Vertebral Fracture Identification in Three Dimensional Spinal Computed Tomography Using High-resolution Fracture Mapping nnU-Net v2 Artificial Intelligence Model: A Promising Tool for Orthopaedic Decision Support

Omar A Abdelaziz, Farid Amirouche, Niharika Belavadi Shekar, Dustin H Massel, Majd Mzeihem

INTRODUCTION:

Accurate detection of vertebral fractures in three dimensional (3D) spinal computed tomography (CT) images is crucial in orthopaedic care, but remains technically challenging due to the complexity of the spine's anatomy and the subtlety of fracture features. For orthopaedic surgeons who rely on imaging to guide operative decisions, rapid and reliable fracture detection is essential. This is especially true in spine surgery, where delayed or missed diagnoses can lead to neurologic deterioration, surgical delays, or long-term complications.

We present an artificial intelligence (AI) solution to this problem using the nnU-Net v2 model. In this retrospective performance diagnostic study, we developed the nnU-Net v2 model to detect CT spinal fractures through a series of varied datasets (ranging from 160 to 1200 images) containing spinal CT scans. Unlike traditional methods for AI model teaching in fracture detection, where scans have every vertebra labeled, we simplified labels to include only “healthy” or “fractured” vertebrae to facilitate increased generalizability and simulate orthopaedic decision-making. Additionally, we applied advanced data augmentation techniques to the scans, such as contrast shift and image noise, to simulate real-world variations in scans. Furthermore, the nnU-Net v2 model itself is a state-of-the-art model that segments at the voxel level, enabling high-resolution fracture mapping (to detect subtle fractures) unlike other models that use bounding boxes.

The model achieved exceptional results, with training Dice of 95.94%, validation Dice of 90.93%, and testing Dice of 88.29%, by utilizing automated configuration and extensive data augmentation. This significantly outperformed traditional AI approaches to fracture detection, highlighting its potential in orthopaedic support. From a clinical perspective, this AI tool can act as an immediate decision-support system, especially in high-volume emergency departments or when access to radiology consultation is delayed. It empowers orthopaedic surgeons with a second set of eyes that are consistently accurate, potentially enhancing diagnostic confidence, reducing interpretation time, and enabling accelerated triage that can reduce the costs of care.

METHODS:

We trained and tested the AI model using 160 3D spinal CT scans from the VerSe 2019 dataset, where each scan was labeled to distinguish between healthy and fractured vertebrae. The images were split into training, validation, and test groups. From these, we then expanded to two additional datasets (480 and 1,200 images, respectively) to train the model by applying advanced data augmentation techniques, such as 3D rotation, contrast adjustments, and simulated anatomical variations. The deep learning model (nnU-Net v2) was trained on each dataset and evaluated using Dice scores, a standard measure of segmentation accuracy. Model predictions were compared to those of other AI models and to expert-labeled fracture segmentations to assess their clinical relevance and performance.

RESULTS:

The number of training images, image quality, and distribution of fracture types were comparable across all datasets. The final model's performance, as measured by Dice similarity coefficients, improved with the addition of dataset augmentation.

The best performance was achieved using the fully augmented dataset (1,200 training images), with nnU-Net v2 reaching a training Dice score of 95.94%, a validation Dice score of 90.93%, and a test Dice score of 87.26%. This significantly outperformed earlier AI models, including MedYOLO (32% accuracy) and 3D CNN (28% accuracy). No significant discrepancies were observed between expert-labeled and AI-predicted fracture locations on qualitative review. All test cases were included in the final analysis.

DISCUSSION AND CONCLUSION:

The nnU-Net v2 model demonstrates exceptional accuracy and robustness in vertebral fracture detection, outperforming traditional AI models and approaching the level of interpretation achieved by radiologists. Its potential for real-time diagnostic support, especially in after-hours or busy clinic environments, makes it an invaluable asset to orthopaedic practice. Its ability to self-adjust and generalize from complex imaging data makes it a promising tool for clinical decision support in orthopaedic spine care and a potential aide in training future residents and fellows in fracture detection. With further validation, the model can evolve to detect even subtle micro-fractures and transform orthopaedic care by enhancing efficiency, reducing diagnostic uncertainty, and improving patient outcomes across diverse orthopaedic settings.

Dataset	Epochs	Training Dice (%)	Validation Dice (%)	Test Dice (%)
Original (120)	250	72.67	63.99	65.29
Original (120)	750	75.04	65.44	66.89
Augmented 1 (480)	250	90.89	79.4	76.57
Augmented 1 (480)	2000	94.24	87.87	83.84
Augmented 2 (1200)	250	93.09	81.9	80.08
Augmented 2 (1200)	1000	95.54	88.31	85.29
Augmented 2 (1200)	2000	95.94	90.93	87.26

Table 1: Summary of nnU-Net v2 Segmentation Performance

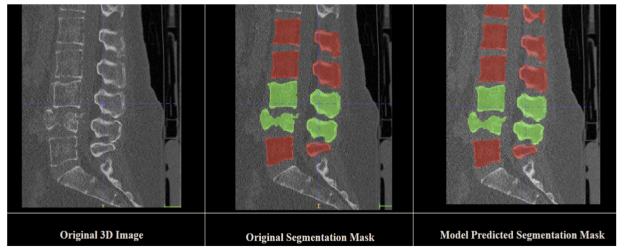

Figure 1: Augmented Data Set 2 Segmentation Masks: 2000 Epochs