Computers and Mathematics with Applications 152 (2023) 16–27

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

Femurs segmentation by machine learning from CT scans combined with autonomous finite elements in orthopedic and endocrinology applications

Zohar Yosibash [a,b,*], Yekutiel Katz [b], Trabelsi Nir [c,b], Amir Sternheim [d]

[a] School of Mechanical Engineering, The Iby and Aladar Fleischman Faculty of Engineering, Tel-Aviv University, Ramat-Aviv, Israel
[b] PerSimiO Ltd, Beer-Sheva, Israel
[c] Department of Mechanical Engineering, Shamoon College of Engineering, Beer-Sheva, Israel
[d] Facutly of Medicine, Tel Aviv University and Sourasky Medical Center, Tel-Aviv, Israel

ARTICLE INFO

Dedicated to our friend, Prof. Leszek Demkowicz, on the occasion of his 70th birthday

Keywords:
Femur
Segmentation
U-Net
Deep learning
Autonomous finite element method

ABSTRACT

Efficient, accurate and reliable segmentation of femurs from CT-scans is of major importance for patient-specific autonomous finite element analysis (AFE) to determine bone's stiffness and strength. We present a fully automated segmentation algorithm for whole and partial femurs with or without tumors, and clinical applications of the AFE [1] in clinical practice.

The segmentation is based on an U-Net convolutional neural network, resulting a 3D mask representing the desired femur in a CT scan. It is robust, independent of the scanning parameters such as slice spacing, pixel size, scanner manufacturer or the femoral length available in the scan. The U-Net was trained on 178 manually segmented femurs (23,721 images) and tested on 43. The performance evaluation resulted in a Dice similarity score (DSC) of 0.9924, intersection over union (IoU) of 0.9849, Hausdorff distance of 4.3315 mm and symmetric average surface distance (ASD) of 0.1326 mm. The algorithm is competitive with the best state-of-the-art femoral segmentation methodologies available.

Based on the segmentation an automatic p-FE mesh is generated and physiological boundary conditions representing sidewise falls or stance are being applied automatically to improve the performance of the AFE described in [1]. New examples of the usage of the AFE in endocrinology and orthopedic oncology demonstrate this disruptive technology in actual clinical practice. We present the use of AFE for predicting hip fracture risk in the elderly population due to a sidewise fall and the identification of patients who require a prophylactic surgery due to metastatic tumors in their femurs.

1. Introduction

Patient specific finite element (FE) analyses were proven to be a powerful non-invasive tool for predicting the mechanical response of human femurs [2–11]. In a FE analysis physiological loads are virtually applied on the model and the mechanical response is computed allowing an assessment of femoral strength and fracture risk. Such patient specific information is valuable in granting suitable clinical treatment to elderly and diabetic patients [12,13] and patients with femoral tumors [11,14–16]. An accurate and automated segmentation algorithm is a prerequisite for such FE analyses, i.e. automatic acquisition of geometrical representation of the femur and inhomogeneous material properties from CT scans. Past femoral segmentation was mostly performed semi-automatically, demanding substantial manual intervention. This includes manual detection of the relevant CT slices, followed by global thresholding and manual boundary corrections. Such procedures are highly inefficient, user dependent and prone to human error. A fully automatic and efficient segmentation algorithm is therefore required to allow the application of autonomous FE simulations in clinical practice on a large scale [1].

Segmentation by global thresholding alone is insufficient for two main reasons. First, bone density varies significantly along the femur, and secondly, thresholding causes the femur and pelvis to appear connected. Many previous studies focused on separating the hip joint into its components using different techniques. Classical image processing algorithms based on image filtering and analysis aimed mainly at identifying the gap between the femoral head and pelvis were suggested by Zoroofi et al. [17] and Cheng et al. [18]. Gangwar et al. [19] introduced the use of phase field models, borrowed from fracture mechanics, to tackle the separation problem. Patch based algorithm was

* Corresponding author.
E-mail address: yosibash@tauex.tau.ac.il (Z. Yosibash).

https://doi.org/10.1016/j.camwa.2023.09.044
Received 28 March 2023; Received in revised form 11 July 2023; Accepted 23 September 2023

proposed by Chang et al. [20] aiming to classify the joint gap pixels into femoral and pelvic. Other common approaches for femoral segmentation were multi atlas (MA) algorithms [21–23] and statistical shape models (SSMs) [21,23–25]. MA algorithms are based on registration of pre labeled set of images on top of the segmented image [26]. SSM algorithms are based on fitting a statistically derived shape model on to the segmented image [27]. The methods described above achieved good results reporting Dice similarity coefficient (DSC) [28] in the range of 0.9036-0.976. Despite the high DSC scores which quantifies the volume overlap between the predicted mask and the ground truth, disagreement could commonly be found on bone's boundaries. These disagreements are particularly important for FE analyses since they contain the bone cortex which highly influences the mechanical behavior.

In recent years, the use of convolutional neural networks (CNNs) gained popularity in femur segmentation [1,29]. In particular the use of V-Net [30] architecture, which is a 3D extension of the U-Net [31]. The new trend perfected the segmentation further, driving the DSC above 0.9800. The use of V-Net for femoral segmentation was first suggested in [32] where a small cohort of only 20 femurs was used. In a recent study [33], a large cohort of 397 femurs was used, resulting in an impressive DSC score of 0.9815. The methods were improved even further in [34] by applying an additional spatial transformation onto the V-Net output (ST-V-Net) and reaching a DSC score of 0.9888. A main limitation of the V-Net models is that they were designed only for CT scans with a defined resolution and proportion (slice spacing of 3 mm and pixel size of 0.8 mm in [33,34]). CT scans with different resolution had to be resampled, thus commonly losing important information due to the large slice spacing. Additionally, the input size of the network was set to 32 slices, limiting the network's efficient application only to a short portion of the femur i.e. mainly to its proximal part.

A more flexible and accurate segmentation approach may be achieved by U-Net models. For example a 3-D U-net was considered in [35], trained on 60 CT scans (120 proximal femurs) and evaluated on 1147 (2294 proximal femurs), a remarkable number of CTs for the test-cohort. The CTs in [35] were for the proximal femur only, all from same Phillips CT scanners with same resolution 512×512, voxels of a size of $0.98 \times 0.98 \times 1\ mm^3$ and the number of slices ranges from 88 to 178. An excellent DSC score of 0.99 was obtained. The 3D segmentation is achieved by stacking the 2D output masks on top of each other thus enabling efficient segmentation of femurs independently of their length or CT slice spacing.

Femur segmentation from MRI scans, was performed by both 2D U-Net models and 3D V-Net models, and compared in [36] (where they were referred as to 2D convolutional neural network (CNN) and 3D CNN respectively). It was concluded that: "*The best performing 2D CNN with post-processing exceeds the precision and DSC of the best performing 3D CNN*", thus suggesting the possibility that U-Net based segmentation algorithms can outperform their 3D extension (V-Net). The flexibility benefits of 2D U-net models encourage to explore this possibility, and to extend the scope also to femurs with tumors. Furthermore, to the best of our knowledge past segmentation algorithms did not account for the automatic determination of slice range within a CT in which the femur is located - the relevant slices were extracted manually from patient's CT scan prior to segmentation. Nor the former algorithms determined anatomical points so to allow the application of forces and boundary conditions in a systematic manner.

Herein we present a fully automated, accurate and efficient segmentation algorithm for intact femurs and for femurs with tumors, be it if the entire femur is present in a CT scan or only a part of it. Since it is intended for an autonomous system, the algorithm first identifies the CT slices in which a femur is present and thereafter performs the segmentation, *independently of the scanning parameters such as slice spacing, pixel size, scanner manufacturer and the femoral length available in the CT scan*. The anatomical points are determined thereafter by an algorithm detailed in [1].

The segmentation algorithm is incorporated into the autonomous FE (AFE) algorithm presented in [1]. We highlight the AFE usage in clinical practice to determine the risk of hip fracture due to metastatic tumors, assisting orthopedic oncologists to determine the need for a prophylactic surgery. Another clinical application of the AFE is the opportunistic determination within the elderly population of the ones at high risk of hip fracture due to a sidewise fall. Clinical examples are presented.

2. The segmentation

The segmentation begins by retrieving the raw CT scan from a PACS database and ends by creating a 3D binary mask which indicates the precise location of the two femurs (left or right) in the scan. It may be applied to CT scans which contain the entire body of the patient where a full length femur is found as well as for CT scans which include only part of the body such as an abdominal CT scan containing only the proximal part of the femur. The segmentation pipeline consists of the following steps: (a) CT data loading and preprocessing. (b) Identification of the most distal femoral slice from which segmentation should start. (c) Segmenting each of the femoral slices from the bottom up until the pelvis is reached. (d) Stacking the 2D masks on top of each other to obtain one 3D mask representing the femur.

2.1. U-net design and architecture

A variation of the U-Net original architecture, firstly proposed by [31] is used as illustrated in Fig. 1. The main differences are (a) zero padding is used before convolution so the output image has the same dimensions as the input. (b) Batch normalization is performed following convolution operations [37]. The obtained network resulted in over 31 million trainable parameters. Basic terminology and principles are available in [38] and [39].

