RESEARCH

Clinical performance of deep learning-enhanced ultrafast whole-body scintigraphy in patients with suspected malignancy

Na Qi[1†], Boyang Pan[2†], Qingyuan Meng[1†], Yihong Yang[1], Jie Ding[1], Zengbei Yuan[1], Nan-Jie Gong[3*] and Jun Zhao[1*]

Abstract

Background To evaluate the clinical performance of two deep learning methods, one utilizing real clinical pairs and the other utilizing simulated datasets, in enhancing image quality for two-dimensional (2D) fast whole-body scintigraphy (WBS).

Methods A total of 83 patients with suspected bone metastasis were retrospectively enrolled. All patients underwent single-photon emission computed tomography (SPECT) WBS at speeds of 20 cm/min (1x), 40 cm/min (2x), and 60 cm/min (3x). Two deep learning models were developed to generate high-quality images from real and simulated fast scans, designated 2x-real and 3x-real (images from real fast data) and 2x-simu and 3x-simu (images from simulated fast data), respectively. A 5-point Likert scale was used to evaluate the image quality of each acquisition. Accuracy, sensitivity, specificity, and the area under the curve (AUC) were used to evaluate diagnostic efficacy. Learned perceptual image patch similarity (LPIPS) and the Fréchet inception distance (FID) were used to assess image quality. Additionally, the count-level consistency of WBS was compared between the two models.

Results Subjective assessments revealed that the 1x images had the highest general image quality (Likert score: 4.40 ± 0.45). The 2x-real, 2x-simu and 3x-real, 3x-simu images demonstrated significantly better quality than the 2x and 3x images (Likert scores: 3.46 ± 0.47, 3.79 ± 0.55 vs. 2.92 ± 0.41, $P < 0.0001$; 2.69 ± 0.40, 2.61 ± 0.41 vs. 1.36 ± 0.51, $P < 0.0001$), respectively. Notably, the quality of the 2x-real images was inferior to that of the 2x-simu images (Likert scores: 3.46 ± 0.47 vs. 3.79 ± 0.55, $P = 0.001$). The diagnostic efficacy for the 2x-real and 2x-simu images was indistinguishable from that of the 1x images (accuracy: 81.2%, 80.7% vs. 84.3%; sensitivity: 77.27%, 77.27% vs. 87.18%; specificity: 87.18%, 84.63% vs. 87.18%. All $P > 0.05$), whereas the diagnostic efficacy for the 3x-real and 3x-simu was better than that for the 3x images (accuracy: 65.1%, 66.35% vs. 59.0%; sensitivity: 63.64%, 63.64% vs. 64.71%; specificity: 66.67%, 69.23% vs. 55.1%. All $P < 0.05$). Objectively, both the real and simulated models achieved significantly enhanced image quality from the accelerated scans in the 2x and 3x groups (FID: 0.15 ± 0.18, 0.18 ± 0.18

[†]Na Qi, Boyang Pan and Qingyuan Meng contributed equally to this work.

*Correspondence:
Nan-Jie Gong
nanjie.gong@gmail.com
Jun Zhao
petcenter@126.com

Full list of author information is available at the end of the article

vs. 0.47 ± 0.34; 0.19 ± 0.23, 0.20 ± 0.22 vs. 0.98 ± 0.59. LPIPS: 0.17 ± 0.05, 0.16 ± 0.04 vs. 0.19 ± 0.05; 0.18 ± 0.05, 0.19 ± 0.05 vs. 0.23 ± 0.04. All $P < 0.05$). The count-level consistency with the 1x images was excellent for all four sets of model-generated images ($P < 0.0001$).

Conclusions Ultrafast 2x speed (real and simulated) images achieved comparable diagnostic value to that of standardly acquired images, but the simulation algorithm does not necessarily reflect real data.

Keywords SPECT, Whole-body scintigraphy, Deep learning, Diagnostic efficacy

Introduction

Whole-body scintigraphy (WBS) is an invaluable tool for examining pathological skeletal conditions, such as those arising from infectious, traumatic, neoplastic, or other aetiologies, that employs phosphates labelled with technetium (Tc) that are rapidly absorbed by and cleared within the skeleton [1, 2]. WBS is highly sensitive to substantial changes in bone metabolism, detecting them weeks to months before evident changes appear on conventional radiographic images [3]. Given its high sensitivity as well as affordability and ability to provide an overview of the entire skeletal system in a single imaging session, WBS is the most common imaging examination in nuclear medicine [4]. In 2019, the number of WBS examinations exceeded the total number of single photon-emission computed tomography (SPECT) scans in China by more than 60% [5].

Two-dimensional (2D) WBS imaging provides a flat, planar view of anatomical structures, which can be useful for initial assessments and broad surveys [2, 3]. The recommended scanning speed for WBS is 10–15 cm/min, as this ensures that both anterior and posterior WBS scans contain more than 1.5 million counts, which is necessary for achieving diagnostic-quality image resolution [1, 2, 6]. This implies that a standard 2D scan WBS typically takes 15–20 min, posing a challenge for patients with bone pain or those unable to maintain the same position for extended periods. Therefore, there is an urgent need to accelerate the scanning speed to increase the examination throughput. Decreasing the radioactive tracer dose would be meaningful for patients requiring follow-up WBS due to suspected or confirmed bone metastasis. However, lowering the scan time and radioactive tracer dose would result in fewer acquired counts, leading to reduced image quality, increased data noise, potential artefacts in the reconstructed images, and deterioration of diagnostic performance [7].

In recent years, deep learning methods have been successfully employed in various medical imaging processing tasks, including automatic lesion detection and anatomical localization [8], structure segmentation [9], automatic diagnosis [10–12], disease classification and image enhancement [13]. By evaluating datasets containing skin injury data and brain MR images, Jafari et al. [14] validated the effectiveness of integrating residual networks and densely connected networks within deep convolutional neural networks. Ansari et al. extensively explored the benefits and challenges of deep learning techniques in ultrasound image segmentation [15]. Concurrently, Han et al. [16] proposed an enhanced medical image segmentation model derived from the classical U-Net architecture, leveraging large convolution kernels and depthwise separable convolutions to significantly streamline model parameters. In recent years, numerous studies have highlighted the significant progress and potential of computer-aided detection technology. Akhtar et al. [17] focused on the indirect impacts of computer-aided detection or diagnostic systems following liver resection (LR) surgery. Additionally, Ansari et al. revealed the potential of generative adversarial networks (GANs) in assessing cancer through the use of elastography ultrasound [18].

Most deep learning methodologies employ a supervised learning approach, wherein paired input and output data are used to improve the effectiveness of model training. Specifically, in the context of accelerating WBS scan, however, the acquisition of perfectly matched pairs of fast-scan images and their corresponding standard scan images is challenging. Existing studies can be broadly classified into two approaches: one that relies on the use of simulated datasets for model training, and another that addresses the inherent mismatch between two distinct, real-world image sets. The simulation approach encompasses techniques such as the generation of low-count images from their high-count counterparts or the creation of simulated standard scan images from rapid acquisition processes [19–21]. Regarding the second approach, our previous work confronted the challenges posed by the inherent discrepancies observed between two distinct real-world image datasets. Specifically, we introduced a residual-in-residual dense block (RRDB)-based network and data preprocessing procedures such as bladder exclusion and patch-focused registration to generate high-quality standard scan images [22].

In this prospective work, we sought to verify the feasibility of deep learning-based fast WBS from a clinical, diagnostic perspective and compare the diagnostic efficacy for images produced by two models trained on simulation datasets and real clinical pairs. In brief, our contributions can be summarized as follows:

- To our knowledge, this is the first study to apply deep learning methods to generate 2x and 3x speed WBS images from real-acquired data.
- An analysis of the reconstruction results from the real and simulated datasets revealed that the simulation algorithm cannot necessarily reflect the real data.

The rest of the paper is structured as follows: in Sect. Materials and methods, we describes the materials and methods of the study, with the sources and acquisition of the datasets discussed in Sect. Patients and Image acquisition, whereas Sect. Deep learning workflow outlines the workflow of the deep learning network. The image analysis metrics and statistical analysis are described in Sect. Image analysis and Statistical analysis. We present the results and performance of the proposed method in Sect. Results. Finally, the results are discussed in Sect. Discussion, and Sect. Conclusion summarizes our work.

