Computer Methods and Programs in Biomedicine 265 (2025) 108724

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

# Predicting strength of femora with metastatic lesions from single 2D radiographic projections using convolutional neural networks

Alexander Synek [a,*], Emir Benca [b], Roxane Licandro [c], Lena Hirtler [d], Dieter H. Pahr [a]

[a] Institute of Lightweight Design and Structural Biomechanics, TU Wien, Gumpendorfer Straße 7, 1060 Vienna, Austria
[b] Department of Orthopedics and Trauma-Surgery, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
[c] Department of Biomedical Imaging and Image-guided Therapy, Computational Imaging Research Lab (CIR), Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
[d] Center for Anatomy and Cell Biology, Medical University of Vienna, Währinger Straße 13, 1090 Vienna, Austria

ARTICLE INFO

*Keywords:*
Metastatic bone disease
Bone lesions
Femur strength
Machine learning
Finite element
Radiograph

ABSTRACT

*Background and objective:* Patients with metastatic bone disease are at risk of pathological femoral fractures and may require prophylactic surgical fixation. Current clinical decision support tools often overestimate fracture risk, leading to overtreatment. While novel scores integrating femoral strength assessment via finite element (FE) models show promise, they require 3D imaging, extensive computation, and are difficult to automate. Predicting femoral strength directly from single 2D radiographic projections using convolutional neural networks (CNNs) could address these limitations, but this approach has not yet been explored for femora with metastatic lesions. This study aimed to test whether CNNs can accurately predict strength of femora with metastatic lesions from single 2D radiographic projections.

*Methods:* CNNs with various architectures were developed and trained using an FE model generated training dataset. This training dataset was based on 36,000 modified computed tomography (CT) scans, created by randomly inserting artificial lytic lesions into the CT scans of 36 intact anatomical femoral specimens. From each modified CT scan, an anterior-posterior 2D projection was generated and femoral strength in one-legged stance was determined using nonlinear FE models. Following training, the CNN performance was evaluated on an independent experimental test dataset consisting of 31 anatomical femoral specimens (16 intact, 15 with artificial lytic lesions). 2D projections of each specimen were created from corresponding CT scans and femoral strength was assessed in mechanical tests. The CNNs' performance was evaluated using linear regression analysis and compared to 2D densitometric predictors (bone mineral density and content) and CT-based 3D FE models.

*Results:* All CNNs accurately predicted the experimentally measured strength in femora with and without metastatic lesions of the test dataset ($R^2 \geq 0.80$, $CCC \geq 0.81$). In femora with metastatic lesions, the performance of the CNNs (best: $R^2 = 0.84$, $CCC = 0.86$) was considerably superior to 2D densitometric predictors ($R^2 \leq 0.07$) and slightly inferior to 3D FE models ($R^2 = 0.90$, $CCC = 0.94$).

*Conclusions:* CNNs, trained on a large dataset generated via FE models, predicted experimentally measured strength of femora with artificial metastatic lesions with accuracy comparable to 3D FE models. By eliminating the need for 3D imaging and reducing computational demands, this novel approach demonstrates potential for application in a clinical setting.

## 1. Introduction

Metastatic bone disease (MBD) is a serious and costly complication in cancer patients [1]. Metastatic lesions can weaken the structural integrity of the affected bone, resulting in its pathological fracture. The femur is among the most affected sites [2] and pathological fractures often occur spontaneously during activities of daily living [3]. Prophylactic surgical fixation is performed in potentially unstable femora [4]. Recognising impending fractures remains challenging and is currently based on clinical experience or clinical scoring systems, such as Mirels' score [5] or axial cortical involvement [6]. However, these methods have limited specificity, which often leads to patient overtreatment [7].

* Corresponding author at: Alexander Synek, Institute of Lightweight Design and Structural Biomechanics, TU Wien, Gumpendorfer Straße 7, 1060 Vienna, Austria.
*E-mail address:* asynek@ilsb.tuwien.ac.at (A. Synek).

https://doi.org/10.1016/j.cmpb.2025.108724
Received 17 January 2025; Received in revised form 2 March 2025; Accepted 14 March 2025
Available online 14 March 2025

Furthermore, once diagnosed with MBD, patients are confronted with the constant fear of fractures, which may reduce their activity levels. This reduction in physical activity can further exacerbate bone resorption and muscle atrophy, negatively impacting their quality of life [8].

The insufficiencies of currently available scoring systems motivated the development of novel approaches based on mechanical models [9], such as patient-specific rigidity analysis [10] or finite element (FE) analysis [11]. While rigidity analysis provides a comparative measure (e.g. with respect to the contralateral healthy bone), FE analysis even allows to predict femoral strength quantitatively. FE-predicted femoral strength was validated in numerous ex vivo studies in femora with [12] and without [13] metastatic lesions. It was recently implemented in a clinical score which showed improved fracture risk assessment compared to conventional methods [14] and was found to be a useful support both for the clinician and patient [15]. In addition to fracture risk assessment, FE models could also help to provide guidelines for permissible sports, activities of daily living, and rehabilitation programs to the patient [16]. Despite its promising results, FE analysis is still challenging to implement in clinical practice, as it relies on 3D imaging data (e.g. computed tomography (CT) scans), requires substantial computational time and is challenging to automate [12,17,18].

One way to facilitate clinical implementation of metastatic femoral strength assessment would be the use of more readily available 2D radiographic projections, such as plain radiographs or dual energy x-ray absorptiometry (DXA) images. Predicting strength from 2D projections has been widely explored for intact femora, with good results using areal bone mineral density (aBMD) or bone mineral content (BMC) from DXA images [19] or using 3D FE models reconstructed from DXA images [20]. Interestingly, only a few studies investigated strength prediction of femora with metastatic lesions from 2D projections. Using BMC led to varying results in femora with simulated lytic lesions, ranging from no significant correlation [21] to an excellent correlation [22]. This discrepancy indicates that aBMD and BMC are limited in fully capturing the relation of a 2D projection with femoral strength in the presence of metastatic lesions.

To better exploit the information available in the 2D projections, machine learning methods could be used. Particularly convolutional neural networks (CNNs) appear suitable for this task, as they have achieved breakthrough improvements in computer vision [23] and have also gained popularity in medical image processing [24,25]. CNNs can be trained to automatically extract relevant image features and allow fast and automated predictions, which makes them attractive for clinical applications. CNNs have been extensively used on plain radiographs to perform classification [26,27] or segmentation [28] tasks, and a few studies explored their potential for quantitative predictions such as subject age [29] or aBMD [30]. Some studies also used CNNs to classify hip fracture risk based on DXA images [31,32]. However, no study is known to the authors that attempted a quantitative prediction of femoral strength from 2D projections using CNNs. The main reason could be that biomechanical studies including experimental testing are usually limited to $<$100 specimens [12,13], which constitutes a challenge to train CNNs. FE models could be used to extend these experimental datasets, but as the CNN would learn the errors inherent to FE modelling, it remains unclear if it could still accurately predict experimentally measured femoral strength.

The goal of this study was to test if CNNs can accurately predict experimentally measured strength of femora with metastatic lesions based on a single 2D radiographic projection. To circumvent the problem of limited data availability, we propose to use a synthetic training dataset, generated via FE models. To assess the accuracy, we applied the CNNs to an independent ex vivo experimental test dataset. Finally, the performance of the CNNs was compared to 2D densitometric predictors (aBMD, BMC) and 3D CT-based FE modelling to evaluate the achieved degree of accuracy.

## 2. Methods

### 2.1. Outline

The outline of this study is shown in Fig. 1. In brief, CNNs were trained using an FE model generated training dataset to predict femoral strength ($F_{max}$) in a one-legged stance configuration from a single anterior-posterior 2D projection (see Section 2.6). The training dataset was based on 36 CT scans of intact femora, which were modified by inserting artificial lytic lesions using image processing, resulting in 36,000 modified CT scans. Anterior-posterior projections were created from each CT scan and femoral strength was predicted using FE models (see Sections 2.3 and 2.4). The CNNs were tested on 31 anatomical femoral specimens (16 intact, 15 with artificial lesions) by comparing the CNN-predicted to the experimentally measured femoral strength (see Section 2.5). Finally, the achieved accuracy was compared to 2D densitometric predictors (aBMD/BMC) and CT-based FE model predictions (see Sections 2.7 and 2.8). An overview of all datasets is provided in Table 1.

### 2.2. Data from previous studies

All CT scans and experimental data were obtained from two previous studies (Dall'Ara et al. [33] and Benca et al. [34]). No additional scans or experiments involving human anatomical specimens were performed for this study. CT scanning and experimental testing were largely consistent in these two studies and performed at identical facilities. In brief, isolated femora from human body donors were scanned in saline solution using a clinical CT scanner (Brilliance 64, Philips, Germany; voxel size: 0.33 mm $\times$ 0.33 mm $\times$ 1 mm), cut to a standardized length, embedded in a polyurethane block with an inclination of 20° to represent femoral alignment during one-legged stance, and uniaxially loaded in a material testing machine (Fig. 2). The measured maximum force was defined as the femoral strength.