The U-Net contraction path is made of convolution blocks which are followed by max pooling. Each block is defined by 3x3 convolution filter applied to obtain 64/128/256/512/1024 channels (feature maps). Next, batch normalization and ReLU activation are applied. The convolution block is performed twice before 2x2 max pooling follows, thereby contracting the spatial dimension by half. The expansion path resembles a mirrored contraction path. The max pooling is replaced by transposed convolution (2x2, stride 2) which restores the spatial dimensions. The resulting outputs are concatenated with a copy of their corresponding pairs from the contraction path, thus forming skip connections (see [40]). Data from the contraction path contains higher resolution information, thus enabling its restoration via the convolution blocks that follow. The expansion path ends when the original input image dimensions are restored. A final 1x1 convolution with sigmoid activation is performed to obtain a 2D one channel probability map in which the values indicate pixels' probability of belonging to the femur. A threshold of 0.5 is thereafter used (during postprocessing) so to transform the output into a binary mask image.

2.1.1. Dataset

Manual masks of individual femoral CT slices were used as the ground truth for training and testing. CT scans from 98 subjects (40 male, 58 female) of age 15 to 94 years old (mean: 64 sd: 17). A total of 128 clinical CT scans (in supine position) were used. These resulted in 221 manually segmented femurs. Note that some patients had only one femur manually segmented and some patients were scanned more than once. The statistical analysis addresses individual femurs (n = 221, 87 male, 134 female) and not patients or CT scans. The CT scans were acquired at various medical centers, different CT scanners from four different manufacturers (Philips: n = 168, Siemens: n = 30, GE: n = 17, Toshiba: n = 6), with a scanning resolution varying among the scans. Slice spacing varied between 0.5 to 3.0 mm with a mean and standard deviation of 1.27 ± 0.41 mm. Pixel size varied between 0.51 to 1.18 mm with a mean and standard deviation of 0.9 ± 0.11 mm. The data included

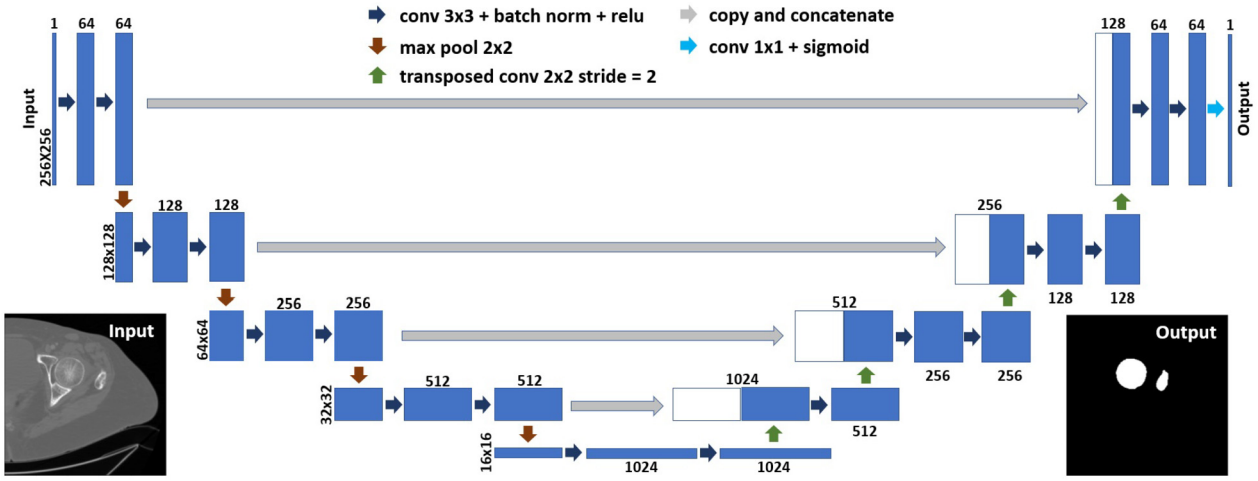

Fig. 1. U-Net architecture.

55 femurs with metastatic tumors of various types and stages. Such tumors commonly distorted the femoral intact geometry. Some of the scans included the entire femur whereas others included only the proximal part up to about 20 mm below the lesser trochanter. Since current FE analyses address the femoral proximal region and shaft, manual segmentation masks of the knee region were rare. All manually segmented femurs did include the proximal part which is most challenging for segmentation.

A testing set of 20% (n = 43) of the femurs was randomly selected. The rest of the data was kept for training. None of the patients from the testing set were included in the training set.

2.1.2. CT image preprocessing

The algorithm is designed so that each femur (left or right) is segmented separately and independently of its collateral. It is assumed that the side of the femur being segmented is known in advance. Before training/testing, the data was preprocessed: First 256x256 images were extracted around the center of mass of each femur at each CT slice. Since the side of the segmented femur is known, the U-Net was trained only on left femurs. Images obtained from the right femur were flipped. This provides two main advantages, first the image variation is reduced, second the amount of training data is doubled thus narrowing and focusing the network's task. To assure no high density objects (as metallic implants, keys, coins wires, etc) or artifacts are present, image CT number values above 3050 were replaced by 0. Pixels were normalized through division by the maximal value in each image (such a normalization with image values between 0 and 1 is a common practice proven to enhance performance). Ground truth masks had binary values (of either 1 or 0).

2.1.3. Training

Image selection

Each femur may be composed of hundreds of 2D images. Most of these are shaft images which usually are very similar. To impose an equal consideration to the proximal femur and to the femoral shaft (i.e. to balance the data), the number of shaft slices were reduced. Pelvic images with corresponding empty masks were also included in the training. These images are useful as indicators that segmentation has reached the pelvis and should be stopped. Knee images were also added when available. The classification of the images and their inclusion rate in training is summarized in Table 1. The different image types are illustrated in Fig. 2. A total number of 23,721 training images were used.

Training algorithm

The model was trained for 200 epochs using Dice loss function (see e.g. [41]). The Dice similarity coefficient (DSC) [28] is defined as the intersection of the masks (manual and predicted) multiplied by 2 and divided by the sum of elements in each mask. It varies between 0 and 1 with the latter indicating a perfect agreement. To define a loss function based on the Dice coefficient (Dice loss), two adjustments were made. First, since during training the minimum of the loss function is sought, the Dice coefficient must be multiplied by -1. Second, 1 should be added to the numerator and denominator. This is done to enable the prediction of an empty mask if for example a pelvic image is considered. In such case the Dice coefficient would be 1 indicating perfect agreement (division by 0 is also prevented). Dice loss is defined by:

$$Dice\ loss = -\frac{2|P \cap M| + 1}{|P| + |M| + 1} \tag{1}$$

Where P and M are the predicted and the manual mask images. Adam optimizer was used to seek the minima of the loss function ([42]). Learning rate was set to 10^{-4} and a batch size of 20 was used. The training data was randomly split into a training set (70%) and a validation set (30%). At each epoch, the model was evaluated on both sets. This was performed to monitor over-fitting. The model with the lowest loss on the validation set (validation loss) was kept as the best model.

To avoid over-fitting, data augmentation was performed. Before exposing the network to an image, the image was randomly transformed. Thus, assuring that the network is not exposed to the same exact image more than once. Image augmentation increases the robustness of the model and improves its generalization. The following transformation parameters and their corresponding bounds were used: rotation range: $\pm 90^o$, vertical and horizontal shift range: ±30%, shear range: ±10%, zoom range: ±50%. The high zoom range is particularly important since it enables the model to generalize for CT scans of different resolutions. During training, the model converged to validation loss of = -0.98927 (obtained after 175 epochs).

Testing

The model was tested on 20% of the manually segmented femurs (n = 43, 18 male, 25 female, 19 with tumor). Unlike the training process where each image was addressed as a separate case, at the testing stage the performance is evaluated for the entire 3D femur (composed of multiple 2D images). Here, all available manually segmented images were used (14,334 images and corresponding masks). The U-net output is a probability map with values between 0 and 1. In practice, values very rarely significantly varied from either 0 or 1. Still, a threshold of 0.5 was used to transform the output into a binary mask image. Following the sequential analysis of individual images, they were stacked on top of each other to construct a continuous 3D volume representing the segmented femur. Only the largest 3D component was kept, any detached outlier pixels were excluded. No additional post processing was performed.

Table 1
Training images definition and selection.