Materials and methods

Patients

The study was approved by the Institutional Review Board of Shanghai East Hospital. All patients signed informed consent forms. Patients with known or suspected bone metastatic lesions who underwent clinical technetium 99 m-methylenediphosphonate (99mTc-MDP) SPECT WBS at Shanghai East Hospital from August 2022 to December 2023 were retrospectively enrolled.

Image acquisition

All patients were injected with 9–11 MBq/kg (0.24–0.29 mCi/kg) 99mTc-MDP, and WBS was performed approximately 3–4 h postinjection. Imaging was performed with a SPECT/CT device (Siemens Symbia Intevo, Erlangen, Germany) equipped with a low-energy, high-resolution parallel-hole collimator. The scan parameters included a matrix size of 256×1024 and an energy peak centred at 140 keV with a 15% window. WBS images were obtained at three scanning speeds: 20 cm/min for standard scans (1x), 40 cm/min for 2x fast scans, and 60 cm/min for 3x fast scans, with scanning durations of 10 min, 5 min, and 3.3 min, respectively. The 3x fast scan was conducted first, followed by the 2x fast scan and the standard scan. The scans were performed consecutively, and all were completed within 20 min per patient.

Deep learning workflow

Two GAN [23]-based deep learning models were developed for this study. One model was trained using real clinical acquisition pairs (the real model). Data preprocessing, including bladder exclusion, count-level consistency and patch-focused registration, were performed

according to the methods described in a previous work [22]. Another model was trained on a simulation dataset (the simulation model). The full count data from the training dataset for the real model were also used as labels. Low-count input images were simulated via the Poisson resample technique as described in a simulation study [19].

The structures of the two models were inspired by a conditional GAN [24] and an enhanced superresolution GAN (ESRGAN) [25]. The generators share the same RRDBNet-based network architecture, which consists of 9 adjacent dense forward connection modules, as shown in Fig. 1. An upsampling efficient subpixel convolutional neural network (ESPCN) and a downsampling convolutional layer were used to obtain refined feature maps.

The loss function consists of four components, including pixelwise L1 loss, binary cross-entropy (BCE) loss, perceptual loss and GAN loss.

$$L = L_p + \lambda \, L_1 + \eta \, L_{BCE} + \gamma \, L_d \tag{1}$$

where L_p represents the perceptual loss, L_1 represents the L1 loss, L_{BCE} represents the BCE loss and L_d represents the GAN loss. λ, η and γ are constant coefficients that reflect the weighting of the respective losses. In the training process, λ, η and γ were set to 0.005, 0.01 and 0.001, respectively. The combination of L_p, L_1 and L_{BCE} reflects pixel-level alignment and structure consistency. The addition of the GAN loss further improves the quality of the generated WBS images since the pretrained perceptual loss lacks adequate a priori information in scintigraphy images.

The networks were trained with the Adam optimizer for 800 epochs with a learning rate of 2e-4. The beta1 and beta2 values of the Adam optimizer were set to 0.9 and 0.999, respectively. The batch size was set to 1, and the patch size was set to 192.

Image analysis

Subjective assessment

In the assessment of the WBS images, two radiologists with 5 and 10 years of experience employed a 5-point Likert scale to subjectively evaluate specific aspects related to image quality and diagnostic reliability. The overall clarity and quality of the images were assessed to ensure that the WBS images adequately displayed certain bone structures and abnormalities. The scoring criteria were as follows: 5 points: excellent image quality with minimal noise and excellent display of anatomical details; 4 points: good image quality with low noise, clear visualization of anatomical structures, and distinct discernability of all structures having no impact on the observation and diagnosis of lesions; 3 points: moderate image quality with an average amount of noise and an average display
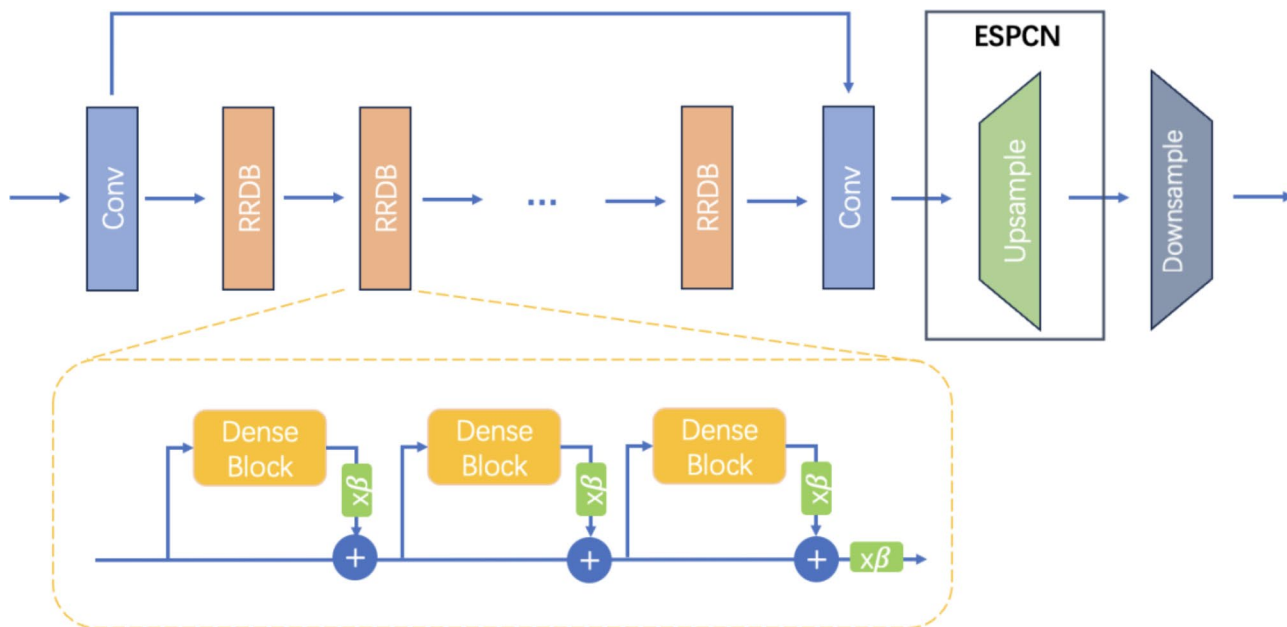
Fig. 1 Illusion of proposed network architecture. The network backbone is deployed using a stacking of 9 RRDB, followed with an up-sampling ESPCN and a down-sampling structure

of anatomical structures minorly impacting lesion observation but still allowing a clinical diagnosis; 2 points: subpar image quality with substantial noise, incomplete visualization of anatomical structures, and poor display of anatomical details notably affecting the observation of lesions; and 1 point: poor image quality with severe noise, inability to yield useful information, and unsuitable for diagnostic purposes. Benign and malignant lesions were determined on the basis of pathological diagnosis, imaging examinations (standard WBS, CT, MRI, PET/CT), and clinical follow-up data (for example, a lesion that was enlarged and had a greater concentration of tracer on imaging follow-up was considered malignant). The 1x, 2x, 3x, 2x-real, 3x-real, 2x-simu, and 3x-simu images were randomly presented to the two radiologists, who independently read the images (while blinded to the details of the diagnoses) and determined whether the lesions were benign (negative) or malignant (positive); an image quality score of 1 was considered negative. If the results were inconsistent, another senior physician was asked to make a judgement. The results of the models were evaluated with the accuracy, sensitivity, specificity, and area under the curve (AUC).

Objective assessment

For the objective assessments, we quantified image quality with learned perceptual image patch similarity (LPIPS) and the Fréchet inception distance (FID). Additionally, the count-level consistency of WBS was compared among the datasets. As described in a previous study [22], traditional metrics such as the peak

signal-to-noise ratio (PSNR) and structure similarity index measurement (SSIM) cannot fully represent consistency among WBS images. In this study, the FID and LPIPS were used to evaluate the performance of different models: the FID is defined as

$$FID = ||\mu - \mu_w||^2 + tr\left(\Sigma + \Sigma_w - 2\left(\Sigma\,\Sigma_w\right)^{\frac{1}{2}}\right) \quad (2)$$

where μ_w, Σ_w and μ, Σ are the mean and covariance of the real data and the model data distributions, respectively. A lower FID indicates smaller distances between the model and real data distributions.