The CT scans and experimental data of both studies [33,34] were pooled to calibrate the FE models prior to CNN training (Table 1). The training dataset was based on the CT scans of 36 intact human femora (age: 76$\pm$12 years; female/male: 19/17; left/right: 18/18) originally scanned and experimentally tested in Dall'Ara et al. [33]. The test dataset was based on 16 intact femora and 15 femora with artificially created lytic lesions in the femoral neck region (age: 80$\pm$11 years; female/male: 9/7; left/right: 16/15) as presented in Benca et al. [34]. The artificial lytic lesions were introduced either in the inferomedial or superolateral region of the femoral neck (inferomedial: 8; superolateral: 7) by physically removing bone tissue (Fig. 1).

### 2.3. FE modelling and calibration

FE models were created from CT scans and used to predict femoral strength in one-legged stance. The loading conditions closely imitated the experimental setup described above [33,34] in order to calibrate the models. The main workflow and modelling approach was already described in previous studies [33,34]. However, a simplified material model was used and calibrated in this study. This was necessary to ensure accurate training data and to enable feasible runtimes for thousands of simulations required to build the training dataset.

Prior to FE model generation, the CT scans were processed as described in [33]. In brief, the CT scans were calibrated using a bone mineral density calibration phantom, femora were aligned with 20° inclination similar to the experimental setup, and cropped distally at the level of the distal embedding. The bone was then masked, resampled to a resolution of 3 mm $\times$ 3 mm $\times$ 3 mm and a proximal embedding layer was added to facilitate load application.

The processed CT scans were used to generate nonlinear, voxel-based FE models (Fig. 2). The mesh was created by converting each voxel to a linear hexahedral element with 3 mm side length. Bone elements were

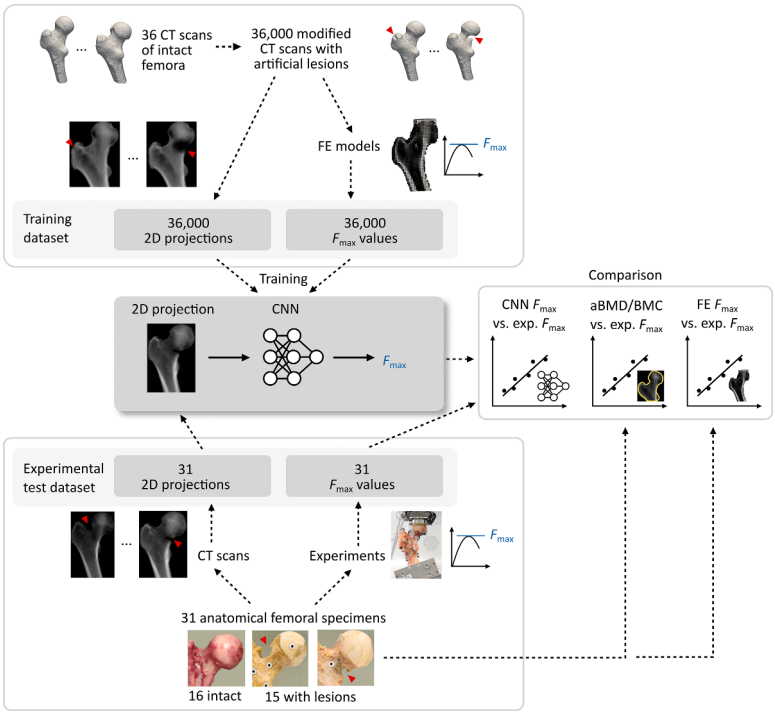


**Fig. 1.** Outline of the study. CNNs were trained with an FE model generated dataset to predict femoral strength from 2D projections. The accuracy of the CNNs was assessed by comparing the predictions to the experimentally measured femoral strength of 31 anatomical femoral specimens. Finally, the CNN performance was compared to other predictors of femoral strength (aBMD/BMC, CT-based FE models).

**Table 1**
Overview of different datasets used for FE model calibration as well as CNN training and testing. The table shows the original study source of the used femora, number of intact specimens and those with metastatic lesions, and whether experimentally assessed or FE-predicted femoral strength were used.

| Dataset | Based on specimens in Dall'Ara et al. [33] | | Based on specimens in Benca et al. [34] | | Total dataset size | Experimental femoral strength | FE-predicted femoral strength |
|---|---|---|---|---|---|---|---|
| | Intact | With lesions[a] | Intact | With lesions[b] | | | |
| FE model calibration | 36 | – | 16 | 15 | 67 | Yes | Yes |
| Training dataset | – | 36,000 | – | – | 36,000 | – | Yes |
| Experimental test dataset | – | – | 16 | 15 | 31 | Yes | – |

[a] Artificial ellipsoidal lytic lesions, introduced through image processing as described in Section 2.4.
[b] Artificial lytic lesions in the femoral neck region, simulated by physical bone removal [34].

assigned a density-dependent isotropic, elasto-plastic material with softening behaviour following Keyak et al. [11] (Table 2). Bone elements inside lesions were modelled using the same material, and the reduced material properties solely resulted from locally attenuated CT grey values as described in Section 2.4. The embedding material was assigned an elastic modulus of 1360 MPa and a Poisson's ratio of 0.3 [33,34]. All nodes at the distal end of the bone were fully constrained. A reference node was positioned in the center of the femoral head and coupled to the most proximal nodes of the embedding. The load was then applied by displacing the reference node distally until a drop of the reaction force in the reference node was detected. The maximum force was defined as femoral strength.

The FE models were calibrated based on the pooled CT scans and experimentally measured strength of all available femora ($n = 67$) (Table 1). Prior to calibration, the correlation between FE model predictions and experimental measurements was already excellent (coefficient of determination $R^2$=0.89). However, a paired $t$-test showed significant quantitative differences ($p < 0.001$) between predicted and experimentally measured femoral strength. Therefore, the ultimate strength ($S$) and the stress plateau after softening ($\sigma_{min}$) of the material model were iteratively scaled with a constant factor $c$ until the root mean squared error reached its minimum. After calibration with a scale factor of 1.55, the coefficient of determination remained nearly unchanged ($R^2$=0.90), but the quantitative agreement improved substantially. A paired $t$-test showed no significant differences ($p = 0.422$) between FE model predicted and experimentally measured femoral strength. The average solving time per FE simulation was 10.8 s (range: 5 to 21 s) on 8 CPUs (see Section 2.9 for details on hard- and software).

### 2.4. Training dataset

Given the limited availability of data, a synthetic training dataset was generated. Conventional data augmentation techniques, such as geometric transformations or image mixing [35], were deemed insufficient for this task, as only 15 CT scans of femora with lesions were available. Instead, the approach focused on locally attenuating the grey values in CT images of intact femora to simulate the physical removal of bone tissue, as performed in the experimental study of Benca et al. [34]. Specifically, the training dataset was created by introducing 1,000 artificial, randomly sized, oriented and located ellipsoidal lytic lesions into each of the 36 intact femora CT scans, originally scanned by Dall'Ara et al. [33] (Table 1, Fig. 3). The size range and orientation of the ellipsoids were selected to cover the size of the lesions created in Benca et al. [34]. The length of the largest semi axis was randomly selected in

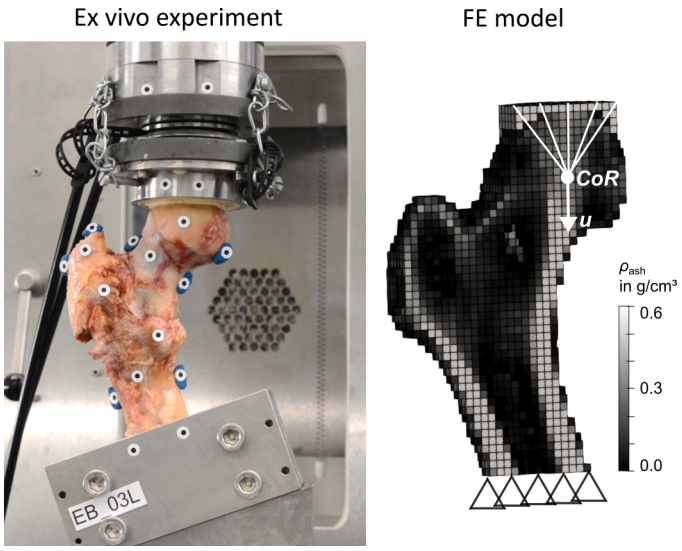


**Fig. 2.** Ex vivo experimental setup (left) and a representative FE model of one femur (right). In the FE model, displacement (***u***) was imposed on the center of rotation (***CoR***), kinematically coupled to the proximal embedding, to mimic the experimental boundary conditions. The material properties of the FE model were assigned element-specific based on local ash density $\rho_{ash}$.

