = 10\, log_{10}\left(\frac{MAX}{MSE}\right), where\ MAX = 255\ for\ Uint8\ image \tag{2}$$

$$Dice = \frac{\sum_{i,j}^{M,N} X(i,j) \cdot X'(i,j)}{\sum_{i,j}^{M,N} X(i,j)^2 + X'(i,j)^2} \tag{3}$$

$$SSIM(X, X') = \frac{(2\mu_X \mu_{X'} + C_1)(2\sigma_{XX'} + C_2)}{({\mu_X}^2 + {\mu_{X'}}^2 + C_1)({\sigma_X}^2 + {\sigma_{X'}}^2 + C_2)} \tag{4}$$

2.4. Finite element (FE) analysis

2.4.1. Finite element (FE) models of axial compression

ABAQUS/Standard 2020 (Dassault Systèmes SE, Vélizy-Villacoublay, France) was employed to simulate axial compression of the digitalized proximal tibiae. Twenty four out of 39 predicted μCT images randomly selected from the test dataset and their paired ground truth images were used in the FE analysis. For each FE model, the volumetric mesh was composed of approximately 1.6 million tetrahedral elements. The bone specimen was assumed to be homogeneous and isotropic. The elastic–plastic properties were defined as Young's modulus of 11 GPa, Poisson's ratio of 0.3, yield stress of 130 MPa, and no loading capacity post yielding based on literature.[39] To assist virtual compression testing of the bone samples, two nondeformable square plates were positioned parallel to the superior and inferior surfaces of the cubical volume of the bone sample. The lower plate was fixed at its original position throughout the test while the upper plate was assigned to move towards the sample till its position reached boundary condition equivalent to a total displacement of 0.5 % of the sample height. A friction coefficient of 0.2 was assigned to

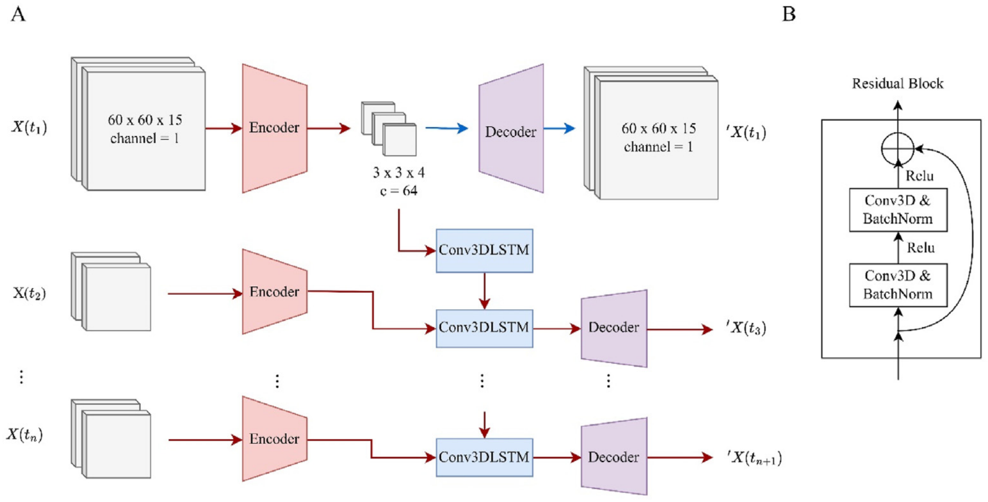

Fig. 1. A) The architecture of the neural network: the encoders extract the lower-dimension feasures from the sequential inputs X (t_i) ($i = 1\ to\ n$), which are reconstructed to forecast the future frame at t_{n+1} by decoders. The hidden variables are parsed to a 3D convolutional LSTM layer to learn the spatial and temporal information. B) Both encoder and decoder are composed of residual blocks of Conv3D layers.

both contact surfaces. To accommodate the nonlinearity during initial contact, a pre-load was used to bring all parts into full contact with about 1 % of the targeted displacement. Automatic stabilization was enabled in Abaqus to facilitate convergence. The pre-load phase used a larger dissipated energy fraction of 0.05, which was reduced to 0.0002 in the subsequent full-load step.

The output ODB files from ABAQUS were postprocessed using ABAQUS Python to extract reaction forces on the moving plate, total displacement, time, total strain energy (ALLSE), and damping energy (ALLSD). The ratio of ALLSD over ALLSE was used to confirm that the fraction of dissipated energy was within the range of 5 % in the full-load step. The contours of von Mises stress (σ_v, Eq. (5)) were directly obtained from Abaqus.

$$\sigma_v = \sqrt{\frac{1}{2}\left[(\sigma_{11}-\sigma_{22})^2+(\sigma_{22}-\sigma_{33})^2+(\sigma_{33}-\sigma_{11})^2\right]+3\left(\sigma_{12}^2+\sigma_{23}^2+\sigma_{31}^2\right)} \tag{5}$$

2.4.2. Comparisons of predicted and ground truth FE models

To compare the results between models generated from the ground truth image (GT) and from the predicted image (Pred), reaction force and displacement were interpolated linearly to extract data at different apparent strains from 1000, 2000, 3000, or 5000 με, which mimicked the range of in vivo strains experienced by bones. Compliances were calculated by finding the slopes of the strain–stress curves. Fifteen GT models and their corresponding Pred models were evaluated. Linear regression and the coefficient of determination (R^2) were used to compare the reaction forces at 4 levels of apparent strain. To further test the concordance between Pred models and GT models, Lin's concordance correlation coefficient (CCC, Eq. (6)) was employed.[51] Mean absolute percentage error (MAPE, Eq. (7)) was also reported as value ± standard deviation.

