



**HAL**  
open science

## Finite element models with automatic computed tomography bone segmentation for failure load computation

Emile Saillard, Marc Gardegaront, Aurélie Levillain, François Bermond, David Mitton, Jean-Baptiste Pialat, Cyrille Confavreux, Thomas Grenier, Hélène Follet

### ► To cite this version:

Emile Saillard, Marc Gardegaront, Aurélie Levillain, François Bermond, David Mitton, et al.. Finite element models with automatic computed tomography bone segmentation for failure load computation. Scientific Reports, 2024, 14 (1), pp.16576. 10.1038/s41598-024-66934-w . hal-04653628

**HAL Id: hal-04653628**

**<https://hal.science/hal-04653628>**

Submitted on 19 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

## Finite element models with automatic computed tomography bone segmentation for failure load computation

Emile Saillard<sup>1,2</sup>, Marc Gardegaront<sup>1,3</sup>, Aurélie Levillain<sup>3</sup>, François Bermond<sup>3</sup>, David Mitton<sup>3</sup>, Jean-Baptiste Pialat<sup>2,4</sup>, Cyrille Confavreux<sup>1,4</sup>, Thomas Grenier<sup>2,5</sup> & Hélène Follet<sup>1,5</sup>✉

Bone segmentation is an important step to perform biomechanical failure load simulations on in-vivo CT data of patients with bone metastasis, as it is a mandatory operation to obtain meshes needed for numerical simulations. Segmentation can be a tedious and time consuming task when done manually, and expert segmentations are subject to intra- and inter-operator variability. Deep learning methods are increasingly employed to automatically carry out image segmentation tasks. These networks usually need to be trained on a large image dataset along with the manual segmentations to maximize generalization to new images, but it is not always possible to have access to a multitude of CT-scans with the associated ground truth. It then becomes necessary to use training techniques to make the best use of the limited available data. In this paper, we propose a dedicated pipeline of preprocessing, deep learning based segmentation method and post-processing for in-vivo human femurs and vertebrae segmentation from CT-scans volumes. We experimented with three U-Net architectures and showed that out-of-the-box models enable automatic and high-quality volume segmentation if carefully trained. We compared the failure load simulation results obtained on femurs and vertebrae using either automatic or manual segmentations and studied the sensitivity of the simulations on small variations of the automatic segmentation. The failure loads obtained using automatic segmentations were comparable to those obtained using manual expert segmentations for all the femurs and vertebrae tested, demonstrating the effectiveness of the automated segmentation approach for failure load simulations.

Bone metastasis can frequently be found on cancer patients, particularly those with primary breast or prostate cancer<sup>1</sup>, and are the cause of a variety of complications, among which pathological fractures can lead to a substantial decrease in the quality of life of the patients. Computed tomography (CT) is the imaging modality of choice for clinicians to evaluate the risk of pathological fractures, and although clinical scores, like Mirels score and SINS exist to evaluate fracture risk or bone instability in the case of femoral<sup>2</sup> and vertebral<sup>3</sup> metastasis respectively. Those scores either lack specificity<sup>4-6</sup> or show limitations when the score is intermediate<sup>7,8</sup>.

To help the clinicians better evaluate the medical response, simulations based on finite element models are adequate to study the mechanical behavior of bones using CT-scans<sup>9</sup>. Failure load computation allows estimation of the fracture risk using information from the CT-scans along with mechanical properties. Several finite element models exist to perform failure load simulations on femurs<sup>10-17</sup> and vertebrae<sup>18-20</sup>, but few on metastatic bones. The quality and precision of bone segmentations are essential to obtain a reliable failure load estimation of the bones under constraint. Having experts manually annotate data is usually a satisfactory way of obtaining high-quality segmentations, but the process, in addition to being dependent on operator reliability, is very tedious and time-consuming. When trying to simulate mechanical loads on the bones, the operator variability on segmentations can impact the results<sup>21</sup> and hinder the reproducibility of the failure load computation. To obtain reproducible experiments and accurate simulations, automating the segmentation task is an important step. Automatic segmentation methods for femurs<sup>17,22-26</sup>, and vertebrae<sup>27-30</sup> allow us to work towards automating

<sup>1</sup>INSERM, LYOS UMR 1033, Université Claude Bernard Lyon 1, 69008 Lyon, France. <sup>2</sup>INSA-Lyon, CREATIS UMR5220, Université Claude Bernard Lyon 1, Villeurbanne, France. <sup>3</sup>Univ Eiffel, LBMC UMR9406, Université Claude Bernard Lyon 1, 69622 Lyon, France. <sup>4</sup>Hospices Civils de Lyon, Lyon, France. <sup>5</sup>These authors contributed equally: Thomas Grenier and Hélène Follet. ✉email: helene.follet@inserm.fr

failure load simulations on those bones. Convolutional neural networks are robust and allow to efficiently carry out segmentation tasks in the medical field<sup>31,32</sup>, showing great reliability and segmentation performance, but heavily depend on the size, annotation precision and image quality in the training dataset.