**Table 2**
Material parameters used for the FE models. The material model was based on the elasto-plastic material with softening behaviour as presented by Keyak et al. [11]. $S$ and $\sigma_{min}$ were scaled by a calibration factor of $c = 1.55$ to improve the 1:1 agreement with experimental results.$\rho_{CHA}$ denotes the bone mineral density in g/cm³ obtained from the CT images. More details about the material model and its parameters are provided in the original publication [11].

| Parameter name | Unit | Value |
|---|---|---|
| Ash density | g/cm³ | $\rho_{ash} = 0.0633 + 0.887 \cdot \rho_{CHA}$ |
| Elastic modulus | MPa | $E = 14,900 \cdot \rho_{ash}^{1.86}$ |
| Poisson's ratio | – | $\nu = 0.4$ |
| Ultimate strength | MPa | $S = c \cdot 102 \cdot \rho_{ash}^{1.80}$ |
| Plastic strain limit before softening | – | $\varepsilon_{AB} = 15/3 \cdot (0.00189 + 0.0241 \cdot \rho_{ash})$ |
| Softening modulus | MPa | $E_p = \frac{3 \cdot E \cdot (-2080 \rho_{ash}^{1.45})}{15 \cdot E - 12 \cdot (-2080 \rho_{ash}^{1.45})}$ |
| Minimum stress after softening | MPa | $\sigma_{min} = c \cdot 43.1 \cdot \rho_{ash}^{1.81}$ |

the range between 10 and 30 mm (semi axis $a$ in Fig. 3), and the length of the remaining two semi axes were randomly selected within a range of 40 to 60 % of the largest semi axis. The ellipsoids were then randomly oriented, such that the largest semi axis lay within a cone with a 15° opening angle centered around the anterior-posterior axis. The center of each ellipsoid was placed in a region delimited proximally by the center of the femoral head and distally 10 mm above the trochanter minor (Fig. 3). To ensure a smooth transition between the lesion and the bone, the grey values were gradually attenuated from the boundary of the ellipsoids towards zero in the center of the ellipsoid, following an exponential function. More specifically, the grey value of a voxel at location $\boldsymbol{x}$ within the ellipsoid was multiplied with a function $f(r) = \left(e^{r/r_s} - 1\right)/\left(e^{1/r_s} - 1\right)$, where $r = x_1^2/a^2 + x_2^2/b^2 + x_3^2/c^2$ with the ellipsoid semi axes $a$, $b$, $c$, and parameter $r_s$ adjusts the smoothness of the transition (here: $r_s = 0.3$).

The 36,000 modified CT scans were resampled to a voxel size of 1 mm × 1 mm × 1 mm and used to create idealized 2D projections by summing up the BMD-calibrated, masked femur images in the anterior-posterior direction (Fig. 3). This resulted in BMD-calibrated 2D images with a pixel size of 1 mm. FE models were then created from each modified CT scan as described in Section 2.3 to predict femoral strength. The variability of location and size of the lesions as well as the impact on femoral strength are shown for one exemplary specimen in Fig. 4.

### 2.5. Experimental test dataset

The experimental test dataset consisted of the 2D projections and experimentally measured femoral strength of 16 intact femora and 15 femora with artificial metastatic lesions which were originally scanned and tested in Benca et al. [34] (Table 1, see also Section 2.2). The 2D projections were created from the CT scans of each specimen as described for the training dataset in Section 2.4. Note that no FE modelling was involved in the experimental test dataset, but femoral strength was directly assessed from the experimental measurements.

### 2.6. CNN architectures, image pre-processing and training

CNNs with four different architectures (custom, VGG16, ResNet50, DenseNet121) were created and trained to predict femoral strength quantitatively from the 2D projections. The idea was to compare the performance of a relatively shallow and simple custom CNN to a classical deep CNN (VGG16) [36], as well as a CNN with shortcut connections (ResNet50) [37] and one with densely connected layers (DenseNet121) [38]. These architectures were selected as they were successfully used to perform regression tasks on plain radiographs in the past [29,30]. All four architectures take a 130×130 pixel image as input, extract features using several convolutional layers, and conclude with fully connected layers to predict femoral strength as a single scalar value (Fig. 5).

The custom CNN architecture (Fig. 5) was created by progressively increasing the level of complexity (e.g. number of layers, number of filters) until satisfactory results were achieved. The custom architecture consisted of four convolutional layers with 64/128/256/512 filters (size 3 × 3, same-size padding, ReLU activation), each followed by max-pooling layers (stride 2), a global average pooling (GAP) layer, two fully connected (FC) layers (512 neurons, ReLU activation) and a final output neuron (no activation function). For the VGG16, ResNet50, and DenseNet121 architectures, the input, GAP and FC layers were kept unchanged, while the convolutional layers were replaced by the baseline networks' architecture. The number of convolutional layers and trainable parameters are compared in Table 3.

Prior to training, all 2D projections were pre-processed by cropping them to the most proximal region with a height of 120 pixels (=120 mm) and placing them into an empty 130×130 pixel image to ensure square size and padding at the proximal end (Fig. 5). The maximum height was limited to 120 pixels given the CT-scanned region of all femoral specimens in this study. The grey values were normalized to a range from 0 to 1 using the maximum grey value of all 2D projections. Images of right femora were flipped.

To train the models and to monitor the training process, the training dataset was split into a training subset (90 %) and a validation subset (10 %). Thus, the CNNs were trained on 32,400 specimens and the training process was monitored by evaluating the performance on the remaining 3,600 specimens. Image augmentation was used to enhance variability by adding small random rotations (±10 °) and random translations (±10 mm). Mean squared error was chosen as the loss function and the Adam algorithm was used for optimization. Initial CNN parameters were set randomly for the custom CNN, whereas VGG16, ResNet50 and DenseNet121 were initialized with parameters obtained from pre-training on ImageNet [39]. To put a stronger emphasis on the under-represented critical specimens during training (i.e. specimens with strong reduction of femoral strength due to a lesion; see Fig. 4), sample weights were introduced to the loss function. The sample weights were computed as

$$w_i = \left( \left| \frac{F_{max,i} - F_{max,intact,i}}{F_{max,intact,i}} \right| + 1 \right)^k$$

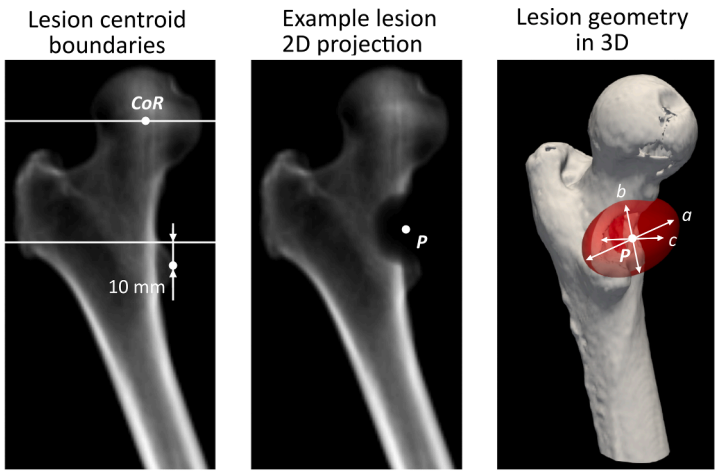


**Fig. 3.** Details of creating the artificial lytic lesions, demonstrated on an exemplary femur. The boundaries for placing a random lesion centroid were defined by the center of rotation (***CoR***) and the lesser trochanter landmarks (left). The 2D projection after inserting a lesion at location ***P*** is shown in the center. The parametrization of the 3D lesion geometry with ellipsoid semi axes *a, b*, and *c* is shown on the right.

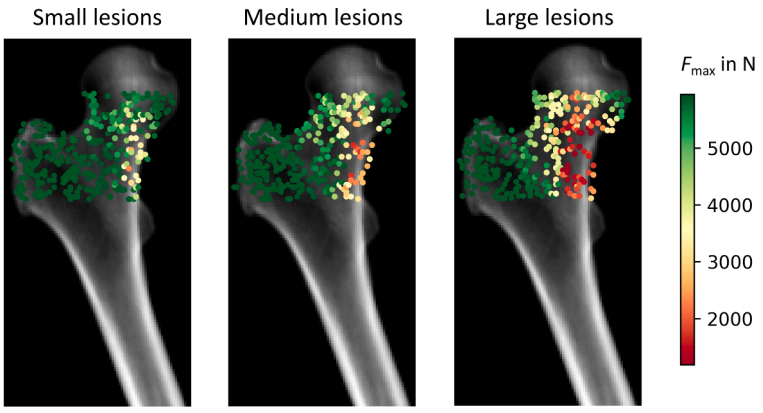


**Fig. 4.** Effect of size and location of the artificial lesions on femoral strength ($F_{max}$). Location and size of 1,000 artificial ellipsoidal lesions and their effect on femoral strength are indicated in one exemplary femoral specimen. Each point represents the centroid of the ellipsoid, and the color indicates the femoral strength. Based on their volume, lesions are separated into small (<33 percentile), medium (≥33 percentile and <66 percentile) and large lesions (≥66 percentile) for better visibility. Larger lesions located in the femoral neck region led to the strongest reduction in strength.

