Generation of Synthetic µCT Images of Rat Lumbar Vertebral Failure via a Deep Generative Model

Allison Tolgyesi[1, 2], Cari M Whyne[1, 2, 3], Michael Hardisty[1, 3]
[1]Sunnybrook Research Institute, Toronto, ON, Canada, [2]Institute of Biomedical Engineering, Faculty of Engineering, University of Toronto, Toronto, ON, Canada, [3]Division of Orthopaedics, Department of Surgery, University of Toronto, Toronto, ON, Canada
Allison.tolgyesi@mail.utoronto.ca

Disclosures: None to disclose.

INTRODUCTION: Modeling bone failure is challenging as bone matrix has heterogeneous material properties, damage nucleation and propagation are stochastic, and bone toughening mechanisms exist on many length scales. To account for this complexity, imaging (histology, micro computed tomography (µCT)), mechanical experiments, and damage mechanics modeling have been used to investigate bone quality. Yet, noise and complex interactions present challenges in fusing these techniques for the inference of bone quality and mechanical behaviour. Bone damage has previously been successfully modeled with micro finite element models (µFE); however, these require extensive setup and are computationally expensive limiting their use for large scale bone quality simulations. Deep learning (DL) presents useful tools for modeling using a data-driven approach. Generative DL models can represent stochastic processes, allow for multi-scale modeling, and incorporate non-linear relationships to input variables. DL models can describe complex interactions and execute quickly, allowing parameter exploration. The present study aimed to develop a generative DL model to simulate the bone failure process through creation of synthetic 3D µCT images of rat lumbar vertebrae.

METHODS: The dataset utilized for this work was a secondary use of sequential µCT image data, originally acquired for load-to-failure analysis of rat lumbar vertebrae. Rats in this dataset either had spine metastases (osteolytic n = 28, osteoblastic n = 3, mixed n = 12) or were healthy controls (n = 29). The L1-L3 lumbar motion segment was potted in bone cement in a µCT compatible loading device (µCT-100, Scanco). µCT images were acquired unloaded, after 1500 µm of axial compressive displacement, and after L2 vertebral fracture. To reduce computation resource requirements, all images were down sampled from their original isotropic voxel spacing of 35 µm to 70 µm. The proposed model for this work was a 3D conditional generative adversarial network (cGAN). This model aims to learn an image-to-image translation task, which learns the mapping between input and output µCT images. The cGAN architecture consisted of two DL models: the generator, which was trained to create synthetic output µCT images, and the discriminator, which was trained to classify output images as real or synthetic. Both models were trained in parallel in a zero-sum game, meaning as an advantageous loss is fed to one model an equivalent disadvantageous loss is fed to the other. The generator had a modified 3D U-Net architecture, which was previously designed in 2D for biomedical image segmentation tasks and performs well with small training datasets. The discriminator model was a four-layer, 3D deep convolutional neural network for conditional-image classification. The training dataset consisted of 64 real µCT images that were augmented by random 3D rotation, increasing the dataset to 1364 images. 8 real µCT images were held out for the test dataset. Three configurations of real input to synthetic output images were tested with the cGAN: unloaded to 1500 µm, 1500 µm to fractured, and unloaded to fractured. For all configurations, the cGAN was trained for 1000 epochs (~ 260s/epoch, run in parallel on four NVIDIA V100-SMX2-32GB GPUs) with a batch size of four and a learning rate of 0.0002. The unloaded to 1500 µm configuration tested the cGANs ability to model the linear elastic component of vertebral deformation. The remaining configurations tested the cGANs ability to model the bone failure process with (1500 µm input) and without (unloaded input) some initial linear deformation. Bone was segmented with thresholding (Otsu) to create 3D binary label fields of the generated and real output images. Dice Similarity Coefficient (DSC) was used to quantify the spatial similarity between label fields. Perceived image quality and realism were assessed with the structural similarity index (SSIM) and the Fréchet inception distance (FID), respectively. Mean differences in quantitative measures between fractured configurations were determined with one-tailed t-tests.
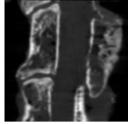
RESULTS SECTION: After 1000 epochs, the generated µCT images from the test dataset were qualitatively and quantitatively assessed for image quality. Qualitatively, the generated 1500 µm images from the unloaded to 1500 µm configuration all had realistic vertebral morphology including metastases when appropriate. The generated fractured images from both input configurations appear slightly less realistic, but still include identifiable vertebral structures and morphology (except for one sample in the unloaded to fractured configuration) and evidence of metastasis when appropriate (Figure 1). The model was able to correctly predict at least one failure location in 6/8 vertebrae with the 1500 µm input data, but only 4/8 when the unloaded scans were used. The DSC for the unloaded to 1500 µm configuration was 0.75 ± 0.08 (with 1 being a perfect overlap), for the 1500 µm to fractured configuration was 0.68 ± 0.08 and for the unloaded to fractured configuration was 0.64 ± 0.07. There was no significant difference in DSC between the two fractured configurations ($p = 0.16$). The SSIM for the unloaded to 1500 µm configuration was 0.66 ± 0.04 (with 1 having identical quality to the real image), for the 1500 µm to fractured configuration was 0.61 ± 0.07 and for the unloaded to fractured configuration was 0.54 ± 0.06, which was significantly lower than the 1500 µm to fractured configuration ($p = 0.03$). The FID for the unloaded to 1500 µm configuration was 117 ± 10 (with 0 being perfectly realistic to a real vertebra), for the 1500 µm to fractured configuration was 140 ± 26 and for the unloaded to fractured configuration was 172 ± 36, which was significantly higher than the 1500 µm to fractured configuration ($p = 0.03$).

DISCUSSION: The cGAN developed for this work demonstrated its ability to generate synthetic µCT images of rat vertebrae that were realistic to real examples. The model predicted linear elastic deformation (unloaded to 1500 µm) more realistically than plastic deformation (fracture) and the latter prediction was improved when the input to the model had already undergone some loading (1500 µm). These findings may be explained by observing that larger deformations may be harder to predict, as required in the unloaded to fractured configuration. Designing a singular recurrent generative model that includes both unloaded and 1500 µm images as input may ultimately provide a more accurate fractured prediction. However, the model may be made more generalized to other datasets by removing the 1500 µm image requirement and predicting the linear elastic deformation from the unloaded image. Introducing simple physical models such as linear elastic µFE may allow for the development of hybrid DL models that can predict the fractured state of a vertebra given an unloaded input. The limitations of the model include the relatively small training dataset compared to other DL models. The quantitative metrics used to assess synthetic image quality have been used in previous studies to evaluate GANs, however, are not specific to medical images.

SIGNIFICANCE/CLINICAL RELEVANCE: This work is motivated by the need to understand the vertebral failure process. A well-trained generative model could be used in clinical research and can be extended as a clinical tool. Generative models can facilitate preclinical research by reducing animal numbers and allowing parameter exploration.
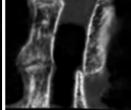
IMAGES AND TABLES:
Figure 1 Sagittal cross-sections of µCT images of real (left) and generated (1500 µm input (center), unloaded input (right)) fractured L2 rat vertebrae.

Real Failed L2   Prediction from 1500 µm   Prediction from Unloaded