LPIPS essentially computes the similarity between the activations of fast scan and standard scan patches for a predefined network at multiple feature map levels and has been shown to match human perception well. A low LPIPS score means that image patches have a smaller feature distance and are more perceptually similar. TorchMetrics was used to calculate the score with the pretrained VGG19 model. Additionally, the count-level consistency of the largest lesion on WBS was also compared among the image sets.

Statistical analysis

Statistical analysis was performed via GraphPad Prism (8.0.0) and SPSS Statistics version 26 (IBM, Armonk, NY, USA). The quantitative parameters (LPIPS and FID) and subjective scores of image quality were compared with the Wilcoxon signed-rank test. The McNemar test was used to assess differences in diagnostic performance.

Table 1 Patient clinical characteristics

N=83	Overall
Age	
Mean(SD)	61 (13)
(Min, Max)	(25, 81)
Gender	
Male	40
Female	43
BMI	
Mean (SD)	22.5 (1.2)
(Min, Max; 95%CI)	(15.9, 26.1; 21.8–24.1)
Diagnosis	
Lung cancer	22
Breast cancer	16
Prostate cancer	15
Gastric cancer	9
Colon cancer	8
Liver cancer	1
Nasopharyngeal carcinoma	5
Bone pain of unknown origin	7

Bland–Altman and Pearson correlation analyses were used to determine the consistency in the counts among the 2x-real, 3x-real, 2x-simu, 3x-simu, and 1x images. Statistical significance was considered at a p value of less than 0.05.

Results

In total, 83 patients (female/male: 43/40; age range: 25–94 years; body mass index (BMI) range: 15.9–26.1) were retrospectively enrolled in this study. Thirty-nine patients were ultimately diagnosed with bone metastasis. The demographic and clinical data of the patients are detailed in Table 1.

As illustrated in Fig. 2, the 1x images had the best general image quality (Likert score: 4.40±0.45), whereas the 3x images had poor quality (Likert score: 1.36±0.51). The 2x-real and 2x-simu and the 3x-real and 3x-simu images had significantly better subjective quality scores than the 2x and 3x images (Likert scores: 3.46±0.47, 3.79±0.55 vs. 2.92±0.41; 2.69±0.40, 2.61±0.41 vs. 1.36±0.51, respectively. $P<0.0001$), respectively. The 2x-real image quality was lower than the 2x-simu image quality (Likert score: 3.46±0.47 vs. 3.79±0.55). $P=0.001$), while there was no significant difference in quality between the 3x-real and 3x-simu images (Likert score: 2.69±0.40 vs. 2.61±0.41. $P=0.95$).

Figures 3 and 4 display the anterior WBS images of representative patients with bone metastasis obtained at different acquisition speeds (1x, 2x, and 3x) as well as the deep learning-enhanced images (2x-real, 3x-real, 2x-simu, and 3x-simu). The 1x images (A) had the best image quality and minimal noise. The 2x-real (D),
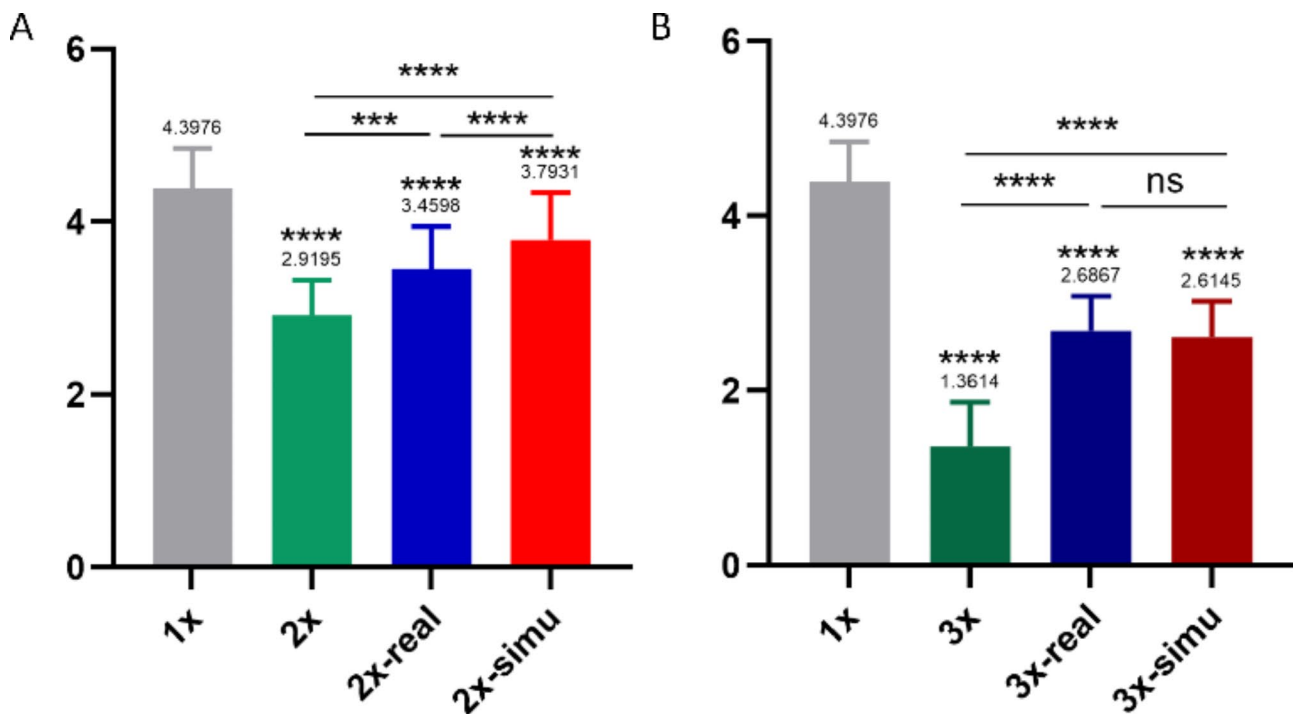