$$CCC = \frac{2\rho\sigma_x\sigma_y}{\left(\mu_x-\mu_y\right)^2+\sigma_x^2+\sigma_y^2}, \tag{6}$$

$$\text{where } \rho \text{ is Pearson's coefficient}$$

$$MAPE = \frac{\left|x_{gt}-x_{pred}\right|}{x_{gt}}*100 \tag{7}$$

3. Results

3.1. Predicted 3D image scans matched well with ground truths

A visual comparison of GT images and Pred outputs by our neural network was conducted and 2D slices from the corresponding 3D stacks are shown (Fig. 2A). Three tibiae with osteolytic lesions were chosen randomly and their scan sequences are shown in three blocks (Fig. 2A). Each tibia underwent two or three trials (columns), in which 2, 3 or 4 scans acquired sequentially (rows) served as inputs (observed) to the neural network and the model outputs (predicted/future frame) were compared with ground truth of the next time point (Fig. 2A). As shown in the figure, the predicted bone shapes were comparable to those of GT images. At time point *t3* (Week 3), the trabecular bone was mostly lost due to tumor growth. In the Pred images, reduction and disappearance of trabecular bone matched well with the GT images but a detailed comparison of the trabecular topology was not made due to the low trabecular bone volumes. Cortical bone integrity and the location of osteolytic lesions were successfully predicted as indicated by the orange arrowheads. However, some lesions (the second trial for tibia #3) were not successfully predicted with three time point inputs. Failures were observed on the third trial of the same tibia, because of the small perforations formed after thinning of the cortical bone (Fig. 2A). These predicted and ground truth images were meshed to build FE models (Fig. 2B). It was noted that the boundaries of lesions were smoother in predicted images than those in GT images, which may alter local stress and yield patterns.

Quantitative metrics showed a general good agreement between the predicted and the ground truth images (Table 2). Overall, the Pred images achieved a Dice score of 0.814 relative to ground truth images. The image quality was assessed by PSNR, SSIM and LPIPS. The metrics were comparable with our previous study,[35] confirming the feasibility of the trained model to quantitatively predict bone structure and mineralization of a tumor bearing skeleton. The grey values of the predicted voxels in Hounsfield units were calibrated with phantoms of a known mineral density.

3.2. Predicted scans recapitulated the load-bearing behaviors of the ground truths

FE simulations of axial compression of the meshes built from the Pred and GT bone structures showed the evolution of the stress fields under

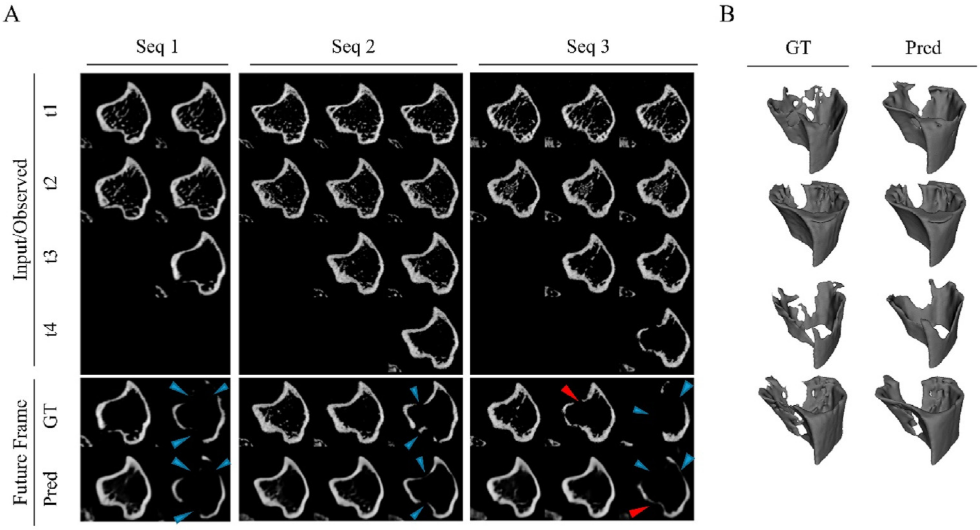

Fig. 2. Representative 2D slices of inputs, ground truth, and predicted images. A) For three randomly chosen tibiae, two to four images captured at prior time points were inputs to the deep learning model, which generated a future frame (Pred) and compared with their corresponding ground truth. Lesions correctly predicted were labeled with blue arrowheads, while those not or incorrectly identified labeled with red arrowheads. B) representative 3D renderings of the GT and Pred images.

Table 2
Evaluation metrics for our deep learning model.

Input length (Seq #)	PSNR	SSIM	LPIPS	Dice
2 (n = 17)	34.50 ± 0.97	0.850 ± 0.081	0.040 ± 0.020	0.852 ± 0.068
3 (n = 17)	35.50 ± 1.57	0.869 ± 0.073	0.039 ± 0.028	0.845 ± 0.087
4 (n = 5)	36.40 ± 1.97	0.848 ± 0.130	0.062 ± 0.07	0.755 ± 0.235
Overall	35.56 ± 1.68	0.857 ± 0.093	0.047 ± 0.044	0.814 ± 0.149
	↑	↑	↓	↑

Note: An up-arrow indicates that higher values are better while a down-arrow indicates that lower values are favorable. Model performance is stratified by the length of input sequence varying from 2, 3, or 4 (Input length of 5 is removed due to insufficient samples).

increasing apparent strains for two samples (Fig. 3). It was not surprising that bone lesions and bone structural geometry greatly impacted the stress distribution over the bone (Fig. 3A). For Sample 1, the larger lesion on the lateral side (yellow arrow) was accurately predicted by the deep learning model, while the two smaller lesions on the posterior side were not captured in the Pred model. Nevertheless, the reaction force–displacement curves were comparable between the GT and Pred models (Fig. 3B). The distributions of stress were also closely matched, as both models showed relatively high stress levels on the tibial ridge and posterior-medial surface than on other locations. However, for Sample 2, the bone spur on the tibial ridge, indicated with the cyan circle, was missing in the Pred model, which adversely affected the initial contact and stability of the FE simulation, leading to deviations in the reaction force–displacement curve compared with that of the GT model (Fig. 3B).

