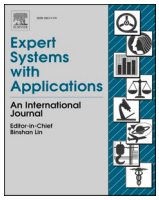
Expert Systems With Applications 264 (2025) 125916

Contents lists available at ScienceDirect

Expert Systems With Applications

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



Machine learning based finite element analysis (FEA) surrogate for hip fracture risk assessment and visualization

Rabina Awal [a,1], Mahmuda Naznin [b], Tanvir R. Faisal [a,*]

[a] Department of Mechanical Engineering, University of Louisiana at Lafayette, LA, USA
[b] Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

ARTICLE INFO

Keywords:
Hip fracture risk
Fracture risk index (FRI)
3D proximal femur
QCT-based FEA
Machine learning
CatBoost model

ABSTRACT

Classical computational biomechanical approaches are costly and require a high level of expertise, often limiting their clinical application. A data-driven Machine Learning (ML) framework can serve as an effective alternative for disease diagnosis and prediction. This study aimed to develop an ML-based modeling approach integrating Quantitative Computed Tomography-based Finite Element Analysis (QCT-based FEA) to assess fall-induced hip fracture probability and probable facture locations in an elderly individual. Specifically, this study focused on predicting patient-specific hip fracture risk using fracture risk index (FRI), calculated on the basis of 1st and 3rd principal strains, and visualizing strain distributions in the proximal femur by generating surrogate FE models via supervised CatBoost model. The training dataset, comprising clinical, anatomical, and mechanical (loading) features, was obtained from QCT image data and QCT-based FEA, which also provided the FRI and strain distributions as targets. The optimized CatBoost model demonstrated 76% accuracy with an AUROC of 0.81 in fracture risk prediction, as well as correlation coefficients of 0.73 for the 1st principal strain and 0.76 for the 3rd principal strain in the surrogate FE models for visualization. Although trained on a limited dataset, this study highlights the efficacy of ML-based surrogate modeling in the QCT-based FEA process for predicting hip fracture risk and visually identifying fracture locations.

1. Introduction

Hip fracture, also known as femoral fracture, is one of the major public health concerns among the elderly population in the United States and worldwide. It can lead to permanent disability as well as substantial mortality (Cooper, et al., 2011). Approximately 1.3 million hip fractures occur annually, and they are associated with nearly 740,000 deaths worldwide (Johnell & Kanis, 2004). This worldwide annual number is projected to exceed 6 million by the year 2050 (Kannus, et al., 1996). In the USA, more than 300,000 patients with hip fractures are hospitalized annually, and 90% of these fractures result from a simple fall (Aschkenasy & Rothenhaus, 2006). Predicting the elevated risk of hip fractures in elderly people is crucial for the development of personalized preventive measures, given the rising incidence of such fractures, and thereby reducing socio economic burden associated with them. Currently, fracture risk assessment tool–FRAX (Kanis, Johnell, Odén, Johansson, & McCloskey, 2008), and Bone Mineral Density (BMD) measurement using Dual-Energy X-ray Absorptiometry (DXA) (Adams, 2013; Albertsson, Mellström, Petersson, Thulesius, & Eggertsen, 2010; Luo, Ferdous, & Leslie, 2011) are the most commonly used predictive methods, but these methods suffer from some limitations. For example, FRAX is a population dependent method, and performs less accurately for inter-racial groups. Moreover, it may lead to overestimation and underestimation of BMD depending on bone size (Ferizi, et al., 2019). Most importantly, these methods ignore the impact of fall and associated mechanisms that mainly trigger hip fracture.

Finite Element Analysis (FEA) has been extensively used as a classical scientific computing tool to model and analyze biomechanical phenomena such as analyzing mechanics of the cardiovascular system, hemodynamics, design and modeling of implants, orthodontics, surgical procedures, and many others (Eshghi, Hojjati, Imani, & Goudarzi, 2011; McCulloch, Guccione, Waldman, & Rogers, 2020; Monteiro, Dal Piva, Tribst, Borges, & Tango, 2018; Post, et al., 2015; Singh, Mogra, Shetty, Shetty, & Philip, 2012). Quantitative Computed Tomography-based FEA (QCT-based FEA) is considered an accurate and effective method for fracture assessment because it considers bone density, femur

* Corresponding author at: Dept. of Mechanical Engineering, 238 E Lewis St., Lafayette, LA 70508, USA.
E-mail addresses: rabina.awal433@gmail.com (R. Awal), mahmudanaznin@cse.buet.ac.bd (M. Naznin), tanvir.faisal@louisiana.edu (T.R. Faisal).
[1] Current Affiliation: SLB, 23400 colonial parkway, Katy, TX 77493, USA.

https://doi.org/10.1016/j.eswa.2024.125916
Received 19 January 2024; Received in revised form 2 November 2024; Accepted 23 November 2024
Available online 26 November 2024

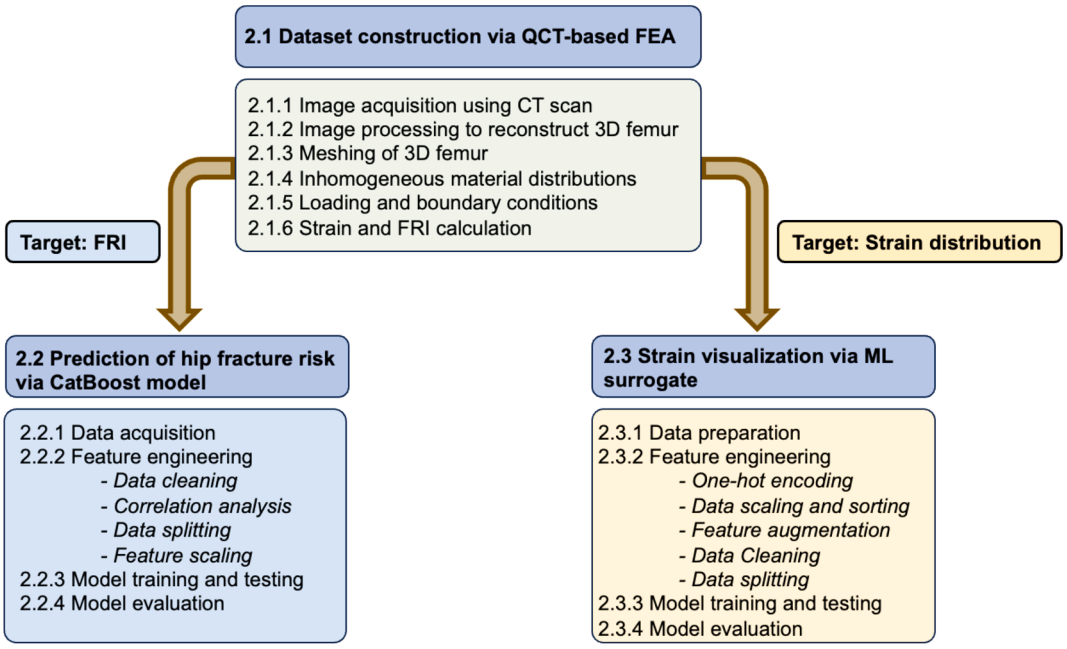


Fig. 1. Schematic of designed ML pipeline.

morphology, and loading effect (Aldieri, Curreli, Szyszko, La Mattina, & Viceconti, 2023; Aldieri, et al., 2022; Bettamer, 2012; Faisal & Luo, 2017; Grassi, et al., 2023; Rui Zhang, 2014; Yang, Palermo, Black, & Eastell, 2014) and provide patient-specific prediction of fracture probability by analyzing the strength of femur (Black, et al., 2008; Dragomir-Daescu, et al., 2011; Engelke, van Rietbergen, & Zysset, 2016; Memiş, Varlı, & Bilgili, 2022). Despite its high fidelity, the application of QCT-based FEA has been limited in the clinical domain due to the high demand for computational resources, expensive software requirements, and the need for skilled experts. However, recent advancements in Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) suggest that ML techniques could be a promising alternative to classical QCT-based FEA. Therefore, we propose using supervised ML to develop data-driven surrogate models that can efficiently and accurately predict subject-specific hip fracture risk.

Data-driven, ML-based models applied clinically for disease diagnosis (Ambale-Venkatesh, et al., 2017; Islam & Nahiduzzaman, 2022; Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Osareh & Shadgar, 2010; Sajjad, et al., 2019; Sultana, Naznin, & Faisal, 2024; Weng, Reps, Kai, Garibaldi, & Qureshi, 2017) and injury prevention (Iliou, Anagnostopoulos, & Anastassopoulos, 2014; Kong, et al., 2020) demonstrated promising results. ML has been used in various orthopedic applications, including fracture detection (Kim & MacKinnon, 2018; Lindsey, et al., 2018; Olczak, et al., 2017; Urakawa, et al., 2019), bone tumor diagnosis (Do, Langlotz, & Beaulieu, 2017), and osteoarthritis grading (Tiulpin, Thevenot, Rahtu, Lehenkari, & Saarakkala, 2018; Xue, Zhang, Deng, Chen, & Jiang, 2017). Liang et al. (Liang, Liu, Martin, & Sun, 2018) developed patient-specific models of stress distribution in the aorta using ML algorithms, with FE model data as the input and aortic wall stress distribution as the output. Madani et al. (Madani, Bakhaty, Kim, Mubarak, & Mofrad, 2019) created a surrogate finite element model using ML to replicate stress distribution in the aortic wall for patients with atherosclerosis. This model learns the correlation between input parameters such as tissue geometry, composition, and arterial pressure, and the resulting stress distribution. These instances in the biomechanical domain highlight the potential of machine learning to develop surrogate FE models, providing patient-specific estimates of system responses.

Krogue et al. (Krogue, et al., 2020) applied ML techniques to radiographs to classify hip fracture categories and automate the identification process to reduce diagnostic error. Artificial Neural Networks (ANNs) have proven effective in constructing predictive models for assessing the risk of hip bone fractures in elderly individuals, both female and male (Liu, et al., 2015). Ferizi et al. (Ferizi, et al., 2019) conducted a comprehensive study to predict osteoporotic bone fractures, comparing fifteen ML-based classifiers using Magnetic Resonance Imaging (MRI) data. However, these ML/DL models have primarily been used to detect and classify fractures that have already occurred. Therefore, the overarching goal of this work is to assess the probability of fracture risk in individuals before a fracture takes place using ML algorithms in conjunction with FEA. In this work, the Categorical Boosting (CatBoost) algorithm was selected due to its more accurate performance, automatic handling of missing values in a dataset, and lower computational time compared to other models such as logistic regression, Support Vector Classifier (SVC), Extreme Gradient Boost Model (XGBM), Random Forest (RF) and Decision Tree (DT), which were tested in-house (Awal et al., 2025) before finalizing the model selection. Additionally, CatBoost can handle both numerical and categorical features without requiring conversion.

To develop the data-driven modeling approach for assessing patient-specific fracture risk, the first objective of this work was to conduct QCT-based FEA to generate target values—strain and Fracture Risk Index (FRI)—for training the CatBoost model. The second objective was to predict fracture risk using a binary classifier, and the final objective was to predictively visualize strain distributions in the femur to qualitatively and/or semi-quantitatively identify potential fracture location.

In this paper, Section 2 outlines the methodology of the study, including the detailed steps of each objective. Section 3 presents the outcomes of the study, demonstrating the performance of the models and their comparison. Section 4 provides a discussion, followed by the conclusion.

2. Materials and methods

The proposed ML pipeline was primarily built upon three major modeling steps (Fig. 1). The detail description of all these steps has been organized in sections 2.1 to 2.3.

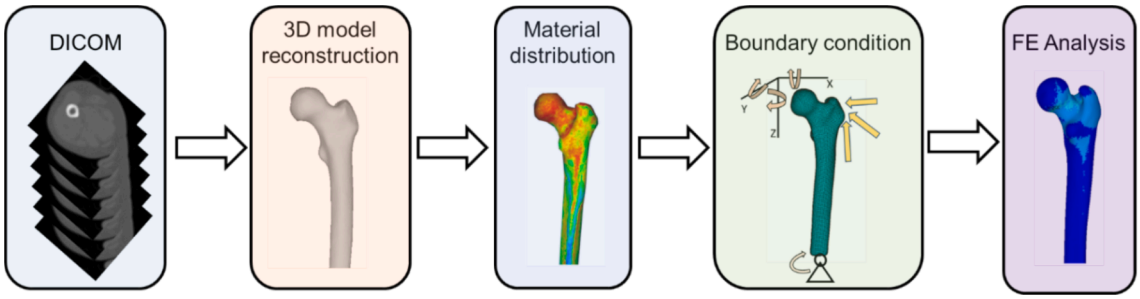


Fig. 2. Workflow of QCT-based FEA from DICOM image dataset to strain distribution.