The objective of our study is to propose automatic segmentation methods for femurs and vertebrae, then evaluate the quality of the automatic segmentations not only in terms of images metrics using DICE score and Hausdorff distance, but also in terms of failure load results compared to expert manual segmentations, when used in our simulation pipeline. We investigated the effect of variations of the initial automatic segmentation on the numerical simulations, for femurs and vertebrae, to better study the sensibility of the finite elements models to the input segmentation.

## Results

### Automatic segmentations

As shown in Table 1, the segmentation scores obtained with our models indicate a good segmentation quality.

The results highlight the importance of dedicated pre-processing when using custom U-Net architectures. For this task, the 3D models (nnUNet, custom 3D U-Net) outperform the 2D models, even when using a multi-axial approach. 3D U-Net and nnUNet perform very similarly, with only a slight improvement of HD for 3D U-Net.

Table 2 contains results of the nnUNet model on the secondary dataset with 16 femurs, with global and local metrics. The metrics obtained are close to the ones obtained previously, with a slightly higher DICE but a higher Hausdorff distance. The shaft is as expected the region segmented more accurately, but the model still achieves quality segmentation of the proximal part of the femur.

For vertebrae segmentation, nnUNet proved very efficient, but the HD is much higher, due to small mislabeling of some vertebrae. The proximal part of the femur is more prone to segmentation errors than the distal part, but inaccuracies remain minimal. On vertebrae, the vertebral body and pedicle are less likely to be mislabeled or mis-segmented than the spinous or transverse process, where the delimitation between two vertebrae is more difficult to assess. The metrics obtained per vertebra are indicated in Supplementary Table S1. Those results show a disparity in the quality of the automatic segmentation depending on the vertebra, with transitional vertebra L6 and thoracic vertebrae T7 to T10 being more frequently mislabeled, causing a decrease of the DICE score for those vertebrae.

Figure 1 shows automatic segmentations obtained with our approach for both femurs and vertebrae, as well as the differences obtained in terms of distance when comparing to expert manual segmentation.

### Failure load simulations

As shown in Table 3 and Fig. 2, the failure loads obtained for femurs, healthy or metastatic, using automatic segmentations are very close to those obtained with expert annotation.

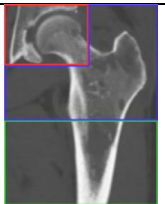
The results obtained when adding morphological operations are very similar to the previous ones. When adding a single erosion or dilation operation, the results are comparable in terms of failure load. When applying erosion operations, the resulting failure loads are lower than when using only the automated segmentation, whereas the failure load value tends to increase when adding dilations. Only the comparison between failure loads obtained with manual and automatic with 2 erosions showed statistical significance. Simulation results obtained for each femur and vertebra are detailed in Supplementary Tables S2 and S3.

The simulation results on vertebrae also show great similarity between the failure loads obtained with automatic and manual segmentations. The failure loads obtained with a single dilation operation are also comparable

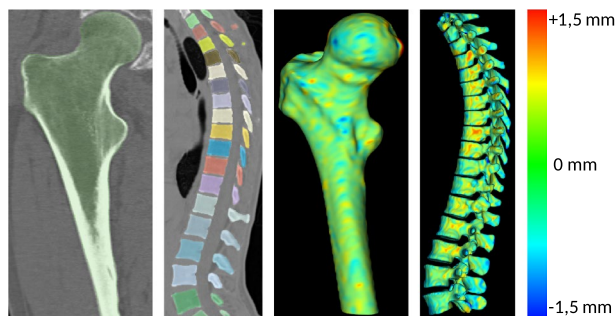
Task/dataset	Algorithm	DSC	HD (mm)
Femur segmentation MEKANOS database	U-Net 2D multi axial*	0.74 ± 0.09	49.86 ± 12.57
	U-Net 2D multi axial	0.93 ± 0.01	2.30 ± 0.82
	U-Net 3D	0.96 ± 0.01	2.20 ± 0.71
	nnUNet 3D fullres	0.96 ± 0.01	2.40 ± 0.84
Vertebrae segmentation VerSe 2019 & 2020	nnUNet 3D fullres—binary	0.95 ± 0.02	19.74 ± 36.54
	nnUNet 3D fullres—multiclass	0.89 ± 0.14	33.89 ± 58.09

**Table 1.** Segmentation results on the primary femur dataset. \*Without pre-processing.

Femur region	DC	HD (mm)
Shaft (green)	0.979 ± 0.004	1.69 ± 0.37
Middle region (blue)	0.975 ± 0.003	2.77 ± 0.71
Head (red)	0.976 ± 0.003	2.66 ± 0.88
Global	0.977 ± 0.002	2.96 ± 0.82



