Hip fracture risk assessment in elderly and diabetic patients: combining autonomous finite element analysis and machine learning

Zohar Yosibash[1,3,*], Nir Trabelsi[2,3,*], Itay Buchnik[4], Kent W Myers[3], Moshe Salai[5], Iris Eshed[6,7], Yiftach Barash[6,7], Eyal Klang[6,7] , Liana Tripto-Shkolnik[7, 8]

[1] School of Mechanical Engineering, The Iby and Aladar Fleischman Faculty of Eng, Tel-Aviv University, Ramat Aviv, Tel-Aviv, Israel
[2] Dept. of Mechanical Engineering, Shamoon College of Eng, Beer-Sheva, Israel
[3] PerSimiO Ltd, Beer-Sheva, Israel
[4] Dept. of Electrical and Computer Engineering, Ben Gurion University, Beer-Sheva, Israel
[5] Orthopedic Dept, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
[6] Department of Diagnostic Imaging, Sheba Medical Center, Tel Hashomer, Israel
[7] Sackler School of Medicine, Tel-Aviv University, Tel Aviv, Israel
[8] Division of Endocrinology, Diabetes and Metabolism, Sheba Medical Center, Tel Hashomer, Israel

Running Title: Hip fracture risk assessment

ORCID number of Zohar Yosibash: 0000-0002-0826-2553

Email and Phone # of Zohar Yosibash: yosibash@tauex.tau.ac.il +972-52-8795751

CRediT Author Statement: Yosibash Z: Conceptualization, Methodology, Supervision, Data curation, Writing- Original draft preparation. **Trabelsi N.**: Methodology, Validation, Writing – reviewing. **Buchnik I**.: Application of SVM algorithms, Methodology. **Myers KW**: Software, Writing – reviewing and editing. **Salai M.**: Conceptualization, Writing-reviewing. **Eshed I.**: Supervision, Writing-reviewing. **Barash Y** and **Klung E**,: Data curation. **Tripto-Shkolnik L.:** Conceptualization, Methodology, Supervision, Writing- reviewing and editing.

Data Availability: Data supporting the results (including all data for all patients in Table 1) is available from the corresponding author upon request.

Disclosure: Z. Yosibash, N. Trabelsi and K.W. Myers are founders and have equity in PerSimiO.

---

[*] Equal contribution

ABSTRACT

Autonomous finite element analyses (AFE) based on CT scans predict the biomechanical response of femurs during stance and sidewise fall positions. We combine AFE with patient data via a machine learning (ML) algorithm to predict the risk of hip fracture. Setting: An opportunistic retrospective clinical study of CT scans. Aim: To develop a ML algorithm with AFE for hip fracture risk assessment in type-2 diabetic Mellitus (T2DM) and non-T2DM patients.

Abdominal/pelvis CT scans of patients who experienced a hip fracture within two years after an index CT scan were retrieved from a tertiary medical center database. A control group of patients without a known hip fracture for at least five years after an index CT scan was retrieved. Scans belonging to patients with/without T2DM were identified from coded diagnoses. All femurs underwent an AFE under three physiological loads. AFE results, patient's age, weight, and height were input to the ML algorithm (Support Vector Machine (SVM)), trained by 80% of the known fracture outcomes, with cross-validation, and verified by the other 20%.

45% of available abdominal/pelvic CT scans were appropriate for AFE (at least 1/4 of the proximal femur was visible in the scan). The AFE success rate in automatically analyzing CT scans was 91%: 836 femurs we successfully analyzed, and the results were processed by the SVM algorithm. 282 T2DM femurs (118 intact and 164 fractured) and 554 non-T2DM (314 intact and 240 fractured) were identified. Among T2DM patients the outcome was: Sensitivity 92%, Specificity 88%, (cross-validation AUC 0.92), and for the non-T2DM patients: Sensitivity 83%, Specificity 84% (cross-validation AUC 0.84).

Combining AFE data with a ML algorithm provides an unprecedented prediction accuracy for the risk of hip fracture in T2DM and non-T2DM populations. The fully autonomous algorithm can be applied as an opportunistic process for hip fracture risk assessment.

Keywords: Diabetes mellitus; Hip fracture; Finite element analysis; Fracture risk assessment, SVM/machine learning.

INTRODUCTION

Hip fractures are among the most common reasons for orthopedic hospitalization in the elderly worldwide, leading to major health and financial burden [1]. The underlying cause of such fractures is most often osteoporosis. Pharmacological treatments are usually prescribed to prevent hip fractures by patients identified to be at high risk. While the strength of the hip is a function of its mechanical material properties, geometry and loading, most risk assessments use bone mineral density as a surrogate for bone strength. Hip fracture risk is usually determined by dual-energy X-ray absorptiometry (DXA) measurement of femoral neck areal bone mineral density (aBMD) or by the Fracture Risk Assessment Tool (FRAX) which is based on eleven clinical factors along with femoral neck aBMD. Neither of these tools is accurate, especially for type 2 diabetic Mellitus patients (T2DM). These patients are at a twofold greater risk of hip fractures and display a "diabetic paradox": increased risk of femoral fractures despite having higher bone mineral density [2-7]. The trabecular bone score (TBS) is an indirect index of trabecular architecture applied to infer information from spine DXA image, but is assessed only for vertebral fracture risk [7, 8] and cannot be applied to the proximal femur.

Finite element analyses of proximal femurs based on computed tomography scans (CTFEA) have been developed for predicting femur stiffness and hip fracture risk. CTFEA has been demonstrated to outperform DXA [9-14]. The practical use of the technology has been hampered by the high patient radiation exposure, the expense of CT scans, and the lack of fully automated FEA calculations. A large number of abdominal and pelvic CT scans are available in hospitals or health maintenance organizations (HMO) picture archiving and communication systems (PACS). These scans also usually include the hip and the lesser tuberosity of the femur. They may therefore be potentially used opportunistically for hip FEA without exposing patients to additional radiation hazards [15].

We have developed Simfini† [16] as an autonomous CTFEA software application for the FEA of femurs. This tool has been shown to provide accurate predictions of pathological hip fractures in patients with metastatic tumors in two retrospective clinical studies [17, 18]. Recently Simfini's performance in predicting hip risk of

† Simfini is a product of PerSimiO, (U.S. Patent 11,449,993).

fracture was also examined in a feasibility retrospective clinical study on a cohort of 51 T2DM patients [19]. This system includes several novel features:

1. It is fully autonomous, with no manual subjective intervention.
2. The two femurs (left and right)are automatically segmented from the CT scan by means of a deep learning (DL) algorithm, and thereafter automatically represented by a mesh of high-order finite elements.
3. Physiological loading conditions are simulated that represent the two common sidewise falls resulting in neck and intertrochanteric fractures.
4. A machine learning (ML) algorithm is employed in the post-AFE stage which accounts for patients' weight, height, gender, and the biomechanical results at different regions along the proximal femur.

We undertook a retrospective clinical study to assess the performance of the Simfini system in predicting the risk of hip fracture in type 2 diabetic and non-diabetic patients, based on opportunistic abdominal and pelvic CT scans obtained from the PACS of a major medical center.

METHODS

Study design

The Sheba Medical Center (MC) database was searched for patients with CT scans of the lower abdomen/pelvis between 2008-2020 who experienced a hip fracture (study group) during the subsequent two years. Both non-contrast and contrast enhanced CT scans were considered. The control group included age and weight-matched patients with CT scans who did not sustain a hip fracture in the subsequent five years (a conservative requirement to make sure that patients indeed are risk-free for a much longer period than compared to the study group) according to the electronic medical record. The CT scans were collected from the hospital's clinic registry at Sheba MC. Approval was granted by the Sheba MC institutional review board (7969-20-SMC). Overall, 974 CT scans were collected for the study.