Table 1
Demographic overview of the patient dataset considered in this study.

	Mean (Min-Max)	Std. Dev.
Number of patients	97	–
Male	52	–
Female	45	–
Age (year)	64.93 (50–86)	8.48
Weight (kg)	83.94 (51.7–148.6)	16.72
Height (cm)	157.2 (145.3–193.2)	6.99

Std. Dev.: Standard Deviation

2.1. Dataset construction via QCT-based FEA

Most of the content in this section revisits our previous works (Awal, Ben Hmida, Luo, & Faisal, 2022; Awal & Faisal, 2024; Awal & Faisal, 2021) to provide a clearer understanding of obtaining ground truth data using classical computational modeling (QCT-based FEA in this context) The workflow of QCT-based FEA started with a Digital Imaging and Communications in Medicine (DICOM) image dataset and culminated in displaying strain distributions in the 3D reconstructed proximal femur under various loading scenarios (Fig. 2) and the corresponding FRI.

2.1.1. Image acquisition using CT scan

A dataset of QCT images of 97 anonymous adults, without any identifiable information was considered in this study. The CT images of the patients in DICOM format was previously obtained from the Great-West Life PET/CT Center located at the Health Science Center, Winnipeg, Canada following ethics approval. The QCT images were obtained by SIEMENS S5VB40B CT scan machine (Siemens Medical Solution, Malvern, USA) with acquisition and reconstruction parameters of 120 kVp and 244 mAs, respectively, and an image matrix of 512 × 512 pixels. The calibration of CT images was performed using a standard procedure, involving the use of a calcium hydroxyapatite calibration phantom (Mindways Inc., Austin, TX, USA) mounted during the scan to ensure accurate estimation of gray value. Table 1 provides a brief demographic overview of the patient dataset.

2.1.2. Image processing to reconstruct 3D femur

The 3D femurs were meticulously reconstructed by semi-automatically segmenting the femur from the pelvis, tibial part, fat, and muscle using 3D Slicer (https://www.slicer.org/), a free open-source software designed for medical image processing and visualization. The reconstruction process utilized the built-in functionalities such as thresholding, smoothing, and segmentation. To reconstruct only the proximal region, we considered approximately half of the femur, about 220 mm from the superior point on the femoral head (Fig. 3). The resulting 3D model was exported to the FEA solver(s) in.STL format for FE simulation.

2.1.3. Meshing of 3D femur

The 3D femur model was meshed using 4-node tetrahedral elements (Fig. 3b) with HyperMesh (Altair, Michigan, USA), a high-performance finite element pre-processor. The FEA of the 3D meshed model was performed using Ansys v19.0 (Ansys, Inc, USA). Mesh convergence was verified, and a maximum edge length of 2 mm was considered (Awal & Faisal, 2024). The FE models generated from the dataset comprised an average of 18,151 nodes (ranging from 16,467 to 23505) and an average of 163,360 elements (ranging from 148,205 to 211,554).

2.1.4. Inhomogeneous material distributions

The study employed an isotropic and inhomogeneous material model. Each voxel in the QCT images was associated with bone density represented in Hounsfield Units (HU) (Faisal & Luo, 2017; Schileo, Taddei, Cristofolini, & Viceconti, 2008). This correlation was achieved through element-wise mapping of HU (Fig. 3c) using the free open-source software Bonemat v3.0 (Taddei, Pancanti, & Viceconti, 2004; Taddei, Schileo, Helgason, Cristofolini, & Viceconti, 2007; Zannoni, et al., 1999). Previous experimental data established a power-law

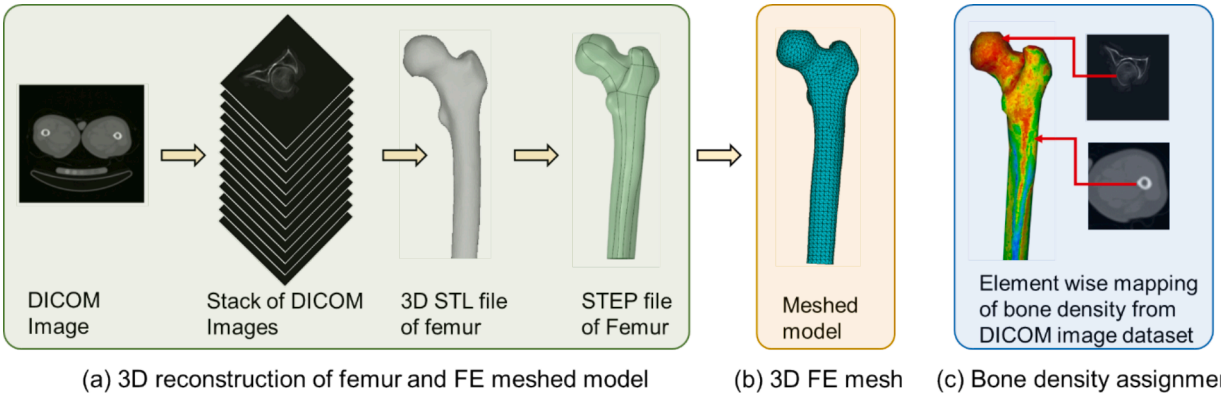


Fig. 3. (a) 3D femur reconstruction from 2D DICOM images, (b) FE meshed model after mesh convergence, and (c) Element-wise mapping of inhomogeneous materials distributions in femur using Bonemat v3.0 (Zannoni, Mantovani, & Viceconti, 1999).

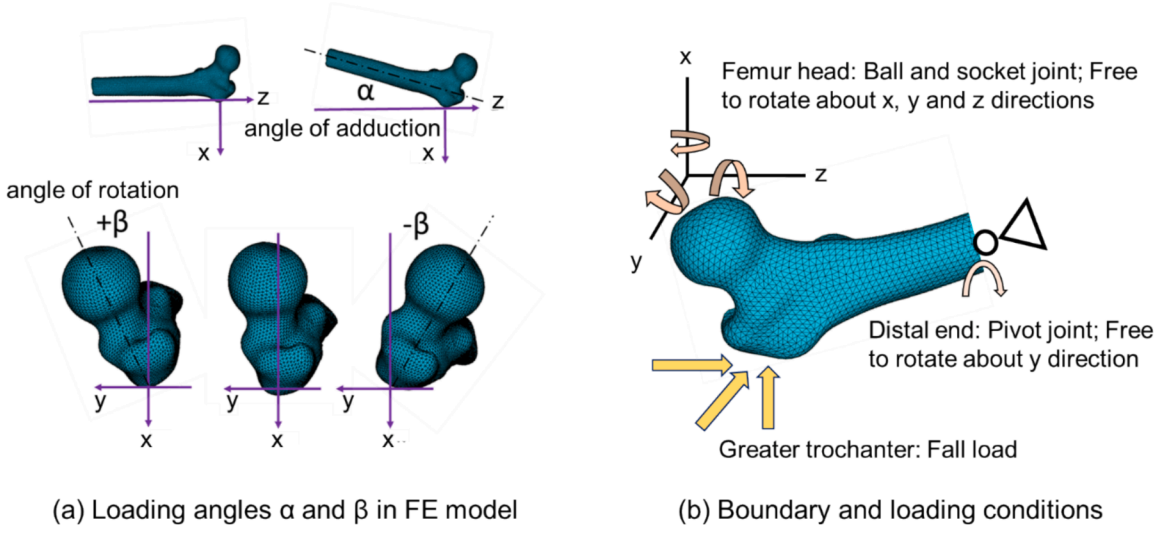


Fig. 4. (a) Representation of loading angles α and β on coronal and transverse planes, respectively, during sideways fall, and (b) Boundary and loading conditions at the distal end, femur head, and greater trochanter of a femur during FEA (Awal & Faisal, 2024).

Table 2
A number of possible fall cases represented by the variations of loading angle (α) on the coronal plane and (β) on the transverse plane.

Sideways fall cases	α (degree)	β (degree)
αβ_0/-15	0	−15
αβ_0/0	0	0
αβ_0/15	0	15
αβ_15/-15	15	−15
αβ_15/0	15	0
αβ_15/15	15	15
αβ_30/-15	30	−15
αβ_30/0	30	0
αβ_30/15	30	15

relationship between Young's modulus and apparent density (Gislason, et al., 2014). The empirical relationships described in Eqs. (1) to (4) exhibited a stronger correlation with experimental data, leading to their adoption for assigning inhomogeneous material properties in this study (Awal & Faisal, 2024; Faisal & Luo, 2016; Marco, Giner, Caeiro-Rey, Miguélez, & Larraínzar-Garijo, 2019).

$$\rho_{QCT} = 0.00079114 \times HU - 0.00382144 \ \left(g/cm^3\right) \tag{1}$$

$$\rho_{ash} = 0.877 \times \rho_{QCT} + 0.0789 \ \left(g/cm^3\right) \tag{2}$$

$$\rho_{app} = \frac{\rho_{ash}}{0.6} \ \left(g/cm^3\right) \tag{3}$$

$$E = 10500 \times \rho_{app}^{2.29} \ (MPa) \tag{4}$$

where HU represents BMD, ρ_{ash} is the ash density, and E is the modulus of elasticity. Poisson's ratio 0.4 was considered for all directions (Awal & Faisal, 2024; Faisal & Luo, 2016).

2.1.5. Loading and boundary conditions

Fall is an uncontrolled event, and a sideways fall can occur in a vareity of orientations. To simulate potential sideways falling scenarios and accommodate various femur positions, nine loading configurations were designed to apply force onto the greater trochanter. These configurations represented different sideway fall postures by simultaneously adjusting the orientation of the loading angle (α) on the coronal plane relative to the shaft axis and the angle (β) on the transverse plane relative to the neck axis, as depicted in Fig. 4a (Awal & Faisal, 2024). The specified angles (Table 2) on each plane were determined based on a typical fall orientation, and the critical angle, where the maximum number of femoral fractures occurred and observed in prior experimental studies (Aldieri, et al., 2022; Bessho, et al., 2004; Ford, Keaveny, & Hayes, 1996; Grassi, et al., 2012; Nishiyama, Gilchrist, Guy, Cripton, & Boyd, 2013; Pinilla, Boardman, Bouxsein, Myers, & Hayes, 1996). The patient-specific load, P_{fall}, required to simulate a sideways fall from a standing height, was then calculated using Eq. (5) (Robinovitch, Hayes, & McMahon, 1991; Yoshikawa, et al., 1994).

$$P_{fall} = 8.25 \times w \times \left(\frac{h}{170}\right)^{1/2} (N) \tag{5}$$

where w and h are the weight and height of a patient, respectively.

In each loading scenario, the distal end of the proximal femur was fixed in all directions except for the rotational degree of freedom along the y-axis, representing the hinge joint (knee) at the femur's distal end (Altai, Qasim, Li, & Viceconti, 2019; Yano, et al., 2022). The translational degrees of freedom at the femur head were constrained, allowing rotation in the x, y, and z directions to replicate the ball-and-socket joint between the femur head and acetabulum. The load due to fall was applied to the greater trochanter (Fig. 4b).