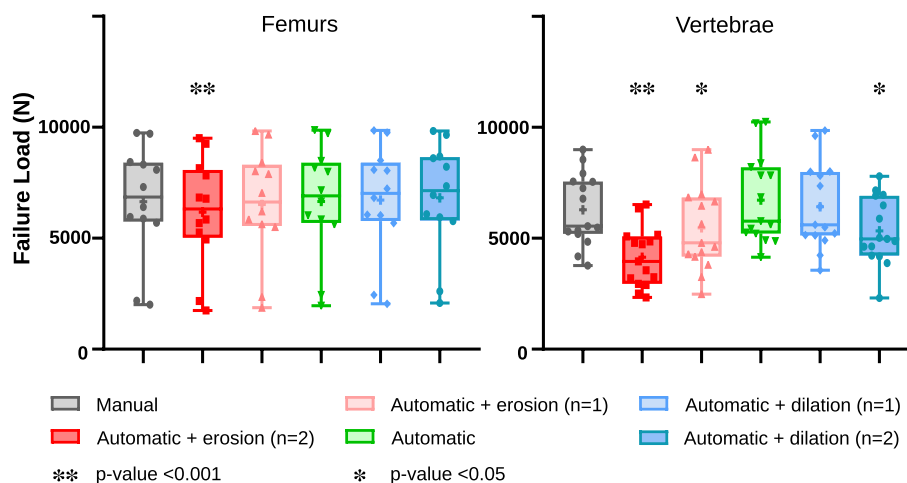
**Table 2.** Segmentation results on our secondary femur dataset depending on the area.



**Figure 1.** Automatic segmentations obtained with nnUNet (left) and pixels distance to the ground-truth (right).

		Automatic segmentation + erosion (2 pixels)	Automatic segmentation + erosion (1 pixel)	Automatic segmentation	Automatic segmentation + dilation (1 pixel)	Automatic segmentation + dilation (2 pixel)
Femurs	DICE	0.882 ± 0.026	0.933 ± 0.012	0.974 ± 0.008	0.959 ± 0.013	0.920 ± 0.026
	Absolute mean of difference with manual FL (N)	473 ± 336	168 ± 105	119 ± 66	132 ± 121	203 ± 201
	Absolute mean of difference with manual FL (%)	7.42 ± 5.00	3.19 ± 2.55	2.25 ± 1.91	2.79 ± 3.37	4.45 ± 5.42
Vertebrae	DICE	0.763 ± 0.055	0.855 ± 0.050	0.929 ± 0.045	0.920 ± 0.043	0.862 ± 0.046
	Absolute mean of difference with manual FL (N)	2320 ± 449	832 ± 420	503 ± 505	500 ± 537	1126 ± 676
	Absolute mean of difference with manual FL (%)	38.69 ± 6.81	16.11 ± 9.11	7.53 ± 8.17	8.63 ± 11.18	19.00 ± 12.89

**Table 3.** Comparison of image metrics and simulation results with manual segmentation.



**Figure 2.** Simulated failure load depending on segmentation.

to those obtained with the manual segmentations. However, with added erosions or when adding a second dilation, the failure load drops and the difference with the ground-truth results is statistically significant in those three cases.

### Discussion

The objective of our approach was to quantify the effect of segmentation variation on failure load simulation. Our segmentation results are visually satisfactory for both femur and vertebrae segmentation, and the image metrics used show that our femur segmentation pipeline has results comparable to state-of-the-arts methods<sup>23–25</sup>, while being trained with a very small dataset. All areas of the femur including the hip joint are segmented with minimal error. Similarly, our vertebrae segmentation method produces results that closely match the performance of the

top-performing techniques on the VerSe datasets<sup>29</sup>. Manual segmentations take at least 30 min to be done while the automatic approach only takes a few minutes, without the need of human intervention.

Our simulation results are in line with the anatomical reality, with the failure load being affected by the position of the metastasis, as well as the structure of the bone, with osteoporosis greatly influencing the simulations. The detailed results in Tables S2 and S3 in supplementary material show that the presence of metastasis tends to decrease substantially the failure load, but in some cases, due to the small size and favorable positioning of the metastasis, it does not impact the simulated failure load. This observation confirms the impact of the location and size of the metastasis on bone strength.

Our simulated failure loads show sensitivity to segmentation variations: When working on femurs, failure loads tend to increase slightly when applying dilation operations and decrease when applying erosions. It is usually obvious that adding material to the object increases strength and removing material decreases it. Indeed, the dilation operation will increase the volume of the area considered as our bone, and potentially add a few voxels corresponding to cortical bone that were not considered in the initial automatic segmentation. The erosion operations reduce the overall volume used in the simulation, and with less cortical bone considered, the failure load obtained decreases. The lower failure loads obtained after erosion were expected, but they were relevant to quantitatively assess the effect of a segmentation error on the simulation results. In the same way, the failure loads obtained after dilation are also very logical but provide additional information on the segmentation errors that can be deemed acceptable for a use in simulation. As seen in Table 3, while the DICE is slightly better with plain automatic segmentation, both the automatic segmentation and the segmentation with added dilation show simulation results really close to the ground-truth simulation. Results also show that an under-segmentation (missing some cortical pixels) impacts the results a lot more than an over-segmentation (more soft tissues are added which do not influence the failure load value). We could therefore consider post-processing methods accordingly, knowing that over-segmenting could help improve the simulation accuracy in some cases.