Image classification	Definition	Slices included	Training images (23,721)
Proximal	70 mm below top slice.	All	10,662
Shaft	All excluding proximal.	Images were skipped so that shaft image number is similar to proximal femur image number.	8,526
Knee	The five distal slices only if the femur is longer than 330 mm.	All except those that were already considered as shaft slices.	632
Pelvis	The 30 mm above the femur.	All	3,901

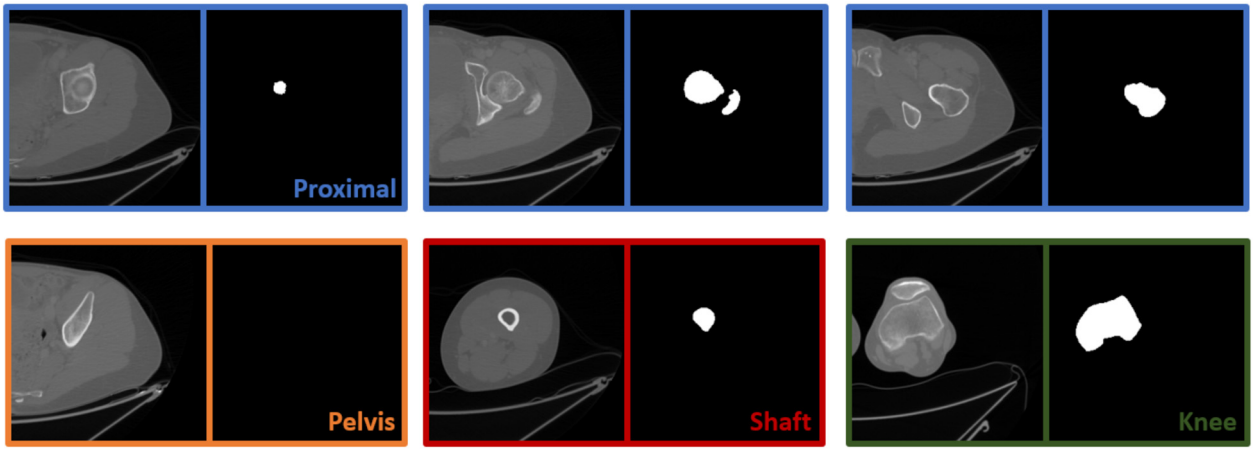

Fig. 2. Images from different areas of the CT scan together with the corresponding ground truth mask, defined by manual segmentation.

Segmentation performance evaluation metrics

Performance quantification is based on evaluation metrics as defined in [43] and [44]: (a) DSC also known as F1 score, (b) intersection over union (IoU), also known as the Jaccard index, (c) sensitivity, (d) specificity, (e) symmetric Hausdorff distance (HD), (f) symmetric and absolute average surface distance (ASD) also known as average symmetric surface distance (ASSD).

The aforementioned metrics vary from slice to slice along the femur and therefore depend on the evaluated volume. Generally, if one considers the femoral proximal regions such as the head and trochanters, the described measures would be inferior compared to the shaft. Thus, the larger the portion of the evaluated femur, the better the metrics become. Here, measures are reported for four different bone volumes defined by the femoral length measured from the most proximal femoral point: 50 mm - including mostly the head, 100 mm - from the top of the head until below the lesser trochanter, 150 mm also includes part of the shaft, and “All” including all the femoral length available (see Fig. 4). The metrics were computed based on the 43 femurs (the dataset): male femurs were compared to female femurs and intact femurs were compared to ones with tumors. Since the group variance was not exactly equal in all cases, the Welch’s t-test was used to examine the difference between the compared groups.

Efficiency evaluation

To evaluate the segmentation efficiency, processing time was measured by both GPU (Nvidia GeForce RTX 2080 Ti) and CPU (Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz).

2.1.4. U-Net imbedded in an automatic clinical CT scan segmentation

For longitudinal AFE clinical applications it is of outmost importance to have an automatic and uniform protocol which is analyst independent. Thus, the U-Net algorithm is imbedded into a comprehensive automated segmentation pipeline of clinical CT scans, as illustrated in Fig. 3. One of the main important steps in such a pipeline is the identification of the distal slice from which segmentation should start. This slice is identified by locating the gap between patient’s legs (femurs) and thereafter identifying the knee. No knee found implies that the femur is not present at its full length thus the most distal slice of the CT scan is chosen.

The U-net algorithm is constructed so to be applied only on patient’s body, thus any dense foreign objects in the CT scan, such as for example the CT bed itself (in some CT scanners), is removed. Such objects are removed by defining a mask which includes only the patients body. A threshold of 1200 CT number (corresponds to a very low Hounsfield Unit) is applied to the data (values below are set to zero). This removes soft tissue, particularly in the contact between the patient and the CT bed. The remaining CT values (corresponding to all bone tissues) are summed over z’ direction (slice out of plane direction) and transformed into a binary mask as shown in Fig. 3 (b). The resulting mask is a projection of the patient and the bed onto a slice sized image. Any holes found in the resulting mask are filled. Binary erosion followed by dilation operations is performed to assure smoothness and separation between the bed and body. The mask is then restricted only to its largest component thereby leaving only the patient’s body. It should also be noted that only the top portion of the patient’s length/hight (1100 cm) is considered to generate the described mask. This is performed to avoid including the feet, which are commonly hanging off the bed thus forming a connection between bed and body.

The obtained mask is applied to all CT slices. To position the patient’s body at the center of the image, all images are uniformly shifted such that the center of the mask coincides with the center of the image. Images are thereafter cropped into half, keeping only the right image side (which includes the left femur). We did not consider any CNN classifier to separate the left and the right femur in the CT images because the position of the patient inside the CT scan is known and patients position registration is an inherent part of the scan protocol. If the right femur must be segmented, the CT slices may be flipped before segmen-

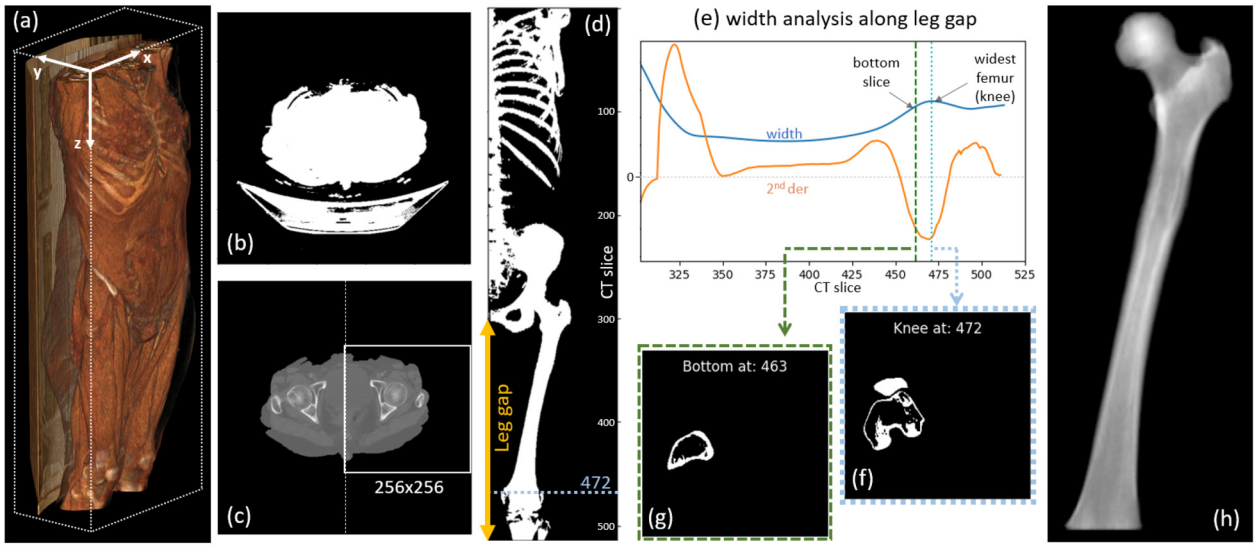

Fig. 3. (a) CT data is loaded into a 3D array. (b) A binary projection only of hard tissue onto the xy plane. After additional processing and CT bed removal, the mask is applied to all CT slices. (c) A masked CT slice with the cropping boundaries is illustrated. (d) The leg gap is illustrated inside the frontal image. The image is obtained through a binary projection onto the xz plane of patients hard tissue forming a map of "bone presence". (e) Bone width along the leg gap is plotted (after moving average). Second derivative is also shown (after scaling by 1000). Knee slice is identified. (f) A binary image of the knee slice with shaft and patella components (threshold CT number > 1200). (g) Bottom slice for segmentation is identified above the knee slice being the first slice without the patella. (h) The Fully segmented 3D femur. Images are processed by the U-Net from the bottom upwards until an empty mask is returned. Femur is obtained from the CT based on the 2D mask outputs.

tation (results are flipped back after segmentation). A masked CT slice with the cropping boundaries is illustrated in Fig. 3 (c). Through a series of thresholding and summation operations, the cropped images are transformed into a 2D "frontal image" binary map, representing bone presence. Note that if the bed would not have been removed, it would have interfered with the resulting image.

The longest consecutive gap (defined by missing bone) at the left side of the "frontal image" (middle of the patient) is identified and measured. This is most likely to be the gap between the patients legs (femurs). If the gap is longer than 150 mm, a "long" bone is identified as illustrated in Fig. 3 (d). Otherwise, bone is classified as short. For short bones the distal slice is initiated at the very bottom of the CT scan. For long bones it is initiated at the bottom of the leg gap and will shift only if a knee is identified. To identify the knee, pixel values are summed over the rows of the frontal image, representing a measure of bone "width". The width (as a function of z') is plotted along with its second derivative as illustrated in Fig. 3 (e). Based on the plotted data, the algorithm identifies the first width peak. This peak corresponds with the widest femoral region which anatomically matches the middle of the knee.