Decent correlations were found between the reaction forces of the GT and Pred models under axial compression (Fig. 4). In the case of 1000 με compression, the correlation of the reaction forces was poor ($R^2 < 0.5$), because of the mismatch of geometries in the contact region as described above. With increasing compression, the reaction forces in the GT and Pred models started to converge with increasing R^2 up to 0.681 (5000 με compression). The reaction forces were approximately 14–19 % larger in Pred models as shown in the slopes of the fitting lines (Fig. 4), possibly due to underestimated lesions in the Pred models than those in the GT

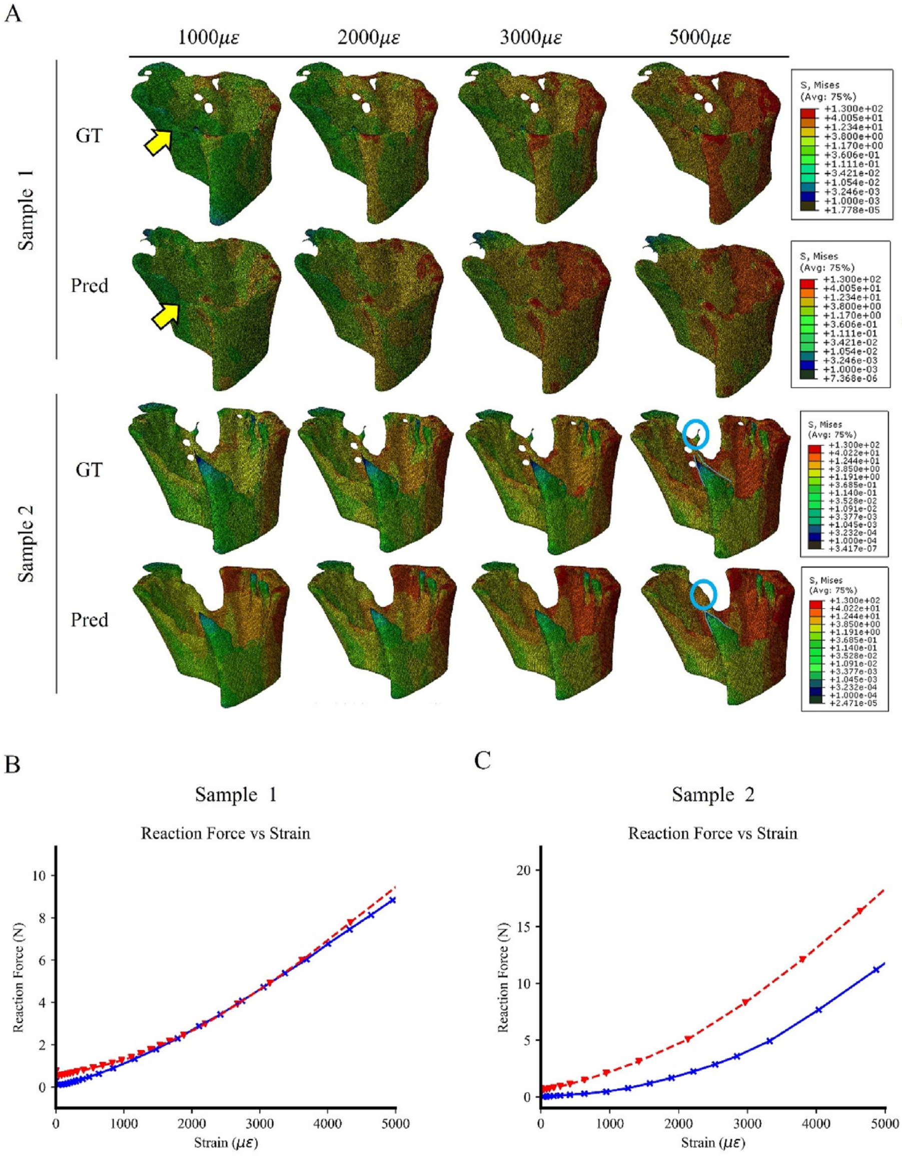

Fig. 3. Results of FE modeling of axial compression of proximal tibiae. A) contours of von Mises stress from Abaqus. B) Force-displacement curves of Pred models (red dashed lines) and GT models (blue solid lines). Yellow arrows: correctly predicted lesion in Sample 1; Cyan circles: missed prediction in Sample 2.