The primary outcome was a binary score of the risk of hip fracture within two years following the CT scan or a non-fracture risk within 5 years following the CT scan.

The results obtained from the combined AFE&ML system were used as a risk factor for sustaining a hip fracture.

Patient population

Inclusion criteria included CT scans with a soft tissue filter, and 120 Peak kiloVoltage (KVP). Exclusion criteria included: (1) Pathologic fractures, subtrochanteric or atypical fractures, high energy fractures, metallic implants, and tumors in the proximal femur; (2) Type 1 diabetes mellitus. Of the 974 CT scans, 507 were excluded because of misfit to the clinical trial protocol. The dataset workflow is presented in Figure 5.

For each patient, clinical data including the weight, age, height, and whether he/she was diagnosed with T2DM, were retrieved from the electronic records.

AFEs

The fully autonomous CTFEA system Simfini was used to perform the strength analysis of all femurs according to the algorithm previously published in [16, 17, 19, 20] and schematically illustrated in Figure 1. Briefly, the geometry of the femurs is automatically segmented from the CT scans by a deep-learing U-Net network to produce a 3D voxel representation of the femur. Inhomogeneous isotropic material properties are assigned to the centroid of each voxel within the femur based on the Hounsfield Unit (HU) in the CT scan. The voxels representing the segmented femur are automatically transformed in a mesh of high order tetrahedral elements[‡]. Three loading configurations were applied as presented in Figure 2 and average maximum principal strains were extracted automatically over a circular region of a diameter of 5mm on the surface of the femur in each region of interest.

[‡] High order elements have shape functions with a polynomial degree increased hierarchically from 1 to 8 (each tetrahedral element has 512 shape functions at p=8), allow for curved edges, and allow the intrinsic estimation of the error in energy norm since 8 hierarchical FE solutions with increasing number of degrees of freedom are obtained. A special numerical integration scheme is used that facilitates exact integration of monomials up to 14th order.

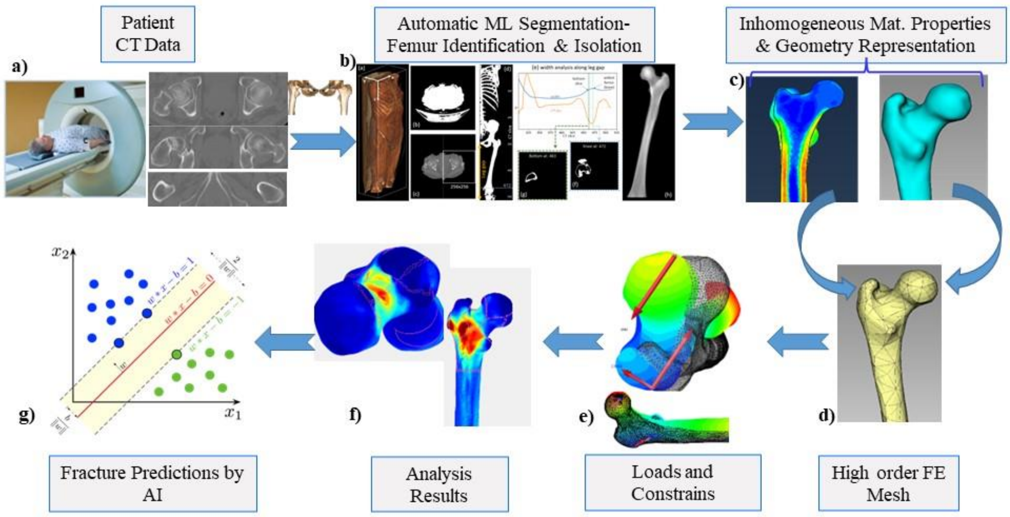


Figure 1 – Schematic description of the Simfini system. a) Retrieval of CT scans from PACS, b) Segmentation of the two femurs by U-Net and identification of anatomical points, c) Generation of the inhomogeneous material data and 3D geometry of both femurs d) Generating a high-order -finite element mesh, e) Application of three different boundary conditions and solution of the FE system, f) Extraction of averaged maximum strains at different locations along the femur, g) Fracture predictions by SVM algorithm.

<u>The three different boundary conditions applied to each femur</u>

A proximal femoral fracture due to a fall on the side is categorized as either a neck or a pertrochanteric fracture, with an almost equal probability to occur [21, 22]. Two different load directions induce two different fracture scenarios. These directions were determined by a former clinical study on 32 patients who experienced a hip fracture and were CT scanned immediately following the fracture. Fourteen patients were diagnosed as having a neck fracture (f=8, m=6) and eighteen were diagnosed as having a pertrochanteric fracture (f=12, m=6) [23]. For the neck fracture group, loading configuration *FallN* (see Figure 2) always stresses the superior and inferior neck with the lowest fracture load and was selected as a good predictor for a femoral neck fracture. For the pertrochanteric fracture, loading configuration *FallP* (see Figure 2) stresses in most of the cases the trochanter but also the anterior and posterior base of the neck. The loading condition was selected as the preferred predictor for trochanteric fracture (see also in-vitro experiments *"…FE models predicted that the fractures initiate under compression on the lateral side of the femoral neck"* [24]). Illustrative examples of the two loading conditions and the maximum compressive strained locations are presented in Figure 2. *FallN* predicts a neck fracture at the superior neck in compression. *FallP* also predicts a pertrochanteric fracture in compression. Therefore, it is conceivable to consider both. The application of multiple loading conditions to best represents a sidewise fall condition has been confirmed by in-vitro experiments *"FE-strength from multiple loading conditions better-classified fracture cases from controls….. Only FE-strength from multiple loading conditions remained significant in age- and aBMD-adjusted models"* [25].

Stance loading (along the vector connecting the head and intercondylar notch) also induces high strains in the superior and inferior neck regardless of the fracture's actual location. AFE results under this loading condition are also considered when determining the risk of fracture. The magnitude of all loads is normalized by the patient's body weight. In the AFE the total magnitude of all applied loads is 2.5 times the patient's weight.

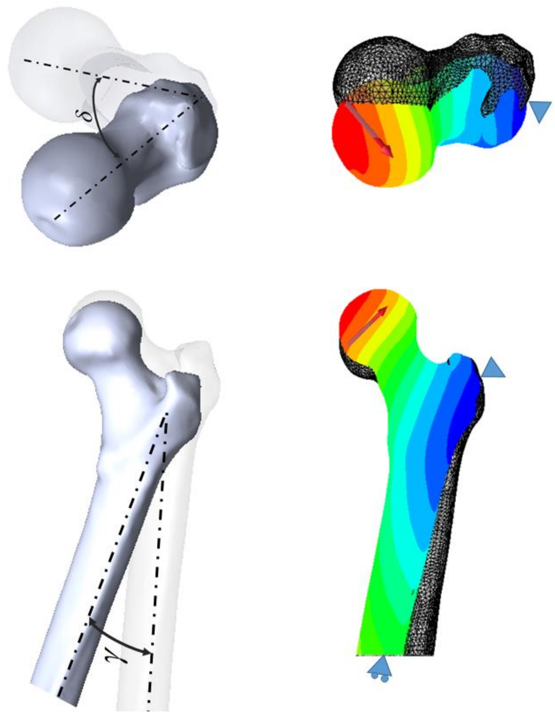


Figure 2 – Definition of boundary conditions for sideways fall configuration; FallN is determined by $\gamma = 10°$ and $\delta = 15°$ and FallP by $\gamma = 30°$ and $\delta = 45°$. Figures with colors representing displacements due to boundary conditions are taken from [26]

Since the γ and δ angles are determined by anatomical points, the algorithm performs best if at least 20 mm below the lesser trochanter is visible in the CT scan. A Borderline case is when only the lesser trochanter is visible. CT scans that do not include the entire lesser trochanter are disqualified from being biomechanically analyzed.