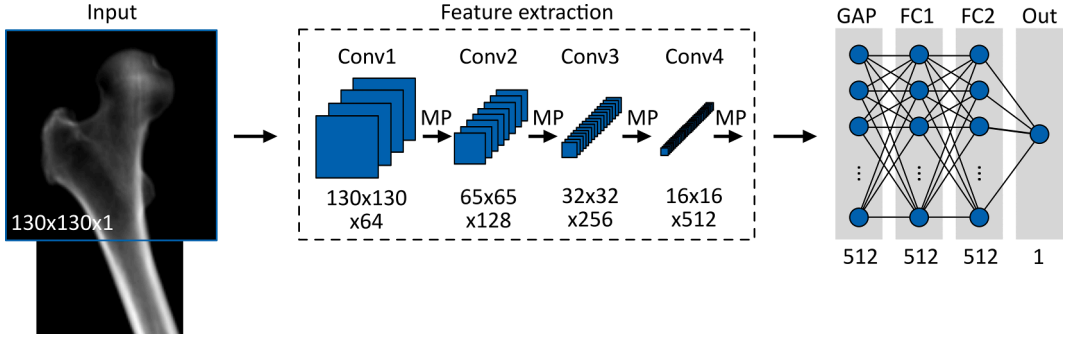


**Fig. 5.** Architecture of the custom CNN to predict femoral strength quantitatively from a 2D projection. Image dimensions and number of filters are indicated for the convolutional (Conv) layers of the custom CNN. MP denotes max pooling operations. The convolutional layers are followed by a global average pooling (GAP) layer and fully connected (FC) layers.

where $F_{max,i}$ is the FE-predicted femoral strength of specimen $i$ with a lesion, $F_{max,intact,i}$ is the femoral strength of the respective intact femur, and $k$ is a hyperparameter. After preliminary testing and manual hyperparameter tuning, all models were trained with $k = 4$, a batch size of 32 and for a maximum of 200 epochs. VGG16 was trained with a constant learning rate of 0.0001 and all other models were trained with a learning rate of 0.001. Other parameters of the Adam optimizer were set to their suggested default values [40]. The training was stopped early

**Table 3**
Comparison of CNN architectures by the number of convolutional layers and trainable parameters including the fully connected layers.

| CNN architecture | Number of convolutional layers | Total number of trainable parameters |
|---|---|---|
| Custom | 4 | 2,075,649 |
| VGG16 | 13 | 15,240,513 |
| ResNet50 | 49 | 24,846,849 |
| DenseNet121 | 120 | 7,741,825 |

if there was no improvement in the validation subset loss for 10 epochs and the best model parameters were restored after stopping. Mean absolute percentage error (*MAPE*) was evaluated as an additional metric during the training process.

### 2.7. Alternative predictors

To assess the performance of the CNNs on the experimental test dataset relative to conventional methods, the ability of 2D densitometric measurements and 3D FE models to predict femoral strength was evaluated. As 2D densitometric measurements, aBMD and BMC were evaluated for the full proximal region of the femur from the 2D projections delimited 20 mm distally to the lesser trochanter. The 3D FE model results were taken directly from the calibration study in Section 2.3.

### 2.8. Statistical analyses

The agreement between variables was evaluated using linear regression analysis. The goodness of fit was quantified using coefficient of determination ($R^2$) as well as Lin's concordance correlation coefficient (*CCC*) [41]. In contrast to $R^2$, *CCC* does not only describe the captured variance but also the extent of quantitative agreement ("1:1 agreement"), ranging from 0 (no agreement) to ±1 (perfect agreement).

### 2.9. Soft- and hardware

Medtool 4.8 (Dr. Pahr Ingenieurs e.U., Pfaffstätten, Austria) and custom Python scripts were used for CT image processing and FE model generation and postprocessing. The FE models were solved using Abaqus 2024 (Dassault Systemes, Velizy-Villacoublay, France) using 8 CPUs (AMD EPYC 7543). The CNNs were created in Python 3.10.14 using Tensorflow 2.14.1 and training was performed on a Nvidia GeForce RTX 4090 GPU. Statistical analyses were performed in Python using SciPy [42].

## 3. Results

### 3.1. Training dataset

CNNs were trained on the training subset ($n$ = 32,400) and the performance was monitored on the validation subset ($n$ = 3,600) of the training dataset. All CNN models were able to predict the femoral strength of the training subset with high accuracy ($R^2$≥0.99; *CCC*≥0.99; *MAPE*<3 %) (Fig. 6, Table 4). Slightly higher errors were observed on the validation subset, but the agreement between CNN and FE model predictions was still excellent ($R^2$≥0.99; *CCC*≥0.99; *MAPE*<4 %) (Fig. 6, Table 4).

Investigating the strength predictions on individual femora of the validation subset showed that the reduction of the femoral strength due to the artificial lesions could be captured very well (Fig. 7, left). Higher errors were tendentially observed for large lesions and lesions located in the femoral neck (Fig. 7, right), which are also the lesions associated with the largest impact on femoral strength (Fig. 4).

### 3.2. Experimental test dataset

Overall, all CNN models predicted the experimental femoral strength from 2D projections with high accuracy ($R^2$≥0.80, *CCC*≥0.81) (Fig. 8, Table 5). Intact femur strength was generally predicted with higher accuracy ($R^2$≥0.89, *CCC*≥0.92) than the strength of femora with metastatic lesions ($R^2$≥0.80, *CCC*≥0.81). As shown in Fig. 8, the CNNs slightly overestimated the femoral strength in femora with metastatic lesions. This observation was consistent for all CNN architectures. DenseNet121 performed best on the pooled experimental test dataset ($R^2$=0.92, *CCC*=0.95), although the custom CNN architecture delivered the best results on femora with metastatic lesions ($R^2$=0.84, *CCC*=0.86).

In comparison to 2D densitometric predictors (aBMD, BMC), the

**Table 4**
Agreement of CNN-predicted and FE-predicted femoral strength in the training and validation subset based on $R^2$, *CCC*, and mean absolute percentage error (*MAPE*).

| CNN architecture | Training subset | | | Validation subset | | |
|---|---|---|---|---|---|---|
| | $R^2$ | *CCC* | *MAPE* in % | $R^2$ | *CCC* | *MAPE* in % |
| Custom | 0.999 | 0.999 | 1.54 | 0.993 | 0.996 | 3.01 |
| VGG16 | 0.997 | 0.998 | 2.21 | 0.995 | 0.997 | 2.74 |
| ResNet50 | 0.997 | 0.998 | 2.35 | 0.994 | 0.997 | 3.15 |
| DenseNet121 | 0.998 | 0.999 | 2.07 | 0.995 | 0.997 | 2.83 |

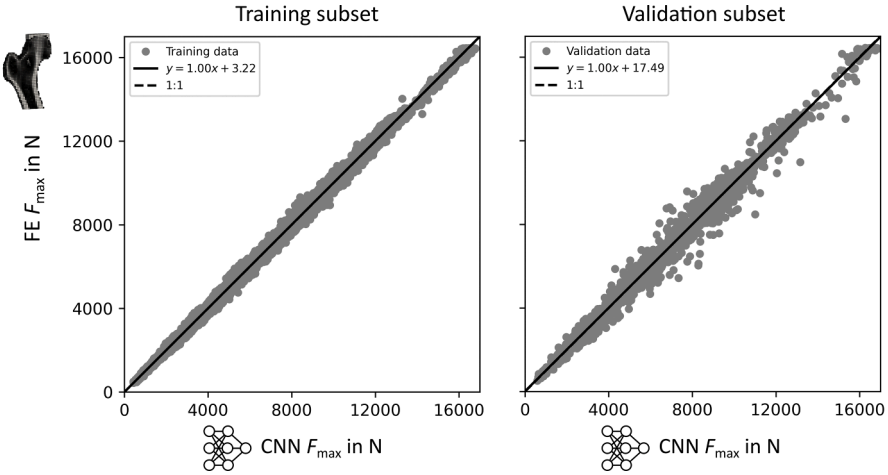


**Fig. 6.** Correlation plots of the custom CNN-predicted and FE-predicted femoral strength ($F_{max}$) on the training dataset. Both the results on the training subset (left; $n$ = 32,400) and the validation subset (right; $n$ = 3,600 are shown).