Fig. 2 Comparison of the subjective scores of the 2x groups (**A**) and 3x groups (**B**). Scores are based on a 5-point Likert scale; the 3x images had poor quality, and the 1x images had excellent quality. The images in the DL groups (2x-real and 2x-simu; 3x-real and 3x-simu) demonstrated significantly higher subjective scores than the real and simulated 2x and 3x data, respectively. Notably, the 2x-real image quality was lower than the 2x-simu image quality, and there was no significant difference in quality between the 3x-real and 3x-simu images. **** indicates $P<0.0001$, *** indicates $P<0.001$; the numbers on the bar graphs represent mean values; **** on these mean values represent the P value with respect to the 1x images
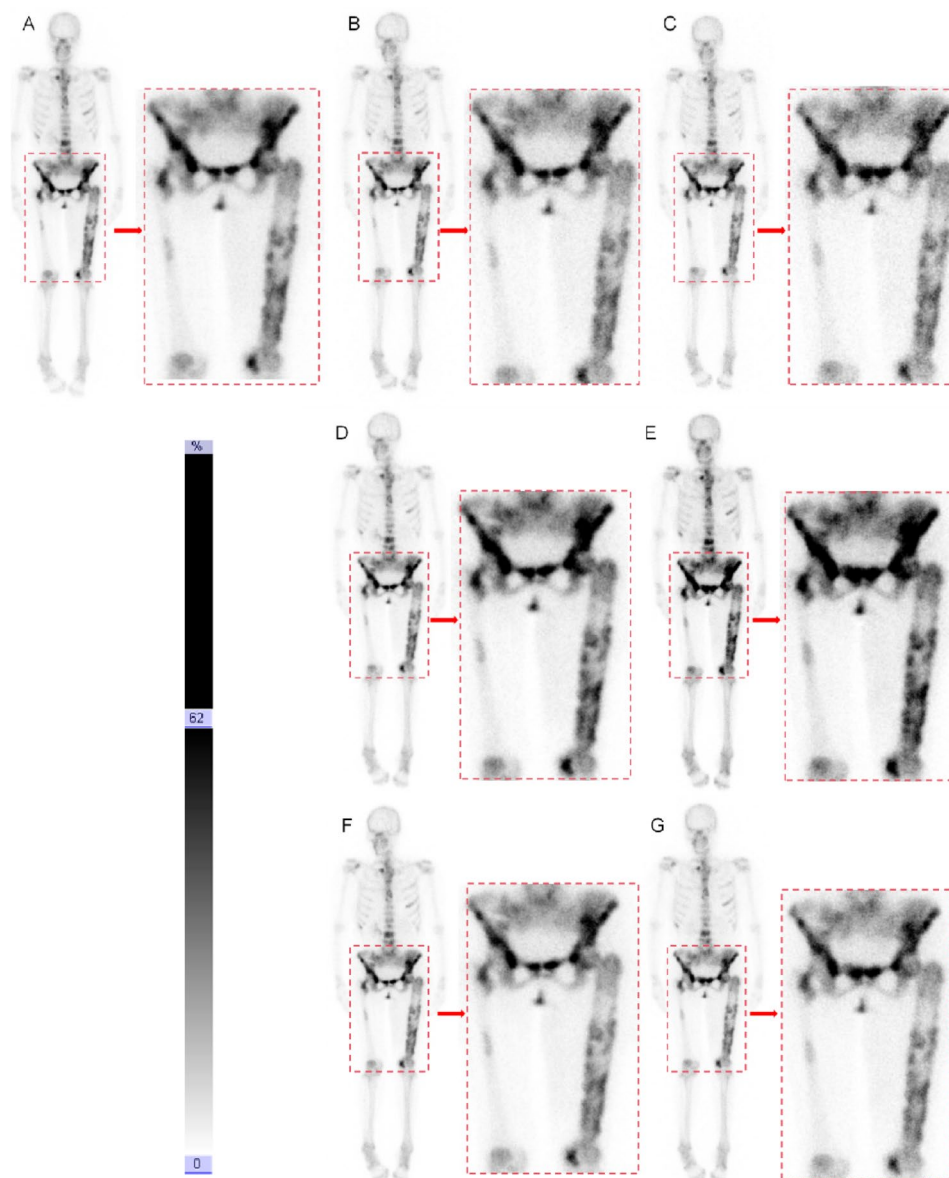
Fig. 3 A female patient with lung cancer presenting with widespread skeletal metastases evident on WBS. **A-G** show example 1x, 2x, 3x, 2x-real, 3x-real, 2x-simu, and 3x-simu images; the red dashed boxes show magnified views of the pelvis and femurs

2x-simu (F), 3x-real (E), and 3x-simu (G) images generated with deep learning exhibited better quality than the original 2x (B) and 3x (C) images.

Table 2 shows the accuracy, sensitivity, and specificity for the 1x, 2x, 2x-real, 2x-simu, 3x, 3x-real, and 3x-simu images. Table 2 shows that the greatest accuracy, sensitivity, and specificity was achieved with the 1x images, followed by the 2x-real and 2x-simu images, for which these values were greater than those of the 2x images. The diagnostic efficacy metrics for the 3x-real and 3x-simu were lower than those for the 2x images but higher than those for the 3x images. Table 3 shows that there was no significant difference in the accuracy, sensitivity, or specificity between the 2x, 2x-real, and 2x-simu images and the 1x

images ($P>0.05$), and there was no significant difference in the three diagnostic parameters between the 2x-real and 2x-simu images ($P>0.05$). The accuracy, specificity, and sensitivity for the 3x images were significantly lower than those of the 1x images ($P<0.05$), whereas the accuracy and specificity of the 3x-real images were significantly lower than those of the 1x images ($P<0.05$). The accuracy for the 3x-simu images was significantly lower than that for the 1x images ($P<0.05$). There was no significant difference in the diagnostic efficacy metrics between the 3x-real and 3x-simu images ($P>0.05$). Figure 5 shows the receiver operating characteristic (ROC) curves of the 7 groups. The data show that the AUC was greatest for the 1x images, followed by the 2x-real,
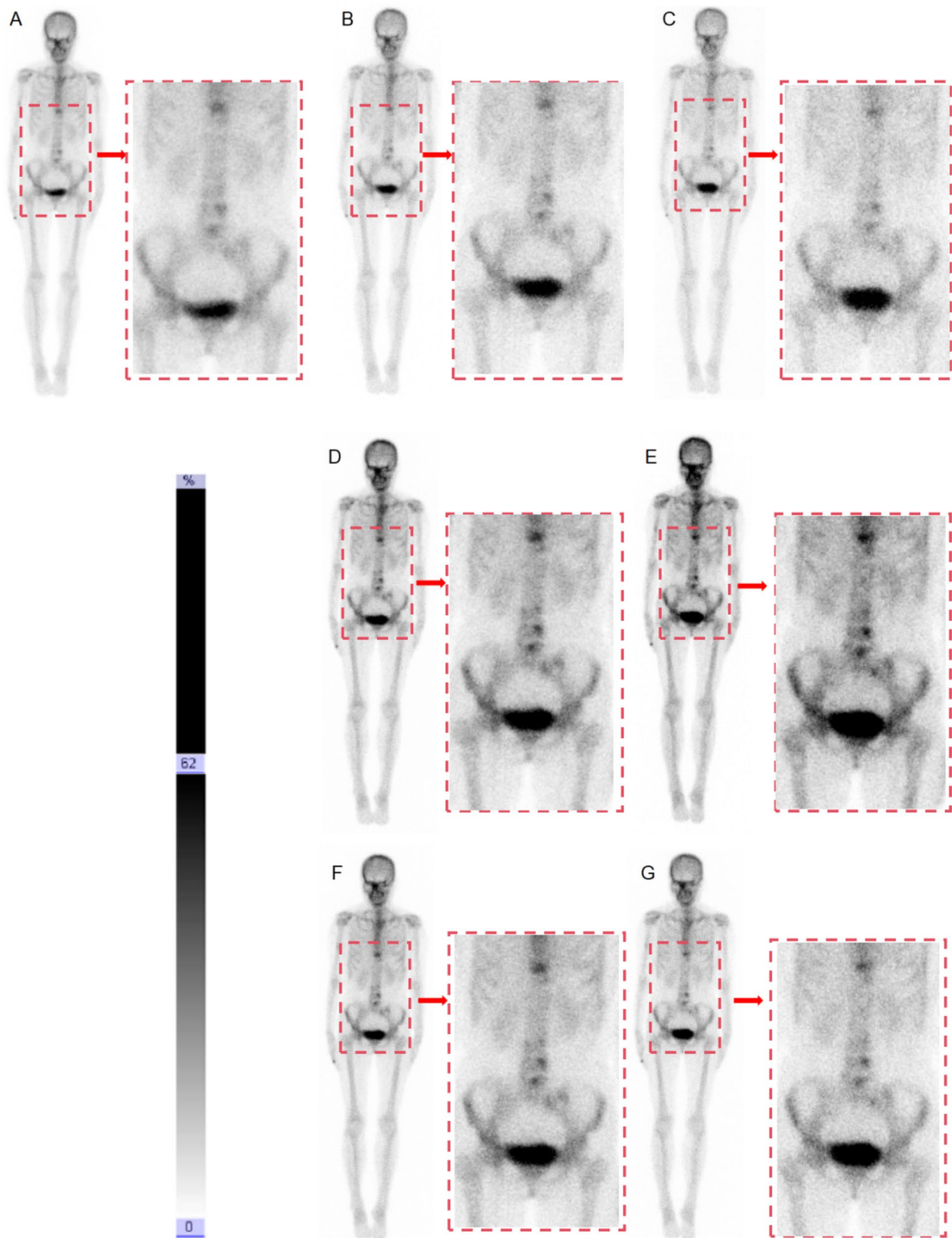
Fig. 4 **A** female patient with lung cancer presenting with thoracolumbar metastasis. **A**-**G** show example 1x, 2x, 3x, 2x-real, 3x-real, 2x-simu, and 3x-simu images; the red dashed boxes show magnified views of the lumbar spine and pelvis. Differentiation between lumbar degenerative changes and bone metastases on the **C** image is challenging; the images shown in **E** and **F**, generated with the deep learning model, have sufficient quality for diagnosing the lumbar metastases