Average maximum principal tensile strains (denoted by E1) and average minimum principal compression strains (denoted by E3) are automatically computed in each of the areas of interest, for each loading condition: Neck superior and inferior, Trochanter posterior and anterior, Head superior and inferior and Lesser Trochanter inferior, see Figure 3. Head movement and bone stiffness (force magnitude divided by head movement) as well as moment applied and maximum and minimum Young's modulus in the femur are also computed.

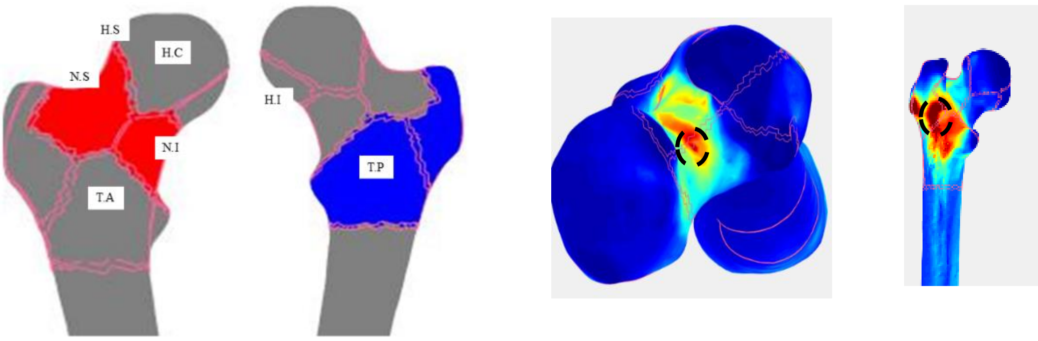


*Figure 3 – The various locations (**H**ead **S**uperior, **H**ead **I**nferior **N**eck **S**uperior, **N**eck **I**nferior, **T**rochanter **P**osterior, **T**rochanter **A**nterior, **H**ead **C**enter) in the proximal femur at which strains are computed by the AFE (left two figures), maximum compressive principal strains at the neck and intertrochanteric regions due to two different sidewise fall loadings (right two figures)*

<u>Combining biomechanical data with patient data and application of machine learning techniques</u>

Statistical learning models, and particularly ML have been recently used to automatically post-process many data combinations [27]. Here, we present a ML model that combines patient data with computational biomechanics results to predict the risk of hip fractures. The ML model was trained separately for the T2DM group and the non-T2DM group.

The available samples were shuffled and split 0.8 for training and 0.2 for testing. Due to the small train set, we used cross-validation over the train set only. Cross-validation is a technique that allows one to estimate the performance of machine learning models on unseen data. We applied the k-fold cross-validation method, where the data was divided into k=6 subsets. The model was then trained on 5 of these subsets and evaluated on the remaining one. This process was repeated 6 times, with each subset being used as the validation set once (Figure 5). We calculated the mean and standard deviation of all statistical metrics (F1, precision, etc) over the left-out subsets to ensure the chosen threshold is a good fit for our model to verify its generalization ability. In that manner, we were able to obtain an estimate of the model's performance that is not affected by the specific data used for training and validating. Then we applied the model, with the chosen threshold, over the independent test set (the remaining 20% of the data).

The available patient dataset is unbalanced thus we had to prevent the ML model from becoming biased toward the predominant class. We used random over-sampling to balance the unbalanced training dataset, i.e. balancing the data by replicating the minority class samples (a method that does not cause any loss of information [28]). Over-sampling wasn’t used either for the folded-out set in each training/validation split, or for the independent testing set that was separated at the preprocessing procedure. Each dataset was normalized by removing the mean and scaling each feature to unit variance. The training samples are given to the model for creating the inference mapping function from the domain of features to the label domain – trying to maximize the number of samples classified correctly but keeping the problem generalized and not overfit. The testing/validation samples are the new cases not used for training the ML process. Based on these, the predicted specificity and sensitivity are computed (thanks to a comparison of the real known labels and the model-predicted ones).

We considered two ML algorithms: Random Forest (RF) and Support Vector Machine (SVM) [29]. Both algorithms are well suited for a mixture of numerical and categorical features. The SVM training algorithm constructs a model that maps training examples to points in space to maximize the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. A detailed discussion on SVM, including the mathematical foundations and the various factors that influence its performance, is provided in [37]. The dominant factor we used is the Nu parameter to control the number of support vectors [30].

RF is an ensemble learning method for classification that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the RF is the class selected by most trees. Random decision forests correct for decision trees' tendency to overfit to their training set.

RF and SVM fracture/non-fracture predictions for the two groups were compared based on the receiver-operating characteristic curve (ROC) and the area under the curve (AUC). The operating point threshold for the inference model was chosen at the point with the highest F1 score for the cross-validation set. Both RF and SVM results are very similar with slightly better performance for the SVM. Therefore, SFM was the chosen method. The sensitivity, specificity, and AUC of the SVM for the T2DM group and the non-T2DM group (computed based on 20% of the CT scans) are presented in the Results section. A total of 41 features were used in the SVM algorithm as detailed in Table 1.

Table 1: List of 41 features used in the SVM algorithm: 37 generated by the AFE and 4 related to patient data.

'Stance Neck Superior E1'	'Stance Trochanter E1'	'Stance Neck Inferior / Sub Capital E3'	'Stance Trochanter E3'	**Stance Head Center Utot**	**Stance Bone K**	'FallN Neck Inferior E1'	'FallN Trochanter Posterior E1'	'FallN Lesser Trochanter Anterior E1'	'FallN Head Superior E1'	'FallN Head Inferior E1'
'FallN Neck Superior E3'	'FallN Neck Inferior E3'	'FallN Trochanter Posterior E3'	'FallN Lesser Trochanter Anterior E3'	'FallN Head Superior E3'	'FallN Head Inferior E3'	**FallN Head Center Utot**	**FallN Bone K**	'FallP Neck Superior E1'	'FallP Neck Inferior E1'	'FallP Trochanter Posterior E1'
'FallP Lesser Trochanter Anterior E1'	'FallP Head Superior E1'	'FallP Head Inferior E1'	'FallP Neck Superior E3'	'FallP Neck Inferior E3'	'FallP Trochanter Posterior E3'	'FallP Lesser Trochanter Anterior E3'	'FallP Head Superior E3'	'FallP Head Inferior E3'	**FallP Bone K**	**FallP Bone K**
'Femoral length mm'	'E max'	'E min'	'Mtot Stance at 80mm below top'	Age	Height	Weight	Gender			

Statistical Analysis and Verification of Results

The predictive performance of the risk of fracture criteria was evaluated for its specificity, sensitivity, and AUC as follows. “Sensitivity” is defined as the percentage of patients for whom fractures were correctly predicted and occurred within two years of the CT scan. “Specificity” is defined as the percentage of patients correctly identified as fracture free for 5 years following the scan. To determine the uncertainty of the estimates of sensitivity and specificity, 95% confidence intervals (CIs) are calculated for the test set according to [31].