2.1.6. Strain and FRI calculation

A linear FEA was performed in this study as the femur bone behaves linearly elastic up to failure during loading conditions (Cristofolini, Juszczyk, Martelli, Taddei, & Viceconti, 2007; Grassi, et al., 2012; Juszczyk, Cristofolini, & Viceconti, 2011). In addition, the femur exhibits brittle behavior that may be better represented by the maximum stress–strain criteria than the magnitude of stress and strain (von Mises stress/strain criteria) (Cristofolini, et al., 2007; Doblaré, Garcıa, & Gómez, 2004). Schileo et al. (Schileo, et al., 2008) and other studies (Altai, et al., 2019; Marco, et al., 2019; Testi, Viceconti, Cappello, & Gnudi, 2002) showed that principal strain-based FEA analysis could correctly estimate fracture risk. Hence, in this study, we have considered the maximum principal strain-based criteria to estimate the FRIs, which are defined based on the absolute maximum tensile and compressive strains (Bayraktar, et al., 2004; Schileo, et al., 2008) as follows:

$$FRI = \frac{\varepsilon^T_{\max}}{0.0073} \tag{6}$$

$$FRI = \frac{|\varepsilon^c_{\max}|}{0.0104} \tag{7}$$

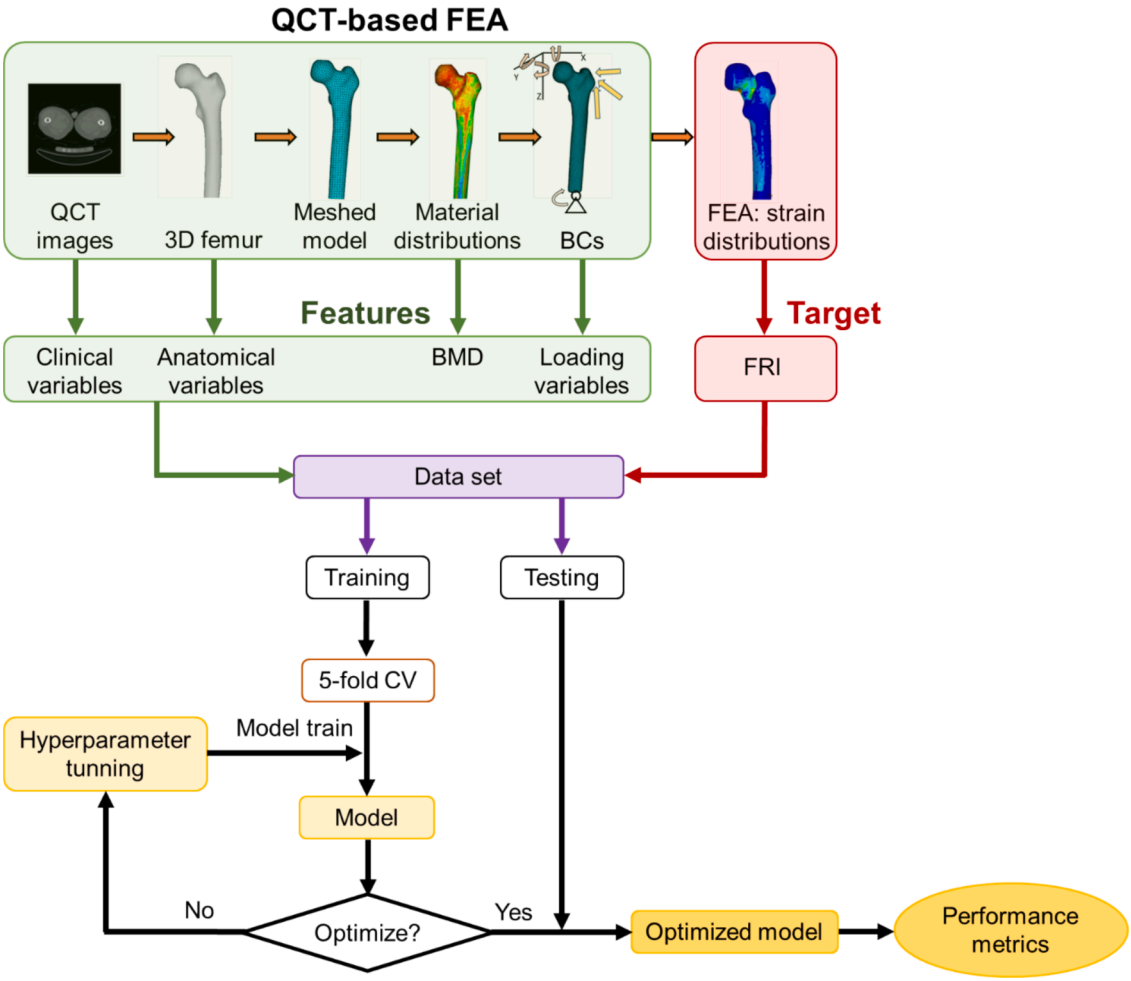


Fig. 5. ML workflow for predicting patient-specific hip fracture risk using FRI as the target variable.

Table 3
Feature and target variable for predicting fracture risk with strain-based FRI as the target variable.

Feature			Target
Clinical and demographic variables	***Anatomical variables***	***Loading variables***	
Age	FNA	α	FRI
Sex	FNW	β	
Weight	TMA		
BMD	FNAL		

where $\varepsilon^{T}_{\max}$ and $\varepsilon^{c}_{\max}$ are the maximum principal strain in tension and compression, respectively. The FRI obtained via our developed QCT-based FEA was validated earlier (secondary validation) (Awal & Faisal, 2024) with previously published FE results by Kheirollahi et al. (Hossein Kheirollahi & Yunhua Luo, 2015; H Kheirollahi & Y Luo, 2015), who used a patient cohort (subset) from the same dataset and full femur models. We showed this validation in our prior work (Awal & Faisal, 2024).

2.2. Prediction of hip fracture risk via CatBoost model

The ML workflow for predicting hip (*a.k.a* femoral) fracture risk was implemented in four phases such as *data acquisition* (extracting features and target variable from QCT-based FEA), *feature engineering, model training and testing,* and *model (performance) evaluation* (Fig. 5).

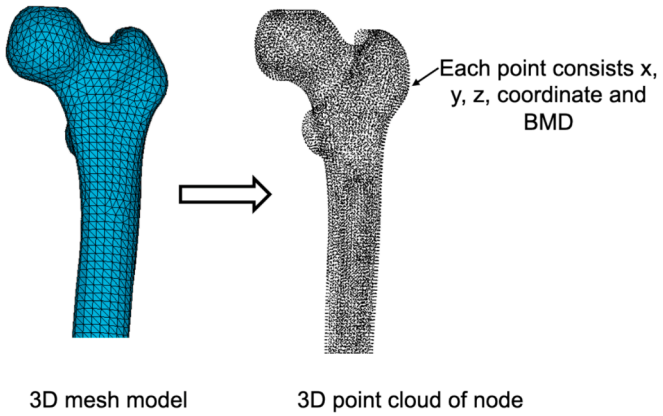


Fig. 6. Representation of a 3D point cloud obtained from a 3D femur mesh.

2.2.1. Data acquisition

The input dataset, consisting of demographic and clinical parameters, bone anatomy and morphology, and loading orientations (Fig. 5), were extracted from QCT dataset, and its corresponding FE model. The key variables influencing the FRI (Awal, et al., 2022; Awal & Faisal, 2021; Ford, et al., 1996; Pinilla, et al., 1996) were prioritized as features. Clinical and demographic data, including age, sex, and weight, were extracted from the QCT dataset in DICOM format, and a weighted-average BMD value was calculated based on the BMD distributions in

Table 4
Feature and target variables for strain visualization via ML surrogate.

Feature			Target
Clinical and demographic variables	***Anatomical variables***	***Loading variables***	1^{st} and 3^{rd} principal strain
BMD (node-wise average values)	x-coordinate y-coordinate z- coordinate	P_{fall} α β	

femur.

Anatomical and morphological features were derived mainly from 2D projections of each femur on the coronal plane such as *Femoral Neck Axis Length* (*FNAL*), *Femoral Neck Width* (*FNW*), *True Moment Arm* (*TMA*), and the horizontal component of FNAL on transverse plane. The *Femoral Neck Angle* (*FNA*), defined as the angle (θ) between the proximal femoral shaft axis and the femoral neck axis, was also considered (Awal & Faisal, 2024; Fajar, Taufan, Syarif, & Azharuddin, 2018).

The target variable, FRI, was obtained from QCT-based FEA (Fig. 5). An FRI value greater than 1 indicates a higher possibility of hip fracture. Therefore, the study used a binary classification "fracture risk" with $FRI > 1$, encoded as 1 and "no fracture risk" with $FRI \leqslant 1$, encoded as 0. Table 3 summarizes all the categorical features and target variables considered in this study.

2.2.2. Feature engineering

The dataset was preprocessed before use in the machine learning model. To improve the performance of the predictive models, feature engineering comprising cleaning, correlation analysis, data splitting, and feature scaling was conducted during the data preprocessing step to reduce the complexity of data and improve model accuracy.

Data cleaning

The FRI values obtained from the QCT-based FEA exhibited a positively skewed distribution and significant dispersion. In this work, the FRI beyond 1.5 times the interquartile range (represented by max) were considered outliers and removed from the dataset. The excluded data includes the instance of two femurs from two patients. Hence, 192 out of 194 femurs were considered in this analysis after removing the outliers.

Correlation analysis

The association between the features was analyzed using the Pearson correlation coefficient to evaluate if the features have any form of association with each other that could affect the target variable.

Data splitting

The dataset was split with an 80:20 ratio for the training and testing of CatBoost model. Additionally, to optimize and increase the robustness of the model by reducing data bias, the training data was further split into 5 folds in the same 80:20 ratio (Little, et al., 2017). To prevent data leakage, splitting was done based on patient's unique identifier.

Feature scaling

Feature scaling is another crucial step towards standardizing the features to increase the predictability of a model. In this step, the MinMax scaler was used to scale the features. MinMax transforms the data

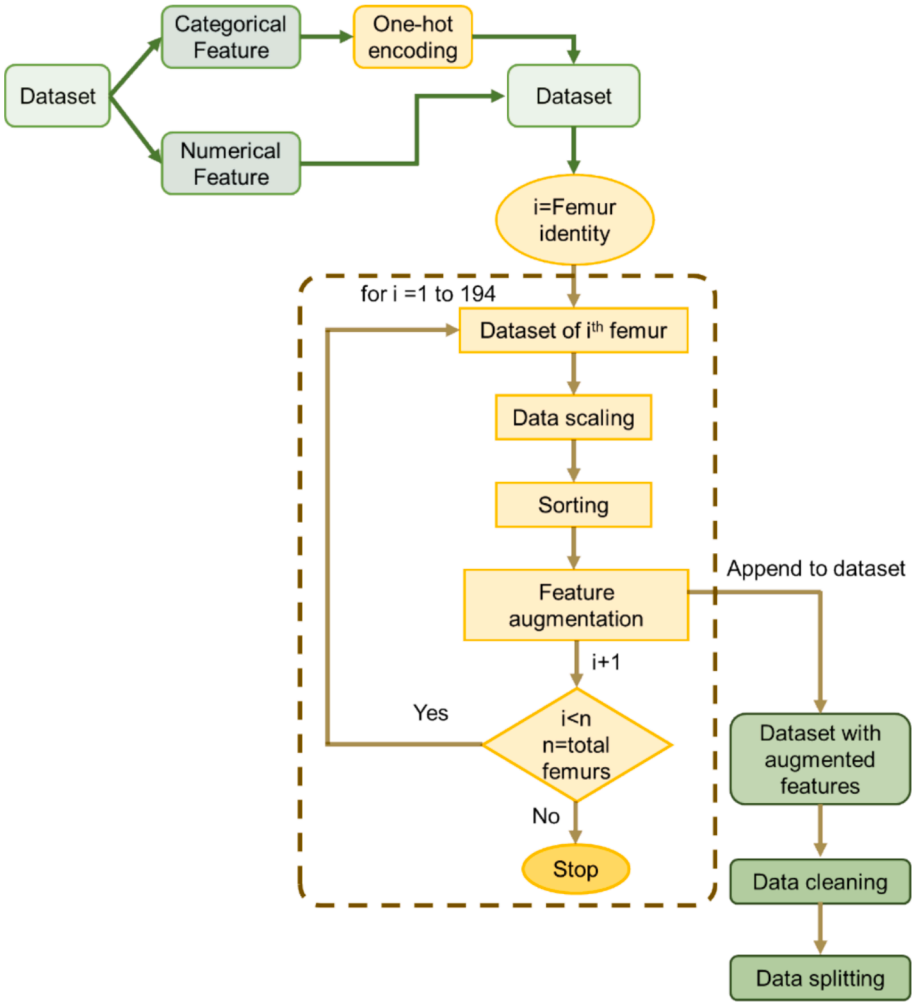


Fig. 7. Workflow of patient-specific feature engineering.