For vertebrae simulations, the resulting failure loads are a lot more sensitive to segmentation variations, but the results obtained with the plain automatic segmentation are very similar to those obtained using the manual segmentation. The obtained failure loads decrease as expected when adding erosion, but they decrease as well when adding dilation operations. This can be explained by the simulation method used for the simulations on vertebrae, with endplate detection being heavily reliant on the outside segmentation pixels. As seen in Table 3, the simulation results are closest to those obtained with the ground-truth when using plain automatic segmentation or automatic segmentation with a single dilation operation.

Failure load estimations could still be improved through optimization of the simulation protocol, and the next step would be to take into consideration specific mechanical properties of the metastasis. However, with only slight variations on the simulations for both vertebrae and femurs, we can assume that the precision attained with our segmentation models is sufficient to convert into accurate failure load simulations. The segmentations obtained with a single added dilation operation can also be of interest for failure load simulations and can provide additional information to the initial results on the considered bone.

## Materials and methods

In the following section, we detail the different datasets used for the two segmentation tasks and the simulations, as well as the simulation parameters and software used for failure load assessment. We also explain the pre-processing pipeline developed to be applied prior to the training phase of the neural networks. We describe our approach for simulation comparisons with manual and automatic segmentations.

The study protocol was approved by the French Ethics Committee (CPP SUD-EST 1 France) under registration number ID-RCB: 2019-A01202-55. All procedures have been conducted in compliance with national and European regulations. All included patients received clear information and provided written consent, and informed consent was obtained from all subjects and/or their legal guardian(s).

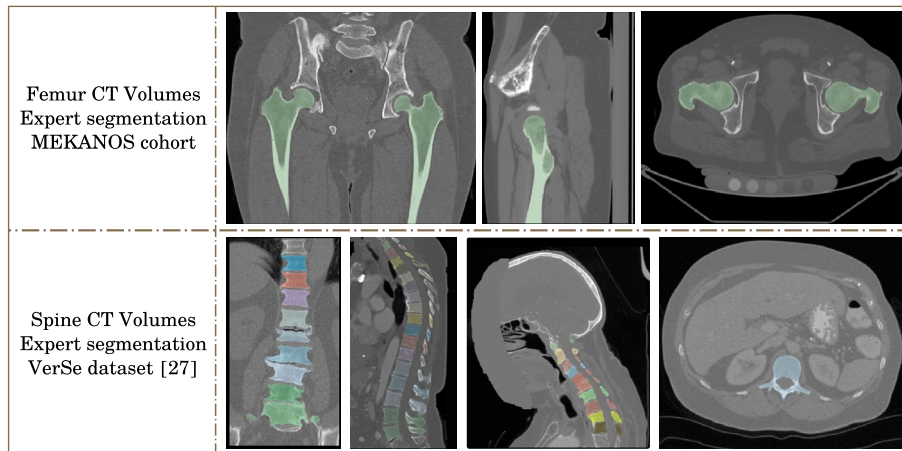
## Datasets

We use two datasets for bone segmentation: one publicly available for the vertebrae and one from the project for the femurs. Existing datasets with CT-scans of femur along with manual segmentation are either not available or lack the accuracy required for the training of robust models, especially on the femoral head. For the femur segmentation task, MEKANOS cohort was used (Hospices civils Lyon, agreement number N. 21 5467, May 28th, 2021, Supplementary Table S5). This cohort consists of eleven in-vivo CT-scans of hips, where both femurs are present. Those scans were acquired in clinical routine following a specific procedure (constant table height, quality phantom QA Mindways, 120 kV, 270 mAs, 1 Pitch, Field of view 360 mm and 200 mm, reconstruction: standard filter B, 512 × 512 matrix, slice thickness 0.7 mm) and with three manufacturers acquisition systems (General Electric, Philips and Siemens). A few femurs have metastatic osteolytic lesions which complicates the segmentation task but allows trained models to segment metastatic bones more efficiently, which is important for our study. From this database, eighteen femurs were manually segmented (4 femurs were not available).

As a secondary test dataset to better assess the robustness of the chosen trained model, additional femurs (n = 16, 9 patients) from four different centers were added from MEKANOS cohort a posteriori, along with manual segmentations. Inter-operator variability was also measured using 6 ex-vivo femurs (Supplementary Table S4). On those femurs, the impact of inter-operator variability on failure load is illustrated in Supplementary Fig. S2).