The receiver-operating characteristic curves (ROC) were generated and the area under the receiver-operating characteristic curve (AUC) was computed and reported.

To further verify the performance once the algorithm was established, the SVM was applied to the 17 additional CT scans for which only one femur was successfully analyzed (due to presence of an implant, pre-existing fracture, etc.). Within this cohort, thirteen patients were non-T2DM, 7 experienced a hip fracture, and 6 with intact femurs. Four patients were T2DM, 2 experienced a hip fracture and 2 with intact femurs.

RESULTS

A total of 974 clinical CT scans were retrieved, generated by several different scanners (manufactured by GE and Phillips). Pixel spacing for the scans was between 0.57 and 0.98 mm. Although slice thickness was between 0.63 and 3 mm, most scans had a 2 mm slice thickness. No duplicate CT scans for any patients were identified in the cohort. Patients of the study group were selected by one of the researchers (EK) who was blind to the content of the scans: A list of CT accession numbers was generated from the Sheba Medical Center radiology department information system. Then, the corresponding CTs were retrieved from the Sheba Medical Center radiology department PACS in Digital Imaging and Communications in Medicine (DICOM) format after anonymization of the DICOMs meta-data fields.

507 CTs were excluded from the study for not complying with the protocol (the majority because the femur was "short"[§]). CT scans in which the lesser trochanter is visible but included less than 20 mm below the trochanter were denoted ``borderline''. Typical examples of short, borderline, and standard CT scans are shown in Figure 4. CTs were excluded if:

a) The CT scan did not include the entire lesser trochanter.

b) A metallic implant was present that resulted in artifacts in the proximal femur.

c) Tumors were clearly visible in the proximal femur.

[§] A "short" CT is defined as a CT which does not contain at least the lesser trochanter of one of the two femurs in the scan.

d) A fracture was reported, but it was either caused by a high-energy trauma or occurred in the distal femur.

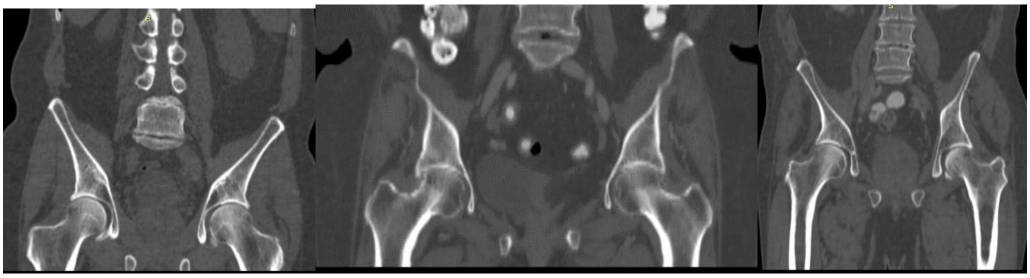


Figure 4 – Illustrative examples of Short (Left), Borderline (Middle), and standard (Right) femurs in CT scans.

467 CTs (Standard & Borderline) were suitable for Simfini analysis (48% of all CT scans collected). Twenty-two of these could not be retrieved successfully from the PACS, Simfini issued an error message for 12 CTs (failed to segment the femur or to generate a finite element mesh), and for 17 CTs the analysis was successful for one femur only. Therefore, the success rate of Simfini was (934-44-24-17)/934 = 91%, resulting in data for 836 femurs representing 418 CT scans. Table 2 summarizes the number of Standard and Borderline femurs in the study and control group. None of the scans had calibration phantoms. Overall, 568 femurs were acquired by GE scanners and 268 femurs by Phillips scanners.

Table 2: Summary of Standard/Borderline femurs for the study and control groups that were successfully analyzed by Simfini.

	Standard	**Border**	**Total**
# of femurs without a fracture within five years after CT	204	274	**478**
# of fractured femurs	104	254	**358**
Total	**308**	**528**	**836**

A flowchart illustrating the femur selection process for the Simfini analysis is presented in Figure 5.

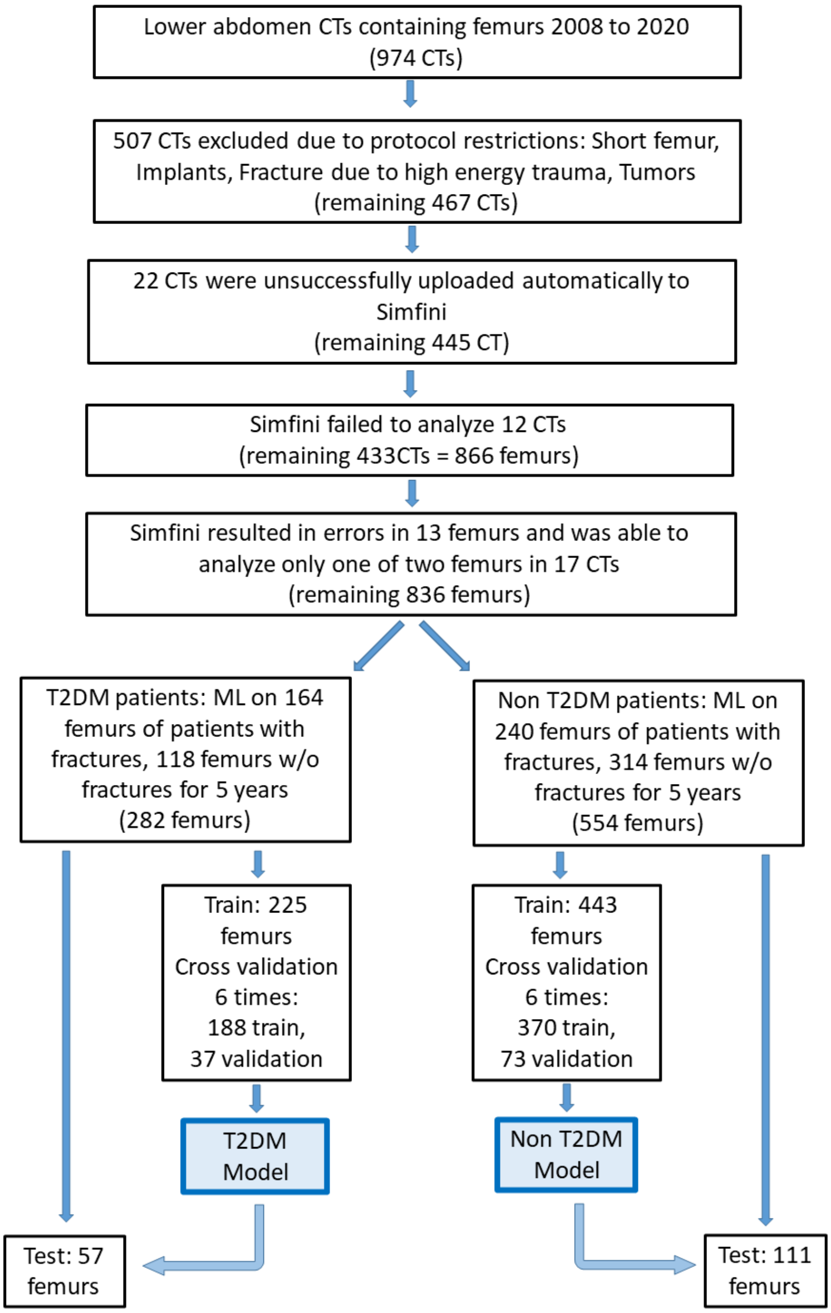


Figure 5 – Case selection process.