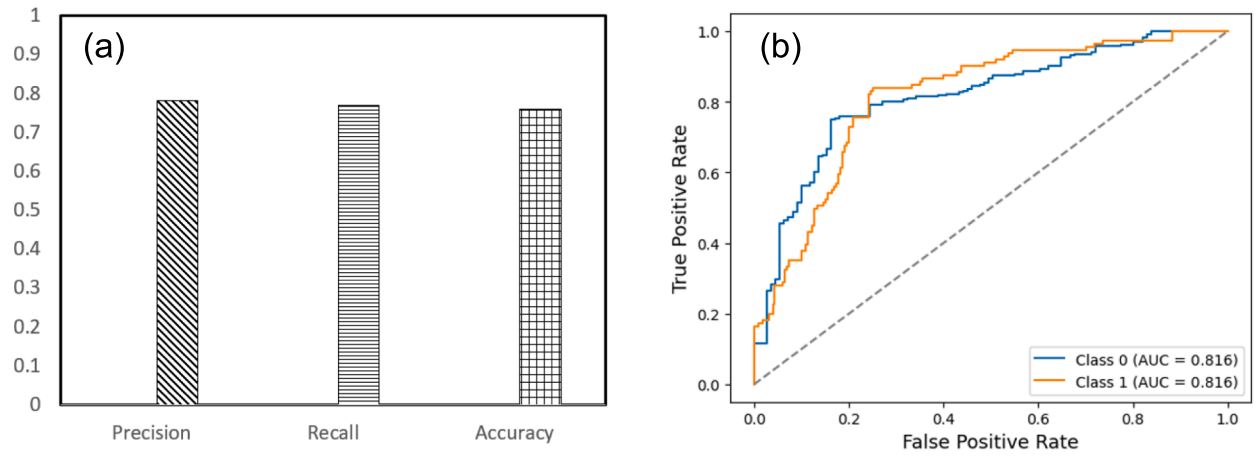


Fig. 8. Performance metrics (precision, recall, and accuracy) (a) and AUROC curve (b) of the CatBoost model in predicting hip fracture risk based on FRI. Class 0 and class 1 represent the possibility of no fracture and the possibility of fracture, respectively.

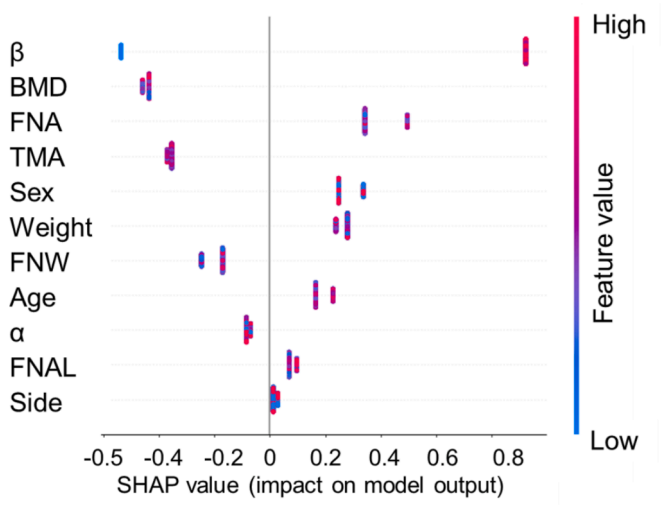


Fig. 9. The waterfall plot of the SHAP values to interpret the features that most contribute to the prediction of hip fracture risk. Red and blue colors mean higher and lower values of a feature, respectively.

Table 5
Mean *R −squared*, *MAE*, and *NMAE*, while predicting (visualizing) strain distributions in a femur bone. All results are shown in *mean* ± *std. dev*.

	R-squared	MAE	NMAE (%)
1st principal strain	0.73 ± 0.13	1.65E-4 ± 8.53E-5	4.14 ± 1.35
3rd principal strain	0.76 ± 0.13	2.85 ± 1.19E-4	2.86 ± 1.18

into a 0 to 1 range, considering the minimum and maximum values of each feature.

2.2.3. Model training and testing

To train the CatBoost model, k-fold cross-validation ($k = 5$) was used, and the model was iterated multiple times, tuning hyperparameters by a random search method until an optimized model was built.

2.2.4. Model (performance) evaluation

To evaluate the performance of the supervised CatBoost model, performance metrics such as *precision*, *recall* (*sensitivity*), *accuracy*, and *Area Under Receiver Operating Characteristics (AUROC)* curve were considered (Aldieri, et al., 2024; Erickson & Kitamura, 2021; Murphy, et al., 2022). In this study, possibility of fracture, when $FRI > 1$, is denoted as 1, and the possibility of no fracture, when $FRI \leqslant 1$, is denoted as 0. Considering QCT-based FE outcomes as the ground truth, the error metrics in this work are as follows,

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

where *True Positive* (*TP*) represents a true fracture risk correctly predicted as fracture risk, and *True Negative* (*TN*) represents the absence of fracture risk correctly predicted as no fracture risk by the ML model. *False Positive* (*FP*) indicates when the ML model falsely predicts risk, while there is no fracture risk in the ground truth, and *False Negative* (*FN*) denotes the ML model falsely predicting no fracture risk when ground truth indicates a fracture risk. Additionally, SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017) was used to explain the contributions of the features to the prediction results. It produces a plot showing the important features and their impacts (sensitivity) on the model's prediction.

2.3. Strain visualization via ML surrogate

2.3.1. Data preparation

The nodal coordinates (x-, y-, and z-coordinates) (Fig. 6) and associated bone material property–BMD (an average BMD value of the shared elements) was obtained for each node of the FE meshed models to accommodate the input features for predictively visualizing the strain distributions in the proximal femur. In addition, both magnitude of load and loading directions (α and β) were considered as input features in this workflow. Table 4 summarizes all the input features and target variables to surrogate strain distributions. For the predictive visualization of strain distributions to identify the probable fracture location in a femur, the nodal strain values obtained from the QCT-based FEA (described in Section 2.1) was considered as the target in this workflow.

2.3.2. Feature engineering

To improve the performance of the visualization, feature engineering was conducted to handle categorical variables using *one-hot encoding*, *data scaling and sorting*, *feature augmentation*, *data cleaning*, and *splitting* (Fig. 7).

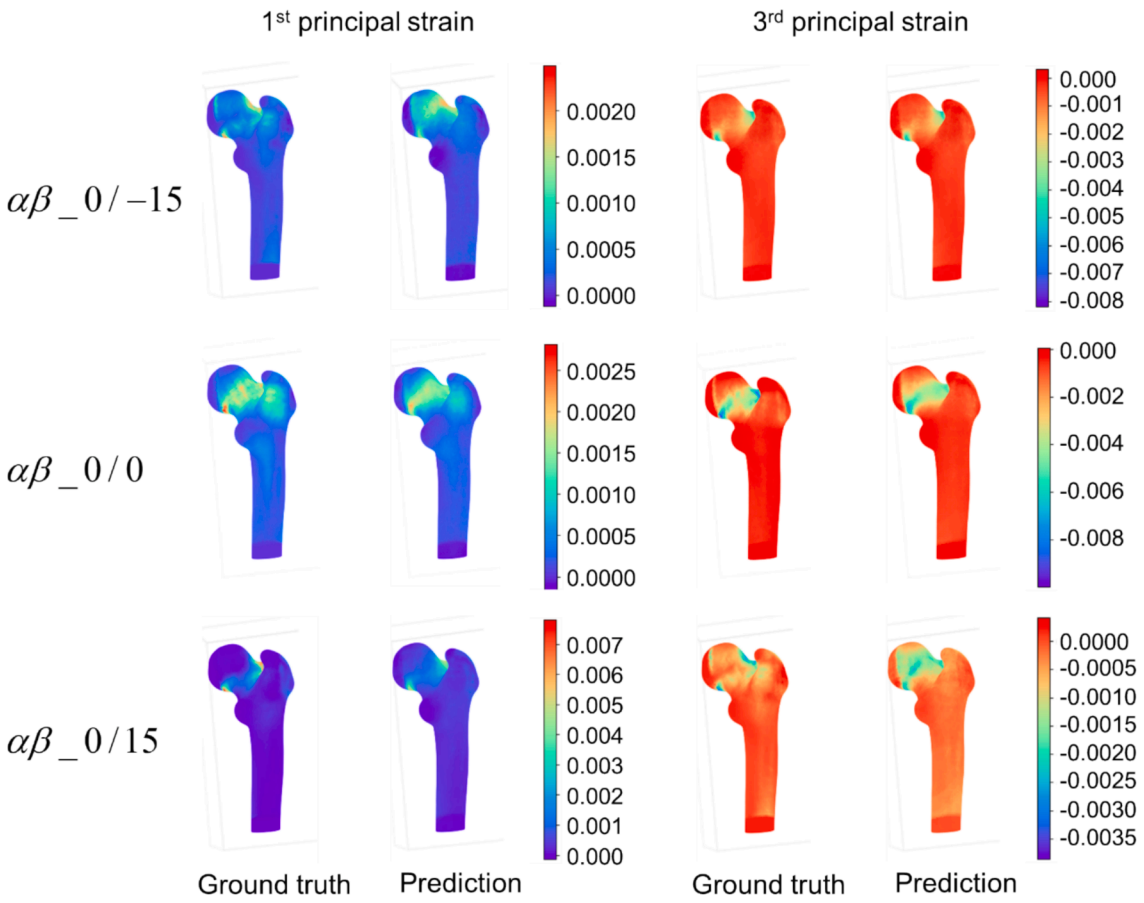


Fig. 10. Comparison of actual strain distributions obtained via QCT-based FEA (ground truth) and predicted strain from the adopted ML method, focusing on the orientation of the loading angle ($\alpha = 0^o$) on the coronal plane with respect to the shaft axis and the angle ($\beta = [-15^o, 0^o, 15^o]$) on the transverse plane with respect to the neck axis.

One-hot encoding

The categorical feature—loading directions in terms of angle—α and β, was converted into numerical values via one-hot encoding. One-hot encoding is considered to be an efficient way to handle categorical variables, leading to higher performance compared to binary and hash encoding (Seger, 2018).

Data scaling and sorting

Even though the dataset was initially organized in a tabular form, the data was further grouped patient-wise by assigning a unique identifier. Applying feature engineering on the complete dataset without making it patient-specific would lead to information sharing among the entire dataset comprising 97 patients, which could result in overfitting of the ML model. Hence, the features were normalized, considering one femur at a time, as shown in Fig. 7. After normalizing the initial feature vector, the feature set was sorted based on the nodal coordinates. The sorting was done sequentially on the z-axis, y-axis, and x-axis, respectively. Sorting the feature based on coordinates arranged them to start sequentially from the lowest point to the highest point of a femur.

Feature augmentation

To enhance performance, several features derived from the existing ones were added. New features, including the mean and standard deviation of a node's x-, y-, and z-coordinates, were incorporated into the feature vector. Furthermore, to analyze the trend, eliminate anomalies, and smooth the data, moving averages of the x-, y-, and z-coordinates, as well as stiffness, were calculated using a window size of 3. The moving standard deviation was also computed for these features. Femur-specific feature augmentation was applied to ensure that the coordinates and stiffness of one femur were not imputed onto another, thereby preventing overfitting of the model.

Data Cleaning

After feature augmentation, *Not-a-Number (NaN)* values, which represent undefined numbers in floating-point calculations, were removed from the dataset. A row containing a NaN value in any column was completely removed. It is important to note that rows containing NaN values comprised less than 0.05 % of the entire dataset.

Data splitting

The augmented and cleaned dataset was randomly split into training and testing datasets in the 80:20 patient ratio. 80% of the patients in training dataset was further divided into 5-fold cross-validation to optimize the model and eliminate data bias. After optimizing the model, the patients in testing dataset were used to test the model performance in visualizing the strain distributions on the unseen dataset.

2.3.3. Model training and testing

Training of the CatBoost model for strain visualization was performed using 5-fold cross-validation using random search method. An optimized CatBoost model was obtained after 7,000 iterations, maximum tree depth of 10, learning rate of 0.055 and *Root Mean Square Error (RMSE)* as a loss function. Hyperparameters were tuned to get an optimized model performance (Hu, et al., 2022; Kaviani, Han, & Sohn, 2022).