Qi *et al. BMC Medical Imaging*   (2024) 24:236

Page 8 of 13

Table 2 Clinical diagnostic efficacy for the 1x, 2x, 3x, 2x-real, 2x-simu, 3x-real, and 3x-simu images

Group	Accuracy	Sensitivity	Specificity
1x	84.3% (70/83)	87.18% (34/39)	81.82% (36/44)
2x	73.4% (61/83)	78.38% (29/37)	69.57% (32/46)
2x-real	81.2% (68/83)	89.47% (34/38)	75.56% (34/45)
2x-simu	80.7% (67/83)	85.00% (34/40)	76.74% (33/43)
3x	59.0% (49/83)	64.71% (22/34)	55.10% (27/49)
3x-real	65.1% (54/83)	70.27% (26/37)	60.87% (28/46)
3x-simu	66.3% (55/83)	69.23% (27/39)	63.64% (28/44)

Table 3 Statistical analysis of the accuracy, sensitivity, and specificity for the 1x, 2x, 3x, 2x-real, 2x-simu, 3x-real, and 3x-simu images ($N=83$)

Group	Accuracy	Sensitivity	Specificity
2x vs. 1x	$P=0.087$	$P=0.308$	$P=0.176$
2x-real vs. 1x	$P=0.678$	$P=0.754$	$P=0.471$
2x-simu vs. 1x	$P=0.361$	$P=0.780$	$P=0.559$
2x-real vs. 2x-simu	$P=0.842$	$P=0.555$	$P=0.896$
3x vs. 1x	$P=0.004$	$P=0.023$	$P=0.006$
3x-real vs. 1x	$P=0.004$	$P=0.071$	$P=0.029$
3x-simu vs. 1x	$P=0.007$	$P=0.055$	$P=0.056$
3x-real vs. 3x-simu	$P=0.870$	$P=0.921$	$P=0.787$

Table 4 Objective quantitative evaluation

	FID		LPIPS	
	2x	3x	2x	3x
Fast Scan	0.465±0.337	0.975±0.585	0.190±0.046	0.231±0.040
Real	0.152±0.183*	0.193±0.227*	0.166±0.046*	0.176±0.046*
Simu	0.177±0.184*	0.200±0.224*	0.163±0.043*	0.186±0.045*†

* significant difference compared to the fast scan

† significant difference compared to the Real

corresponding fast scans ($P<0.0001$). No difference was found between the 2x-real and 2x-simu images and between the 3x-real and x3-simu images in terms of the FID ($P=0.4599$). However, the 3x-real images achieved better LPIPS than the 3x-simu images did ($P=0.0016$).

Bland−Altman analysis revealed a maximum count value difference of −0.0055. The 95% limits of agreement for the differences in the 1x versus 2x-real images, containing 37/39 of the differences, were (-19.26, 34.70) (Fig. 6A); those for the 1x versus 2x-simu images, containing 38/39 of the differences, were (-26.22, 24.68) (Fig. 6B); those of the 1x versus 3x-real images, containing 38/39 of the differences, were (-36.92, 52.92) (Fig. 6C); and those of the 1x versus 3x-simu images, containing 36/39 of the differences, were (-40.25, 31.94) (Fig. 6D). Pearson correlation analysis revealed a notable association between the 2x-real and 1x image counts, with a correlation coefficient of $r=0.999$ (Fig. 7A); a significant correlation between the counts of the 2x-simu and 1x images, with $r=1.017$ (Fig. 7B), a significant correlation between
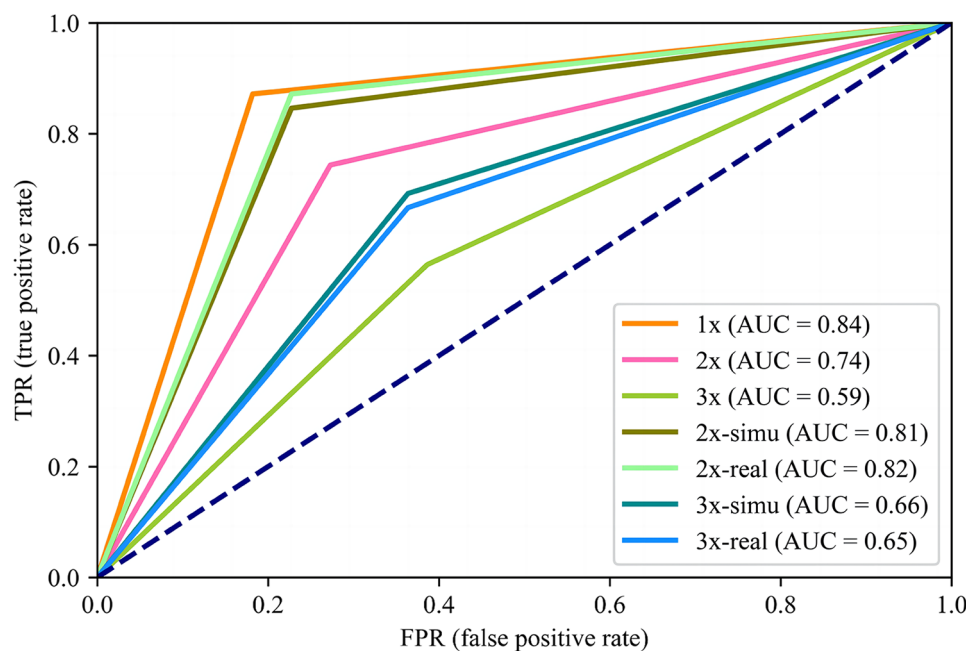
2x-simu, 2x, 3x-simu, 3x-real, and 3x images; furthermore, the AUCs for the 1x, 2x-real, and 2x-simu images were all above 0.80.

An objective assessment of model performance is shown in Table 4. Both the real model and the simu model resulted in significant improvements over the

Fig. 5 Diagnostic performance for the 1x, 2x, 3x, 2x-real, 2x-simu, 3x-real, and 3x-simu images as assessed by ROC curve analysis. ROC: receiver operating characteristic, AUC: area under the curve, TPR: true positive rate, FPR: false positive rate
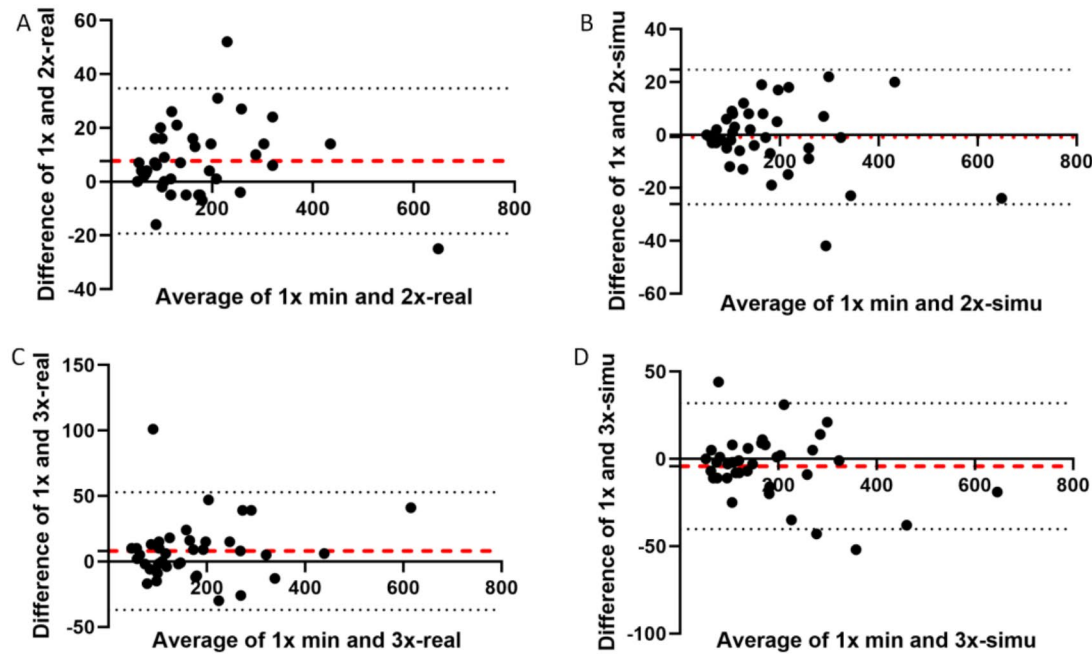
Fig. 6 Bland–Altman plot for the maximum count values of the 1x images versus the 2x-real (**A**), 2x-simu (**B**), 3x-real (**C**), and 3x-simu (**D**) images ($n=39$), with the limits of agreement depicted as dotted lines plotted at -1.96 SD and $+1.96$ SD

the 3x-real and 1x image counts, with $r=0.953$ (Fig. 7C), and a significant correlation between the 3x-simu and 1x image counts, with $r=1.042$ (Fig. 7D) ($P<0.0001$).