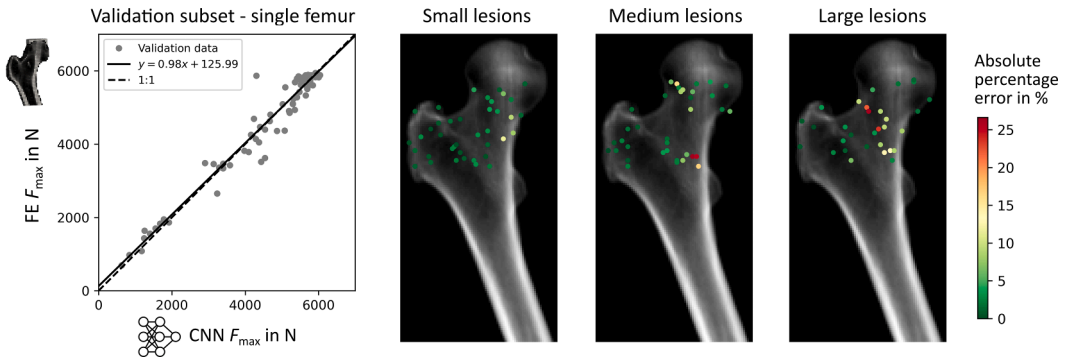


**Fig. 7.** Results on a single femoral specimen of the validation subset using the custom CNN. A correlation plot of the CNN-predicted and FE-predicted femoral strength (left) and relative errors associated with different locations of the metastatic lesions (right) are shown.

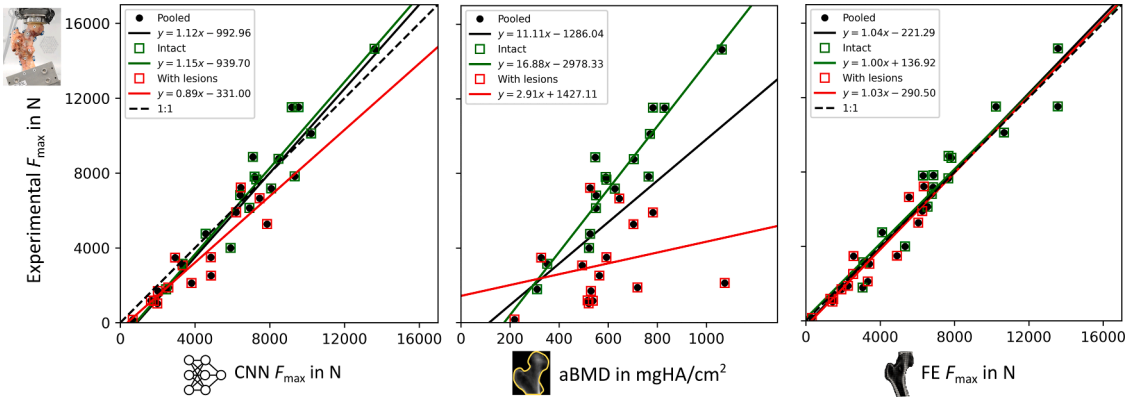


**Fig. 8.** Correlation plots of experimentally measured femoral strength ($F_{max}$) with the custom CNN-predicted $F_{max}$, (left), aBMD (center) and 3D FE model predicted $F_{max}$ (right).

**Table 5**
Correlations of CNN-predicted and experimentally measured femoral strength in the experimental test dataset based on $R^2$ and *CCC*.

| Model | | Intact | | With lesions | | Pooled | |
|---|---|---|---|---|---|---|---|
| **Category** | **Type** | $R^2$ | *CCC* | $R^2$ | *CCC* | $R^2$ | *CCC* |
| 2D CNN | Custom | 0.89 | 0.92 | 0.84 | 0.86 | 0.91 | 0.94 |
| | VGG16 | 0.92 | 0.95 | 0.82 | 0.81 | 0.90 | 0.93 |
| | ResNet50 | 0.92 | 0.95 | 0.83 | 0.81 | 0.91 | 0.93 |
| | DenseNet121 | 0.95 | 0.97 | 0.80 | 0.83 | 0.92 | 0.95 |
| 2D Densitometric | aBMD | 0.88 | – | 0.07 | – | 0.34 | – |
| | BMC | 0.84 | – | 0.04 | – | 0.31 | – |
| 3D FE | Calibrated | 0.90 | 0.95 | 0.90 | 0.94 | 0.94 | 0.97 |

CNNs delivered substantially better strength predictions in femora with metastatic lesions (aBMD/BMC: $R^2 \leq 0.07$; CNNs: $R^2 \geq 0.80$; Fig. 8, Table 5). In intact femora, 2D densitometric predictors and CNNs performed equally well (aBMD/BMC: $R^2 \geq 0.84$; CNNs: $R^2 \geq 0.89$; Fig. 8, Table 5). In comparison to the calibrated 3D FE models, the CNNs performed equally on intact femora, but inferior on the femora with metastatic lesions and slightly inferior on the pooled data (Fig. 8, Table 5).

## 4. Discussion

The CNNs could predict the experimentally measured femoral strength from 2D projections with high accuracy ($R^2 \geq 0.80$, $CCC \geq 0.81$), both in intact femora and those with artificial metastatic lesions. This result was achieved using a large training dataset based on modified CT scans with artificial lesions and FE models for femoral strength predictions. The CNNs outperformed standard 2D densitometric measurements (aBMD/BMC) on femora with metastatic lesions and showed only slightly lower accuracy compared to 3D CT-based FE models. The choice of CNN architecture did not have a substantial impact on the results.

To the best of the authors' knowledge, this is the first study that quantitatively predicted the strength of femora with metastatic lesions from 2D projections using CNNs. Previous studies on 2D projections only used aBMD or BMC measured from real or simulated DXA to predict the strength of femora with metastatic lesions using linear regression analysis [22,34]. They found varying results, ranging from $R^2$=0.10 to $R^2$=0.78 for the linear relationships between strength and densitometric quantities. In the present study, aBMD and BMC delivered even lower correlations for femora with metastatic lesions ($R^2 \leq 0.07$). These results suggest that aBMD and BMC may indeed be too simple image features to capture the location- and size-dependent impact of lesions on femoral strength (see also Fig. 4). In clinical routine, physicians also do not predict impending fractures based on densitometric values, but rather a number of factors, such as site, size or nature of the lesion. The superior performance of CNNs, on the other hand, may be explained by their ability to automatically extract image features, which include grey value averages in larger regions of interest (similar to aBMD), but also a local

reduction of grey values in a critical region (e.g. in the femoral neck region). These capabilities of CNNs are important for strength predictions in femora with metastatic lesions, but may only play a minor role for strength prediction in intact femora. Previous studies already showed that aBMD and BMC are good predictors of intact femur strength [19] and methods that reconstruct 3D FE models from 2D projections even deliver results comparable to conventional CT-based FE models [20]. This is in line with the results of the present study, which showed that the CNNs performed only slightly better than aBMD or BMC to predict intact femur strength. Thus, overall, the CNN approach could provide a considerable improvement in accuracy for strength prediction from 2D projections in femora with metastatic lesions, but simple measurements such as aBMD or BMC may be sufficient for intact femora.

Although the CNNs achieved high accuracy in femoral strength prediction, there is still room for improvement. In particular, the CNNs showed a very good performance on the validation subset of the training data ($R^2 \geq 0.99$, $CCC \geq 0.99$), but a considerably lower accuracy on the experimental test dataset ($R^2 \geq 0.80$, $CCC \geq 0.81$). To further investigate and delineate error sources, the CNNs were tested on additional FE generated datasets in Appendix A. The results on the validation subset (Table 4) show that the CNNs excellently predict the strength reduction due to ellipsoidal lesions with arbitrary location and size, as long as the respective femur is included in the training data. However, the accuracy dropped markedly when the CNN was applied to femora not included in the training data (compare results in Table 4 with Table A.2). Thus, a larger cohort of specimens may be beneficial to improve the CNN performance. Only a minor difference in accuracy was observed in the CNN predictions of FE-based rather than experimental femoral strength values (compare results in Table 5 with Table A.2). This indicates that using FE-predicted femoral strength for CNN training may be sufficient to obtain good agreement with experimental measurements, despite errors inherent to FE modelling itself. In addition to reduced accuracy in the experimental test dataset, there was a systematic overestimation of the strength in femora with metastatic lesions in all CNNs. This may be explained by the simplified shape of the ellipsoidal lesions inserted into the CT images in the training dataset, which is only a rough approximation of the true lesion shape in the experimental test dataset. In particular, the physical bone tissue removal in the experiments could have caused notches or other irregular geometrical features which are not well represented by the smooth ellipsoidal lesions used in training dataset. Thus, in order to further improve the overall accuracy and to combat remaining systematic errors of the CNNs, the training data should be enhanced by including more variable metastatic lesions and a larger cohort of specimens in the training dataset.