For the vertebrae segmentation task, we used two publicly available datasets: VerSe2019 and VerSe2020<sup>29</sup>. Those datasets contain 374 CT-scans of various sizes all with manual segmentation. The number of vertebrae on each scan ranges from 3 to 25, with all types of vertebrae present in the dataset. In this study, we retained 363 patients, excluding those with additional transitional T13 scanned. Figure 3 shows examples of axial, coronal and sagittal slices taken from those datasets. The data used for simulation tests consists of 12 femurs from 6 patients (6





**Figure 3.** Example of CT data used for vertebrae and femurs segmentation.

scans), among which 6 are healthy and 6 with metastasis, as well as 15 vertebrae from 2 patients (2 scans) among which 13 are healthy and 2 with metastasis (1 thoracic and 1 lumbar), all taken from the MEKANOS database.

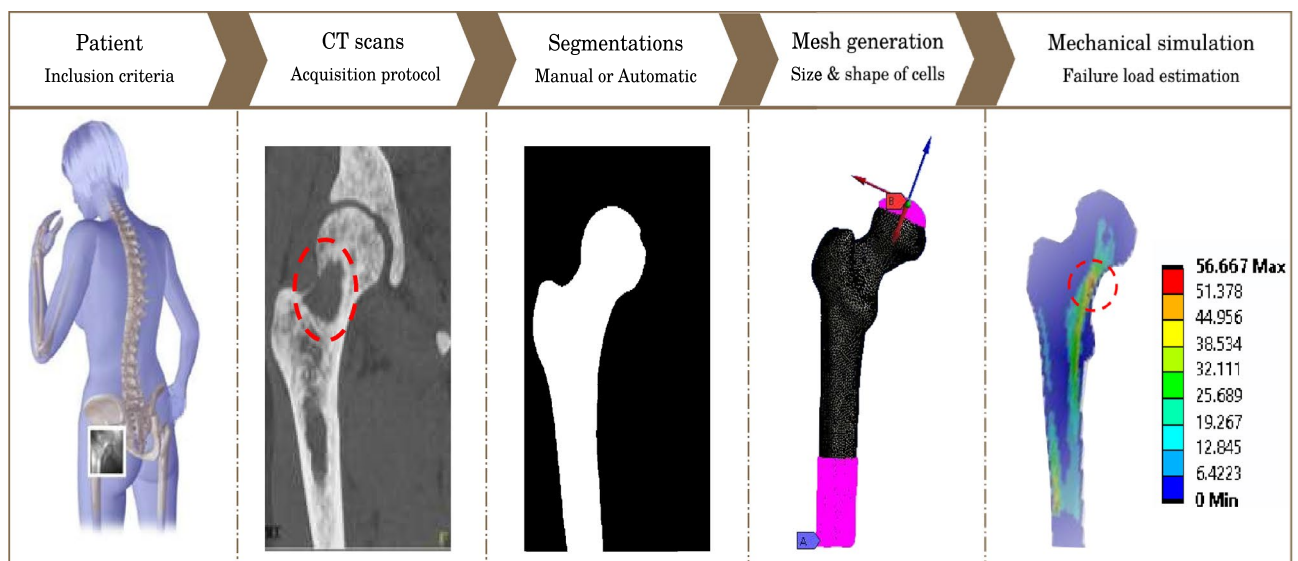
### Simulation pipeline

For both vertebrae and femurs failure load simulations, a custom finite elements simulation pipeline, described in Fig. 4 was used with dedicated parameters in order to compute the failure load of the considered bones. When working on femurs, we used a published model from Sas et al. based on voxel-based hexahedral meshes<sup>10</sup>. These meshes were either obtained using a manual CT-segmentation using the software 3D Slicer, or using an automatic segmentation from a neural network described in part D. The intensity values in the CT-scan were converted to bone density using the calibration phantom included in the acquisition<sup>10</sup>. Each element of the mesh was attributed a bone density corresponding to the voxels bone densities. The bone density of each element was then used to compute the mechanical parameters of the non linear constitutive law<sup>10</sup> (cf. Supplementary Fig. S1).

The grey density values were converted to Young’s modulus using the calibration phantom included in the acquisitions<sup>10</sup>:

$$E(MPa) = 14900\rho QCT^{1.86} \tag{1}$$

Using Ansys software (version 2021 R1), an axial compression was applied on the femurs, mimicking a standing position. An incremental displacement was applied on the top nodes of the femoral head (quasi-static) until it reached a maximal displacement. The nodes at the distal end of the diaphysis were constrained by a null displacement. The failure is defined as the maximum load occurring during the simulation<sup>10</sup>.



**Figure 4.** Simulation pipeline for fracture risk assessment.

For vertebrae simulations, a quadratic tetrahedral mesh (10 nodes) with a volume of element of  $1 \text{ mm}^3$  was used based on experimental data from<sup>33</sup>, and for the numerical model, we used the elasticity-density relationship from<sup>34</sup> using the calibration phantom:

$$E(\text{MPa}) = 3230\rho QCT - 34.7 \quad (2)$$

We used a linear elastic—perfectly plastic constitutive law, with a yield strain of 1.5% strain<sup>35</sup>. The failure criteria consisted in considering a strain of 1.9% of the total vertebral height reduction<sup>36</sup>. All the simulations were also run using Ansys software.