2.3.4. Model (performance) evaluation

After training, the CatBoost model was tested with patients from the testing dataset to measure the performance of the ML model on unseen data. After obtaining the optimized CatBoost model, the prediction was done by taking one femur at a time from the test dataset. Datapoint belonging to single femur was fed into optimized ML model and strain

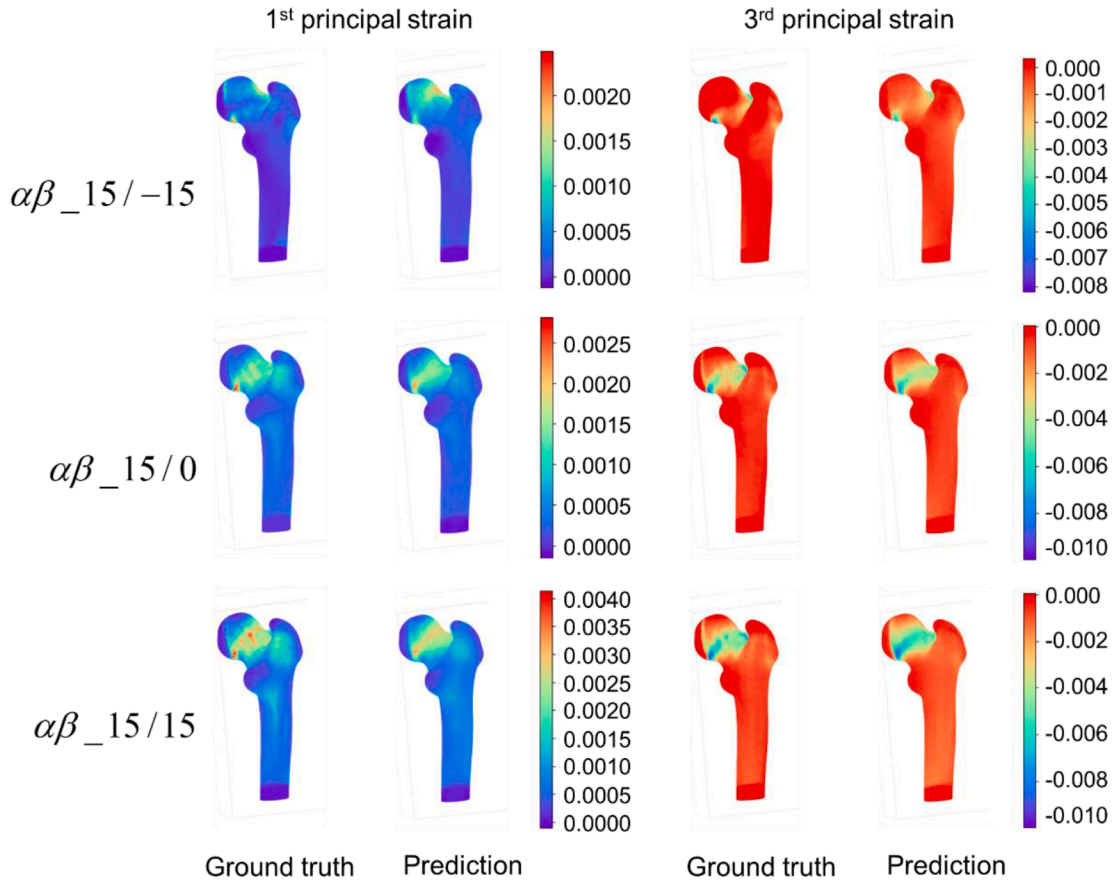


Fig. 11. Comparison of actual strain distributions obtained via QCT-based FEA (ground truth) and predicted strain from the adopted ML method, focusing on the orientation of the loading angle ($\alpha = 15^o$) on the coronal plane with respect to the shaft axis and the angle ($\beta = [-15^o, 0^o, 15^o]$) on the transverse plane with respect to the neck axis.

values were predicted. The ML model provided the 1st and 3rd principal strains along with *R-squared value*, *Mean Absolute Error (MAE)* and *Normalized Mean Absolute Error (NMAE)*. The prediction and evaluation were done for all patients in the test dataset, and an average was taken to analyze the model performance.

2.4. Testbed description

The testing environment for our experiments involved a cluster computer with an Intel ®Xeon ® E5-2600 v4 CPU running at 2.00 GHz, 1 Tesla GPU with 16 GB of RAM, and 256 GB of RAM. We chose to use the Python3 (Van Rossum & Drake, 2009) programming language due to its ease of use and the availability of relevant libraries. For the implementation of our model, several Python libraries were incorporated in this work. We used TensorFlow (Abadi, et al., 2016) and the SciPy (Virtanen, et al., 2020) library for scientific and technical computing. We also used Skimage, which is a collection of algorithms for image processing and computer vision, as well as the libraries associated with Google Colab, the cloud-based platform for data analysis and machine learning, and Visual Studio Code, which is widely used for Python development.

3. Results

3.1. Performance evaluation of hip fracture risk assessment

To evaluate the performance of the CatBoost model for predicting hip fracture risk based on FRI, the performance metrics—precision, recall, and accuracy—were determined. As shown in Fig. 8a, the precision, recall (sensitivity), and accuracy of the model are 0.78, 0.77, and 0.76, respectively. The model's performance indicates its ability to predict fracture risk with considerable accuracy as well as a low risk of missing actual hip fracture cases. To compare the overall performance of the model, AUROC curves were also depicted in Fig. 8b. It illustrates the balance between the true positive rate and the false positive rate anticipated by the model, aiding in the assessment of its performance. The higher AUROC value of 0.82 for CatBoost demonstrates its feasibility for fracture risk prediction.

While the performance metrics help us understand the overall performance of the adopted model, we need more insights into how different features impact the models' predictability. The SHAP plot (Fig. 9) shows the important features and their impacts on the CatBoost model's predictions, identifying β and BMD as the most important features in predicting fracture risk, followed by the other features.

3.2. Performance evaluation strain visualization

After predicting the fracture risk, a new CatBoost model was trained and optimized with 7,000 iterations, a learning rate of 0.055, a maximum tree depth of 10, and *RMSE* as the loss function. Table 5 shows the mean *R −squared*, *MAE*, and *NMAE*, while predicting strain distributions in a femur bone.

4D scatter plots were constructed to visualize the strain distributions in the proximal femur and to visually identify probable fracture locations. Figs. 10-12 compare the strain distributions in the proximal femur obtained via QCT-based FEA (ground truth) and those obtained via the supervised CatBoost model (prediction). It is evident that the strain visualization qualitatively exhibits a high correlation between the

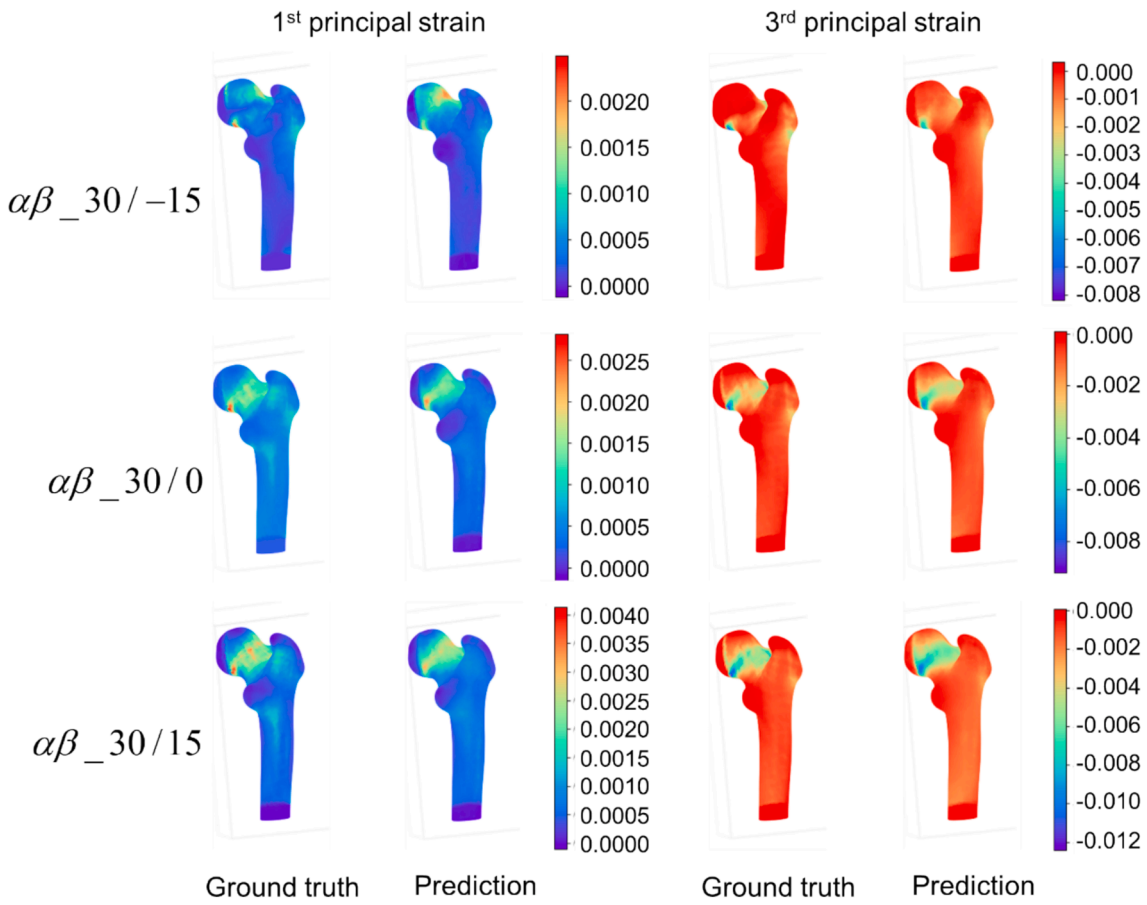


Fig. 12. Comparison of actual strain distributions obtained via QCT-based FEA (ground truth) and predicted strain from the adopted ML method, focusing on the orientation of the loading angle ($\alpha = 30^o$) on the coronal plane with respect to the shaft axis and the angle ($\beta = [-15^o, 0^o, 15^o]$) on the transverse plane with respect to the neck axis.

ground truth and the prediction.

To further evaluate the performance of the regression (CatBoost) model, data visualization was performed via scatter plot to compare the actual strain and predicted strain. Fig. 13 shows the comparison of actual vs. predicted 1st principal strain, whereas Fig. 14 shows similar plots for 3rd principal strain for nine different loading conditions. The moderately high coefficient of determinant (R^2) values show the model's ability to predict the strain close to the actual strain.

4. Discussion

This research focuses on developing ML-based FEA surrogate to assess subject-specific hip fracture risk due to sideways fall from standing height, which typically happened to geriatric people. The research objective was primarily motivated to elude the complex and time-consuming scientific computing process—QCT-based FEA as well as the requirement for expensive software. In this work, QCT-based FEA (Awal & Faisal, 2024) was used to construct the training dataset and obtain the ground truth, because scientific computation via classical computational models has proven its ability to develop accurate ground truth data, reducing the need for conducting experiments. In particular, there exists a high correlation between the QCT-based FEA and experiments, while evaluating femur strength (Black, et al., 2008; D. Dragomir-Daescu, 2011; Engelke, van Rietbergen, Zysset, & metabolism, 2016).

In this work, CatBoost model was applied separately to predict hip fracture risk and to visualize strain distributions due to the distinctive input features of the training datasets and different target variables (Table 3 and Table 4). FRI was considered the target variable to assess the hip fracture risk, whereas 1st (tensile) and 3rd (compressive) principal strains were the target variables for visually identifying the probable locations of fracture by generating an ML surrogate of strain distributions. A femur inherently exhibits brittle behavior that may be better represented by the maximum stress–strain criteria than by the magnitude of stress and strain (von Mises stress/strain criteria) (Cristofolini, et al., 2007; Doblaré, et al., 2004). Schileo et al. (Schileo, et al., 2008) and other studies (Altai, et al., 2019; Marco, et al., 2019; Testi, et al., 2002) showed that principal strain-based FEA analysis could correctly estimate fracture risk. Hence, in this study, we considered the maximum principal strain-based criteria to estimate the FRIs, which are defined based on the absolute maximum tensile and compressive strains (Bayraktar, et al., 2004; Schileo, et al., 2008). Furthermore, the adopted ML model was optimized using 5-fold cross-validation to ensure the model's robustness by reducing data bias. For both cases (fracture risk prediction and strain visualization), the performance of the optimized CatBoost model was tested with an unseen cohort of patients (testing dataset) that had never been introduced during the training phase. This signifies that the developed ML pipeline is robust, and we argue that the pipeline is able to work with any dataset with convincing performance.