Table 3 summarizes the distribution of the 836 femurs of T2DM and non-T2DM patients that were successfully analyzed by Simfini.

Table 3: Summary of the number of femurs for T2DM and non-T2DM patients successfully analyzed by Simfini.

	Intact	**Fractured**	**Total**
T2DM patients	118	164	**282**
Non-T2DM patients	314	240	**554**
Total	**432**	**404**	**836**

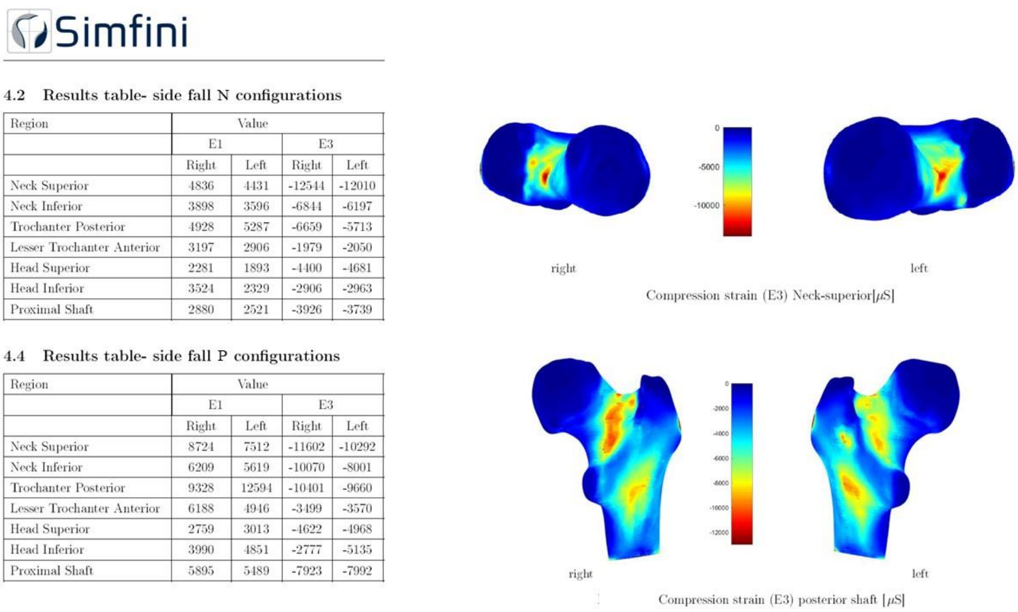
Table 4 summarizes the average age, weight, and height of the patients for which Simfini analyses were successfully performed.

Table 4: Summary of age, weight, and height (with standard deviation) and gender for patients successfully analyzed by Simfini. 418 CT scans (836 femurs)

	# of CTs	Male/Female	Avg Age [years]	Avg Weight [kg]	Avg Height [cm]
Fx T2DM	82	35M/47F	75.8±8.4	71.5±15.8	164±8.6
Intact T2DM	59	19M/40F	77.5±9.4	69.3±16.3	163±9.7
Fx non-T2DM	157	46M/111F	75.8±9.4	66.0±17.4	163±8.8
Intact non-T2DM	120	36M/82F	75.9±9.3	67.8±13.3	162±8.6

There were no statistically significant differences between the study and the control groups regarding age, weight, and height (Table 4).

For each patient, the strains computed by Simfini under the different loading conditions were extracted and shown as an example for *FallN* and *FallP* in Figure 6.


Simfini

4.2 Results table- side fall N configurations

Region	Value			
	E1		E3	
	Right	Left	Right	Left
Neck Superior	4836	4431	-12544	-12010
Neck Inferior	3898	3596	-6844	-6197
Trochanter Posterior	4928	5287	-6659	-5713
Lesser Trochanter Anterior	3197	2906	-1979	-2050
Head Superior	2281	1893	-4400	-4681
Head Inferior	3524	2329	-2906	-2963
Proximal Shaft	2880	2521	-3926	-3739

4.4 Results table- side fall P configurations

Region	Value			
	E1		E3	
	Right	Left	Right	Left
Neck Superior	8724	7512	-11602	-10292
Neck Inferior	6209	5619	-10070	-8001
Trochanter Posterior	9328	12594	-10401	-9660
Lesser Trochanter Anterior	6188	4946	-3499	-3570
Head Superior	2759	3013	-4622	-4968
Head Inferior	3990	4851	-2777	-5135
Proximal Shaft	5895	5489	-7923	-7992

Figure 6 – Simfini computed strains (tensile E1 and compressive E3) under FallN and FallP loadings (2.5 body weights) for a typical patient.

The SVM cross-validation performance is summarized in Table 5a and the corresponding receiver operating characteristic (ROC) curves are presented in Figure 7. The areas under the curve (AUC) values for the ROC curves are also reported in Table 5a. The SVM test set predictions are summarized in Table 5b. The p-value of all dataset configurations was less than 0.01.

Table 5a: SVM cross-validation predictions mean values and standard deviation (in parenthesis) of the sensitivity, specificity, precision, and AUC.

Mean (Std)	F1 Score	Sensitivity (STD)	Specificity (STD)	Precision	AUC for the cross-validation set
T2DM Cross-validation	0.81 (0.03)	0.77 (0.09)	0.82 (0.04)	0.89 (0.06)	0.92
Non-T2DM Cross-validation	0.78 (0.04)	0.81 (0.08)	0.80 (0.05)	0.79 (0.05)	0.84
Combined T2DM and Non-T2DM Cross-validation	0.78 (0.02)	0.8 (0.03)	0.78 (0.06)	0.83 (0.04)	0.88