### Pre-processing pipeline

For femur segmentation, we propose a fully automated segmentation method with a pre-processing pipeline as illustrated in Fig. 5 to facilitate the deep learning training. Our dataset contains only few annotated data, and dedicated pre-processing is important to ensure the robustness of the proposed models.

The pre-processing pipeline is made of several steps: after selection and manual expert annotation of the femurs, the volumes are all resampled to the median voxel size ( $0.78 \times 0.78 \times 0.67 \text{ mm}$ ), then cropped when both femurs are present in order to separate them into two distinct volumes. The split left femurs are then flipped to obtain a comparable orientation for all volumes. To further increase the homogeneity of the dataset, especially the spatial orientation of the femurs, the flipped left femurs and right femurs are registered together, using affine transforms. Adding the co-registration step between femurs makes the global spatial orientation more similar between femurs. Co-registration ensures the robustness of the network despite few training data, while preserving the automatic aspect of our pipeline. The resulting volumes are then all normalized before being used as input of the convolutional neural network.

In addition to ensuring the proper training of the neural network, the pre-processing pipeline allows to increase the size of the dataset, thanks to the splitting of the initial volumes.

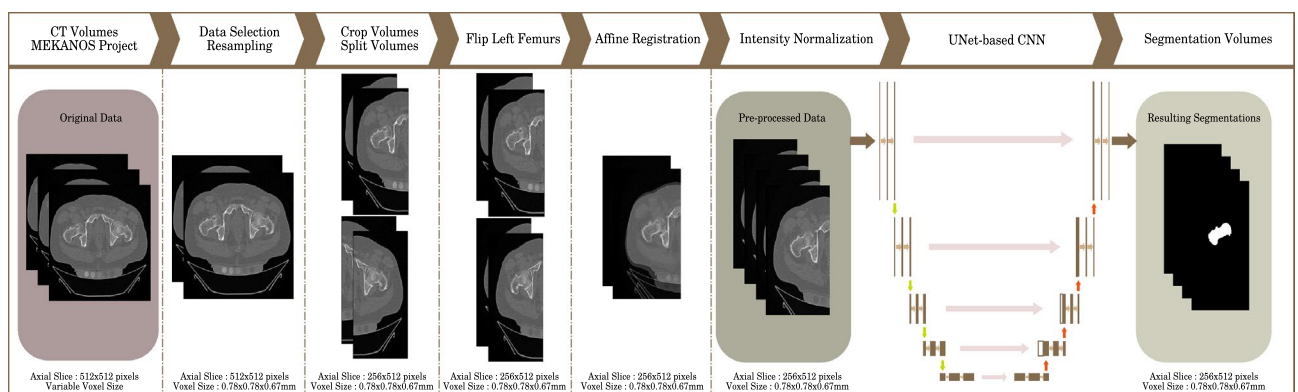
### Neural networks

We used several convolutional neural networks all based on the U-Net architecture<sup>31</sup>. We implemented a 2D multi-planar U-Net, as well as a 3D U-Net for femur segmentation. We compared the results with nnUNet, the state-of-the-art convolutional neural network for medical image segmentation<sup>32</sup>.

Three 2D-U-Net were trained on axial, coronal and sagittal slices for 500 epochs. The 3 resulting segmentations were then fused using majority voting. The 3D U-Net model was trained using random patches of size  $64 \times 64 \times 64$  for 300 epochs. Data augmentation, such as random rotations, translations, shearing and scaling was used on-the-fly to prevent overfitting. All custom UNets were trained using Adam optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ) and a DICE loss, with a learning rate  $\alpha = 2 \times 10^{-4}$  and a batch size of 16 for 2D U-Net,  $\alpha = 3 \times 10^{-5}$  and a batch size of 4 for 3D U-Net. We also added morphological post-processing operations based on binary dilation and erosion to remove small unwanted islands and improve segmentation results.

The nnUNet architecture used is the ‘3d fullres’, with patch sizes automatically selected ( $238 \times 196 \times 208$  for femur segmentation and  $205 \times 205 \times 205$  for vertebrae segmentation) and default parameters, and was trained for 1000 epochs. The optimizer used is stochastic gradient descent with an initial learning rate of 0.01. The batch size was set to 2 for both trainings. We only used this architecture for vertebrae segmentation as the results for femur segmentation were comparable to our custom 3D U-Net and the amount of training data was sufficient to avoid the need for dedicated pre-processing, and the only pre-treatments operations were automatically made with nnUNet.

The models were trained on a Nvidia P100 GPU with 16 GB VRAM. The total training time was 12 h for 2D U-Net per axis, 16 h for 3D U-Net and 48 h for nnUNet on the femur dataset. This substantial difference is also present during inference, where nnUNet takes up to 30 min for a prediction when standard models only take up to 3 min.