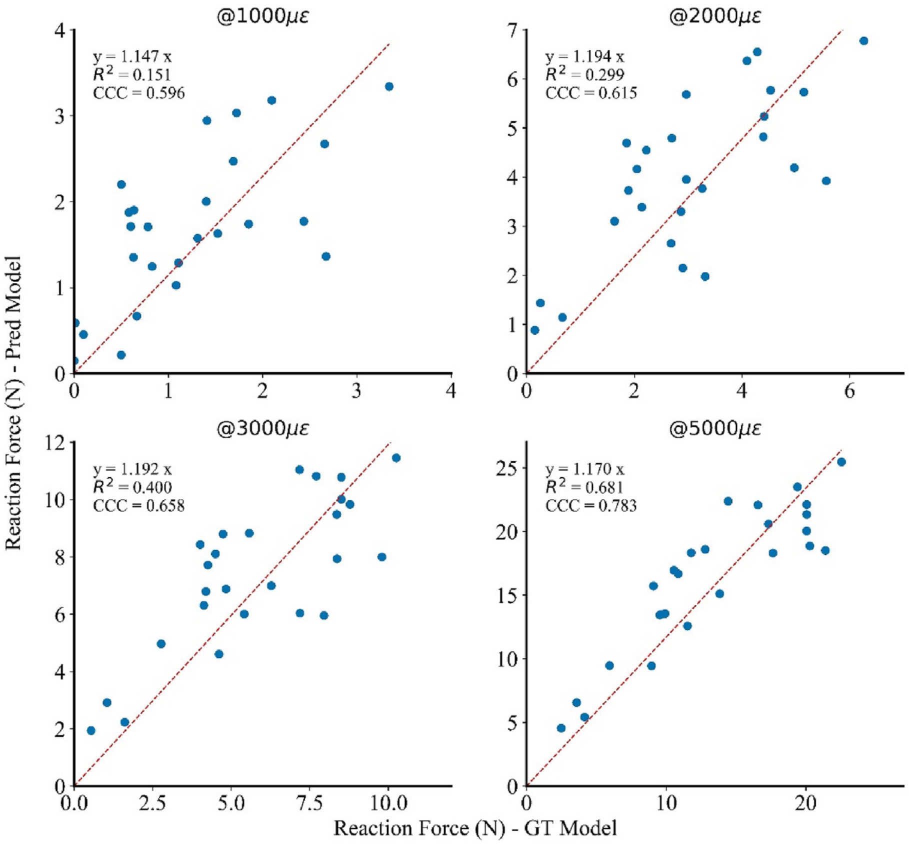

Fig. 4. Comparisons of reaction forces in GT and Pred models at 4 different strain levels. Goodness of linear fitting was indicated by R^2 and CCC.

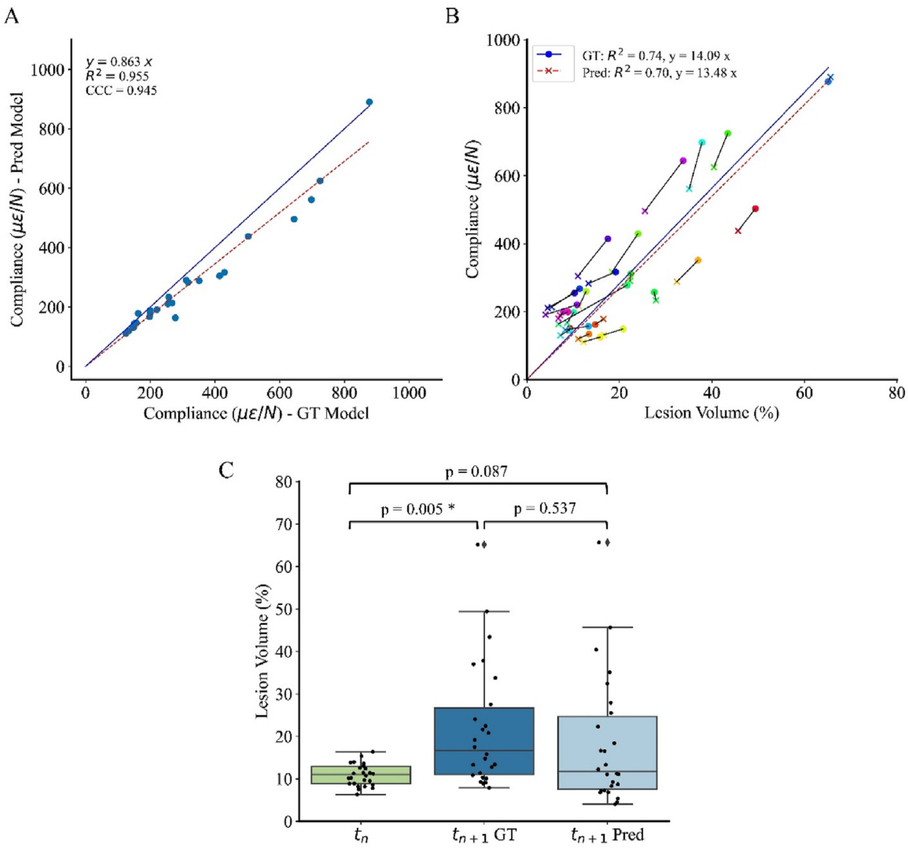

Fig. 5. A) Compliances of Pred models were highly correlated with GT models. The red dashed line represents the fitted curve while the blue solid line indicates perfect regression (GT = Pred). B) Bone c*ompliances are positively correlated with the lesion volumes*. GT-Pred pairs, linked with short lines, are differentiated by colors. Solid blue line and dashed red line are fitting curves for GT and Pred, respectively. C) Comparing the osteolytic lesions at current time (t_n) and future time (t_{n+1}) showed that lesion volumes significantly increased with time in the ground truth sequences (p = 0.005) and that the predicted lesion volumes did not differ from the ground truth (p = 0.537), indicating the diagnostic potentials of the predictive model.

models.

The compliances of 15 pairs of Pred and GT bone models were calculated using the linear portion of their force–displacement curves (Fig. 5A). Tibiae with larger lesions, not surprisingly, had higher compliances, while those with small or no lesions showed lower compliances, which were comparable to our previously reported values from strain gauging tests.[39] The Pred models' compliances agreed well with those of the GT models ($y = 0.864\ x$, $R^2 = 0.956$ and CCC = 0.945, Fig. 5A), although the former was 13.6 % lower, which might be caused by an underestimation of small lesions in the predicted models.