The identified knee slice includes both the femur and the patella (Fig. 3 (f)). It is used as the lower bound to seek the distal slice for segmentation. Here we defined the distal image from which segmentation should start as the first image above the knee slice which has only one component (femoral shaft without patella). The bounding box of the single component in the 'x' direction (horizontal) must be greater than in the 'y' (vertical) direction by at least 10% illustrated in Fig. 3 (g). The object count alone is an insufficient criterion since the patella may appear connected to the femur. Note that the pipeline described here is adjusted to produce an output fit for finite element model construction without the knee.

Segmentation is performed from the bottom slice upwards until an empty mask is returned indicating that the pelvis is reached and segmentation may be stopped. A segmented 3D femur is shown in Fig. 3 (h).

2.2. Results

The average metrics for the 43 "test" femurs are reported in Table 2 for different femoral portions defined by their length. Seven of the 43 femurs were segmented from an abdominal CT where the femur was shorter than 150 mm. These were excluded from the statistics for the proximal 150 mm segment. When comparing male to female or intact femurs to ones with tumors, metrics were found to be very similar. No statistically significant differences were observed (based on the student's *t*-test). An example of fully segmented long and short femurs together with their evaluation metrics are presented in Figs. 4 and 5. Individual image masks are shown in Fig. 6. These include slice images of healthy typical bones as well as images of bones with tumors.

2.2.1. Algorithm efficiency

The U-Net processes a single image in less than 0.4 seconds on a CPU (Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz). GPU (Nvidia GeForce RTX 2080 Ti) process time was about 0.05 seconds (up to 8 times faster). A full length femur having 1 mm slice spacing (about 500 slices) would be processed in 25 seconds on the GPU and in 185 seconds on the CPU. The rest of the operations in the pipeline (besides U-Net processing) were performed on CPU and added 0.05 seconds per slice on average (see the pipeline running times provided in Figs. 4 and 5).

2.3. The segmentation algorithm within the framework of AFE in a real clinical environment

An important evaluation of a segmentation algorithm is its robustness in a real clinical environment. The robustness was tested on preliminary data of a clinical trial aimed at determining the risk of fracture in femurs of the elderly population. The segmentation algorithm has been implemented into the AFE solver *Simfini*.[1] The automatic identification and segmentation of the femurs are the first two steps in the entire AFE procedure, followed by identification of anatomical points, application of a fall on the side loading condition, meshing the femur

[1] Simfini is a trademark of PerSimiO, Beer-Sheva, Israel.

Table 2
Evaluation metrics considering 43 "test" femurs are reported for the different femoral portions (7 femurs that were taken from abdominal CT scans containing only the proximal shaft were excluded from "150 mm statistics"). Mean and standard deviation (parentheses) are reported. Hausdorff distance (HD) and symmetric and absolute average surface distance (ASD) are given in mm.

length mm	DSC		IoU		Sensitivity		Specificity		HD		ASD	
50	0.9882	(0.0043)	0.9766	(0.0083)	0.9875	(0.0054)	0.9996	(0.0003)	3.9473	(1.407)	0.2211	(0.0684)
100	0.9908	(0.003)	0.9817	(0.0058)	0.9891	(0.0046)	0.9997	(0.0002)	4.1894	(1.5293)	0.1748	(0.0474)
150	0.9913	(0.0026)	0.9827	(0.0052)	0.9892	(0.0049)	0.9998	(0.0001)	4.2133	(1.444)	0.1526	(0.0546)
full	0.9924	(0.0021)	0.9849	(0.0042)	0.9889	(0.0045)	0.9999	(0.0001)	4.3315	(1.5596)	0.1326	(0.0332)

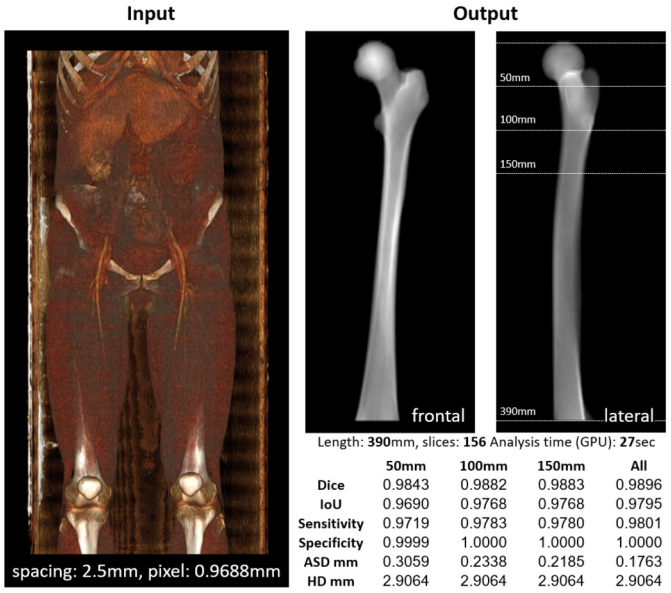

Fig. 4. Segmentation performance is demonstrated on a CT scan of a 41 years old female. A long left femur is segmented. Evaluation metrics are reported for different femoral volumes. The total analysis time is 27 seconds on a GPU (Nvidia GeForce RTX 2080 Ti).

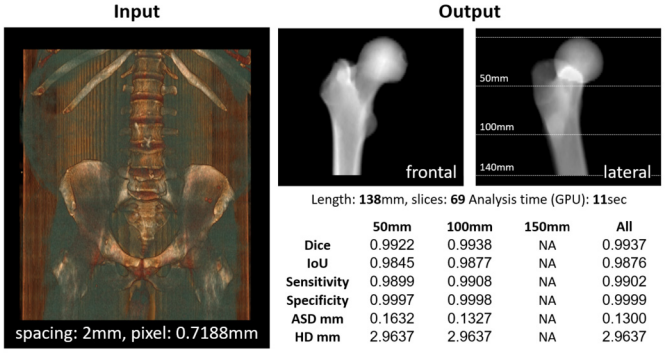

Fig. 5. Segmentation performance is demonstrated on a CT scan of a 60 years old male. A short left femur is segmented. Evaluation metrics are reported for different femoral volumes. The total analysis time is 11 seconds on a GPU (Nvidia GeForce RTX 2080 Ti).

domain, perform a finite element analysis and post-process the results to identify the risk of fracture. Details on the AFE are provided in [1,11] and a flowchart is shown in Fig. 7. The AFE solver was applied to 484 CT scans (968 femurs) of lower abdomen (most of which contained only the proximal part of the femur), from very different scanners (different models of GE and Philips) with different pixel size resolutions (ranging from 0.6 to 3.0 mm slice thickness), and a wide range of patients age and weight. A successful outcome of the AFE was obtained for over 91% of the CT scans [45]. Since most of AFE analyses failed after a successful segmentation (in the meshing stage or if only one of the femurs was successfully segmented), the segmentation success rate is better than 95%, which is considered to be a high robustness rate (compared for example with 87% in [46] - there vertebrae were considered on a larger cohort).

2.4. Examples for the use of AFE in clinical practice

The presented improved segmentation algorithm replaced the convolution neural network segmentation algorithm of the AFE system *Simfini* presented in [1]. The AFE in [1] is briefly summarized with an emphasis on enhancements enabling endocrinologists to determine risk of hip fracture in the elderly population. Once femurs segmentation is available, a finite element mesh consisting of tetrahedrons with curved faces is automatically generated for the patient, followed by an efficient high-order linear elastic FE solver that solves the system of finite element equations and generates the data of interest.

Linear elasticity well represents femur's response under physiological loading, and although the bone at the macroscopic level is orthotropic, excellent predictions were obtained using isotropic inhomogeneous relations (see [47] for stance position loading and [48] for sideways fall loading). Pointwise inhomogeneous mechanical properties at each integration point within each element are determined by the Hounsfield Units (HU) in the closest voxel in the CT scan. The relationships between Young's modulus and HU, validated in experimental settings [47], are used. Verification of the numerical errors is assured by automatic monitoring of the error in energy norm and the maximum and minimum principal strains at the locations of interest as the polynomial degree over the elements is increased from 1 to 6 or 8.

The AFE determines anatomical points (center of femur's head, intercondylar notch, and center of shaft 20 mm below the lesser trochanter), required for the application of the different boundary conditions. In addition to a stance position load that acts along the vector connecting the head center and intercondylar notch, two other load directions were implemented to mimic sidewise falls, defined by the angles γ and δ shown in Fig. 8. The angle γ represents the angle between the shaft and the ground during impact and is related to the amount of knee flexion present if the foot is on the ground, and δ reflects the amount of internal or external rotation of the femur relative to the ground. It has been found that neck fractures are associated with one set of specific angles, (γ_N, δ_N), where $\gamma_N = 10^o$ and $\delta_N = 15^o$, while pertrochanteric fractures are associated with another set of specific angles, (γ_P, δ_P), where $\gamma_P = 30^o$ and $\delta_P = 45^o$. The distal shaft is constrained for displacements along shaft axis and at the contralateral location of the load, in the load direction. Almost half of the hip fractures due to a sidewise fall are neck and half are pertrochanteric.