**Figure 5.** Segmentation pipeline used for femur segmentation.

## Segmentations and simulations comparison

To quantify the segmentation results, we used the Sørensen-Dice score (noted DSC) to evaluate the similarity between the ground-truth and the automatic segmentations ([0;1] where 1 is the best), as well as the Hausdorff distance (noted HD) to evaluate the maximum errors of the automatic segmentations (in mm, smaller the better). All metrics are computed on 3D volumes. We used a fivefold cross-validation to quantify more accurately the results. Among the 18 available femurs, 12 were used for training, 4 for validation and 2 for testing. For vertebrae segmentation, 242 scans were used for training, 61 for validation and 60 for testing.

To compare the influence of the segmentation on the failure load simulations, we computed the failure load on 12 femurs, using automatic segmentations and using expert manual annotation for comparison. We also compared results using automatic and manual segmentation on 15 thoracic and lumbar vertebrae. In both cases, we also applied simple morphological operations (dilation/erosion), with either one or two iterations to the automatic segmentation as a way to introduce variability to the automatic segmentations. The objective is to investigate the effect of slight segmentation variations on the resulting failure load.

## Statistical analysis

Statistical tests were performed using SPSS software (SAS Institute, Cary, NC). Differences among groups were evaluated using non-parametric test (Friedman test). When a significant overall F value ( $P < 0.05$ ) was present, differences between individual group means were tested using Dunn's multiple comparison post-hoc tests. Only comparisons with the manual segmentation are presented. For all tests,  $P < 0.05$  was considered statistically significant. Data are presented as mean  $\pm$  standard error.

## Conclusion

In this paper we proposed a dedicated pre-processing pipeline for femur segmentation as well as deep learning based segmentation methods for femurs and vertebrae segmentation. From our experiments, we showed that it is primordial to use pre-processing in order to improve the segmentation results. U-Net architectures are efficient and can serve as primary tools to perform automated bone segmentations. We showed that failure load simulations depend on the initial segmentations, and that automatic segmentations yield similar simulation results as simulations made with expert segmentations. The variations on segmentations when adding dilation or erosion operations impact the simulations, with the closest results to the simulations using expert simulations being those using the plain automatic segmentation and the automatic segmentation with one dilation operation. The results obtained allow us to envision the use of this approach in a broader pipeline in biomechanical simulations on patients with metastatic lesions.

## Data availability

Data is not made publicly available due to ethical or privacy reasons. Data used for this study is available on request to the corresponding author.

Received: 21 December 2023; Accepted: 5 July 2024

Published online: 17 July 2024

## References

- Macedo, F. *et al.* Bone metastases: An overview. *Oncol. Rev.* **11**(1), 321 (2017).
- Mirels, H. Metastatic disease in long bones a proposed scoring system for diagnosing impending pathologic fractures. *Clin. Orthop. Relat. Res.* (1976–2007) **249**, 256–264 (1989).
- Fisher, C. G. *et al.* A novel classification system for spinal instability in neoplastic disease: An evidence-based approach and expert consensus from the spine oncology study group. *Spine* **35**(22), E1221–E1229 (2010).
- Damron, T. A. *et al.* Critical evaluation of Mirels' rating system for impending pathologic fractures. *Clin. Orthop. Relat. Res.* **415**, S201–S207 (2003).
- Piccioli, A., Spinelli, M. S. & Maccauro, G. Impending fracture: A difficult diagnosis. *Injury* **45**, S138–S141 (2014).
- Benca, E. *et al.* The insufficiencies of risk analysis of impending pathological fractures in patients with femoral metastases: A literature review. *Bone Rep.* **5**, 51–56 (2016).
- Lenschow, M. *et al.* Impact of spinal instrumentation on neurological outcome in patients with intermediate spinal instability neoplastic score (SINS). *Cancers* **14**(9), 2193 (2022).
- Shi, D. D. *et al.* Assessing the utility of the spinal instability neoplastic score (SINS) to predict fracture after conventional radiation therapy (RT) for spinal metastases. *Pract. Radiat. Oncol.* **8**(5), e285–e294 (2018).
- Pahr, D. H. & Zysset, P. K. From high-resolution CT data to finite element models: Development of an integrated modular framework. *Comput. Methods Biomech. Biomed. Eng.* **12**(1), 45–57 (2009).
- Sas, A. *et al.* Nonlinear voxel-based finite element model for strength assessment of healthy and metastatic proximal femurs. *Bone Rep.* **12**, 100263 (2020).
- Keyak, J. H. *et al.* Prediction of femoral fracture load using automated finite element modeling. *J. Biomech.* **31**(2), 125–133 (1997).
- Duchemin, L. *et al.* An anatomical subject-specific FE-model for hip fracture load prediction. *Comput. Methods Biomech. Biomed. Eng.* **11**(2), 105–111 (2008).
- Eggermont, F. *et al.* Patient-specific finite element computer models improve fracture risk assessments in cancer patients with femoral bone metastases compared to clinical guidelines. *Bone* **130**, 115101 (2020).
- Tanck, E. *et al.* Pathological fracture prediction in patients with metastatic lesions can be improved with quantitative computed tomography based computer models. *Bone* **45**(4), 777–783 (2009).
- Dall'Ara, E. *et al.* QCT-based finite element models predict human vertebral strength in vitro significantly better than simulated DEXA. *Osteoporos. Int.* **23**, 563–572 (2012).
- Yosibash, Z. *et al.* Hip fracture risk assessment in elderly and diabetic patients: Combining autonomous finite element analysis and machine learning. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **38**(6), 876–886 (2020).
- Björnsson, P. A. *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), 253–265 (2023).