3.3. *Predicted lesion volumes correlated with bone compliances*

To connect the lesion volume with bone's mechanical competence, we investigated the relationship between bone compliances and the percentages of osteolytic lesion volume in the cortical bone region (Fig. 5B). The differences between the FE models derived from the GT and Pred images were indicated by short black linkers (Fig. 5B). Positive relationships between tibial compliances and cortical lesion volumes were observed for both the Predicted and GT models, with comparable fitting slopes and R^2 values for GT ($y = 14.09x$, $R^2 = 0.74$) and Pred ($y = 13.48x$, $R^2 = 0.7$).

3.4. *ML-based prediction of lesions in sequential scans showed prognostic potential*

To answer the questions (1) if predicting cancer lesion volume is necessary and (2) if our predictions are reasonably well for prognostic purpose, we plotted the lesion volumes at t_n and t_{n+1} of the ground truth sequences as well as the t_{n+1} predictions using our ML method (Fig. 5C). As the disease progresses, the bone lesions in our preclinical dataset become significantly larger from t_n (10.8 % ± 6.5 %) to t_{n+1} (22.1 % ± 14.8 %, p = 0.005), demonstrating the need of forecasting of the bone lesions to track the decline of bone's loading capacity. The predicted lesions at t_{n+1} (18.5 % ± 15.3 %), despite being 3.6 % lower than the ground truth values at the t_{n+1}, were superior to the ground truth at t_n in representing the distribution of lesions (Fig. 5C). Thus, the predicted bone structures at t_{n+1} were expected to provide a more accurate assessment of bones' loading capacity during the progress of metastasis (Fig. 5B). There is no significant difference between the predicted lesions and ground truths (Fig. 5C).

4. Discussion

Assessment of fracture risk in cancer metastatic diseases is challenging but critical for patient outcome.[3] Bone metastases can remain asymptomatic until SREs occur, such as severe pain, spinal cord compression, and pathological fractures.[3,13,22] Bone fractures, which lead to a variety of complications, impact not only the patients' physical functions but also survival rates.[6,8,52–54] Prophylactic treatments such as curative resection, prosthetic reconstruction of impending fracture, and intramedullary fixation with bone cement injections have been reported to improve survival rates.[8,11,54–56] The surgical interventions, unfortunately, increase the risk of complications like lung embolism, aspiration pneumonia, deep vein thrombosis and infections.[9,10,57] Medical professionals must carefully balance the risks and benefits of prophylactic surgeries over conservative treatments. Timely assessment of fracture risk is therefore paramount in the management of bone metastases and rational design of treatment plans.[3–5] To this end, many studies have been conducted towards comprehensive screening and prognosis of impending pathological fractures[2,58–61] as well as developing analytical methods for fracture assessment such as rigidity analysis and FE modeling.[19–23,62,63]

This study aimed not to replace the above-mentioned methods but rather to introduce a data-driven augmentation to the existing methods. We combined neural network and finite element modeling to test if we could predict cancer lesions and the associated declining of load bearing capacity at future time points. The knowledge of disease trajectory is important for clinicians to determine the time window of the surgical interventions. As mentioned earlier, aggressive surgical interventions could also have negative effects. The tools of predicting the lesion progression and mechanical properties could be highly valuable in not only the planning of breast cancer surgical treatment but also in evaluating the effects of other interventions. In the animal studies that provided our datasets, some mice received treadmill running and localized tibial loading and thus showed nonuniform progression of the osteolytic lesions relative to nonloaded controls.[39] This variability reflected the mechanosensitive nature of the skeleton.[64] The predicted images were then used in the simulation of axial compression by finite element method. The mean absolute percentage errors (MAPEs) of elastic compliances of the Pred models were small (14.2 %) and there was a high correlation of mechanical behaviors between the Pred and GT FE models. Although in the current work, homogenous Young's modulus were assigned in the FE analysis, the predicted voxel intensities can be used to estimate bone mineral density and build heterogenous FE models with calibrated Hounsfield and Young's modulus conversion as successfully demonstrated in the literature.[65] This study presents a proof-of-concept to predict bone structure by deep learning method and forecast the bone's mechanical properties as the disease evolves.

Deep learning has dramatically boosted biomedical image analysis in many aspects such as classification, segmentation, and diagnosis.[24,25] In orthopedics, progression has been made in the detection and classification of bone fractures,[66] osteoarthritis,[67,68] osseous metastases,[69–71] and degenerative disc disease.[72] Prediction of tumor growth is not new and interest in predicting spatial and temporal changes of pathological conditions is growing rapidly.[26,27,30,32,34] Our research group first proposed prediction of tumor-induced osteolysis using the temporal variational autoencoders.[35] In this study, we further demonstrated the potential of the ML-based assessment of bone's fragility and fracture risk by performing virtual mechanical testing.