We present in Fig. 9 two examples of femurs of two patients automatically segmented from lower-abdomen CT scans, with the three different loadings presented (stance and two fall on the side) that are solved sequentially. In Table 3 the three different boundary conditions are summarized.

Each model has about 9000-10,000 high order finite elements resulting in about 900,000 degrees of freedom at p = 6. The entire simulation time including the pre and post processing for both femurs for one patient is about one hour on a standard PC.

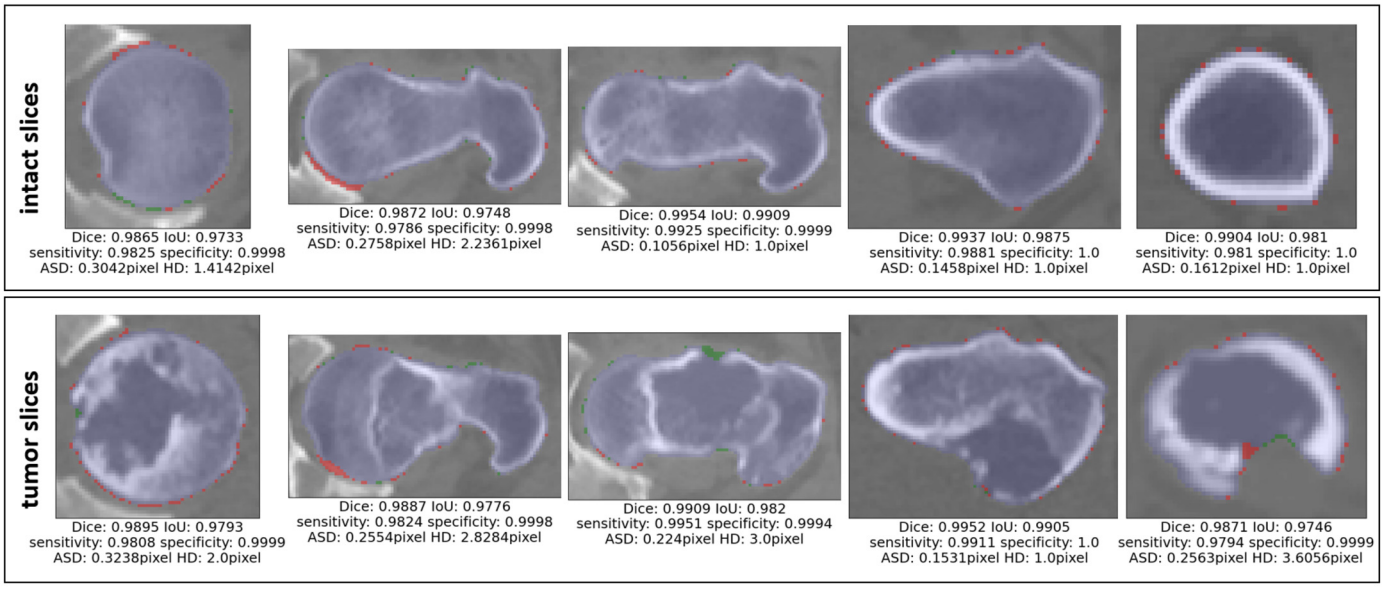

Fig. 6. The agreement between manual segmentation and U-net predictions is demonstrated on slices of intact femurs (top) and femurs with tumors (bottom). Blue pixels represent the intersection of both masks. Red pixels are exclusive to the manual segmentation and green are exclusive to U-Net predictions. Evaluation metrics are reported for each image. ASD and HD are reported in pixels.

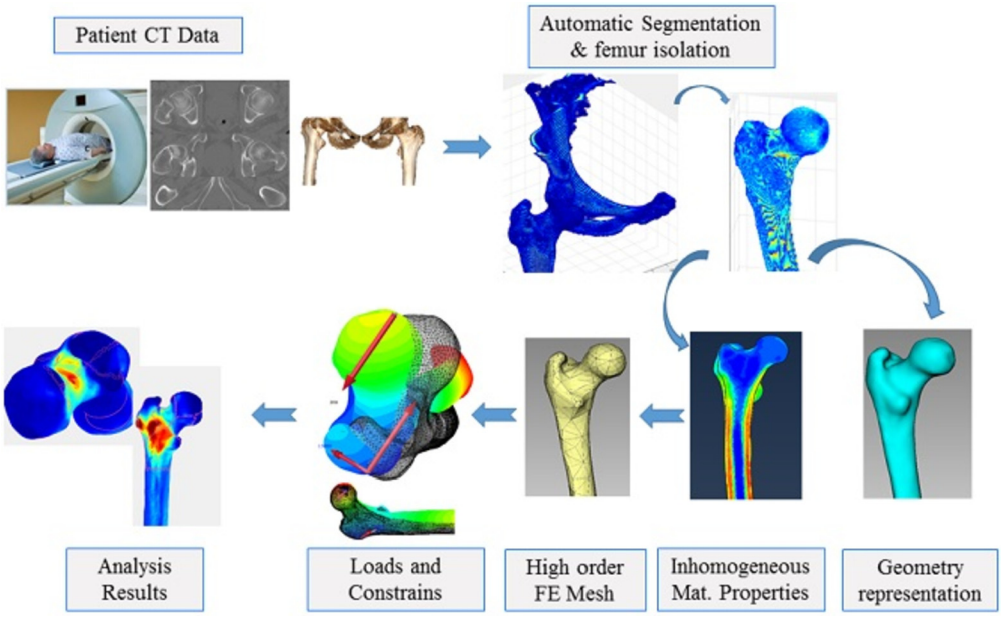

Fig. 7. Schematic algorithm (from [13]) of the patient-specific AFE system *Simfini*. a) CT scan of patient's femurs, b) Automatic segmentation, c) Geometry and material representation, d) High-order FE mesh, e) Application of loading boundary conditions, f) Principal strains extraction.

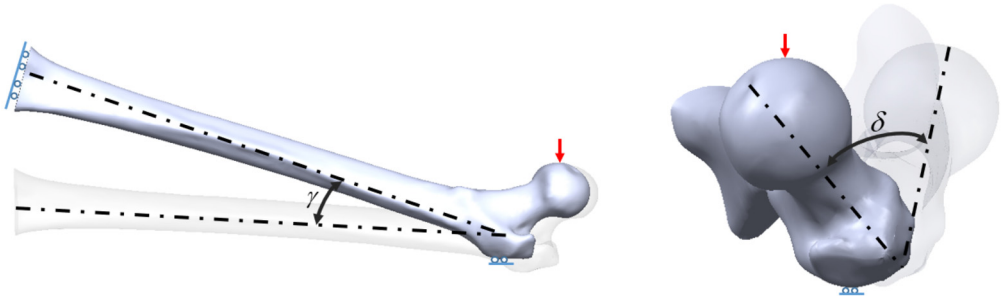

Fig. 8. Boundary conditions on the femur that determine sidewise falls.

Table 3
Summary of the three physiological boundary conditions.

Physiologic load	Magnitude	Load application	Load Direction	Displacements BCs.	Details
Stance	2.5 body weight	Femoral head on top	Vector connecting head center and intercondylar notch.	Clamped at distal Shaft	
Sidewise Fall	2.5 body weight	Femoral head	$\gamma_N = 10^o$, $\delta_N = 15^o$ $\gamma_P = 30^o$, $\delta_P = 45^o$	Clamped at distal shaft, and zero displacement along load direction at greater trochanter	Fig. 8, 9

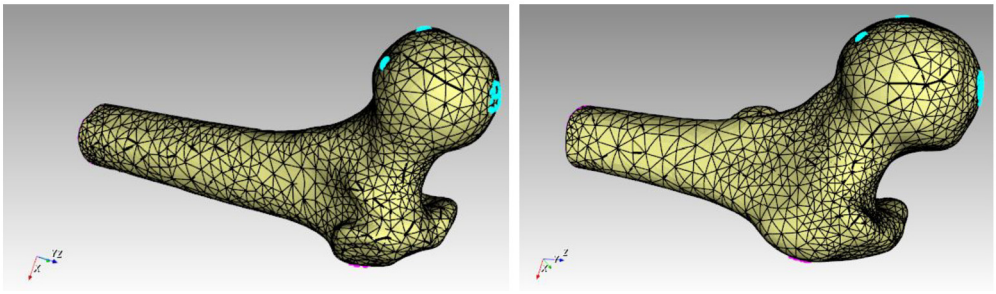

Fig. 9. Stance and two sidewise fall loads (light blue) and displacements boundary conditions (red) on the FE meshes representing two different patients.