The study also showed that there are only minor differences in predictive performance between CNN architectures. The four selected architectures (custom, VGG16, ResNet50, DenseNet121) were selected as they incorporate different levels of complexity. Interestingly, even the simple custom CNN with four convolutional layers could capture femoral strength with comparable accuracy to DenseNet121, which comprises 120 convolutional layers and various shortcut connections. In a clinical environment, using a simpler architecture may be beneficial as the computation processes can be explained and visualized more intuitively than in very deep CNNs. A minor impact of the chosen CNN architecture for regression tasks on 2D images was also reported in other studies, with no clear preference for larger or deeper models [29,43]. However, it must be emphasized that the present study included several simplifications, such as using an ideal 2D projection from a CT scan of an isolated femoral specimen. More complex CNN architectures such as DenseNet121 may be advantageous when using 2D projections obtained from real clinical modalities, such as radiographs or DXA, which can involve various image artifacts and overlapping anatomical structures.

While the results of this study are promising, it is important to note that the training and test datasets were limited to artificial lytic lesions with a simplified shape. Real metastatic lesions can be lytic, blastic, or mixed [44], and their shape is rather irregular [45]. This also means that the herein presented approach for data augmentation by randomly inserting ellipsoidal lesions into intact bones could be insufficient to capture the variety of real metastatic lesions. Deep learning-based augmentation techniques, such as deep generative models [46], could be used to create more realistic synthetic training datasets in future studies. Bone tissue with real metastatic lesions may also differ in its material properties [47,48]. Although previous studies suggest that FE models predict the bones' structural mechanical properties accurately without modification of material models [11,47], the performance of CNN predictions based on 2D projections of femora with real metastatic lesions remains to be investigated.

It is also noteworthy that this study focused on conventional CNNs, rather than including other emerging machine learning approaches such as ensemble methods [49] or vision transformers [50]. The decision to use CNNs in this study was intentional, aiming to assess the performance of a well-established machine learning approach that has previously proven successful in regression tasks on radiographs [29,30]. Indeed, CNNs achieved good results on the test cases presented in this study. However, when including real metastatic lesions and real clinical images rather than artificial lesions and idealized 2D projections, other methods should be implemented and compared to conventional CNNs. Particularly vision transformers show potential for improved performance [51], but their need for even larger training datasets [52] could exacerbate the challenge of data scarcity.

Several limitations of the study remain to be mentioned. First, the experimental test dataset only included femora with artificial lytic lesions. Future studies should include specimens with real metastatic lesions, as well as blastic or mixed lesions. Second, an idealized BMD-calibrated and size-scaled anterior-posterior 2D projection from a CT of an isolated femur specimen was used in this study. This idealised image ignores artifacts that may occur in real clinical 2D imaging, such as distortion, field inhomogeneities, overlap with the pelvis, and soft tissues. Third, the training dataset was tailored to the experimental test dataset to some extent. Although this dataset may serve the pre-training of future CNN models, more training data should be generated to better represent the true variety of metastatic lesions. Fourth, CNN hyperparameter tuning was conducted manually. Systematic or automated methods for hyperparameter optimization [53] could be used to further enhance the performance. Finally, this study aimed to predict femoral strength in a simplified one-legged stance configuration, measured in an experimental setup ex vivo. This experimental setup has several inherent limitations, such as neglecting forces induced by muscles and other soft tissues. Still, we considered the comparison to experimental measurements as an important first step to show the capabilities and limitations of CNNs to predict femoral strength from 2D projections.

In conclusion, CNNs could accurately predict experimentally measured femoral strength from 2D projections, both for intact femora and femora with artificial metastatic lesions. The problem of data scarcity could be overcome by creating a large synthetic training dataset using image processing and FE models. The CNNs outperformed 2D densitometric predictors of femoral strength (aBMD, BMC) in femora with metastatic lesions and delivered comparable performance on intact femora. The predictions with CNNs are fast, require little image preprocessing and are easy to automate. This makes CNNs a promising tool for clinical assessment of safety during activities of daily living, sports, and physical therapy, or to estimate pathologic fracture risk, especially when 3D imaging data is not available. Future studies must further investigate their applicability to real metastatic lesions and clinical 2D images such as plain radiographs or DXA images.

## Ethical approval

No new experiments or scans involving human donor bones were performed in this study. Instead, this study reused previously collected ex vivo CT scans and experimental data from human donor bones, obtained in two prior studies [29,30]. These studies were approved by the

ethics committee of the Medical University of Vienna and the donors had provided written consent for their bodies to be used for research and education.

## Funding

The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

## CRediT authorship contribution statement

**Alexander Synek:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Emir Benca:** Writing – review & editing, Resources, Investigation, Data curation. **Roxane Licandro:** Writing – review & editing, Methodology, Investigation. **Lena Hirtler:** Writing – review & editing, Resources, Data curation. **Dieter H. Pahr:** Writing – review & editing, Supervision, Software, Resources.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Dieter H. Pahr is CEO of Dr. Pahr Ingenieurs e.U., which develops and distributes Medtool. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

During the preparation of this work the authors used ChatGPT (GPT-4-turbo) for language editing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2025.108724.

## Appendix A. Additional tests on FE model generated data

Two additional FE test datasets (Table A.1) were created to gain a deeper understanding of the CNN performance and to delineate error sources.

**Table A.1**
Overview of the additional FE test datasets. The table shows the original study source of the used femora, number of intact specimens and those with metastatic lesions, and if experimentally assessed or FE-predicted femoral strength were used.

| Dataset | Based on specimens in Benca et al. [34] | | Total dataset size | Experimental femoral strength | FE-predicted femoral strength |
|---|---|---|---|---|---|
| | Intact | With lesions | | | |
| FE test dataset 1 | 16 | 15[a] | 31 | – | Yes |
| FE test dataset 2 | – | 16,000[b] | 16,000 | – | Yes |

[a] Artificial lytic lesions in the femoral neck region, simulated by physical bone removal [34].
[b] Artificial ellipsoidal lytic lesions, introduced through image processing as described in Section 2.4.

FE test dataset 1 used the same 31 femora as the experimental test dataset, but femoral strength values were predicted by FE models. This would reveal how well the CNN agrees with FE model predictions, which it was trained for, and could be directly compared to the performance on the experimental test dataset. Indeed, the CNNs provided similar performance both for predicting the FE-model based femoral strength and the experimentally measured femoral strength (compare Table 5 and A.2). Some CNN architectures showed more pronounced outliers in the FE test dataset than in the experimental test dataset, particularly for intact femora (Fig. A.1, left). This resulted in lower $R^2$ values for intact femora compared to the experimental test dataset (compare Table 5 and A.2). As in the experimental test dataset, the CNNs still slightly overestimated the strength of femora with metastatic lesions (Fig. A.1, left).

FE test dataset 2 was created by introducing 1,000 artificial ellipsoidal lytic lesions into each of the 16 intact femora originally scanned and tested in Benca et al. [34]. The same methodology as described in Section 2.4 was applied, resulting in 16,000 2D projections and FE-predicted femoral strength values. This dataset allowed to show the abilities of the CNN on a dataset that used the same ellipsoidal lesion geometries as in the training dataset, but femora that were not included in the training dataset. In FE test dataset 2, the scatter increased substantially (Fig. A.1, center), but the overall accuracy did not decrease considerably ($R^2 \geq 0.85$; $CCC \geq 0.92$; Table A.2). Note that the accuracy mainly reflected the accuracy in predicting the intact femur strength of FE test dataset 1 and that the scatter was mainly driven by inaccuracies in predicting intact femur strength (Fig. A.1, Table A.2). The reduction in strength due to different lesions in a single femur was still captured well (Fig. A.1, right), but a larger systematic error and scatter was observed when compared to the specimens in the validation subset (compare Fig. A.1, right, to Fig. 7, left).