There are several limitations of this study. The present study was based on a murine dataset and the translation to clinics needs to overcome several challenges. Mouse models, however, are valuable tools in elucidating disease mechanisms, identifying diagnostic biomarkers, and testing treatments.[73] The small-sized mouse skeleton allows quick three-dimensional CT scans at multiple time points over the span of breast cancer metastasis with control of the age, the type of cancer, and site of inoculation, as well as interventions. With rigorous experimental design incorporating various controls, the acquired murine image datasets could reflect the complexity of human breast cancer progression with superior image quality and labeling over available patient data. The mouse dataset thus provides an idealized testbench for us to test and optimize AI tools prior to tackling more challenging human datasets. Currently, due to the concern of overexposure of radiation in repeated imaging, human datasets of sequential CT images are lacking, a critical limiting factor in unlocking the power of AI in treating cancer bone metastasis. It is worth noting that sequential volumetric CT images may not be desired.[74] A baseline CT image with follow-up planar radiographs can be used to synthesize time series of 3D volumes by utilizing 2D/3D registration[75,76] and statistical shape models[77,78] of metastatic bones. In studies of treatment response in metastatic bone disease, the frequency of imaging ranges from weeks to 2–6 months.[13,79] In the current study, temporal information is implicitly embedded as the chronological order of images since the frequency of imaging is fixed (weekly). In contrast, the interval of clinical CT imaging is highly varied, depending on the clinical design and possible missed exams. However, this complication could be addressed by explicitly encoding time-related input parameters.[80] The second limitation of this study was the idealized boundary conditions used in the FE simulations (axial compression), and the impacts of different physical activities were not investigated. Finally, our datasets were obtained in a murine model of osteolytic bone metastasis, in which cancer treatments such as radiation therapy, chemotherapy or anti-resorptive therapeutics were not applied. The progression of bone lesion in breast cancer patients would be much more variable, given the different phenotypes associated with cancer subtypes and the highly personalized responses to cancer treatments. With the accumulation of more clinical data and the advance of machine learning, we anticipate that data-driven biomechanical prediction of bone strengths and fracture risk assessment will be further improved.

5. Conclusion

This study provides more insights on reducing the disparity between the uncertainties in existing scoring systems of bone strength and urgent needs of improving risk management of pathological fractures for bone metastases patients. Our new approach combining spatial–temporal neural network and FE modeling exhibits the ability to predict bone structure and loading capacity in an animal model of bone metastases. This study demonstrates the importance of collecting sequential human imaging data and the potential of generative approaches in enhancing clinical datasets, which will improve diagnosis, prognosis, and treatment evaluation of cancer bone metastasis.

CRediT authorship contribution statement

Shubo Wang: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tiankuo Chu:** Resources, Data curation. **Murtaza Wasi:** Data curation. **Rosa M. Guerra:** Writing – review & editing. **Xu Yuan:** Writing – review & editing, Resources. **Liyun Wang:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Data availability

The microCT data are stored on a data server at the University of Delaware. Free download is available upon request.

Ethical approval

The animal experiments were approved by the IACUC (AUP 1316) of University of Delaware.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank the OLAM staff at the University of Delaware for assisting the animal studies. The study was partially supported by the Core Access Awards from the Delaware Center for Musculoskeletal Research funded by NIH COBRE (P20 GM139760). M. Wasi and R.M. Guerra received support from University of Delaware (Graduate Scholar Awards).