Discussion

In the image preprocessing procedure described in the methods section, we introduced additional steps to improve the robustness of the model to noise present in the input images. Previous studies have shown that preprocessing steps not only help reduce the impact of noise in the input data but also improve model generalizability and robustness, ensuring consistent and reliable performance when processing input images in real-world scenarios [26, 27].

The results of the present study validate the efficacy of deep learning methods in enhancing the quality of 2x-speed images, as demonstrated by both subjective and objective assessments, producing images that meet clinical diagnostic standards. However, the Likert scores of the 2x-real and 2x-simu images were significantly different from those of standard 1x images, which can reflect subjectivity and reliance on physician experience for interpretation. Nevertheless, objective evaluations using image quality metrics such as LPIPS and FID showed no significant differences in quality from 1x images. The FID quantifies the disparities between synthetic and authentic data distributions, whereas the LPIPS gauges the feature distance and perceptual similarity of image patches,

potentially indicating consistency in the whole-body bone images compared here. Moreover, the observed differences in 2x-real and 2x-simu images did not impact diagnostic efficacy, as metrics including diagnostic accuracy (81.2%, 80.7%), sensitivity (89.47%, 85.00%), specificity (75.56%, 76.74%), and AUC values (0.82, 0.81) were comparable to those of the 1x images. The results are also similar to those reported in a previous meta-analysis [3], which reported a sensitivity of 86.5% and specificity of 79.9% for diagnosing bone metastases with WBS, confirming the ability to discern malignant bone lesions effectively with the model-enhanced images.

Notably, in this study, the diagnostic performance for the 2x images, obtained with a doubled scanning speed, was slightly lower than that for the standard 1x images. However, this difference was not statistically significant, possibly because of the relatively small sample size and potential subjective bias in the physicians' interpretations, given the reliance of bone scintigraphy interpretation on individual knowledge and experience. Nonetheless, the AUC value was less than 0.8 for these faster-acquired images, indicating a diminished ability to differentiate malignant bone lesions with respect to both the 1x images and the 2x-real and 2x-simu images [28]. The image quality of the 3x-real and 3x-simu images enhanced by the deep learning methods was notably superior to that of the 3x images according to both subjective and objective assessments. However, the accuracy,
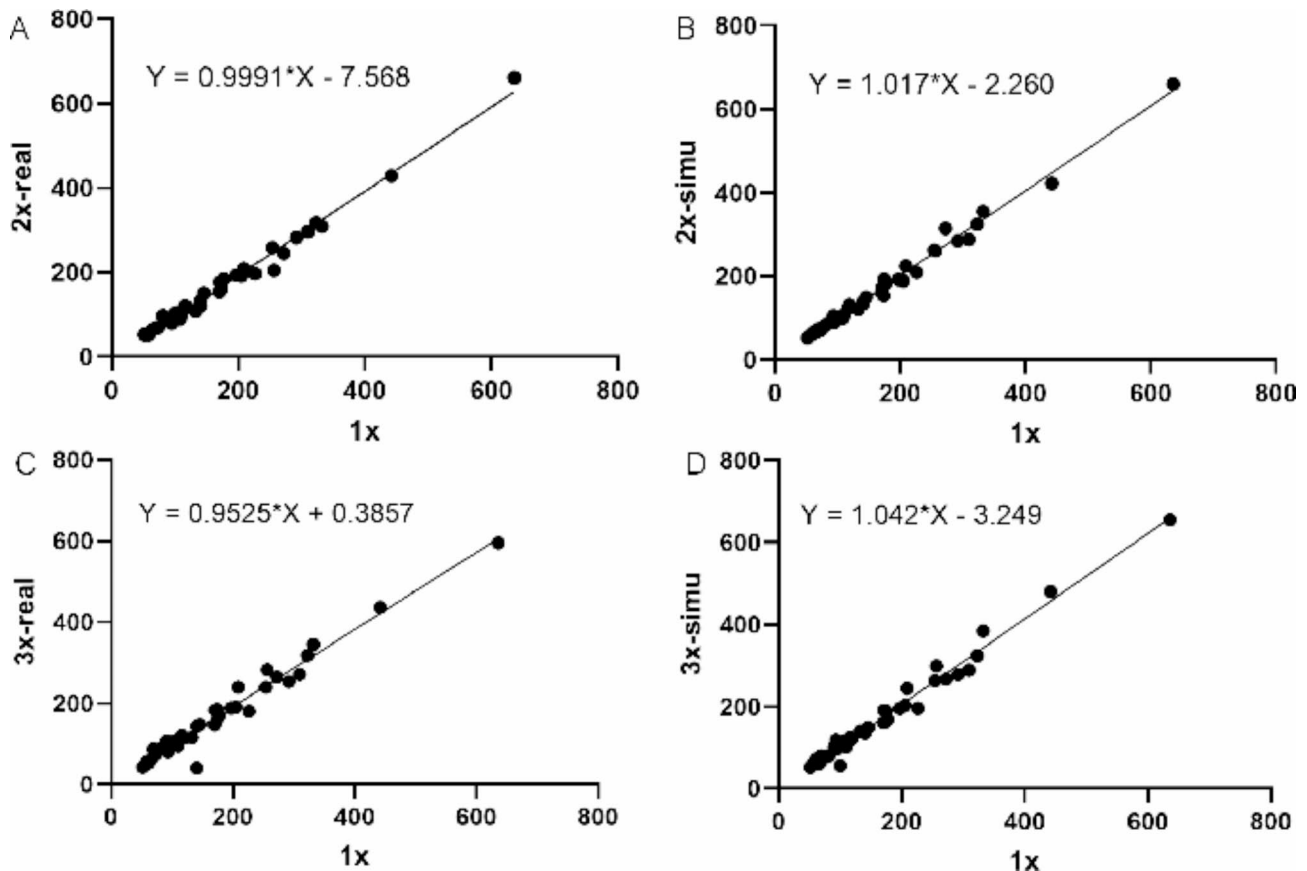
Fig. 7 Pearson correlation analysis revealed significant correlations between the counts of 2x-real and 1x, 2x-simu and 1x, 3x-real and 1x, and 3x-simu and 1x images, with correlation coefficients of $r = 0.999$, 1.017, 0.953, and 1.042, respectively ($P < 0.0001$)

sensitivity, and specificity were significantly lower than those of the 1x images, with all the AUC values being less than 0.7, making them unsuitable for clinical diagnostic purposes.

A previous study developed a deep-learning noise reduction (DLNR) algorithm for whole-body cadmium and zinc telluride (CZT) SPECT images using data from 19 patients and demonstrated that image quality remained at a good-to-excellent level even when the acquisition time was reduced to 60% [29]. In the present study, a dual-head gamma camera, which is more commonly used in clinical practice [30], was used in the development of the deep learning models for generating WBS images. This study, which boasts a relatively large sample size, showed that ultrafast imaging at 50% of the acquisition time (using both real and simulated datasets) can achieve comparable diagnostic efficacy to standard acquisition protocol, as determined by subjective and objective assessments. Similarly, Minarik et al. conducted a visual evaluation to determine the presence of bone metastases in deep learning-enhanced images with 50% counts and reported no significant difference in diagnostic performance with respect to the reference images [31]. A previous study indicated that despite the use of

enhanced planar processing, the noise-reducing Pixon algorithm—based on the principle that the ideal image is represented by the lowest possible number of parameters that correctly represent the raw data image—could not fully compensate for the loss of counts associated with reducing the scan time by half for WBS images [32]. In this study, two supervised deep learning models were trained on resampled simulation datasets or acquired real clinical pairs. Our results suggest that advancements in deep learning technology allow the corresponding models to continue improving the quality of low-dose scan images.