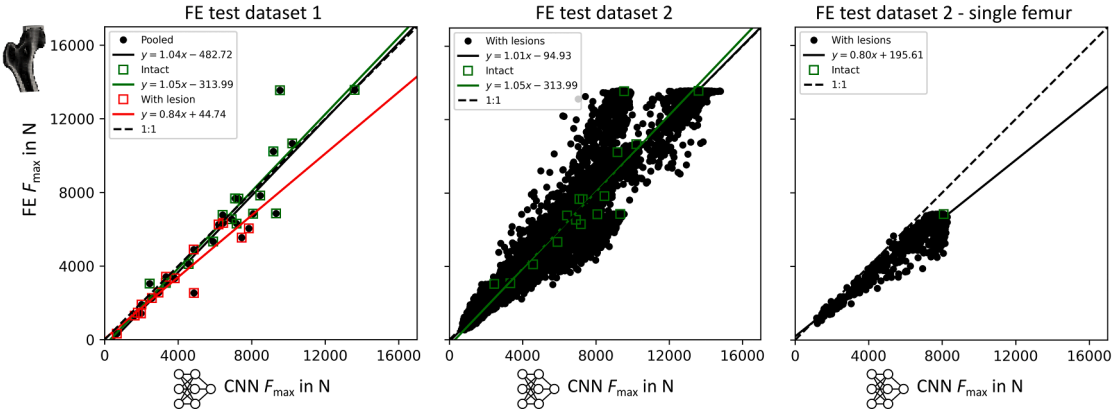


**Fig. A.1.** Results of the custom CNN on the additional FE test datasets. Correlation plots of CNN-predicted and FE-predicted femoral strength ($F_{max}$) on FE test dataset 1 (left), FE test dataset 2 (center), and a single femoral specimen of FE test dataset 2 (right) are shown.

**Table A.2**
Correlations of CNN-predicted and FE-predicted femoral strength on the additional FE test datasets based on $R^2$ and *CCC*.

| CNN architecture | FE test dataset 1 | | | | | | FE test dataset 2 | |
|---|---|---|---|---|---|---|---|---|
| | Intact | | With lesions[a] | | Pooled | | With lesions[b] | |
| | $R^2$ | *CCC* | $R^2$ | *CCC* | $R^2$ | *CCC* | $R^2$ | *CCC* |
| Custom | 0.82 | 0.89 | 0.89 | 0.90 | 0.89 | 0.93 | 0.85 | 0.92 |
| VGG16 | 0.86 | 0.92 | 0.88 | 0.85 | 0.89 | 0.93 | 0.89 | 0.94 |
| ResNet50 | 0.91 | 0.95 | 0.94 | 0.88 | 0.92 | 0.95 | 0.92 | 0.96 |
| DenseNet121 | 0.92 | 0.96 | 0.93 | 0.90 | 0.93 | 0.96 | 0.94 | 0.97 |

[a] Artificial lytic lesions in the neck region introduced physically to the donor bones as described in [34].
[b] Artificial ellipsoidal lytic lesions, introduced through image processing as described in Section 2.4.

## References


[1] K.L. Schulman, J. Kohles, Economic burden of metastatic bone disease in the U.S, Cancer 109 (2007) 2334–2342, https://doi.org/10.1002/cncr.22678.
[2] M. Ratasvuori, R. Wedin, J. Keller, M. Nottrott, O. Zaikova, P. Bergh, A. Kalen, J. Nilsson, H. Jonsson, M. Laitinen, Insight opinion to surgically treated metastatic bone disease: scandinavian sarcoma group skeletal metastasis registry report of 1195 operated skeletal metastasis, Surg. Oncol. 22 (2013) 132–138, https://doi.org/10.1016/j.suronc.2013.02.008.
[3] D.J. Jacofsky, G.J. Haidukewych, Management of pathologic fractures of the proximal femur: state of the art, J. Orthop. Trauma. 18 (2004) 459–469, https://doi.org/10.1097/00005131-200408000-00013.
[4] T.A. Damron, K.A. Mann, Fracture risk assessment and clinical decision making for patients with metastatic bone disease, J. Orthop. Res. 38 (2020) 1175–1190, https://doi.org/10.1002/JOR.24660.
[5] H. Mirels, Metastatic disease in long bones: a proposed scoring system for diagnosing impending pathologic fractures. 1989, Clin. Orthop. Relat. Res. (2003) 415, https://doi.org/10.1097/01.BLO.0000093045.56370.DD.
[6] Y.M. Van Der Linden, H.M. Kroon, S.P.D.S. Dijkstra, J.J. Lok, E.M. Noordijk, J.W. H. Leer, C.A.M. Marijnen, Simple radiographic parameter predicts fracturing in metastatic femoral bone lesions: results from a randomised trial, Radiother. Oncol. 69 (2003) 21–31, https://doi.org/10.1016/S0167-8140(03)00232-9.
[7] E. Benca, J.M. Patsch, W. Mayr, D.H. Pahr, R. Windhager, The insufficiencies of risk analysis of impending pathological fractures in patients with femoral metastases: a literature review, Bone Rep 5 (2016) 51–56, https://doi.org/10.1016/j.bonr.2016.02.003.
[8] E.M. Guinan, K. Devenney, C. Quinn, G. Sheill, C. Mac Eochagáin, M.J. Kennedy, R. McDermott, L. Balding, Associations among physical activity, skeletal related events, and patient reported outcomes in patients with bone metastases, Semin. Oncol. Nurs. 38 (2022), https://doi.org/10.1016/j.soncn.2022.151274.
[9] M.V. Nguyễn, C. Carlier, C. Nich, F. Gouin, V. Crenn, Fracture risk of long bone metastases: a review of current and new decision-making tools for prophylactic surgery, Cancers (Basel) 13 (2021) 1–18, https://doi.org/10.3390/cancers13153662.
[10] T.A. Damron, A. Nazarian, V. Entezari, C. Brown, W. Grant, N. Calderon, D. Zurakowski, R.M. Terek, M.E. Anderson, E.Y. Cheng, A.J. Aboulafia, M. C. Gebhardt, B.D. Snyder, CT-based structural rigidity analysis is more accurate than Mirels scoring for fracture prediction in metastatic femoral lesions, Clin. Orthop. Relat. Res. 474 (2016) 643–651, https://doi.org/10.1007/s11999-015-4453-0.
[11] J.H. Keyak, T.S. Kaneko, J. Tehranzadeh, H.B. Skinner, Predicting proximal femoral strength using structural engineering models, Clin. Orthop. Relat. Res. 437 (2005) 219–228, https://doi.org/10.1097/01.blo.0000164400.37905.22.
[12] A. Sas, E. Tanck, A. Sermon, G.H. van Lenthe, Finite element models for fracture prevention in patients with metastatic bone disease. A literature review, Bone Rep 12 (2020) 100286, https://doi.org/10.1016/J.BONR.2020.100286.
[13] C. Falcinelli, C. Whyne, Image-based finite-element modeling of the human femur, Comput. Methods Biomech. Biomed. Engin. (2020) 1138–1161, https://doi.org/10.1080/10255842.2020.1789863.
[14] F. Eggermont, G. van der Wal, P. Westhoff, A. Laar, M. de Jong, T. Rozema, H. M. Kroon, O. Ayu, L. Derikx, S. Dijkstra, N. Verdonschot, Y. van der Linden, E. Tanck, Patient-specific finite element computer models improve fracture risk assessments in cancer patients with femoral bone metastases compared to clinical guidelines, Bone 130 (2020) 115101, https://doi.org/10.1016/J.BONE.2019.115101.
[15] F. Eggermont, Y. van der Linden, N. Verdonschot, E. Dierselhuis, S. Ligthert, T. Bitter, P. Westhoff, E. Tanck, A patient-specific fracture risk assessment tool for femoral bone metastases: using the bone strength (BOS) score in clinical practice, Cancers (Basel) 14 (2022) 5904, https://doi.org/10.3390/CANCERS14235904/S1.
[16] C.B. Confavreux, H. Follet, D. Mitton, J.B. Pialat, P. Clézardin, Fracture risk evaluation of bone metastases: a burning issue, Cancers (Basel) (2021) 13, https://doi.org/10.3390/cancers13225711.
[17] Z. Yosibash, K. Myers, N. Trabelsi, A. Sternheim, Autonomous FEs (AFE) - A stride toward personalized medicine, Comput. Math. with Appl. 80 (2020) 2417–2432, https://doi.org/10.1016/J.CAMWA.2020.03.012.
[18] I. Benemerito, W. Griffiths, J. Allsopp, W. Furnass, P. Bhattacharya, X. Li, A. Marzo, S. Wood, M. Viceconti, A. Narracott, Delivering computationally-intensive digital patient applications to the clinic: an exemplar solution to predict femoral bone strength from CT data, Comput. Methods Programs Biomed. 208 (2021) 106200, https://doi.org/10.1016/J.CMPB.2021.106200.
[19] E. Dall'Ara, B. Luisier, R. Schmidt, M. Pretterklieber, F. Kainberger, P. Zysset, D. Pahr, DXA predictions of human femoral mechanical properties depend on the load configuration, Med. Eng. Phys. 35 (2013) 1564–1572, https://doi.org/10.1016/J.MEDENGPHY.2013.04.008.
[20] L. Grassi, I. Fleps, H. Sahlstedt, S.P. Väänänen, S.J. Ferguson, H. Isaksson, B. Helgason, Validation of 3D finite element models from simulated DXA images for biofidelic simulations of sideways fall impact to the hip, Bone 142 (2021) 115678, https://doi.org/10.1016/J.BONE.2020.115678.
[21] E. Benca, A. Reisinger, J.M. Patsch, L. Hirtler, A. Synek, S. Stenicka, R. Windhager, W. Mayr, D.H. Pahr, Effect of simulated metastatic lesions on the biomechanical behavior of the proximal femur, J. Orthop. Res. 35 (2017), https://doi.org/10.1002/jor.23550.