2.4.1. The need of a prophylactic surgery - femurs with tumors

Patients with tumors in their femur may be at risk of a pathological fracture under daily activities. For these patients an AFE is performed under a stance position loading. The "typical median principal strains" for 5 regions (neck, trochanter, proximal shaft, middle shaft and distal shaft) in a "healthy femur" were computed based on 12 femurs as detailed elsewhere [11]. The ratio between the absolute maximum principal strain in the diseased femur and the median strain in the same anatomical region of the disease-free femur is calculated and labeled the "strain fold ratio" (SFR). A SFR value of 1.48 was used as the determinant of the threshold for a pathological femoral fracture [11]. The SFR is computed for both femurs of each patient. The location at which the highest SFR (larger than 1.48) is obtained in the AFE is estimated to be the location of the expected pathological fracture (in some cases there were more than one region at which SFR>1.48). This algorithm has been shown to well predict the risk of a pathological fracture, and is being used in clinical practice to assist orthopedic oncologists to determine the need of a prophylactic surgery. Recently, AFE has been extended to monitor patients who undergo radiation therapy. As an example we present the case of a seventy-four years old male (weight of 80 kg) with multiple myeloma lytic bone lesions who was recommended radiation therapy. Two months following the radiation therapy treatment, he complained about pain in his right femur. The orthopedic oncology surgeon, by inspecting visually the CT scans after the radiation therapy and two months later, could not detect any differences and performed an AFE analysis. In Fig. 10 one may see there is no much differences in the CT slices at the location of interest taken two months apart. However, the AFE analysis detected a high SFR at the right femur well above 1.48, and furthermore, exactly at the location of interest the SFR increased by 30% at one location between the two CT-scans. The ration between the SFRs computed using the two CT scans remained around 1 or less in the left femur and in all other locations in the right femur. This indicated the need for a prophylactic surgery which the orthopedic oncologist surgeon performed immediately following the AFE analysis, as shown in the right bottom corner of Fig. 10.

2.4.2. The need of a prophylactic surgery - young patient with irregular shaped femur

By using the SFR, there are cases where an orthopedic surgeon has to decide on the need of a prophylactic surgery to stiffen the femur if weakening is assumed. For example, a thirty-three years old female (weight of 74 kg) had in her past two surgeries to both left and right femurs, and in a recent CT scan an anomalous configuration was noticed in her left femur during a monitoring CT scan, see left of Fig. 11. The surgeon could not decide whether to perform a prophylactic surgery or not, therefore an AFE was performed. The results clearly show that the SFR in the left femur is well bellow 1 in the area of interest. No surgery was performed, and so far the patient does not complain of any discomfort or problem.

2.4.3. Risk of hip fracture in the elderly as a result of a sidewise fall

Post processing the AFE results obtained for all three boundary conditions (stance and two sidewise falls) by a machine-learning algorithm (support vector machine (SVM)), considering also the gender, height, weight and age, can provide an unprecedented ability to differentiate between the elderly population at a high risk of hip fracture as a result of a sidewise fall. To this end we analyzed by AFE abdominal/pelvis CT scans of patients who experienced a hip fracture within two years after acquiring a CT scan together with a control group of patients without a known hip fracture for at least five years after acquiring a CT scan. Overall 836 femurs were analyzed (432 intact femurs and 404 fractured ones). The results were processed by the SVM algorithm. These CT scans only contained the proximal part of the femurs, usually having no more than 2-3 cm distal to the lesser trochanter. Details are given in [45]. The sensitivity and specificity of the SVM based on AFE to determine fracture risk are 86% (with a 95% CI interval of 73-89%) and 79% (with a 95% CI interval of 75-82%). Comparing to the gold-standard DXA for example used nowadays in clinical practice, having a sensitivity in women of 52% (with a 95% CI interval of 47-56%) and specificity of 77% (with a 95% CI interval of 73-81%) and having a sensitivity in men of 43% (with a 95% CI interval of 37-50%) and specificity of 83% (with a 95% CI interval of 77-88%) [51], our proposed AFE measures are by far better.

Combining AFE data with a ML algorithm provides an unprecedented prediction accuracy for the risk of hip fracture in the elderly populations. The fully autonomous algorithm can be applied as an opportunistic process on lower abdomen CT scans for hip fracture risk assessment.

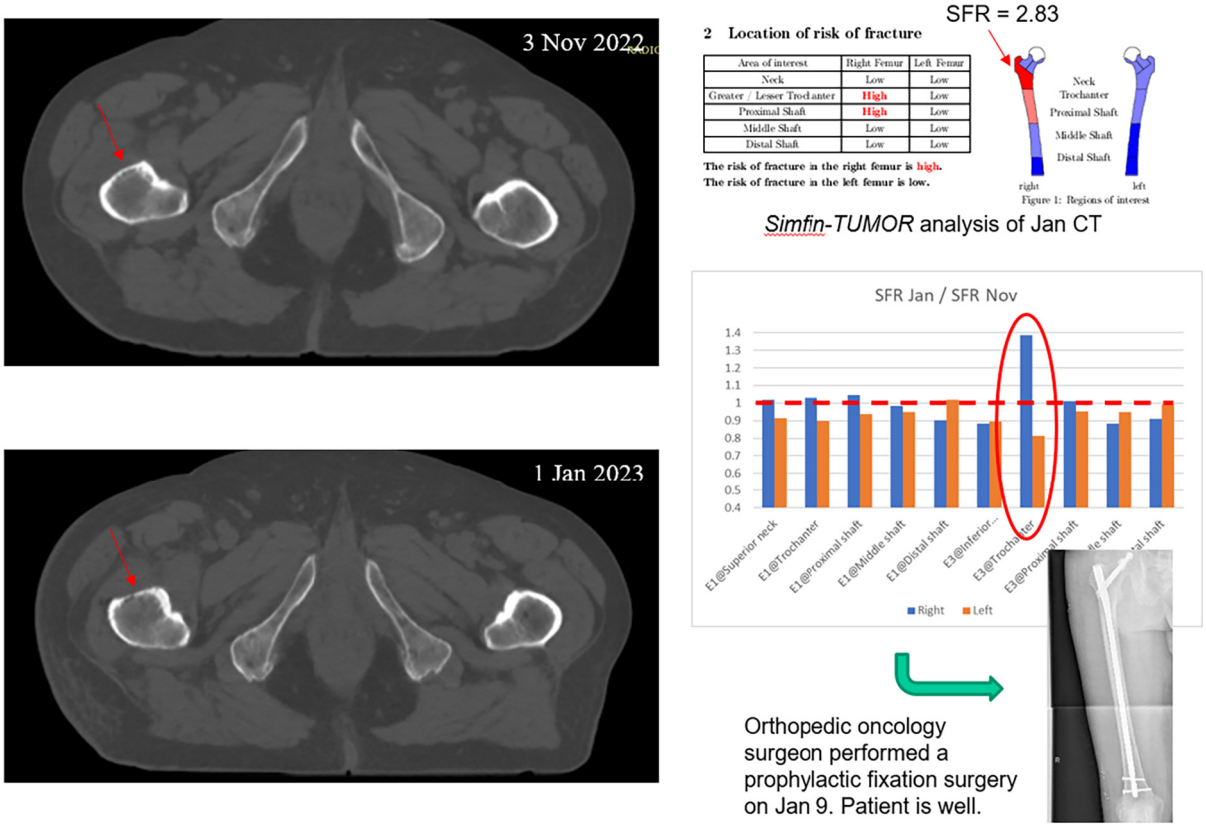

Fig. 10. Seventy-four years old male patient with multiple myeloma lytic bone lesions. Left - two CT scans two months apart. Right - AFE results, the increase in SFR at a specific location and X-ray following the prophylactic surgery that strengthen the right femur.

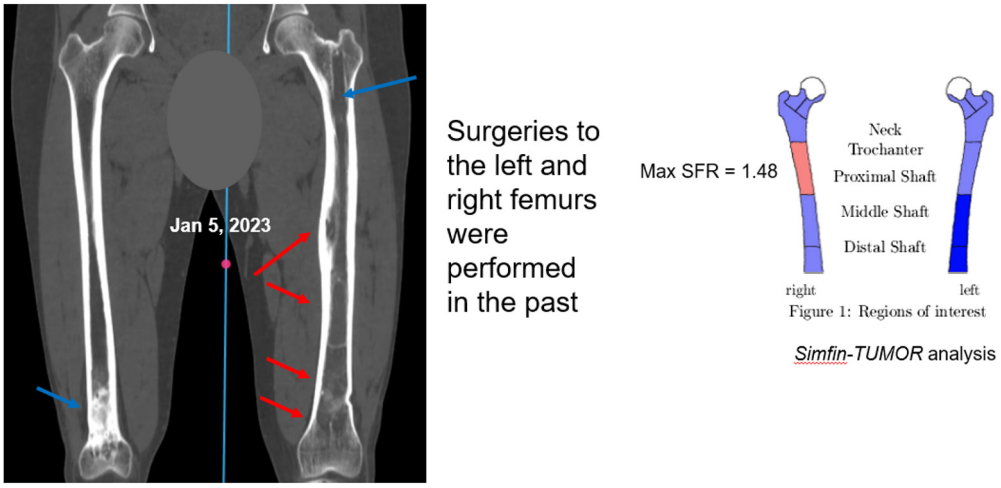

Fig. 11. Thirty-three years old female patient with anomalous left femur due to metabolic bone disease. Left - CT scan. Right - AFE results, demonstrating sufficient strengthen of the left femur.