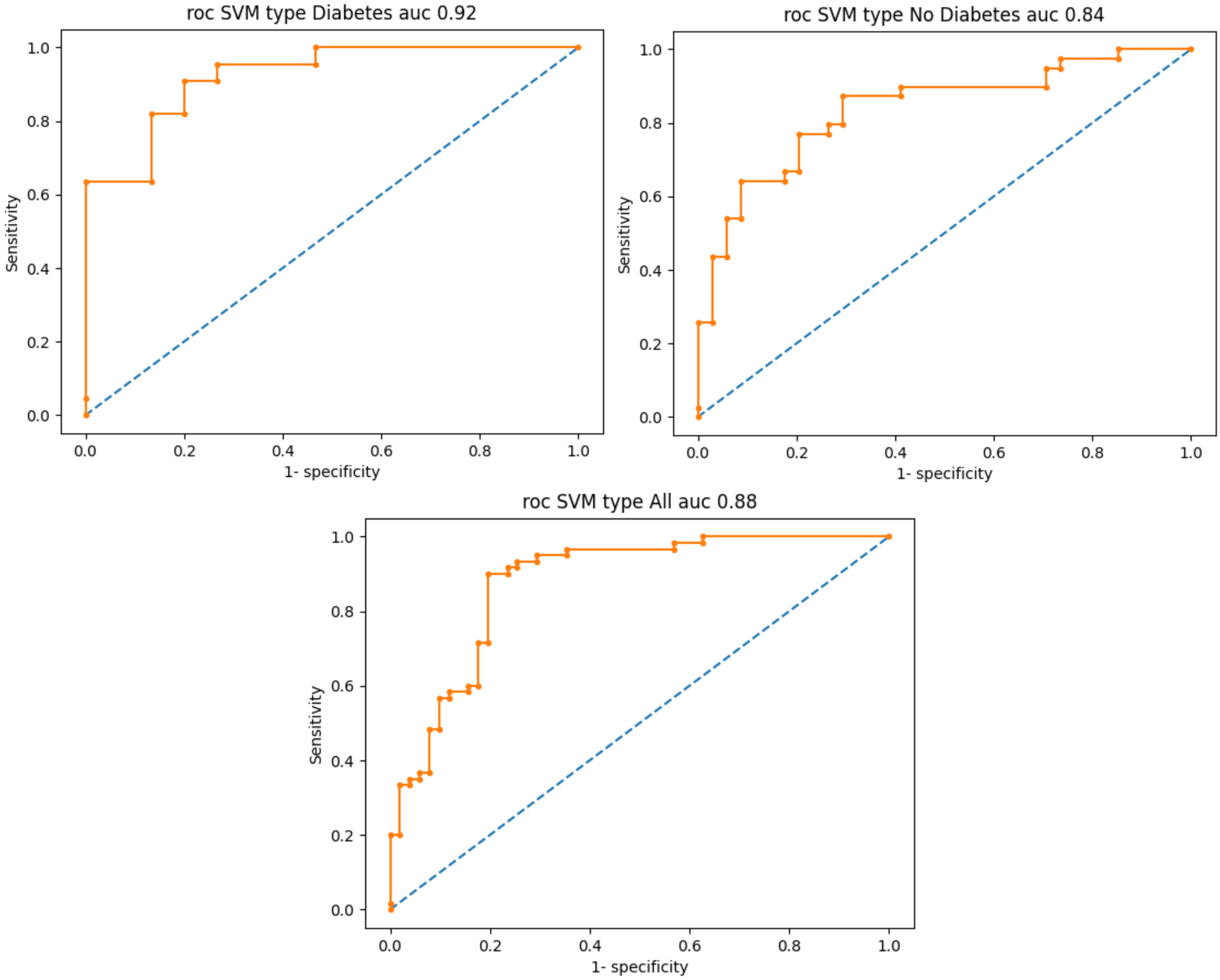


Figure 7 – Receiver operating characteristic (ROC) curves for the T2DM population (Upper-left), the non-T2DM population (Upper-right), and the Combined T2DM&Non-T2DM population (Lower-middle) for the cross-validation set, demonstrating an area under the curve (AUC) of 0.92, 0.84 and 0.88 correspondingly.

Table 5b: SVM test set predictions in terms of sensitivity, specificity (with 95% CI), and precision.

	F1 Score	Sensitivity (95% CI)	Specificity (95% CI)	Precision
T2DM- Test set	0.84	92% (85-99%)	88% (80-97%)	0.85
Non-T2DM- Test set	0.81	83% (76-90%)	84% (77-91%)	0.86
Combined T2DM and Non-T2DM Cross-validation	0.82	86% (73-89%)	79% (75-82%)	0.85

It is important to emphasize that no attempt was made to optimize the outcome of the SVM algorithm by including or excluding input features.

Further verification

The seventeen patients for which the AFE failed to analyze both femurs that were not included in the SVM analysis were used for further verification of the accuracy in predicting hip fracture risk. Using the AFE results for one femur and the trained SVM algorithm, the following statistics were obtained: for the four T2DM patients, the sensitivity was 100% and the specificity was 67%. For the non-T2DM patients (13 patients) the sensitivity was 75% and the specificity was 80%.

DXA data

Only eleven of the 418 patients who were AFE analyzed had available DXA scores in the Sheba MC database: Two T2DM patients, one who fractured and one who did not, both had a T-score of -1.5 at the proximal femur and -1.9. -2.0 at the lower neck. Among the nine non-T2DM patients, three fractured with a T-score of -1.5, -1.9, -2.2 at the proximal femur and -1.6, -2.2 at the neck. Six non-T2DM patients who did not fracture had a T-score between 0.5 and -1.4 at the proximal femur and -0.4 to -2.2 at the neck. None of the 4 who fractured had a T-score below -2.5, i.e. diagnosed as osteoporotic. The DXA data is too limited for statistical analysis, however it demonstrates that none of those who fractured had a densitometric diagnosis of osteoporosis (the average age was 75 years old, similar to the AFE cohort).

DISCUSSION

Simfini is a fully autonomous finite element system that can be easily used for opportunistic biomechanical analysis of abdomen or pelvis CT scans of the femur. The biomechanical analysis is fused to an ML (SVM) algorithm and provides highly accurate hip fracture risk prediction in elderly T2DM and non-T2DM populations. Forty-eight percent of the abdominal and pelvic CT scans evaluated were appropriate for the AFE (very similar to the percentage reported in [14]) out of which ninety-one percent were successfully analyzed by the AFE (i.e. forty-three percent of the available lower abdomen and pelvic CT scans were successfully analyzed). An excellent prediction of hip fractures within the next two years for both T2DM patients (a group that possesses a special challenge) as well as non-T2DM ones was demonstrated.

The further verification on seventeen patients for which the AFE was able to analyze only one femur showed that the outcome corresponds well with the statistical data presented.

The CT utilization rate in our study is on par with other published studies using opportunistic screening tools: Dagan et al reported an 83.6% utilization rate [32] and Adams et al reported an 86% utilization rate [13], both in very large and diverse populations.

During the past five years, several studies have shown the feasibility of using opportunistic CT scans to predict osteoporotic fractures [33], specifically hip fractures [9, 34]. The only autonomous algorithm (based entirely on ML) [32] was trained and verified on over 48,000 CT scans to assess the 5-year risk of osteoporotic fractures. The ML predictions for a hip fracture were shown to be the same as the FRAX performance without BMD input. The ML algorithm relies mostly on BMD assessment from CT scans. A sensitivity of 92.6%, specificity of 36.9%, and AUC of 0.76 were achieved, which were almost identical to FRAX performance [32].

FEA determination of femoral strength has been shown to better predict hip fracture than hip BMD [35, 36]. Several previous studies have demonstrated the use of femoral strength measurement derived from existing CT scans to predict hip fracture risk [11, 13, 14]. In [13], 1,959 patients aged 65 or older who sustained a hip fracture and who had a prior pelvic or abdominal CT scan and a DXA, were compared to a sex-matched group. The study population included 30% diabetic patients, but there was no sub-analysis to determine the validity of this method specifically in those patients. In

[14] 490 lower abdomen CT scans out of 1158 were suitable for FEA (43.2%) out of which 123 suffered a fracture within 5 years of the CT scan date. Fracture prediction by combining both BMD and FE-estimated bone strength was not statistically different than using either BMD or FE-estimated bone strength alone. Predicting fractures in women determined the greatest AUC of 0.710 by using both BMD and FEA (sensitivity 48% and specificity 84%). The study reported in [11] used very uniform CT scans, all resulting from a single CT scanner with a slice thickness of 1mm and all having calibration phantoms. This database was unusual because typical clinical scans are from a variety of CT scanners, have lower resolution and none use calibration phantoms.

CTFEA accurately predicts one of the most important components required to determine the risk of femoral fracture – the bone strength under a load that is believed to represent a sidewise fall. One of the reasons CTFEA is not commonly used in clinical practice is the manual labor and expertise required to set up the analysis and interpret the output – which may be a lengthy and subjective process. Also, the patient's weight was not taken into consideration in former CTFEAs, which in the authors' opinion is an important component.