Our previous study demonstrated that the quality of images acquired with ultrafast SPECT/CT with a 1/7 acquisition time could be improved with deep learning methods to achieve diagnostic value comparable to that of standard acquisition protocols [33, 34]. SPECT/CT and WBS differ in a number of technical aspects, however. SPECT/CT is a 3D quantitative multimodal imaging technique that provides functional information on SPECT and high-resolution structural information on CT, whereas whole-body bone scintigraphy scans are 2D images with high noise levels. This leads to the use of different image preprocessing techniques and

normalization functions and influences the design of the network structure. Furthermore, with the xSPECT reconstruction technique, the smooth structure in the generated SPECT images is considered acceptable by radiologists, but the smooth structure in the whole-body bone scan is considered unreal and unsuitable for diagnosis; this was the impetus for employing the GAN loss. While WBS and bone SPECT imaging both rely on single-photon counting principles, adjacent layers and the corresponding images from tomography provide more a priori image information; thus, images obtained with a 1/7 standard acquisition time can be "accelerated" via deep learning techniques to meet diagnostic requirements. For example, gathering additional stratified standard scan times at intervals of 48%, 46%, and 44% could help refine the minimum scan time for accelerated imaging, ensuring that both suitable image quality is achieved and that the images meet clinical requirements. However, in clinical practice, acquiring more low-count images is a labour-intensive process. A study in which low-count original images (75%, 50%, 25%, 10%, and 5% counts) were generated from reference images (100% counts) via Poisson resampling indicated that the use of deep learning methods improved image quality and bone metastasis detection accuracy for low-count bone scintigraphy [19]. In our study, the subjective scores of the 2x-simu (low-count original images from reference images generated via Poisson resampling) and 2x-real images were still different, but their objective LPIPS and FID values did not differ significantly, whereas the objective LPIPS values between the 3x-simu (low-count original images from reference images via Poisson resampling) and 3x-real images were significantly different. It is worth considering whether the objective image quality differences between simulated and real images increase with a reduction in the simulated count, emphasizing the need for further investigations with larger sample sizes to validate these observations. As shown in Fig. 3, the quality of the 2x-real (D), 2x-simu (F), 3x-real (E), and 3x-simu (G) images was greater than that of the 2x (B) and 3x (C) images, with reduced noise and increased radiopharmaceutical counts. Furthermore, there appeared to be no significant difference in image quality between the 2x-real (D) and 2x-simu (F) images or between the 3x-real (E) and 3x-simu (G) images. However, by zooming in on the images (the enlarged images within the red dashed boxes), the range of bone metastatic lesions in the 2x-simu (F) and 3x-simu (G) images appears to be closer that seen on the standard 1x (A) images, whereas in the 2x-real and (D) 3x-real (E) images, the lesion range is slightly larger, and the boundaries between the lesions and normal bone are less distinct. These observations suggest that the simulation images may not accurately simulate real images. While the two models share the same network architecture and training parameters, the differences in the generated images are due to discrepancies between the simulated low-count data and the real low-count data. Possible explanations are that the drug are continuously metabolized during the examination and that the noise is not fully consistent with the Poisson downsampling process for atomic decay.

Advancements in SPECT/CT technology have facilitated quantitative evaluations in bone imaging [35–37]. Several studies have shown strong interobserver agreement in quantitative analysis and established a significant correlation between standardized uptake values (SUVs) derived from SPECT images of bone metastases and those from PET images [38–40]. The SUV in SPECT/CT imaging is based on single-photon counting; therefore, in WBS, single-photon counts also have a certain quantitative reference value. In this study, we sought to improve image quality by increasing the radiopharmaceutical counts in images acquired with reduced scan times with deep learning. Consequently, the "accelerated" counts of 39 lesions in the 2x-real, 2x-simu, 3x-real, and 3x-simu images were highly consistent with those of the original images. This confirms the applicability of deep learning-accelerated planar bone imaging in clinical practice from an objective perspective. These findings also align with those of a previous study [19], in which the simulated lesion counts were strongly correlated with those of the original standard images, albeit decreased by 10%, possibly due to the risk of signal loss due to smoothing observed in that study. The high consistency in the maximum lesion count between the deep learningenhanced images and the reference WBS images could be the result of the count consistency preprocessor. Hence, the scale gap between low-count scans and full-count scans in different cases was normalized well. Our proposed method also contributes to achieving good alignment in the maximum lesion count.

In previous research, parallel computing technology was applied to handle computationally intensive tasks and achieved notable results. Zhai et al. [41] implemented real-time automatic segmentation technology for cerebral aneurysms on a Zynq system chip and subsequently designed and implemented an efficient hardware architecture based on the lattice Boltzmann method [42]. To achieve real-time visualization and guidance for cerebral blood flow image acquisition, Esfahani et al. [43] developed and employed an improved version of HemeLB as the main computational core of the pipeline. Zhai et al. [44] also developed and evaluated a version of HemeLB on multiple, heterogeneous system chip platforms and visualized simulation results in real time, achieving a significant improvement in algorithm performance after acceleration. In the future, we plan to explore ways to

optimize parallel computing strategies for addressing potential computationally intensive challenges.

Although this research explored the clinical performance of a deep learning method in WBS, it has several limitations. First, the prospective nature of the data collection from a single centre and the relatively small sample size may have led to biased results. Pathological verification of some bone lesions was challenging, and the gold standard for diagnosis, which is based on the clinical history and other imaging studies, may have caused bias in the results. Second, this study exclusively utilized 99mTc-MDP. Future studies should explore the use of other radiopharmaceuticals to increase the generalizability and robustness of the deep learning models. Finally, our study only acquired images at 2x and 3x scanning speeds and assessed the quality of these images after they were subjected to deep learning processing. The results indicated that the 2x-real and 2x-simu images met clinical diagnostic requirements. However, an investigation of scanning speeds between 2x and 3x is needed to identify the shortest scanning time that can meet clinical diagnostic needs following the use of deep learning methods.

Conclusion

Our study demonstrated that the image quality of deep learning-enhanced ultrafast 2x- and 3x-speed images, based on both real and simulated data, was markedly enhanced in both subjective and objective evaluations. Notably, the 2x speed images enhanced based on both real and simulated data were diagnostically equivalent to images obtained with standard acquisition protocols in terms of subjective and objective assessments. The quality of the 2x-real images was subjectively inferior to that of the 2x-simu images, and the quality of 3x-simu images was objectively inferior to that of 3x-real images, indicating that the simulation algorithm could not necessarily reflect real data. The findings of this study suggest that the application of deep learning enhancement techniques may contribute to the implementation of ultrafast and highly reliable ultralow-dose WBS scans in future routine clinical practice, offering the potential to significantly increase patient safety and diagnostic accuracy while minimizing scanning time.

Declarations

Author details
[1]Department of Nuclear Medicine, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China
[2]RadioDynamic Healthcare, Shanghai, China
[3]Tsinghua Cross-Strait Research Institute, Laboratory of Intelligent Medical Imaging, Beijing, China