While predicting the fracture risk, the precision exhibits the model's ability to predict the true fracture risk out of all predicted fracture risk, and the CatBoost model demonstrated good prediction ability in assessing fracture risk. The recall (sensitivity) value of the ML model suggests that the prediction of false negatives is low, indicating the model has a low chance of falsely predicting fracture risk. Accuracy was also obtained to ensure the overall performance of the ML model, with its value reflecting the prediction of true negative and true positive cases. The accuracy of CatBoost model implies its ability to correctly

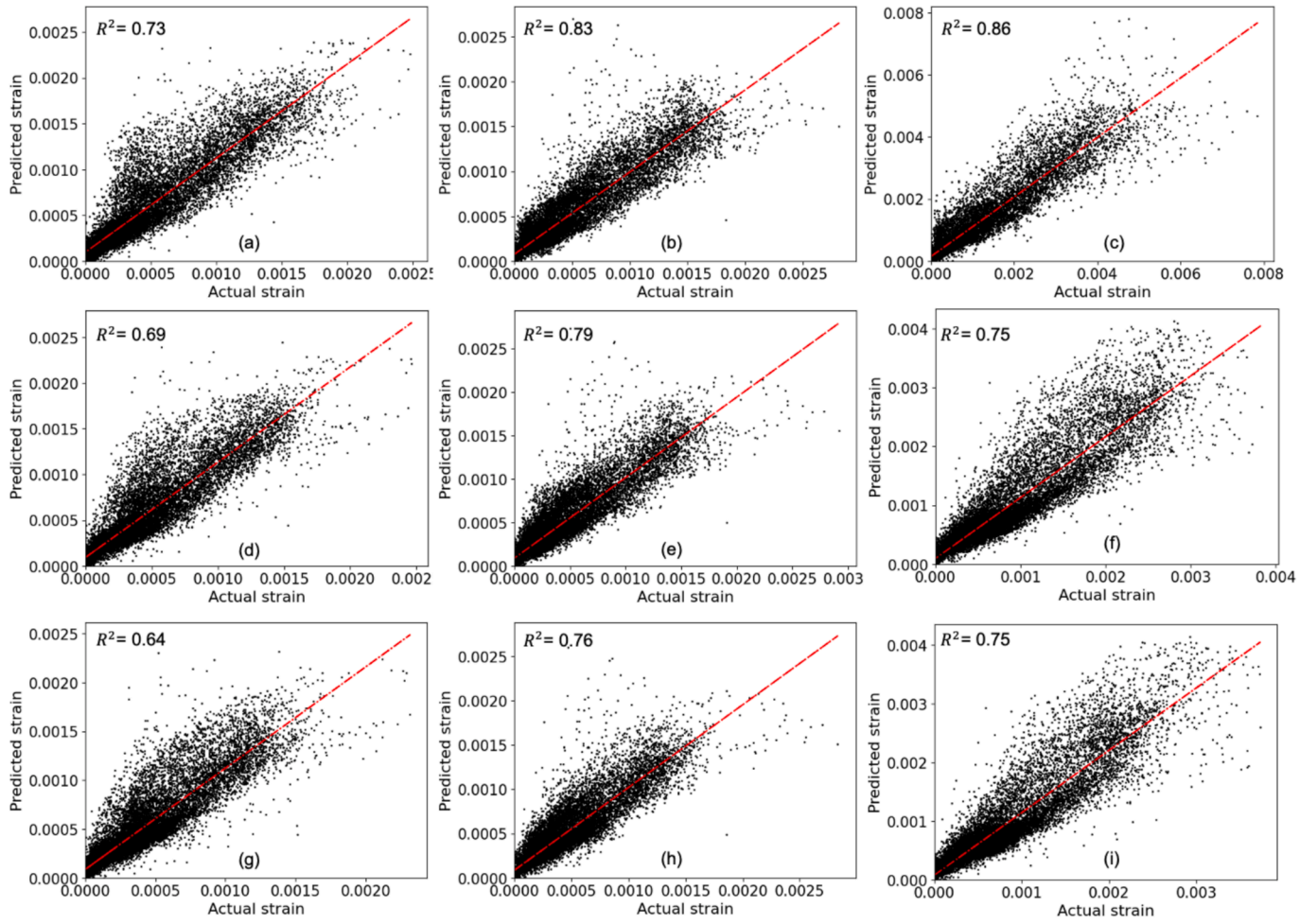


Fig. 13. Scatter plot comparing actual 1st principal strain obtained via QCT-based FEA (ground truth) vs predicted strain from the adopted ML method at nine different loading conditions: (a) αβ_0/-15, (b) αβ_0/0, (c) αβ_0/15, (d) αβ_15/-15, (e) αβ_15/0, (f) αβ_15/15, (g) αβ_30/-15, (h) αβ_30/0, (i) αβ_30/15.

predict 76% of instances of the possibility of fracture or no fracture. This observation is further supported by the AUROC curve (Fig. 8b). The AUC value of 0.81 indicates that this ML model is capable of distinguishing the probability of fracture (positive cases) from no fracture (negative cases). Therefore, this study suggests that CatBoost model trained with QCT-based FE results can be considered as an effective data-driven predictive tool for assessing the future possibility of hip fracture. It should be noted that, prior to selection, this CatBoost model was compared with other ML models (such as logistic regression, SVC, XGBM, RF, and DT) (Awal et al., 2025). The greater performance of the CatBoost model is likely attributed to its inherent ability to handle categorical variables like sex, α, and β.

SHAP plot was considered to analyze the effect of features or sensitivity of the features that impact fracture risk prediction the most (Fig. 9). The SHAP analysis of the CatBoost model demonstrates that the model learns significantly from the variable β, the loading angle/direction on the transverse plane with respect to the neck axis. Additionally, the global interpretability of SHAP values also indicates whether a feature has a positive or negative influence on predictions. Accordingly, the positive impact of high β value and negative impact of low β value show that increasing the angle of rotation increases the hip fracture risk.

The primary objective of strain visualization is to visually identify the fracture-prone region. A qualitative (visual) comparison between the ground truth (obtained via QCT-based FEA) and predicted strain distributions (Figs. 10-12) demonstrates that our optimized CatBoost model is able to convincingly predict the strain distributions (visually) with an acceptable accuracy. This will further help identify the probable fracture location along with fracture risk assessment and enable informed planning for taking preventive measures to prevent hip fracture if someone falls.

Furthermore, we also used performance metrics (Table 5) to quantitatively assess the performance of strain visualization. Accordingly, the $R-squared$ values of 0.73 and 0.76, while predicting 1st and 3rd principal strain, respectively, resembles the moderate performance of the CatBoost model in predicting strain distributions within a 3D proximal femur. The median $R-squared$ values of 0.76 for 1st principal strain and 0.81 for 3rd principal strain delineate that the strain distributions are accurately surrogated for most of the femurs of unseen patients. The minimal mean errors between the actual and predicted principal strains (Table 5) demonstrate that the predicted strain values are close to the actual values. Additionally, less than 5 % of mean *NMAE* describes an overall high reliability on predicted strain values. However, the objective of showing the loss function is to quantitively evaluate the performance of strain visualization, even though this error metric may not truly reflect the primary objective of strain visualization. Nevertheless, high $R-squared$, low MAE and NMAE values, and the qualitative similarities in strain distributions (visualization) between the predicted and actual strain distributions delineate that an optimized CatBoost model can generate a surrogate model of QCT-based FEA in assessing hip fracture risk. Similar interpretation can be perceived from the scatter plots obtained from the actual and predicted 1st and 3rd principal strains. Points clustered tightly around the diagonal line indicate a strong correlation between predictions and actual values, suggesting good model performance. However, a low coefficient of determination on predicting

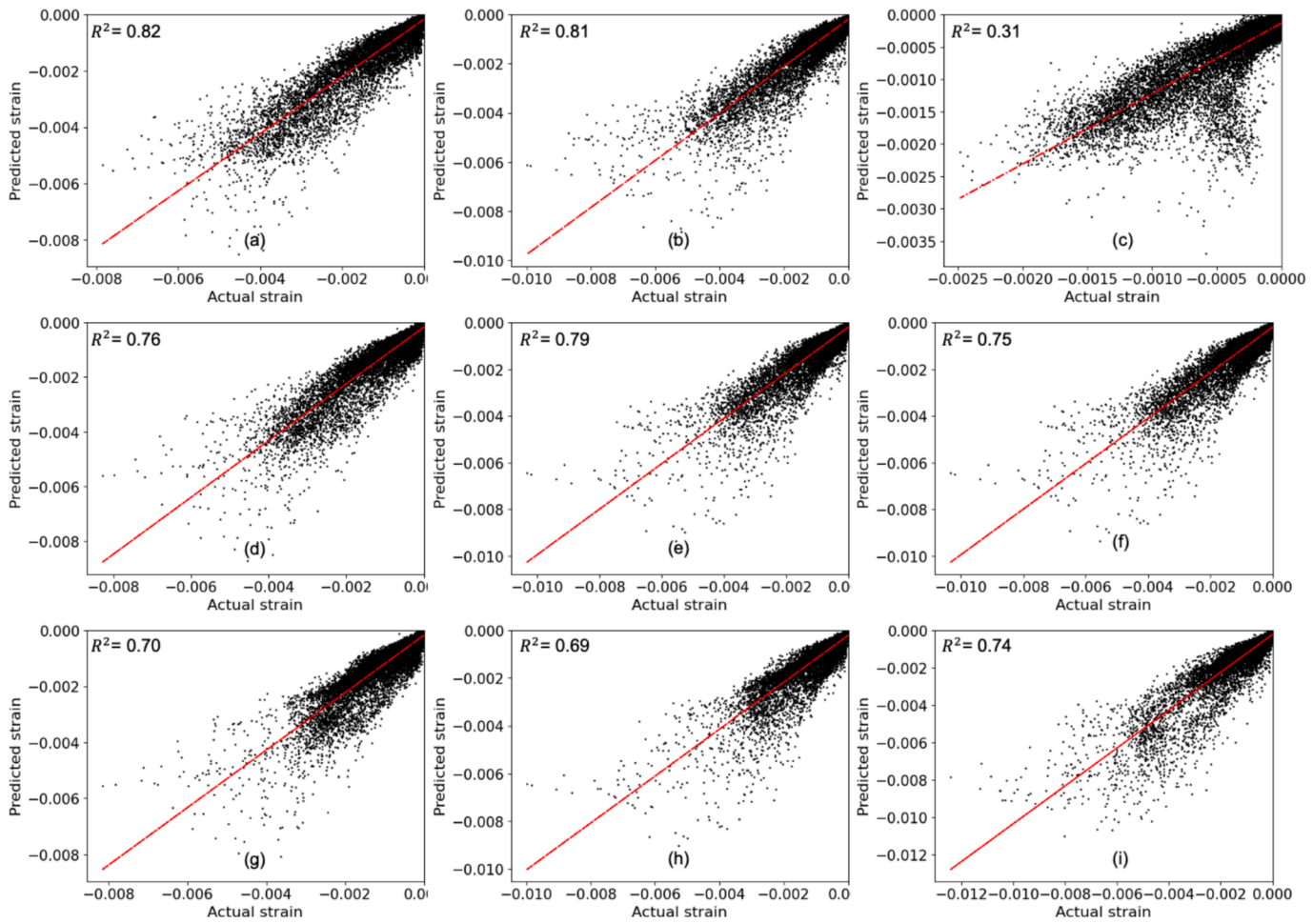


Fig. 14. Scatter plot comparing actual 3rd principal strain obtained via QCT-based FEA (ground truth) vs predicted strain from the adopted ML method at nine different loading conditions: (a) αβ_0/-15, (b) αβ_0/0, (c) αβ_0/15, (d) αβ_15/-15, (e) αβ_15/0, (f) αβ_15/15, (g) αβ_30/-15, (h) αβ_30/0, (i) αβ_30/15.