In order to address the perceived need for improved fracture risk assessment, we developed the fully AFE system [16] that automatically retrieves CT scans from a hospital's PACS, segments the femurs by a DL algorithm, automatically performs FE analyses with physiological loads, and applies a SVM post-processing algorithm. We found the most influential factor over the post processing performance is the Nu parameter that controls the number of support vectors. The fully autonomous system demonstrated unprecedented identification of hip fracture risk within 2 years following the CT scan. In Table 6 we summarize the current system's performance compared to the performance reported in former publications.

Table 6: Summary of the performance of recent methods for identifying risk of hip fractures: number of CTs considered, sensitivity, specificity, and AUC.

Method (ref)	**# CTs**	**Sensitivity**	**Specificity**	**AUC**
Current CTFEA&ML T2DM (cross-validation set)	**141**	**92% (77%)**	**88% (82%)**	**(0.92)**
Current CTFEA&ML Non-T2DM (cross-validation set)	**277**	**83% (81%)**	**84% (80%)**	**(0.84)**
Current CTFEA&ML Combined (cross-validation set)	**418**	**86% (80%)**	**79% (78%)**	**(0.88)**

[14] CTFEA+aBMD	490	48%	84%	0.71
[11] CTFEA	601			0.71-0.80
[13] Women CTFEA	~1900	66%	66%	0.70-0.73
[13] Men CTFEA	~860	56%	76%	0.75
[32] ML	~48,000	92.6%	36.9 %	0.76
[19] CTFEA T2DM	51	89%	76%	0.9

In conclusion, this clinical study demonstrates a high accuracy achieved when predicting the risk of fracture due to a sidewise fall by combining AFE and machine learning in both T2DM and non-T2DM populations. Since there is a significant clinical need to develop a reliable risk assessment tool for the T2DM population, implementing such a tool as an opportunistic measure on a large scale could contribute significantly to the prevention of osteoporosis-related complications in diabetic patients, specifically hip fractures.

The proposed AFE may be used in many other clinical applications by assessing bones' strength in longitudinal studies to monitor, for example, radiation therapy influence, medication efficacy, over/under stress, etc. Application of the methodology to other bones such as the humerus, vertebra, and tibia is another promising outcome of the presented methodology.

This study has several limitations: (1) Results were not compared with current commonly used methods to measure bone strength or assess fracture risk, namely a DXA or a FRAX score, since the hospital registry in Israel has very limited data on these for most patients; (2) CTs which do not include the entire lesser trochanter are excluded from the AFE (about 50% of the overall lower abdomen CT), (3) Data on the first diagnosis of T2DM for these patients is missing.

The encouraging results pave the path to further clinical and scientific enhancements. A follow-on research study is planned that will include AFEs of CT scans in which only a part of the lesser trochanter is visible. Although this approach is expected to considerably increase the number of usable femur scans in the study, it will likely see a decrease in sensitivity and specificity of the fracture risk assessment. Optimal input features to the SVM algorithm will also be investigated, and a prospective study is planned to use opportunistic CT scans with corresponding DXA scores to allow direct comparison with the AFE performance.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the partial funding by the Israel Innovation Authority for the clinical trial.

Appendix – A Summary of the AFE System (Based on [16,19])

The femur's response under physiological loading is well described by the linear theory of elasticity and although the bone at the macroscopic level is orthotropic, excellent predictions were obtained using isotropic inhomogeneous relations[*] (see [20] for stance position loading and [38] for sideway fall loading). Thus, a linear finite element analysis was performed by Simfini. Verification of the numerical errors was assured by monitoring the error in energy norm and the maximum and minimum principal strains at the locations of interest as the polynomial degree over the elements was increased from 1 to 6 or 8.

To realize an autonomous FE analysis, several components are combined. The automatic identification of the femur's starting and ending CT slices and the femur's segmentation is obtained by a deep learning algorithm (a U-net algorithm). The U-net algorithm was trained on 178 femurs and tested on 43 femurs, resulting in a Dice score of 0.99. Another important component of the AFE is the determination of the anatomical points (center of femur's head, intercondylar notch, and center of shaft 20 mm below the lesser trochanter), for the application of the different boundary conditions.

Pointwise inhomogeneous mechanical properties are then computed at each voxel in the CT scan. The relationships between Young's modulus and ash density[†] for cortical and trabecular bone tissue, validated in experimental settings [20], were used:

$$\rho_{K_2HPO_4} = 10^{-3}\,(a \times HU + b) \qquad [grm/cm^3] \tag{A.1}$$

$$\rho_{ash} = 0.877 \times 1.21 \times \rho_{K2HPO4} + 0.08 \qquad [grm/cm^3] \tag{A.2}$$

$$E_{cort} = 10200 \times \rho_{ash}^{2.01}\ [MPa], \qquad \rho_{ash} \geq 0.486\ [grm/cm^3] \tag{A.3}$$

$$E_{trab} = 2398\ [MPa], \qquad 0.3 < \rho_{ash} < 0.486\ [grm/cm^3] \tag{A.4}$$

$$E_{trab} = 33900 \times \rho_{ash}^{2.2}\ [MPa], \qquad \rho_{ash} \leq 0.3\ [grm/cm^3] \tag{A.5}$$

Since most clinical CT scans are phantomless, *a* and *b* in (1) are estimated by an

---

[*] An isotropic material has an equal mechanical response when stretched in any direction. An inhomogeneous material has a different mechanical response at different locations within the domain.

[†] These relationships are for a soft tissue CT scan with 120 KVP (as all collected CT scan) and validated by a set of experiments on fresh frozen femurs [21,28]

algorithm that involves HU=0 in air and a histogram of HU in the femurs, using the 0.1% highest HU that is associated with a Young modulus of 20GPa (details are given in [16]). The Poisson ratio was set to the constant value of $\nu = 0.3$.

An automatic algorithm is applied which generates a finite element mesh consisting of tetrahedrons having curved faces followed by an efficient high-order FE algorithm that solves the system of finite element equations and generates the data of interest. We present in Figure A.1 two examples of femurs from two patients (which have a relatively long part of the shaft visible in the CT scan), with the three different loadings presented (stance and two fall on the side) that are solved sequentially. Each model has about 9000-10,000 finite elements resulting in about 900,000 degrees of freedom at p=6. The entire simulation time including the pre and post processing for two femurs for a patient is about one hour on a standard PC.

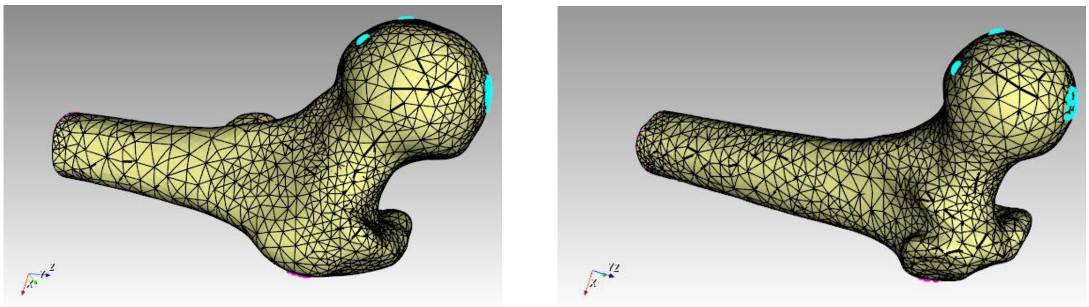


Figure A.1 – Two finite element models of the left femur of two randomly selected patients. The three locations of the applied stance, FallN and FallP loadings on the head are shown by blue (in the web publication) and displacement boundary conditions at the lateral greater trochanter shown in pink (in the web publication).