References

1. American Cancer Society. *Breast Cancer Facts & Figures 2022-2024*. Atlanta: American Cancer Society, Inc; 2022.
2. O'Sullivan GJ, Carty FL, Cronin CG. Imaging of bone metastasis: an update. *World J Radiol*. 2015;7(8):202–211.
3. Confavreux CB, Follet H, Mitton D, Pialat JB, Clezardin P. Fracture risk evaluation of bone metastases: a burning issue. *Cancers (Basel)*. 2021;13(22):5711.
4. Damron TA, Mann KA. Fracture risk assessment and clinical decision making for patients with metastatic bone disease. *J Orthop Res*. 2020;38(6):1175–1190.
5. Nguyễn M-V, Carlier C, Nich C, Gouin F, Crenn V. Fracture risk of long bone metastases: a review of current and new decision-making tools for prophylactic surgery. *Cancers*. 2021;13(15):3662.
6. Ye C, Leslie WD, Al-Azazi S, et al. Fractures and long-term mortality in cancer patients: a population-based cohort study. *Osteoporos Int*. 2022;33(12):2629–2635.
7. Anract P, Biau D, Boudou-Rouquette P. Metastatic fractures of long limb bones. *Orthop Traumatol Surg Res*. 2017;103(1S):S41–S51.
8. Philipp TC, Mikula JD, Doung Y-C, Gundle KR. Is there an association between prophylactic femur stabilization and survival in patients with metastatic bone disease? *Clin Orthop Relat Res*. 2020;478(3):540.
9. Aneja A, Jiang JJ, Cohen-Rosenblum A, et al. Thromboembolic disease in patients with metastatic femoral lesions: a comparison between prophylactic fixation and fracture fixation. *JBJS*. 2017;99(4):315–323.
10. Choong PF. Cardiopulmonary complications of intramedullary fixation of long bone metastases. *Clin Orthop Relat Res*. 2003;(415):245–253.
11. Zickel RE, Mouradian WH. Intramedullary fixation of pathological fractures and lesions of the subtrochanteric region of the femur. *J Bone Joint Surg Am*. 1976;58(8): 1061–1066.
12. Vassiliou V, Andreopoulos D, Frangos S, Tselis N, Giannopoulou E, Lutz S. Bone metastases: assessment of therapeutic response through radiological and nuclear medicine imaging modalities. *Clin Oncol*. 2011;23(9):632–645.
13. Cook GJ, Azad GK, Goh V. Imaging bone metastases in breast cancer: staging and response assessment. *J Nucl Med*. 2016;57:27–33.
14. Heindel W, Gübitz R, Vieth V, Weckesser M, Schober O, Schäfers M. The diagnostic imaging of bone metastases. *Deutsches Ärzteblatt International*. 2014;111(44):741.
15. Bäuerle T, Semmler W. Imaging response to systemic therapy for bone metastases. *Eur Radiol*. 2009;19:2495–2507.
16. Mirels H. Metastatic disease in long bones A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop Relat Res*. 1989;249:256–264.
17. Howard E, Shepherd K, Cribb G, Cool P. The validity of the Mirels score for predicting impending pathological fractures of the lower limb. *Bone Joint J*. 2018; 100(8):1100–1105.
18. Damron TA, Nazarian A, Entezari V, et al. CT-based structural rigidity analysis is more accurate than Mirels scoring for fracture prediction in metastatic femoral lesions. *Clin Orthop Relat Res*. 2016;474:643–651.
19. Lin Y, Ma L, Zhu Y, et al. Assessment of fracture risk in proximal tibia with tumorous bone defects by a finite element method. *Microsc Res Tech*. 2017;80(9):975–984.
20. Sternheim A, Giladi O, Gortzak Y, et al. Pathological fracture risk assessment in patients with femoral metastases using CT-based finite element methods. A retrospective clinical study. *Bone*. 2018;110:215–220.
21. Kawabata Y, Matsuo K, Nezu Y, Kamiishi T, Inaba Y, Saito T. The risk assessment of pathological fracture in the proximal femur using a CT-based finite element method. *J Orthop Sci*. 2017;22(5):931–937.
22. Sternheim A, Traub F, Trabelsi N, et al. When and where do patients with bone metastases actually break their femurs? A CT-based finite element analysis. *Bone & Jt J*. 2020;102(5):638–645.
23. Anez-Bustillos L, Derikx LC, Verdonschot N, et al. Finite element analysis and CT-based structural rigidity analysis to assess failure load in bones with simulated lytic defects. *Bone*. 2014;58:160–167.
24. Haque IRI, Neubert J. Deep learning approaches to biomedical image segmentation. *Inform Med Unlocked*. 2020;18:100297.
25. Isensee F, Jaeger PF, Kohl SaA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–211.
26. Elazab A, Wang C, Gardezi SJS, et al. GP-GAN: brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images. *Neural Netw*. 2020;132:321–332.
27. Petersen J, Jäger PF, Isensee F, et al. Deep probabilistic modeling of glioma growth. InMedical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. Springer International Publishing; 2019:806–814.
28. Lipková J, Angelikopoulos P, Wu S, et al. Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and bayesian inference. *IEEE Trans Med Imag*. 2019;38(8):1875–1884.
29. Domschke P, Trucu D, Gerisch A. Mathematical modelling of cancer invasion: implications of cell adhesion variability for tumour infiltrative growth patterns. *J Theor Biol*. 2014;361:41–60.
30. Elazab A, Bai H, Abdulazeem YM, et al. Post-surgery glioma growth modeling from magnetic resonance images for patients with treatment. *Sci Rep*. 2017;7(1): 1–13.
31. Swan A, Hillen T, Bowman JC, Murtha A, et al. A patient-specific anisotropic diffusion model for brain tumour spread. *Bull Math Biol*. 2018;80(5):1259–1291.
32. Zhang L, Lu L, Wang X, et al. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. *IEEE Trans Med Imag*. 2020; 39(4):1114–1126.
33. Zhang L, Lu L, Summers RM, Kebebew E, Yao J. Personalized pancreatic tumor growth prediction via group learning. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017. Cham: Springer International Publishing; 2017.
34. Wang C, S RA, Zhang S, et al. Predicting spatial esophageal changes in a multimodal longitudinal imaging study via a convolutional recurrent neural network. *Phys Med Biol*. 2020;65(23):235027.
35. Xiong W, Yeung N, Wang S, Liao H, Wang L, Luo W, et al. Breast cancer induced bone osteolysis prediction using temporal variational autoencoders. *BME Front*. 2022; 2022:9763284. https://doi.org/10.34133/2022/9763284. PMID: 37850158; PMCID: PMC10521666.
36. Lind JS, Postmus PE, Smit EF. Osteoblastic bone lesions developing during treatment with erlotinib indicate major response in patients with non-small cell lung cancer: a brief report. *J Thorac Oncol*. 2010;5(4):554–557.