References
1. Van den Wyngaert T, Strobel K, Kampen WU, et al. The EANM practice guidelines for bone scintigraphy. Eur J Nucl Med Mol Imaging. 2016;43(9):1723–38.
2. Bombardieri E, Aktolun C, Baum RP, et al. Bone scintigraphy: procedure guidelines for tumour imaging. Eur J Nucl Med Mol Imaging. 2003;30(12):BP99–106.
3. Yang HL, Liu T, Wang XM, Xu Y, Deng SM. Diagnosis of bone metastases: a meta-analysis comparing (1)(8)FDG PET, CT, MRI and bone scintigraphy. Eur Radiol. 2011;21(12):2604–17.
4. Kosuda S, Kaji T, Yokoyama H, et al. Does bone SPECT actually have lower sensitivity for detecting vertebral metastasis than MRI? J Nucl Med. 1996;37(6):975–8.
5. Wang JLS. A brief report on the results of the national survey of nuclear medicine in 2020. Chin J Nucl Med Mol Imaging. 2020;49:747–9.
6. Bartel TB, Kuruva M, Gnanasegaran G, et al. SNMMI Procedure Standard for Bone Scintigraphy 4.0. J Nucl Med Technol. 2018;46(4):398–404.
7. Schaefferkoetter J, Yan J, Moon S, et al. Deep learning for whole-body medical image generation. Eur J Nucl Med Mol Imaging. 2021;48(12):3817–26.
8. Huang K, Huang S, Chen G et al. An end-to-end multi-task system of automatic lesion detection and anatomical localization in whole-body bone scintigraphy by deep learning. Bioinformatics 2023, 39(1).
9. Motegi K, Miyaji N, Yamashita K, Koizumi M, Terauchi T. Comparison of skeletal segmentation by deep learning-based and atlas-based segmentation in prostate cancer patients. Ann Nucl Med. 2022;36(9):834–41.
10. Liu S, Feng M, Qiao T, et al. Deep learning for the Automatic diagnosis and analysis of bone metastasis on bone scintigrams. Cancer Manag Res. 2022;14:51–65.
11. Wuestemann J, Hupfeld S, Kupitz D et al. Analysis of bone scans in various tumor entities using a deep-learning-based Artificial neural network algorithm-evaluation of diagnostic performance. Cancers (Basel) 2020, 12(9).
12. Han S, Oh JS, Seo SY, Lee JJ. Performance of deep learning models for response evaluation on whole-body bone scans in prostate cancer. Ann Nucl Med. 2023;37(12):685–94.
13. Hajianfar G, Sabouri M, Salimi Y et al. Artificial intelligence-based analysis of whole-body bone scintigraphy: the quest for the optimal deep learning

algorithm and comparison with human observer performance. Z Med Phys 2023.

14. Jafari M, Auer D, Francis S, Garibaldi J, Chen X. DRU-Net: an efficient deep convolutional neural network for medical image segmentation. In2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) 2020 Apr 3 (pp. 1144–1148). IEEE.

15. Ansari MY, Mangalote IA, Meher PK, Aboumarzouk O et al. Advancements in Deep Learning for B-Mode Ultrasound Segmentation: a Comprehensive Review. IEEE Trans Emerg Top Comput Intell. 2024 Apr 2.

16. Meng HZ, Jian MW, Wang GG. ConvUNeXt: an efficient convolution neural network for medical image segmentation. Knowl Based Syst. 2022;253:109512.

17. Akhtar Y, Dakua SP, Abdalla A, et al. Risk assessment of computer-aided diagnostic software for hepatic resection. IEEE Trans Radiation Plasma Med Sci. 2021;6(6):667–77.

18. Ansari MY, Qaraqe M, Righetti R, Serpedin E, Qaraqe K. Unveiling the future of breast cancer assessment: a critical review on generative adversarial networks in elastography ultrasound. Front Oncol. 2023;13:1282536.

19. Murata T, Hashimoto T, Onoguchi M, et al. Verification of image quality improvement of low-count bone scintigraphy using deep learning. Radiol Phys Technol. 2024;17(1):269–79.

20. Ito T, Maeno T, Tsuchikame H, et al. Adapting a low-count acquisition of the bone scintigraphy using deep denoising super-resolution convolutional neural network. Phys Med. 2022;100:18–25.

21. Ichikawa S, Sugimori H, Ichijiri K, Yoshimura T, Nagaki A. Acquisition time reduction in pediatric (99m) Tc-DMSA planar imaging using deep learning. J Appl Clin Med Phys. 2023;24(6):e13978.

22. Pan Z, Qi N, Meng Q et al. Fast SPECT/CT planar bone imaging enabled by deep learning enhancement. Med Phys 2024.

23. Ian G, Jean PA, Mehdi M et al. Generative Adversarial Networks. COMMUNICATIONS OF THE ACM. 2020, 63(11):139–144.

24. Mehdi M. Nov. Conditional Generative Adversarial Nets. arXiv:1411.1784v1 [cs.LG] 6 2014.

25. Wang XT, Yu K, WuSX et al. ESRGAN: enhanced Super-resolution Generative Adversarial Networks. 4arXiv:1809.00219[cs.CV]. 17 Sep 2018.

26. Ansari MY, Yang Y, Meher PK, Dakua SP. Dense-PSP-UNet: a neural network for fast inference liver ultrasound segmentation. Comput Biol Med. 2023;153:106478.

27. Mohanty S, Dakua SP. Toward computing cross-modality symmetric non-rigid medical image registration. IEEE Access. 2022;10:24528–39.

28. De Hond AAH, Steyerberg EW, Van Calster B. Interpreting area under the receiver operating characteristic curve. Lancet Digit Health. 2022, 4(12):e853-e855.

29. Bahloul A, Verger A, Lamash Y, et al. Ultra-fast whole-body bone tomoscintigraphies achieved with a high-sensitivity 360 degrees CZT camera and a dedicated deep-learning noise reduction algorithm. Eur J Nucl Med Mol Imaging. 2024;51(5):1215–20.

30. Ritt P. Recent developments in SPECT/CT. Semin Nucl Med. 2022;52(3):276–85.

31. Minarik D, Enqvist O, Tragardh E. Denoising of scintillation camera images using a deep convolutional neural network: a Monte Carlo Simulation Approach. J Nucl Med. 2020;61(2):298–303.

32. Ardenfors O, Svanholm U, Jacobsson H, et al. Reduced acquisition times in whole body bone scintigraphy using a noise-reducing Pixon(R)-algorithm-a qualitative evaluation study. EJNMMI Res. 2015;5(1):48.

33. Qi N, Pan B, Meng Q et al. Deep learning enhanced ultra-fast SPECT/CT bone scan in patients with suspected malignancy: quantitative assessment and clinical performance. Phys Med Biol 2023, 68(13).

34. Pan B, Qi N, Meng Q, Wang J, et al. Ultra high speed SPECT bone imaging enabled by a deep learning enhancement method: a proof of concept. EJNMMI Phys. 2022;9(1):43.

35. Dickson JC, Armstrong IS, Gabina PM, et al. EANM practice guideline for quantitative SPECT-CT. Eur J Nucl Med Mol Imaging. 2023;50(4):980–95.

36. Halim F, Yahya H, Jaafar KN, Mansor S. Accuracy Assessment of SUV measurements in SPECT/CT: a Phantom Study. J Nucl Med Technol. 2021;49(3):250–5.

37. 37 Kaneta T, Ogawa M, Daisaki H, et al. SUV measurement of normal vertebrae using SPECT/CT with Tc-99m methylene diphosphonate. Am J Nucl Med Mol Imaging. 2016;6(5):262–8.

38. Beck M, Sanders JC, Ritt P, et al. Longitudinal analysis of bone metabolism using SPECT/CT and (99m)Tc-diphosphono-propanedicarboxylic acid: comparison of visual and quantitative analysis. EJNMMI Res. 2016;6(1):60.

39. Arvola S, Jambor I, Kuisma A, et al. Comparison of standardized uptake values between (99m)Tc-HDP SPECT/CT and (18)F-NaF PET/CT in bone metastases of breast and prostate cancer. EJNMMI Res. 2019;9(1):6.

40. Qi N, Meng Q, You Z, et al. Standardized uptake values of (99m)Tc-MDP in normal vertebrae assessed using quantitative SPECT/CT for differentiation diagnosis of benign and malignant bone lesions. BMC Med Imaging. 2021;21(1):39.

41. ZhaiXJ, Eslamib M, Hussein ES, et al. Real-time automated image segmentation technique for cerebral aneurysm on reconfigurable system-on-chip. J Comput Sci. 2018;27:35–45.

42. Zhai XJ, Amira A, Bensaali F et al. Zynq SoC based acceleration of the lattice boltzmann method. Concurrency Computation: Pract Experience 31.17 (2019): e5184.

43. Esfahani SS, Zhai X, Chen M, et al. Lattice-boltzmann interactive blood flow simulation pipeline. Int J Comput Assist Radiol Surg. 2020;15(4):629–39.

44. Zhai SJ, Chen M, Esfahani ss et al. Heterogeneous system-on-chip-based Lattice-Boltzmann visual simulation system. IEEE Syst Journal14.2 (2019): 1592–601.

Publisher's note