Finally, a post-processing algorithm extracts from the finite element solutions (three different solutions that correspond to three different boundary conditions) strains in five different anatomical locations along the femur. The maximum and minimum averaged principal strains on the bone's surface are then processed and reported in a file.

REFERENCES

1. Hernlund, E., et al., Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). Arch Osteoporos, 2013. 8: p. 136.

2. Fan, Y., et al., Diabetes mellitus and risk of hip fractures: a meta-analysis. Osteoporos Int, 2016. 27(1): p. 219-28.

3. Botella Martinez, S., et al., The diabetic paradox: Bone mineral density and fracture in type 2 diabetes. Endocrinol Nutr, 2016. 63(9): p. 495-501.

4. Schwartz, A.V., et al., Association of BMD and FRAX score with risk of fracture in older adults with type 2 diabetes. JAMA, 2011. 305(21): p. 2184-92.

5. Giangregorio, L.M., et al., FRAX underestimates fracture risk in patients with diabetes. J Bone Miner Res, 2012. 27(2): p. 301-8.

6. Ferrari, S.L., et al., Diagnosis and management of bone fragility in diabetes: an emerging challenge. Osteoporos Int, 2018. 29(12): p. 2585-2596.

7. Schacter, G.I. and W.D. Leslie, DXA-Based Measurements in Diabetes: Can They Predict Fracture Risk? Calcif Tissue Int, 2017. 100(2): p. 150-164.

8. Rubio, J.B., et al., Review of the scientific evidence regarding clinical use of the Trabecular Bone Score (TBS) SEIOMM official position (2018). Revista De Osteoporosis Y Metabolismo Mineral, 2018. 10(4): p. 149-159.

9. Johannesdottir, F., B. Allaire, and M.L. Bouxsein, Fracture Prediction by Computed Tomography and Finite Element Analysis: Current and Future Perspectives. Curr Osteoporos Rep, 2018. 16(4): p. 411-422.

10. Johannesdottir, F., B. Allaire, and M.L. Bouxsein, Correction to: Fracture Prediction by Computed Tomography and Finite Element Analysis: Current and Future Perspectives. Curr Osteoporos Rep, 2022. 20(5): p. 364.

11. Fleps, I., et al., Finite element derived femoral strength is a better predictor of hip fracture risk than aBMD in the AGES Reykjavik study cohort. Bone, 2022. 154: p. 116219.

12. van den Munckhof, S. and A.A. Zadpoor, How accurately can we predict the fracture load of the proximal femur using finite element models? Clinical Biomechanics, 2014. 29(4): p. 373-380.

13. Adams, A.L., et al., Osteoporosis and Hip Fracture Risk From Routine Computed Tomography Scans: The Fracture, Osteoporosis, and CT Utilization Study (FOCUS). J Bone Miner Res, 2018. 33(7): p. 1291-1301.

14. Michalski, A.S., et al., Opportunistic CT screening predicts individuals at risk of major osteoporotic fracture. Osteoporos Int, 2021. 32(8): p. 1639-1649.

15. Aggarwal, V., et al., Opportunistic diagnosis of osteoporosis, fragile bone strength and vertebral fractures from routine CT scans; a review of approved technology systems and pathways to implementation. Ther Adv Musculoskelet Dis, 2021. 13: p. 1759720X211024029.

16. Yosibash, Z., et al., Autonomous FEs (AFE) - A stride toward personalized medicine. Computers & Mathematics with Applications, 2020. 80(11): p. 2417-2432.

17. Sternheim, A., et al., Pathological fracture risk assessment in patients with femoral metastases using CT-based finite element methods. A retrospective clinical study. Bone, 2018. 110: p. 215-220.

18. Sternheim, A., et al., When and where do patients with bone metastases actually break their femurs? Bone Joint J, 2020. 102-B(5): p. 638-645.

19. Rotman, D., et al., Assessing hip fracture risk in type-2 diabetic patients using CT-based autonomous finite element methods : a feasibility study. Bone Joint J, 2021. 103-B(9): p. 1497-1504.

20. Yosibash, Z., N. Trabelsi, and C. Milgrom, Reliable simulations of the human proximal femur by high-order finite element analysis validated by experimental observations. Journal of Biomechanics, 2007. 40(16): p. 3688-3699.

21. Filipov, O., Epidemiology and social burden of the femoral neck fractures. Journal of IMAB - Annual Proceedings (Scientific Papers), 2014. 20(4): p. 516-518.

22. Mokawem, M., P. Bobak, and J. Aderinto, The management of pertrochanteric fractures of the hip. Orthopaedics and Trauma, 2012. 26(2): p. 112-123.

23. Yosibash, Z. and N. Trabelsi, Assessment of probability of bone fracture, in Provisional Application for Patent, Application number 63040549. 2020.

24. Kok, J., et al., Femoral strength and strains in sideways fall: Validation of finite element models against bilateral strain measurements. J Biomech, 2021. 122: p. 110445.

25. Falcinelli, C., et al., Multiple loading conditions analysis can improve the association between finite element bone strength estimates and proximal femur fractures: a preliminary study in elderly women. Bone, 2014. 67: p. 71-80.

26. Rotman, D., et al., Assessing hip fracture risk in type-2 diabetic patients using CT-based autonomous finite element methods A FEASIBILITY STUDY. Bone & Joint Journal, 2021. 103b(9): p. 1497-1504.

27. Badgeley, M.A., et al., Deep learning predicts hip fracture using confounding patient and healthcare variables. Npj Digital Medicine, 2019. 2.

28. Ghazikhani, A., H.S. Yazdi, and R. Monsefi. Class imbalance handling using wrapper-based random oversampling. in 20th Iranian Conference on Electrical Engineering (ICEE2012). 2012.

29. Tin Kam, H. Random decision forests. in Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995.

30. Crisp, D.J. and C.J.C. Burges, A geometric interpretation of nu-SVM classifiers. Advances in Neural Information Processing Systems 12, 2000. 12: p. 244-250.

31. Ying, G.S., et al., Calculating Sensitivity, Specificity, and Predictive Values for Correlated Eye Data. Invest Ophthalmol Vis Sci, 2020. 61(11): p. 29.

32. Dagan, N., et al., Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization. Nat Med, 2020. 26(1): p. 77-82.

33. Lenchik, L., et al., Opportunistic Screening for Osteoporosis Using Computed Tomography: State of the Art and Argument for Paradigm Shift. Curr Rheumatol Rep, 2018. 20(12): p. 74.

34. Lee, S.J., P.A. Anderson, and P.J. Pickhardt, Predicting Future Hip Fractures on Routine Abdominal CT Using Opportunistic Osteoporosis Screening Measures: A Matched Case-Control Study. AJR Am J Roentgenol, 2017. 209(2): p. 395-402.

35. Kopperdahl, D.L., et al., Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. J Bone Miner Res, 2014. 29(3): p. 570-80.

36. Orwoll, E.S., et al., Finite element analysis of the proximal femur and hip fracture risk in older men. J Bone Miner Res, 2009. 24(3): p. 475-83.

37. Cristianini, N. and Shawe-Taylor J. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000

38. Altai Z, Qasim M, Li XS, Viceconti M. The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture. Clin Biomech. 2019, 68: p. 137-43