3. Discussion and conclusions

A fully automatic and robust segmentation for femurs from clinical CT scans was presented implemented and used in an AFE. An excellent performance for intact femurs and femurs with tumors was demonstrated. It is appropriate for a large range of image resolutions and is independent of the slice spacing and femoral length. The methodology may be easily generalized to other bones as vertebrae, humeri, etc. The advantage of the presented algorithm compared to state-of-the-art algorithms, see [34] for example, is its flexibility and robustness. It is not restricted to a specific slice spacing or pixel size, was shown to perform well for femurs with or without tumors, and tested on CT scans from different CT scanners. It is fully autonomous and it accounts also for the identification of the femoral slices inside the CT scan (most former studies did not consider this important ingredient for a fully automatic algorithms, see e.g. [17–25,29,32–36,49]).

Because performance metrics improve as larger parts of the femur are considered (including more shaft domain), a metric which distinguishes between the different femoral portions was proposed (Table 2). A comparison to former publications is presented in Table 4. To emphasize the performance of our algorithm we adopted a conservative approach (metrics considering the proximal femoral portion, top

Table 4
Comparison of performance to former segmentation algorithms. A conservative femoral length of 100 mm was chosen for comparison. Best performance in boldface.

Study	Femur length	Method	DSC/ Dice	IoU/ Jaccard	Sensitivity/ recall	Specificity	HD (mm)	ASSD/ ASD (mm)
[17], evaluated in [18]	Proximal	Classic methods	0.9036 (0.0531)	NA	NA	NA	NA	1.31 (1.21)
[24], evaluated in [18]	Proximal	SSM	0.9014 (0.0195)	NA	NA	NA	NA	1.49 (1.04)
[25]*	Full length	SSM	NA	NA	NA	NA	NA	1.21 (0.35)
[18]	Proximal	Classic methods	0.9155 (0.0482)	NA	NA	NA	NA	1.22 (0.98)
[21]*	Proximal	MA	0.950 (0.024)	NA	0.956 (0.041)	0.997 (0.001)	NA	0.846 (0.456)
[21]*	Proximal	SSM	0.946 (0.013)	NA	0.964 (0.020)	0.998 (0.001)	NA	0.910 (0.309)
[50]	Full length	SSM	0.87 (0.026)	NA	NA	NA	10.53 (3.19)	1.48 (0.28)
[22]	Proximal	MA	0.976 (0.006)	0.953 (0.011)	0.97 (0.01)	NA	NA	0.203 (0.057)
[23]	Proximal	MA SSM combined	0.9735 (0.0085)	NA	NA	NA	NA	0.465 (0.900)
[49]	Full length	SSM	0.94 (0.016)	NA	NA	NA	4.336 (0.861)	1.014 (0.474)
[32]*	Proximal	V-Net	0.987 (NA)	0.974 (NA)	0.982 (NA)	NA	6.4 (NA)	0.22 (NA)
[36]*	Proximal	V-Net	0.953 (0.016)	NA	0.953 (0.030)	NA	7.88 (4.33)	0.39 (0.19)
[36]*	Proximal	U-Net	0.960 (0.022)	NA	0.943 (0.036)	NA	8.18 (5.87)	0.39 (0.44)
[19]	Mid length	Phase-field	0.9339 (0.0287)	NA	0.9339 (0.0287)	0.9855 (0.0115)	NA	NA
[20]	Proximal	Patch based	0.9490 (0.0070)	NA	NA	NA	NA	0.52 (0.08)
[29]	Full length	FCN	0.9688 (0.0095)	NA	NA	NA	NA	0.41 (0.11)
[1]	Full length	CNN	0.98 (0.003)	NA	NA	NA	NA	0.36 (0.05)
[33], evaluated in [34]	Proximal	V-Net	0.9815 (0.0009)	NA	0.9906 (0.0033)	0.9990 (0.0000)	9.144 (2.096)	NA**
[34]	Proximal	ST-V-Net	0.9888 (0.0047)	NA	0.9966 (0.0013)	0.9988 (0.0001)	5.917 (1.412)	NA**
[35]	Proximal	U-net	0.990 (0.008)	NA	NA	NA	5.9-10.6***	NA
Current study	**Proximal (100 mm)**	**U-Net**	**0.9908 (0.0030)**	**0.9817 (0.0058)**	0.9891 (0.0046)	**0.9997 (0.0002)**	**4.19 (1.53)**	**0.1748 (0.0474)**

* Magnetic resonance imaging (MRI) used (not CT).
** Although reported as ASD in the papers, distances were addressed as relative (not absolute) and only from predictions to ground truth (not symmetric). Positive and negative distances cancel out each other. Thus providing a misleadingly better measure which can't be considered ASD.
*** Only the 95th percentile is reported HD95 = 1 mm, but two cases are shown with HD95 5.9 mm and 10.6 mm.

100 mm). A test set of 43 femurs was used, similar to other recent high performing algorithms such as [33,34] (n = 40) and [1] (n = 41). To the best of our knowledge, the performance metrics obtained for the presented algorithm are close to optimal for femoral segmentation and also to [35] tested on over 1000 femurs. Under the constraint of the limited sample size, no evidence for difference between the compared groups was observed.

Since 3D images contain more details, it is conceivable that 3D V-Net models, such as used in previous deep learning based algorithms [32–35] might be more suitable for femoral segmentation. On the other hand, a major downside of 3D inputs is their large size which demands a significant amount of memory. Consequently, the large input size comes at the expense of the depth and capacity of the network. In other words, it will always be possible to construct deeper U-Net models than V-Net models. For example, the V-Net suggested in [33] and [34] had half the number of convolution filters in each convolution layer compared to the U-Net, and only 3 down/up steps in the contraction/expansion paths (four such steps in the U-Net). The V-Net resulted in 8.6 million trainable parameters whereas the U-Net had over 31 million (GPUs with similar memory were used). The large 3D inputs also affect the efficiency of the training procedure, imposing smaller training batches sizes. The amount of training data is also smaller, each femur is equivalent to one 3D image. With 2D inputs, hundreds of images may be extracted from the same femur. U-Nets, heaving inputs smaller in size and less complex, allow the network to adapt for larger image variability, and construct deeper networks with greater capacity and flexibility. Thereby, enabling segmentation of a larger variety of image types i.e. images of different resolutions, obtained from different bone regions and with different bone anomalies caused by various diseases. Based on

the numerous advantages, we advocate the use of 2D U-Net models for femoral segmentation instead on 3D V-Net models.

The current algorithm is also more efficient and more accurate than the CNN algorithm presented in [1]. The latter was designed to classify hip joint voxels into pelvic and femoral. The hip joint region includes hundreds of thousands of such voxels that must be classified. Thus, the CNN segmentation becomes very time consuming and strongly dependent on GPU resources for efficiency improvement. The U-Net processes an entire slice at each iteration, thus substantially increasing efficiency enabling the segmentation to be performed within a reasonable amount of time even on a CPU (below 0.4 seconds per slice).

We noticed that in numerous cases the mask predicted by the U-Net was more accurate than the manual mask, several such cases may be found in Fig. 6. This is not surprising since manual segmentation is a long exhausting procedure which may include human errors. The U-Net on the other hand, adapts itself to recognize the patterns common to the majority of cases thus neglecting human random errors.

Combining the improved automatic segmentation in the AFE presented in [1] resulted in an improved AFE that allows surgeons to use advanced computational mechanics in daily clinical practice for patient specific medicine. We presented two recent real life applications of the use of AFE that allowed orthopedic surgeons to scientifically determine the need of a prophylactic surgery. In one case a patient was identified just before a pathological fracture and in the other case an unnecessary surgery was avoided in a young patient.

We also enhanced the AFE to consider the elderly population by including sidewise fall loading conditions. This enhancement, together with a clinical study on 836 different femurs and an SVP ML algorithm showed to predict with high sensitivity and specificity these with a high risk of hip fracture within the next 2 years as a result of a sidewise fall. The AFE opens the path for accurate and efficient fracture risk prediction by opportunistic analysis of lower abdomen CT scans in the elderly population.

Data availability

Data will be made available on request.

Acknowledgements

YK acknowledges the partial funding by a “Pilot study grant” awarded by the Israel Innovation Authority. ZY and AS acknowledge the support of this research by the Israel Ministry of Science and Technology under the Tenth Call of Israel-Italy Scientific collaboration.