37. Bersanelli M, Bini P, Rabaiotti E, et al. Osteoblastic progression during EGFR tyrosine kinase inhibitor therapy in mutated non-small cell lung cancer: a potential blunder. *Tumori J*. 2017;103(1):66–71.
38. Garfield D. Increasing osteoblastic lesions as a manifestation of a major response to gefitinib. *J Thorac Oncol*. 2006;1(8):859–860.
39. Wang S, Pei S, Wasi M, et al. Moderate tibial loading and treadmill running, but not overloading, protect adult murine bone from destruction by metastasized breast cancer. *Bone*. 2021;153:116100.
40. Yaniv Z, Lowekamp BC, Johnson HJ. SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. *J Digit Imag*. 2018;31(3):290–303.
41. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*. 2012; 30(9):1323–1341.
42. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8): 1735–1780.
43. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770–778.
44. Li P, Wu W, Liu L, Serry FM, Wang J, Han H. Automatic brain tumor segmentation from Multiparametric MRI based on cascaded 3D U-Net and 3D U-Net++. *Biomed Signal Process Control*. 2022;78:103979.
45. Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*. IEEE; 2010.
46. Bertels J, Eelbode T, Berman M, et al. Optimizing the dice score and jaccard index for medical image segmentation: theory and practice. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2019:92–100.
47. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600–612.
48. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
49. Krizhevsky A. One weird trick for parallelizing convolutional neural networks. arXiv preprint arXiv:1404.5997, 2014.
50. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
51. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989:255–268.
52. Behnke NK, Baker DK, Xu S, Niemeier TE, Watson SL, Ponce BA. Risk factors for same-admission mortality after pathologic fracture secondary to metastatic cancer. *Support Care Cancer*. 2017;25:513–521.
53. Mavrogenis AF, Pala E, Romagnoli C, Romantini M, Calabro T, Ruggieri P. Survival analysis of patients with femoral metastases. *J Surg Oncol*. 2012;105(2):135–141.
54. Saad F, Lipton A, Cook R, Chen YM, Smith M, Coleman R. Pathologic fractures correlate with reduced survival in patients with malignant bone disease. *Cancer*. 2007;110(8):1860–1867.
55. Hardman PD, Robb JE, Kerr GR, Rodger A, Macfarlane A. The value of internal fixation and radiotherapy in the management of upper and lower limb bone metastases. *Clin Oncol*. 1992;4(4):244–248.
56. Arvinius C, Parra JL, Mateo LS, Maroto RG, Borrego AF, Stern LL. Benefits of early intramedullary nailing in femoral metastases. *Int Orthop*. 2014;38:129–132.
57. Ristevski B, Jenkinson RJ, Stephen DJ, et al. Mortality and complications following stabilization of femoral metastatic lesions: a population-based study of regional variation and outcome. *Can J Surg*. 2009;52(4):302–308.
58. Van Der Linden YM, Kroon HM, Dijkstra SP, et al. Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial. *Radiotherapy Oncol*. 2003;69(1):21–31.
59. Van Der Linden Y, Dijkstra P, Kroon H, et al. Comparative analysis of risk factors for pathological fracture with femoral metastases: results based on a randomised trial of radiotherapy. *J Bone & Jt Surg Br Volume*. 2004;86(4):566–573.
60. Riaz S, Bashir H, Niazi IK, Butt S, Qamar F. 99mTc MDP SPECT-CT–based modified mirels classification for evaluation of risk of fracture in skeletal metastasis: a Pilot study. *Clin Nucl Med*. 2018;43(6):180–183.
61. Ulaner GA, Zindman AM, Zheng J, Kim TW, Healey JH. FDG PET/CT assesses the risk of femoral pathological fractures in patients with metastatic breast cancer. *Clin Nucl Med*. 2017;42(4):264.
62. Sas A, Tanck E, Wafa H, Van Der Linden Y, Sermon A, Van Lenthe GH. Fracture risk assessment and evaluation of femoroplasty in metastatic proximal femurs. An in Vivo CT-based Finite Element Study. *Journal of Orthopaedic Research*. 2022;41(1): 225–234.
63. Eggermont F, Van Der Wal G, Westhoff P, et al. Patient-specific finite element computer models improve fracture risk assessments in cancer patients with femoral bone metastases compared to clinical guidelines. *Bone*. 2020;130: 115101.
64. Wang C. Fu R, Yang H. Toward a clear relationship between mechanical signals and bone adaptation. *Mechanobiology in Medicine*. 2025;3(1):100115.
65. Yang H, Butz KD, Duffy D, Niebur GL, Nauman EA, Main RP. Characterization of cancellous and cortical bone strain in the in vivo mouse tibial loading model using microCT-based finite element analysis. *Bone*. 2014;66:131–139.
66. Tanzi L, Vezzetti E, Moreno R, Moos S. X-ray bone fracture classification using deep learning: a baseline for designing a reliable approach. *Appl Sci*. 2020;10(4): 1507.
67. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep*. 2018;8(1):1–10.
68. Xue Y, Zhang R, Deng Y, Chen K, Jiang T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PloS one*. 2017;12(6):e0178992.
69. Wang J, Fang Z, Lang N, Yuan H, Su MY, Baldi P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. *Comput Biol Med*. 2017;84:137–146.
70. Chmelik J, Jakubicek R, Walek P, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Med Image Anal*. 2018;49:76–88.
71. Lang N, Zhang Y, Zhang E, et al. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magnetic resonance imaging*. 2019;64:4–12.
72. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus*. 2018;45(5):E6.
73. Fantozzi Anna, Christofori Gerhard. Mouse models of breast cancer metastasis. *Breast Cancer Res*. 2006;8:1–11.
74. Schmidt GP, Schoenberg SO, Schmid R, et al. Screening for bone metastases: whole-body MRI using a 32-channel system versus dual-modality PET-CT. *Eur Radiol*. 2007; 17:939–949.
75. Gendrin C, Furtado H, Weber C, et al. Monitoring tumor motion by real time 2D/3D registration during radiotherapy. *Radiotherapy Oncol*. 2012;102(2):274–280.
76. Weese J, Buzug TM, Lorenz C, Fassnacht C. An approach to 2d/3d registration of a vertebra in 2d x-ray fluoroscopies with 3d ct images. Springer; 1997:119–128.
77. Sarkalkan N, Weinans H, Zadpoor AA. Statistical shape and appearance models of bones. *Bone*. 2014;60:129–140.
78. Zhang J, Malcolm D, Hislop-Jambrich J, Thomas CDL, Nielsen PM. An anatomical region-based statistical shape model of the human femur. *Comput Methods Biomech Biomed Eng Imaging & Visualization*. 2014;2(3):176–185.
79. Woolf D, Padhani A, Makris A. Assessing response to treatment of bone metastases from breast cancer: what should be the standard of care? *Ann Oncol*. 2015;26(6): 1048–1057.
80. Antipov G, Baccouche M, Dugelay J-L. Face aging with conditional generative adversarial networks. 2017 IEEE International Conference on Image Processing (ICIP). IEEE; 2017:2089–2093.