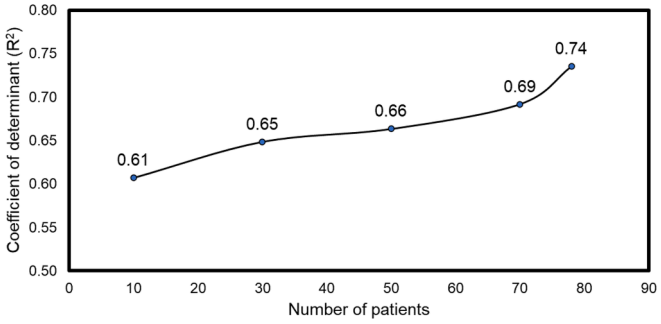


Fig. 15. Correlation of data size (number of patients) and the performance of the CatBoost ML model.

3rd principal strain at loading condition αβ_0/15 showing ML model inability to predict strain to outliers. However, it could be minimized by training ML models with a larger dataset and with a wide range of strain distributions.

Assessing hip fracture risk considering stress- or strain-dependent FRI by QCT-based FEA is a time-consuming and tedious process that typically requires highly skilled human resources and expensive software, thereby limiting the application of this physics-based modeling in clinical and pre-clinical domains. Data-driven ML-based predictive modeling can be a potential alternative to FEA without compromising accuracy, which can be significantly enhanced by adding variability in patient dataset, input features, and increasing data size (big data) in the training dataset. Performance analysis of the CatBoost model with varying the size of training dataset clearly substantiates the fact that increasing the number of patients in the training dataset will further improve the performance of the ML model (Fig. 15).

The proposed data-driven framework exhibits encouraging performance but certainly with some limitations. The ML-driven framework was built upon data obtained from one geographical region, which limits the model's versatility. For robustness, the model needs to be trained with data from different ethnicities and geographical locations. The current dataset may not truly reflect the elderly population as it includes patients well below 65 years of age. A dataset of older patients aged 65 and above may train the ML model more appropriately for the population group. However, the inclusion of younger adults increases the variability, which might be beneficial for developing a generalized ML model. Most importantly, a dataset with clinically diagnosed osteoporosis or information on prior fractures will increase the efficiency of model training. The major limitation of this work is the smaller dataset, which restricts achieving higher accuracy. Data augmentation could solve the limitation of limited dataset, increasing the accuracy of model. However, data augmention was avoided in this study considering that synthetic data may not represent true fall case scenario. On predicting strain distributions, an ML model trained on dense point cloud could increase the accuracy but could cost more computational resource for training and testing the model.

In conclusion, we have presented an ML-based pipeline integrated with QCT-based FEA (a classical computational modeling tool) for

predicting hip fracture risk and visualizing strain as a surrogate FE model to identify the probable fracture location. The sequential process of QCT-based FEA typically involves 3D femur reconstruction using high-end medical image processing software and commercial FE solver, making the process expensive and time-consuming. Therefore, computational models are often limited in clinical set up. Hence, this proposed study shows the potential of ML-based modeling to predict the risk of hip fracture, bypassing all complicated steps involved in traditional computational modeling approaches. ML models have been widely used in a variety of biomedical and clinical applications including cancer screening, genomics classification, DNA sequencing, and structural analysis of proteins (Bartoszewicz, Seidel, Rentzsch, & Renard, 2020; Yue Cao, Geddes, Yang, & Yang, 2020; Yongchun Cao, et al., 2023; Che, Liu, Rasheed, & Tao, 2011; Hu, et al., 2022; Lorente, et al., 2017; Tan & Gilbert, 2003), but they have been very limitedly applied in bone fracture. The convincing performance of CatBoost model with such limited dataset shows an enormous feasibility of ML-based modeling in predicting hip fracture risk accurately in a cost-effective way. Lastly, the viability of ML models for predicting hip fracture risk and visualization suggests the enormous potential of ML-based modeling as an alternative to complex and expensive QCT-based FEA.

Funding

No funding was available for this study.

Declaration of competing interest

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

Acknowledgments

The authors greatly acknowledge undergraduate research student, Sarah Doll of Mechanical Engineering at UL Lafayette for assisting in 3D femur reconstruction. The authors also acknowledge Dr. Raju Gottumukkla, Associate Professor and Director of Research, Informatics Research Institute, UL Lafayette, for providing access to the high performance computing cluster.

Data availability

The datasets presented during the current study are not publicly available due to privacy and ethical restrictions but might be available on reasonable request from the corresponding author.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., & Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Adams, J. E. (2013). Advances in bone imaging for osteoporosis. *Nature Reviews Endocrinology, 9*, 28.

Albertsson, D., Mellström, D., Petersson, C., Thulesius, H., & Eggertsen, R. (2010). Hip and fragility fracture prediction by 4-item clinical risk score and mobile heel BMD: A women cohort study. *BMC musculoskeletal disorders, 11*, 55.

Aldieri, A., Curreli, C., Szyszko, J. A., La Mattina, A. A., & Viceconti, M. (2023). Credibility assessment of computational models according to ASME V&V40: Application to the Bologna biomechanical computed tomography solution. *Computer Methods and Programs in Biomedicine, 240*, Article 107727.

Aldieri, A., Paggiosi, M., Eastell, R., Bignardi, C., Audenino, A. L., Bhattacharya, P., & Terzini, M. (2024). DXA-based statistical models of shape and intensity outperform aBMD hip fracture prediction: A retrospective study. *Bone, 182*, Article 117051.

Aldieri, A., Terzini, M., Audenino, A. L., Bignardi, C., Paggiosi, M., Eastell, R., Viceconti, M., & Bhattacharya, P. (2022). Personalised 3D assessment of trochanteric soft tissues improves HIP fracture classification accuracy. *Annals of Biomedical Engineering, 50*, 303–313.

Altai, Z., Qasim, M., Li, X., & Viceconti, M. (2019). The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture. *Clinical Biomechanics, 68*, 137–143.

Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., Gomes, A. S., Folsom, A. R., Shea, S., & Guallar, E. (2017). Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis. *Circulation research, 121*, 1092–1101.

Aschkenasy, M. T., & Rothenhaus, T. C. (2006). Trauma and falls in the elderly. *Emergency Medicine Clinics, 24*, 413–432.

Awal, R., Ben Hmida, J., Luo, Y., & Faisal, T. (2022). Study of the significance of parameters and their interaction on assessing femoral fracture risk by quantitative statistical analysis. *Medical & Biological Engineering & Computing, 60*, 843–854.

Awal, R., & Faisal, T. (2024). QCT-based 3D finite element modeling to assess patient-specific hip fracture risk and risk factors. *Journal of the Mechanical Behavior of Biomedical Materials, 150*, Article 106299.

Awal, R., Doll, S. C., Naznin, M., & Faisal, T. R. (2025). Interpretable machine learning classifiers for the reliable prediction of fall induced hip fracture risk. *Machine Learning for Computational Science and Engineering, 1*, 2.

Awal, R., & Faisal, T. R. (2021). Multiple Regression Analysis of Hip Fracture Risk Assessment Via Finite Element Analysis. *Journal of Engineering and Science in Medical Diagnostics and Therapy, 4*, Article 011006.

Bartoszewicz, J. M., Seidel, A., Rentzsch, R., & Renard, B. Y. (2020). DeePaC: Predicting pathogenic potential of novel DNA with reverse-complement neural networks. *Bioinformatics, 36*, 81–89.

Bayraktar, H. H., Morgan, E. F., Niebur, G. L., Morris, G. E., Wong, E. K., & Keaveny, T. M. (2004). Comparison of the elastic and yield properties of human femoral trabecular and cortical bone tissue. *Journal of biomechanics, 37*, 27–35.

Bessho, M., Ohnishi, I., Okazaki, H., Sato, W., Kominami, H., Matsunaga, S., & Nakamura, K. (2004). Prediction of the strength and fracture location of the femoral neck by CT-based finite-element method: A preliminary study on patients with hip fracture. *Journal of Orthopaedic science, 9*, 545–550.

Bettamer, A. (2012). prediction of proximal femur fracture: finite element modeling based on mechanical damage and experimental validation.

Black, D. M., Bouxsein, M. L., Marshall, L. M., Cummings, S. R., Lang, T. F., Cauley, J. A., Ensrud, K. E., Nielson, C. M., Orwoll, E. S. J. J. o. B., & Research, M. (2008). *Proximal femoral structure and the prediction of hip fracture in men: a large prospective study using QCT., 23*, 1326–1333.

Cao, Y., Geddes, T. A., Yang, J. Y. H., & Yang, P. (2020). Ensemble deep learning in bioinformatics. *Nature Machine Intelligence, 2*, 500–508.

Cao, Y., Liu, L., Chen, X., Man, Z., Lin, Q., Zeng, X., & Huang, X. (2023). Segmentation of lung cancer-caused metastatic lesions in bone scan images using self-defined model with deep supervision. *Biomedical Signal Processing and Control, 79*, Article 104068.

Che, D., Liu, Q., Rasheed, K., & Tao, X. (2011). Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Adv Exp Med Biol, 696*, 191–199.

Cooper, C., Cole, Z., Holroyd, C., Earl, S., Harvey, N. C., Dennison, E. M., Melton, L., Cummings, S. R., Kanis, J. A., Epidemiology, I. C. W. G., & o. F.. (2011). Secular trends in the incidence of hip and other osteoporotic fractures. *Osteoporosis international, 22*, 1277–1288.

Cristofolini, L., Juszczyk, M., Martelli, S., Taddei, F., & Viceconti, M. (2007). In vitro replication of spontaneous fractures of the proximal human femur. *Journal of biomechanics, 40*, 2837–2845.

Dragomir-Daescu, D., & J. O. D. B., S. McEligot, Y. Dai, R. C. Entwistle, C. Salas, L. J. Melton, K. E. Bennet, S. Khosla, and S. Amin.. (2011). Robust QCT/FEA models of proximal femur stiffness and fracture load during a sideways fall on the hip. *Annals of biomedical engineering, 39*, 742–755.

Do, B. H., Langlotz, C., & Beaulieu, C. F. (2017). Bone tumor diagnosis using a naïve Bayesian model of demographic and radiographic features. *Journal of digital imaging, 30*, 640–647.

Doblaré, M., Garcıa, J., & Gómez, M. (2004). Modelling bone tissue fracture and healing: A review. *Engineering Fracture Mechanics, 71*, 1809–1840.

Dragomir-Daescu, D., Op Den Buijs, J., McEligot, S., Dai, Y., Entwistle, R. C., Salas, C., Melton, L. J., Bennet, K. E., Khosla, S., & Amin, S. (2011). Robust QCT/FEA models of proximal femur stiffness and fracture load during a sideways fall on the hip. *Annals of biomedical engineering, 39*, 742–755.

Engelke, K., van Rietbergen, B., & Zysset, P. (2016). FEA to measure bone strength: A review. *Clinical reviews in bone and mineral metabolism, 14*, 26–37.

Engelke, K., van Rietbergen, B., Zysset, P. J. C., & r. i. b., & metabolism, m.. (2016). *FEA to measure bone strength: a review., 14*, 26–37.

Erickson, B. J., & Kitamura, F. (2021). Magician's corner: 9. Performance metrics for machine learning models. In (Vol. 3, pp. e200126): Radiological Society of North America.

Eshghi, N., Hojjati, M., Imani, M., & Goudarzi, A. (2011). Finite element analysis of mechanical behaviors of coronary stent. *Procedia Engineering, 10*, 3056–3061.

Faisal, T. R., & Luo, Y. (2016). Study of stress variations in single-stance and sideways fall using image-based finite element analysis. *Bio-medical materials and engineering, 27*, 1–14.

Faisal, T. R., & Luo, Y. (2017). Study of the variations of fall induced hip fracture risk between right and left femurs using CT-based FEA. *BioMedical Engineering OnLine, 16*, 116.

Fajar, J. K., Taufan, T., Syarif, M., & Azharuddin, A. (2018). Hip geometry and femoral neck fractures: A meta-analysis. *Journal of orthopaedic translation, 13*, 1–6.