References

[1] Z. Yosibash, K. Myers, N. Trabelsi, et al., Autonomous FEs (AFE) - a stride toward personalized medicine, Comput. Math. Appl. 80 (11, SI) (2020) 2417–2432.
[2] J.H. Keyak, T.S. Kaneko, S.A. Rossi, et al., Predicting the strength of femoral shafts with and without metastatic lesions, Clin. Orthop. Relat. Res. 439 (2005) 161–170.
[3] M. Bessho, I. Ohnishi, J. Matsuyama, et al., Prediction of strength and strain of the proximal femur by a CT-based finite element method, J. Biomech. 40 (8) (2007) 1745–1753.
[4] J.H. Keyak, S. Sigurdsson, G. Karlsdottir, et al., Male-female differences in the association between incident hip fracture and proximal femoral strength: a finite element analysis study, Bone 48 (2011) 1239–1245.
[5] E. Schileo, L. Balistreri, L. Grassi, et al., To what extent can linear finite element models of human femora predict failure under stance and fall loading configurations?, J. Biomech. 47 (14) (2014) 3531–3538.
[6] K. Engelke, B. van Rietbergen, P. Zysset, FEA to measure bone strength: a review, Clin. Rev. Bone Miner. Metab. 14 (2016) 26–37.
[7] M.J. de Bakker, W.J. Tseng, L. Yihan, et al., Clinical evaluation of bone strength and fracture risk, Curr. Osteoporos. Rep. 15 (2017) 32–42.
[8] G. Iori, L. Peralta, A. Reisinger, et al., Femur strength predictions by nonlinear homogenized voxel finite element models reflect the microarchitecture of the femoral neck, Med. Eng. Phys. 79 (2020) 60–66.
[9] N. Trabelsi, Z. Yosibash, C. Wutte, et al., Patient-specific finite element analysis of the human femur - a double-blinded biomechanical validation, J. Biomech. 44 (2011) 1666–1672.
[10] Y. Katz, Z. Yosibash, New insights on the proximal femur biomechanics using Digital Image Correlation, J. Biomech. 101 (2020).
[11] A. Sternheim, O. Giladi, Y. Gortzak, et al., Pathological fracture risk assessment in patients with femoral metastases using CT-based finite element methods. A retrospective clinical study, Bone 110 (2018) 215–220.
[12] Y. Lee, N. Ogihara, T. Lee, Assessment of finite element models for prediction of osteoporotic fracture, J. Mech. Behav. Biomed. Mater. 97 (2019) 312–320.
[13] D. Rotman, G. Ariel, J. Rojas Lievano, et al., Assessing hip fracture risk in type-2 diabetic patients using CT-based autonomous finite element methods, Bone Joint J. 103-B (9) (2021) 1497–1504.
[14] Z. Yosibash, R. Plitman Mayo, G. Dahan, et al., Predicting the stiffness and strength of human femurs with realistic metastatic tumors, Bone 69 (2014) 180–190.
[15] H. Schermann, Y. Gortzak, Y. Kollender, et al., Patient-specific computed tomography-based finite element analysis: a new tool to assess fracture risk in benign bone lesions of the femur, Clin. Biomech. 80 (2020) 105155.
[16] A. Sternheim, F. Traub, N. Trabelsi, et al., When and where do patients with bone metastases actually break their femurs? A CT-based finite element analysis, Bone Joint J. 102B (5) (2020) 638–645.
[17] R.A. Zoroofi, Y. Sato, T. Sasama, et al., Automated segmentation of acetabulum and femoral head from 3-D CT images, IEEE Trans. Inf. Technol. Biomed. 7 (4) (2003) 329–343.
[18] Y. Cheng, S. Zhou, Y. Wang, et al., Automatic segmentation technique for acetabulum and femoral head in CT images, Pattern Recognit. 46 (11) (2013) 2969–2984.
[19] T. Gangwar, J. Calder, T. Takahashi, et al., Robust variational segmentation of 3D bone CT data with thin cartilage interfaces, Med. Image Anal. 47 (2018) 95–110.
[20] Y. Chang, Y. Yuan, C. Guo, et al., Accurate pelvis and femur segmentation in hip CT with a novel patch-based refinement, IEEE J. Biomed. Health Inform. 23 (3) (2019) 1192–1204.
[21] Y. Xia, J. Fripp, S.S. Chandra, et al., Automated bone segmentation from large field of view 3D MR images of the hip joint, Phys. Med. Biol. 58 (20) (2013) 7375–7390.
[22] J. Carballido-Gamio, S. Bonaretti, I. Saeed, et al., Automatic multi-parametric quantification of the proximal femur with quantitative computed tomography, Quant. Imag. Med. Surg. 5 (4) (2015) 552–568.
[23] C. Chu, C. Chen, L. Liu, et al., FACTS: fully automatic CT segmentation of a hip joint, Ann. Biomed. Eng. 43 (5) (2015) 1247–1259.
[24] F. Yokota, T. Okada, M. Takao, et al., Automated segmentation of the femur and pelvis from 3D CT data of diseased hip using hierarchical statistical shape model of joint structure, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5762 LNCS (PART 2), 2009, pp. 811–818.
[25] J. Schmid, J. Kim, N. Magnenat-Thalmann, Robust statistical shape models for MRI bone segmentation in presence of small field of view, Med. Image Anal. 15 (1) (2011) 155–168.
[26] J.E. Iglesias, M.R. Sabuncu, Multi-atlas segmentation of biomedical images: a survey, Med. Image Anal. 24 (1) (2015) 205–219.
[27] T. Heimann, H.-P. Meinzer, Statistical shape models for 3D medical image segmentation: a review, Med. Image Anal. 13 (4) (2009) 543–563.
[28] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[29] F. Chen, J. Liu, Z. Zhao, et al., Three-dimensional feature-enhanced network for automatic femur segmentation, IEEE J. Biomed. Health Inform. 23 (1) (2019) 243–252.
[30] F. Milletari, N. Navab, S.A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565–571.
[31] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, arXiv:1505.04597, 2015.
[32] G. Zeng, X. Yang, J. Li, et al., 3D U-net with Multi-level Deep Supervision: Fully Automatic Segmentation of Proximal Femur in 3D MR Images, vol. 10541, Springer International Publishing, 2017.
[33] C. Zhao, J.H. Keyak, J. Tang, et al., A deep learning-based method for automatic segmentation of proximal femur from quantitative computed tomography images, arXiv:2006.05513, 2020.
[34] C. Zhao, J.H. Keyak, J. Tang, et al., ST-V-Net: incorporating shape prior into convolutional neural networks for proximal femur segmentation, Complex Intell. Syst. 0123456789 (2021).
[35] P. Bjornsson, A. Baker, I. Fleps, et al., Fast and robust femur segmentation from computed tomography images for patient-specific hip fracture risk screening, Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 11 (2) (2023) 253–265.
[36] C.M. Deniz, S. Xiang, R.S. Hallyburton, et al., Segmentation of the proximal femur from MR images using deep convolutional neural networks, Sci. Rep. 8 (1) (2018) 16485.
[37] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 37, PMLR, Lille, France, 2015, pp. 448–456.
[38] A. Patil, M. Rane, Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition, Smart Innovation, Systems and Technologies, vol. 195, 2021, pp. 21–30.

[39] V. Dumoulin, F. Visin, A guide to convolution arithmetic for deep learning, arXiv: 1603.07285v2, 2018.
[40] M. Drozdzal, E. Vorontsov, G. Chartrand, et al., The importance of skip connections in biomedical image segmentation, in: G. Carneiro, D. Mateus, L. Peter, et al. (Eds.), Deep Learning and Data Labeling for Medical Applications, Springer International Publishing, Cham, 2016, pp. 179–187.
[41] S. Jadon, A survey of loss functions for semantic segmentation, in: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2020, pp. 1–7.
[42] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015.
[43] A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, BMC Med. Imaging 15 (1) (2015).
[44] V. Yeghiazaryan, I. Voiculescu, Family of boundary overlap metrics for the evaluation of medical image segmentation, J. Med. Imag. 5 (01) (2018).
[45] Z. Yosibash, N. Trabelsi, I. Buchnik, et al., Hip fracture risk assessment in elderly and diabetic patients: combining autonomous finite element analysis and machine learning, J. Bone Miner. Res. 38 (6) (2023) 876–886.
[46] N. Dagan, E. Elnekave, N. Barda, et al., Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization, Nat. Med. 26 (2020) 77–82.
[47] Z. Yosibash, N. Trabelsi, C. Milgrom, Reliable simulations of the human proximal femur by high-order finite element analysis validated by experimental observations, J. Biomech. 40 (2007) 3688–3699.
[48] Z. Altai, M. Qasim, X. Li, et al., The effect of boundary and loading conditions on patient classification using finite element predicted risk of fracture, Clin. Biomech. 68 (2019) 137–143.
[49] D.F. Almeida, R.B. Ruben, J. Folgado, et al., Fully automatic segmentation of femurs with medullary canal definition in high and in low resolution CT scans, Med. Eng. Phys. 38 (12) (2016) 1474–1480.
[50] L. Ben Younes, Y. Nakajima, T. Saito, Fully automatic segmentation of the Femur from 3D-CT images using primitive shape recognition and statistical shape models, Int. J. Comput. Assisted Radiol. Surg. 9 (2) (2014) 189–196.
[51] A.L. Adams, H. Fischer, D.L. Kopperdahl, D.C. Lee, D.M. Black, M.L. Bouxsein, S. Fatemi, S. Khosla, E.S. Orwoll, E.S. Siris, T.M. Keaveny, Osteoporosis and Hip Fracture Risk From Routine Computed Tomography Scans: The Fracture, Osteoporosis, and CT Utilization Study (FOCUS), JBMR 33 (7) (2018) 1291–1301.