18. Stadelmann, M. A. *et al.* Conventional finite element models estimate the strength of metastatic human vertebrae despite alterations of the bone's tissue and structure. *Bone* **141**, 115598 (2020).
19. Palanca, M., Cavazzoni, G. & Dall'Ara, E. The role of bone metastases on the mechanical competence of human vertebrae. *Bone* **173**, 116814 (2023).
20. Imai, K. *et al.* In vivo assessment of lumbar vertebral strength in elderly women using computed tomography-based nonlinear finite element model. *Spine* **33**(1), 27–32 (2008).
21. Levillain, A. *et al.* Inter-operator variability of a finite element model of femurs with and without simulated metastatic defects. In *ESB2021, 26th Congress of the European Society of Biomechanics* 1 (2021).
22. Chu, C. *et al.* MASCG: Multi-atlas segmentation constrained graph method for accurate segmentation of hip CT images. *Med. Image Anal.* **26**(1), 173–184 (2015).
23. Liang, K. *et al.* Accurate and automatic 3D segmentation of femur and pelvis from CT images of the hip based on deep learning. *J. Imaging Sci. Technol.* **65**(3), 1–6 (2021).
24. Deng, Y. *et al.* A deep learning-based approach to automatic proximal femur segmentation in quantitative CT images. *Med. Biol. Eng. Comput.* **60**(5), 1417–1429 (2022).
25. Chen, F. *et al.* Three-dimensional feature-enhanced network for automatic femur segmentation. *IEEE J. Biomed. Health Inform.* **23**(1), 243–252 (2017).
26. Gardegaront, M. *et al.* *Inter-Laboratory Replicability and Sensitivity Study of a Finite Element Model to Quantify Human Femur Failure Load: Case of Metastases*. JMBBM-D-23-01657, submitted Nov 2023.
27. Klinder, T. *et al.* Automated model-based vertebra detection, identification, and segmentation in CT images. *Med. Image Anal.* **13**(3), 471–482 (2009).
28. Forsberg, D. Atlas-based registration for accurate segmentation of thoracic and lumbar vertebrae in CT data. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging. Lecture Notes in Computational Vision and Biomechanics* (eds Yao, J. *et al.*) 49–59 (Springer International Publishing, 2015).
29. Sekuboyina, A. *et al.* VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images. *Med. Image Anal.* **73**, 102166 (2021).
30. Allard, V. *et al.* Comparison of two models to predict vertebral failure loads on the same experimental dataset. In *Computer Methods in Biomechanics and Biomedical Engineering in Springer book*, accepted.
31. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*, Vol. 18 234–241 (Springer, 2015).
32. Isensee, F. *et al.* nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021).
33. Wegrzyn, J. *et al.* Determinants of the mechanical behavior of human lumbar vertebrae after simulated mild fracture. *J. Bone Miner. Res.* **26**(4), 739–746 (2011).
34. Kopperdahl, D. L., Morgan, E. F. & Keaveny, T. M. Quantitative computed tomography estimates of the mechanical properties of human vertebral trabecular bone. *J. Orthop. Res.* **20**(4), 801–805 (2002).
35. Rincón-Kohli, L. & Zysset, P. K. Multi-axial mechanical properties of human trabecular bone. *Biomech. Model. Mechanobiol.* **8**, 195–208 (2009).
36. Wang, X. *et al.* Prediction of new clinical vertebral fractures in elderly men using finite element analysis of CT scans. *J. Bone Miner. Res.* **27**(4), 808–816 (2012).

## Acknowledgements

The authors thank master student (Audrey Auvinet and Océane Hovasse) who served as “segmentation beginners” for the femurs. This work was partly funded by LabEx Primes, France (ANR-11-LABX-0063) and MSDAV-ENIR Research Grant. The images were selected and acquired in clinical routine thanks to the MEKANOS clinical research assistants.

## Author contributions

E.S.: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing; M.G.: Methodology, Software, Writing—review & editing; A.L.: Methodology; Validation; F.B.: Writing—review & editing; D.M.: Writing—review & editing; J.-B.P.: Writing—review & editing; C.C.: Writing—review & editing; T.G.: Conceptualization, Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing—review & editing; H.F.: Formal analysis, Funding acquisition, Supervision, Writing—review & editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-66934-w>.

**Correspondence** and requests for materials should be addressed to H.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024