[22] D.A. Michaeli, K. Inoue, W.C. Hayes, J.A. Hipp, Density predicts the activity-dependent failure load of proximal femora with defects, Skeletal Radiol 28 (1999) 90–95, https://doi.org/10.1007/s002560050480.
[23] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans. Neural Networks Learn. Syst. 33 (2022) 6999–7019, https://doi.org/10.1109/TNNLS.2021.3084827.
[24] R. Yamashita, M. Nishio, R.K.G. Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights Imaging 9 (2018) 611–629, https://doi.org/10.1007/S13244-018-0639-9.
[25] S.M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, M.K. Khan, Medical image analysis using convolutional Neural networks: a review, J. Med. Syst. 42 (2018) 1–13, https://doi.org/10.1007/S10916-018-1088-1.
[26] N. Yamamoto, S. Sukegawa, A. Kitamura, R. Goto, T. Noda, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka, K. Kawasaki, Y. Furuki, T. Ozaki, Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates, Biomol 10 (2020) 1534, https://doi.org/10.3390/BIOM10111534, 2020Page 1534. 10.
[27] A. Sarmadi, Z.S. Razavi, A.J. van Wijnen, M. Soltani, Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images, Sci. Reports 141 (14) (2024) 1–17, https://doi.org/10.1038/s41598-024-69119-7, 2024.
[28] K.C. Kim, H.C. Cho, T.J. Jang, J.M. Choi, J.K. Seo, Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation, Comput. Methods Programs Biomed. 200 (2021) 105833, https://doi.org/10.1016/J.CMPB.2020.105833.
[29] C.F. Sabottke, M.A. Breaux, B.M. Spieler, Estimation of age in unidentified patients via chest radiography using convolutional neural network regression, Emerg. Radiol. 27 (2020) 463–468, https://doi.org/10.1007/s10140-020-01782-5.
[30] C.S. Ho, Y.P. Chen, T.Y. Fan, C.F. Kuo, T.Y. Yen, Y.C. Liu, Y.C. Pei, Application of deep learning neural network in predicting bone mineral density from plain X-ray radiography, Arch. Osteoporos. 16 (2021), https://doi.org/10.1007/s11657-021-00985-8.
[31] D. Senanayake, S. Seneviratne, M. Imani, C. Harijanto, M. Sales, P. Lee, G. Duque, D.C. Ackland, Classification of fracture risk in fallers using dual-energy X-ray absorptiometry (DXA) images and deep learning-based feature extraction, JBMR Plus 7 (2023), https://doi.org/10.1002/JBM4.10828/7612360.
[32] A. Shaik, K. Larsen, N.E. Lane, C. Zhao, K.-J. Su, J.H. Keyak, Q. Tian, Q. Sha, H. Shen, H.-W. Deng, W. Zhou, A staged approach using machine learning and uncertainty quantification to predict the risk of hip fracture, Bone Rep. 22 (2024) 101805, https://doi.org/10.1016/J.BONR.2024.101805.
[33] E. Dall'Ara, B. Luisier, R. Schmidt, F. Kainberger, P. Zysset, D. Pahr, A nonlinear QCT-based finite element model validation study for the human femur tested in two configurations in vitro, Bone 52 (2013) 27–38, https://doi.org/10.1016/j.bone.2012.09.006.
[34] E. Benca, A. Synek, M. Amini, F. Kainberger, L. Hirtler, R. Windhager, W. Mayr, D. H. Pahr, QCT-based finite element prediction of pathologic fractures in proximal femora with metastatic lesions, Sci. Rep. 9 (2019), https://doi.org/10.1038/s41598-019-46739-y.
[35] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data. 6 (2019) 1–48, https://doi.org/10.1186/s40537-019-0197-0.
[36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 3rd int. Conf. Learn. Represent. ICLR 2015 - conf, Track Proc. (2014). https://arxiv.org/abs/1409.1556v6 (accessed August 13, 2024).
[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2015) 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016-Decem.
[38] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017-Janua, 2016, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.
[39] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009, 2009, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.
[40] D.P. Kingma, J.L. Ba, Adam: a method for stochastic optimization, 3rd int. Conf. Learn. Represent. ICLR 2015 - conf, Track Proc. (2014). https://arxiv.org/abs/1412.6980v9 (accessed August 13, 2024).
[41] L.I.-K. Lin, A concordance correlation coefficient to evaluate reproducibility, Biometrics 45 (1989) 255, https://doi.org/10.2307/2532051.
[42] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C.N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D.A. Nicholson, D.R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G.A. Price, G.L. Ingold, G.E. Allen, G.R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J.T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J.V. de Miranda Cardoso, J. Reimer, J. Harrington, J.L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, P. Lee, R.T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T.J. Pingel, T.P. Robitaille, T. Spura, T.R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y.O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 173 (17) (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.
[43] C.I. Cira, A. Díaz-Álvarez, F. Serradilla, M.Á. Manso-Callejo, Convolutional neural networks adapted for regression tasks: predicting the orientation of straight arrows on marked road pavement using deep learning and rectified orthophotography, Electron 12 (2023), https://doi.org/10.3390/electronics12183980.
[44] F. Macedo, K. Ladeira, F. Pinho, N. Saraiva, N. Bonito, L. Pinto, F. Gonçalves, Bone metastases: an overview, Oncol. Rev. 11 (2017), https://doi.org/10.4081/oncol.2017.321.
[45] B.M. Rothschild, I. Hershkovitz, O. Dutour, Clues potentially distinguishing lytic lesions of multiple myeloma from those of metastatic carcinoma, Am. J. Phys. Anthropol. 105 (1998) 241–250, https://doi.org/10.1002/(SICI)1096-8644(199802)105:2<241::AID-AJPA10>3.0.CO;2-0.
[46] A. Kebaili, J. Lapuyade-Lahorgue, S. Ruan, Deep learning approaches for data augmentation in medical imaging: a review, J. Imaging. 9 (2023) 81, https://doi.org/10.3390/jimaging9040081.
[47] M.A. Stadelmann, D.E. Schenk, G. Maquer, C. Lenherr, F.M. Buck, D.D. Bosshardt, S. Hoppe, N. Theumann, R.N. Alkalay, P.K. Zysset, Conventional finite element models estimate the strength of metastatic human vertebrae despite alterations of the bone's tissue and structure, Bone 141 (2020) 115598, https://doi.org/10.1016/J.BONE.2020.115598.
[48] T.S. Kaneko, M.R. Pejcic, J. Tehranzadeh, J.H. Keyak, Relationships between material properties and CT scan data of cortical bone with and without metastatic lesions, Med. Eng. Phys. 25 (2003) 445–454, https://doi.org/10.1016/S1350-4533(03)00030-4.
[49] A. Mohammed, R. Kora, A comprehensive review on ensemble deep learning: opportunities and challenges, J. King Saud Univ. - Comput. Inf. Sci. 35 (2023) 757–774, https://doi.org/10.1016/J.JKSUCI.2023.01.014.
[50] R. Azad, A. Kazerouni, M. Heidari, E.K. Aghdam, A. Molaei, Y. Jia, A. Jose, R. Roy, D. Merhof, Advances in medical image analysis with vision transformers: a comprehensive review, Med. Image Anal. 91 (2024) 103000, https://doi.org/10.1016/J.MEDIA.2023.103000.
[51] S. Takahashi, Y. Sakaguchi, N. Kouno, K. Takasawa, K. Ishizu, Y. Akagi, R. Aoyama, N. Teraya, A. Bolatkan, N. Shinkai, H. Machino, K. Kobayashi, K. Asada, M. Komatsu, S. Kaneko, M. Sugiyama, R. Hamamoto, Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review, J. Med. Syst. 48 (2024) 1–22, https://doi.org/10.1007/S10916-024-02105-8/METRICS.
[52] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, M. Shah, Transformers in Vision: a survey, ACM Comput. Surv. 54 (2022), https://doi.org/10.1145/3505244.
[53] M.A.K. Raiaan, S. Sakib, N.M. Fahad, A. Al Mamun, M.A. Rahman, S. Shatabda, M. S.H. Mukta, A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks, Decis. Anal. J. 11 (2024) 100470, https://doi.org/10.1016/J.DAJOUR.2024.100470.