Ferizi, U., Besser, H., Hysi, P., Jacobs, J., Rajapakse, C. S., Chen, C., Saha, P. K., Honig, S., & Chang, G. (2019). Artificial intelligence applied to osteoporosis: A performance comparison of machine learning algorithms in predicting fragility fractures from MRI data. *Journal of Magnetic Resonance Imaging, 49*, 1029–1038.

Ford, C. M., Keaveny, T. M., & Hayes, W. C. (1996). The effect of impact direction on the structural capacity of the proximal femur during falls. *J. Bone Miner. Res., 11*.

Gislason, M. K., Ingvarsson, P., Gargiulo, P., Yngvason, S., Guðmundsdóttir, V., Knútsdóttir, S., & Helgason, Þ. (2014). Finite element modelling of the femur bone of a subject suffering from motor neuron lesion subjected to electrical stimulation. European journal of translational myology, 24.

Grassi, L., Schileo, E., Taddei, F., Zani, L., Juszczyk, M., Cristofolini, L., & Viceconti, M. (2012). Accuracy of finite element predictions in sideways load configurations for the proximal human femur. *Journal of biomechanics, 45*, 394–399.

Grassi, L., Väänänen, S. P., Jehpsson, L., Ljunggren, Ö., Rosengren, B. E., Karlsson, M. K., & Isaksson, H. (2023). 3D Finite Element Models Reconstructed From 2D Dual-Energy X-Ray Absorptiometry (DXA) Images Improve Hip Fracture Prediction Compared to Areal BMD in Osteoporotic Fractures in Men (MrOS) Sweden Cohort. *Journal of Bone and Mineral Research, 38*, 1258–1267.

Hu, Y., Zhao, L., Li, Z., Dong, X., Xu, T., & Zhao, Y. (2022). Classifying the multi-omics data of gastric cancer using a deep feature selection method. *Expert Systems with Applications, 200*, Article 116813.

Iliou, T., Anagnostopoulos, C.-N., & Anastassopoulos, G. (2014). Osteoporosis detection using machine learning techniques and feature selection. *International Journal on Artificial Intelligence Tools, 23*.

Islam, M. R., & Nahiduzzaman, M. (2022). Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach. *Expert Systems with Applications, 195*, Article 116554.

Johnell, O., & Kanis, J. (2004). An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. *Osteoporosis International, 15*, 897–902.

Juszczyk, M. M., Cristofolini, L., & Viceconti, M. (2011). The human proximal femur behaves linearly elastic up to failure under physiological loading conditions. *Journal of biomechanics, 44*, 2259–2266.

Kanis, J., Johnell, O., Odén, A., Johansson, H., & McCloskey, E. (2008). FRAX™ and the assessment of fracture probability in men and women from the UK. *Osteoporosis international, 19*, 385–397.

Kannus, P., Parkkari, J., Sievänen, H., Heinonen, A., Vuori, I., & Järvinen, M. (1996). Epidemiology of hip fractures. *Bone, 18*, S57–S63.

Kaviani, S., Han, K. J., & Sohn, I. (2022). Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications, 198*, Article 116815.

Kheirollahi, H., & Luo, Y. (2015). Assessment of hip fracture risk using cross-section strain energy determined by QCT-based finite element modeling. BioMed research international, 2015.

Kheirollahi, H., & Luo, Y. (2015). Identification of high stress and strain regions in proximal femur during single-leg stance and sideways fall using QCT-based finite element model. *International Journal Biomedical and Biological Engineering, 9*, 541–548.

Kim, D., & MacKinnon, T. (2018). Artificial intelligence in fracture detection: Transfer learning from deep convolutional neural networks. *Clinical radiology, 73*, 439–445.

Kong, S. H., Ahn, D., Kim, B., Srinivasan, K., Ram, S., Kim, H., Hong, A. R., Kim, J. H., Cho, N. H., & Shin, C. S. (2020). A novel fracture prediction model using machine learning in a community-based cohort. *JBMR plus, 4*.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal, 13*, 8–17.

Krogue, J. D., Cheng, K. V., Hwang, K. M., Toogood, P., Meinberg, E. G., Geiger, E. J., Zaid, M., McGill, K. C., Patel, R., & Sohn, J. H. (2020). Automatic hip fracture identification and functional subclassification with deep learning. Radiology. *Artificial Intelligence, 2*.

Liang, L., Liu, M., Martin, C., & Sun, W. (2018). A deep learning approach to estimate stress distribution: A fast and accurate surrogate of finite-element analysis. *Journal of The Royal Society Interface, 15*.

Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., & Hotchkiss, R. (2018). Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences, 115*, 11591–11596.

Little, M. A., Varoquaux, G., Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2017). Using and understanding cross-validation strategies. Perspectives on Saeb et al. Gigascience, 6, 1-6.

Liu, Q., Cui, X., Chou, Y.-C., Abbod, M. F., Lin, J., & Shieh, J.-S. (2015). Ensemble artificial neural networks applied to predict the key risk factors of hip bone fracture for elders. *Biomedical Signal Processing and Control, 21*, 146–156.

Lorente, D., Martínez-Martínez, F., Rupérez, M. J., Lago, M., Martínez-Sober, M., Escandell-Montero, P., Martínez-Martínez, J. M., Martínez-Sanchis, S., Serrano-López, A. J., & Monserrat, C. (2017). A framework for modelling the biomechanical behaviour of the human liver during breathing in real time using machine learning. *Expert Systems with Applications, 71*, 342–357.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (p. 30).

Luo, Y., Ferdous, Z., & Leslie, W. (2011). A preliminary dual-energy X-ray absorptiometry-based finite element model for assessing osteoporotic hip fracture risk. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 225*, 1188–1195.

Madani, A., Bakhaty, A., Kim, J., Mubarak, Y., & Mofrad, M. R. (2019). Bridging finite element and machine learning modeling: Stress prediction of arterial walls in atherosclerosis. *Journal of biomechanical engineering, 141*, Article 084502.

Marco, M., Giner, E., Caeiro-Rey, J. R., Miguélez, M. H., & Larraínzar-Garijo, R. (2019). Numerical modelling of hip fracture patterns in human femur. *Computer methods and programs in biomedicine, 173*, 67–75.

McCulloch, A., Guccione, J., Waldman, L., & Rogers, J. (2020). Large-scale finite element analysis of the beating heart. *High-performance computing in biomedical research*, 27–49.

Memiş, A., Varlı, S., & Bilgili, F. (2022). Image based quantification of the proximal femur shape deformities in 3D by using the contralateral healthy shape structure: A preliminary study. *Biomedical Signal Processing and Control, 71*, Article 103079.

Monteiro, J. B., Dal Piva, A. M., & d. O., Tribst, J. P. M., Borges, A. L. S., & Tango, R. N.. (2018). The effect of resection angle on stress distribution after root-end surgery. *Iranian Endodontic Journal, 13*, 188.

Murphy, E., Ehrhardt, B., Gregson, C. L., von Arx, O., Hartley, A., Whitehouse, M., Thomas, M., Stenhouse, G., Chesser, T., & Budd, C. (2022). Machine learning outperforms clinical experts in classification of hip fractures. *Scientific Reports, 12*, 2058.

Nishiyama, K. K., Gilchrist, S., Guy, P., Cripton, P., & Boyd, S. K. (2013). Proximal femur bone strength estimated by a computationally fast finite element analysis in a sideways fall configuration. *Journal of biomechanics, 46*, 1231–1236.

Olczak, J., Fahlberg, N., Maki, A., Razavian, A. S., Jilert, A., Stark, A., Sköldenberg, O., & Gordon, M. (2017). Artificial intelligence for analyzing orthopedic trauma radiographs: Deep learning algorithms—are they on par with humans for diagnosing fractures? *Acta orthopaedica, 88*, 581–586.

Osareh, A., & Shadgar, B. (2010). In *Machine learning techniques to diagnose breast cancer* (pp. 114–120). IEEE.

Pinilla, T. P., Boardman, K. C., Bouxsein, M. L., Myers, E. R., & Hayes, W. C. (1996). Impact direction from a fall influences the failure load of the proximal femur as much as age-related bone loss. *Calcified Tissue International, 58*, 231–235.

Post, A., Kendall, M., Koncan, D., Cournoyer, J., Hoshizaki, T. B., Gilchrist, M. D., Brien, S., Cusimano, M. D., & Marshall, S. (2015). Characterization of persistent concussive syndrome using injury reconstruction and finite element modelling. *Journal of the Mechanical Behavior of Biomedical Materials, 41*, 325–335.

Robinovitch, S., Hayes, W. C., & McMahon, T. (1991). Prediction of femoral impact forces in falls on the hip.

Rui Zhang, H. G., Fang, J., & Zhu, Z. G. A. D. (2014). Prediction Of Proximal Femoral Fracture In Sideways Falls Using Nonlinear Dynamic Finite Element Analysis. *Journal of Mechanics in Medicine and Biology, 14*.

Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of computational science, 30*, 174–182.

Schileo, E., Taddei, F., Cristofolini, L., & Viceconti, M. (2008). Subject-specific finite element models implementing a maximum principal strain criterion are able to estimate failure risk and fracture location on human femurs tested in vitro. *Journal of biomechanics, 41*, 356–367.

Seger, C. (2018). An investigation of categorical variable encoding techniques in machine learning: Binary versus one-hot and feature hashing. *In.*

Singh, S., Mogra, S., Shetty, V. S., Shetty, S., & Philip, P. (2012). Three-dimensional finite element analysis of strength, stability, and stress distribution in orthodontic anchorage: A conical, self-drilling miniscrew implant system. *American Journal of Orthodontics and Dentofacial Orthopedics, 141*, 327–336.

Sultana, J., Naznin, M., & Faisal, T. R. (2024). SSDL-an automated semi-supervised deep learning approach for patient-specific 3D reconstruction of proximal femur from QCT images. *Medical & Biological Engineering & Computing*.

Taddei, F., Pancanti, A., & Viceconti, M. (2004). An improved method for the automatic mapping of computed tomography numbers onto finite element models. *Medical engineering & physics, 26*, 61–69.

Taddei, F., Schileo, E., Helgason, B., Cristofolini, L., & Viceconti, M. (2007). The material mapping strategy influences the accuracy of CT-based finite element models of bones: An evaluation against experimental measurements. *Medical engineering & physics, 29*, 973–979.

Tan, A. C., & Gilbert, D. (2003). Ensemble machine learning on gene expression data for cancer classification.

Testi, D., Viceconti, M., Cappello, A., & Gnudi, S. (2002). Prediction of hip fracture can be significantly improved by a single biomedical indicator. *Annals of Biomedical Engineering, 30*, 801–807.

Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P., & Saarakkala, S. (2018). Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. Sci Rep 2018; 8 (1): 1727. In.

Urakawa, T., Tanaka, Y., Goto, S., Matsuzawa, H., Watanabe, K., & Endo, N. (2019). Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal radiology, 48*, 239–244.

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual: (Python Documentation Manual Part 2): CreateSpace Independent Publishing Platform.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature methods, 17*, 261–272.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one, 12*.

Xue, Y., Zhang, R., Deng, Y., Chen, K., & Jiang, T. (2017). A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLOS ONE, 12*.

Yang, L., Palermo, L., Black, D. M., & Eastell, R. (2014). Prediction of incident hip fracture with the estimated femoral strength by finite element analysis of DXA scans in the study of osteoporotic fractures. *Journal of Bone and Mineral Research, 29*, 2594–2600.

Yano, S., Matsuura, Y., Hagiwara, S., Nakamura, J., Kawarai, Y., Suzuki, T., Kanno, K., Shoda, J., Tsurumi, Y., & Ohtori, S. (2022). Determinants of fracture type in the

proximal femur: Biomechanical study of fresh frozen cadavers and finite element models. *Bone, 158*, Article 116352.
Yoshikawa, T., Turner, C., Peacock, M., Slemenda, C., Weaver, C., Teegarden, D., Markwardt, P., & Burr, D. (1994). Geometric structure of the femoral neck measured using dual-energy X-ray absorptiometry. *Journal of Bone and Mineral Research, 9*, 1053–1064.
Zannoni, C., Mantovani, R., & Viceconti, M. (1999). Material properties assignment to finite element models of bone structures: A new method. *Medical Engineering & Physics, 20*, 735–740.