A Novel Deep Learning Model for Automated Identification of Spinal Fusion Hardware and Vertebral Levels on Lateral Thoracolumbar Radiographs

Eugenia Lee[1], Carmelo Burgio[2], Karlos Zepeda[2], Delano Trenchfield[2], Seong Jang[2], Eytan Debbi[2], David Mayman[2], Jonathan Vigdorchik[2], Seth Jerabek[2], Peter Sculco[2]
[1]Weill Cornell Medicine, New York City, NY, [2]Hospital for Special Surgery, New York City, NY
Email of Presenting Author: burgioc@hss.edu

Disclosures: EL (N), CB (N), KZ (N), DT (N), SJ (N), ED (3B: DePuy, J&J; 4: Think Surgical), DM (4:Cymedica, Imagen, MiCare Path, Wishbone; 7A: OrthAlign, S&N, Stryker; 8: Hip Society, Knee Society), JV (3B: Stryker, DePuy, J&J; 4: Intellijoint, Aware, Motion Insights, Ortho AI, Polaris, Corin; 8: JBJS, AAHKS), SJ (3B: Stryker; 4: Stryker), PS (3B: Depuy, Intellijoint, EOS, Enovis, Zimmer Biomet; 4: Intellijoint, Parvizi Surgical Innovation; 5: Intellijoint).

INTRODUCTION: The application of artificial intelligence (AI) in medical imaging is rapidly transforming modern healthcare, offering solutions that enhance diagnostic speed, accuracy, and standardization [1–3]. In orthopaedics, deep learning (DL) has demonstrated promising results in complex tasks such as implant identification and coronal alignment analysis in knee arthroplasty [1-4]. However, its application in other subspecialties has not progressed at the same pace. Spinal instrumentation is a well-established procedure for stabilizing the spine in patients with structural deformities or degenerative pathology. Imaging is indispensable throughout diagnosis, planning, and follow-up in spinal care, yet manually identifying hardware on radiographs is labor-intensive and error-prone. With spinal fusion volumes climbing sharply, this workflow places significant strain on radiology services and hinders effective clinical decision-making, highlighting the need for faster, standardized solutions [4–6]. This study aims to develop and validate a DL model capable of automatically detecting spinal fusion hardware in the thoracic and lumbar spine using standard radiographs.

METHODS: We conducted a retrospective study involving patients who underwent thoracolumbar spinal instrumentation. Institutional review board approval was obtained. Eligible cases were selected from the hospital's picture archiving and communication system (PACS) database based on availability of radiographs of patients who have undergone spinal fusion. All images were anonymized and manually annotated using 3D Slicer (v5.1.0) to delineate vertebrae, sacrum, and spinal instrumentation to create a ground truth dataset for training and validation. A convolutional neural network (CNN)-based segmentation model was developed using a U-Net architecture, optimized with custom masking layers and Dice-based loss functions. The dataset was divided into a training set, validation set, and an unseen hold-out test set in an 8:2:2 ratio, which is a balanced and widely accepted split for datasets of similar sizes [7]. Three models were created to detect spinal instrumentation, sacrum, and vertebrae. These models utilized an Adam optimizer and a learning rate scheduler with a batch size of 16. Additional data augmentation, including the watershed segmentation algorithm, brightness shifts, and Gaussian noise, was used to standardize real-world variability. Model performance was assessed using the Dice similarity coefficient and visual comparison of predicted segmentation overlays with ground truth values.

RESULTS: The final model demonstrated excellent performance in detecting spinal instrumentation and localizing fusion levels. A total of 435 lateral radiographs were collected, and with the 8:2:2 ratio, the training set consisted of 291 images, the validation set consisted of 72 images, and the unseen hold-out set consisted of 72 images. Dice coefficients for spinal instrumentation, sacrum and vertebral segmentation were excellent with values of 0.940, 0.943 and 0.953, respectively, and minimum predicted loss values were also excellent, yielding values of 0.0151, 0.0239, and 0.0398, respectively. The watershed segmentation algorithm distinctly labeled separate objects, and it successfully identified separate vertebral levels allowing us to identify the sacrum and L1-L5 without needing to mask each level separately (Figure 1). Identification of the vertebral levels allowed for the determination of corresponding spinal hardware levels on radiographs, enabling the classification of images according to their instrumented levels (Figure 2). Misclassification of spinal fusion levels was primarily associated with incomplete imaging coverage or errors from masking predictions.

DISCUSSION: We present the first deep learning model capable of automatically identifying spinal fusion hardware and associated vertebral levels on standard thoracolumbar radiographs. Not only is this model a pioneering advancement in the spinal fusion domain, but its performance is also comparable to well-established AI models for deep learning analysis of the knee [3,6,8,9]. Future development will focus on further analysis of prediction results, extending functionality and generalizability, improving usability, and supporting integration with PACS for real-time clinical deployment.

SIGNIFICANCE/CLINICAL RELEVANCE: This represents the first deep learning model capable of automatically identifying both spinal fusion hardware and associated vertebral levels on standard radiographs. The technology addresses critical workflow inefficiencies in spine care by reducing manual interpretation time, standardizing hardware identification, and enabling real-time clinical decision support when integrated with PACS systems.

REFERERENCES: [1] Lubowitz et al, Arthroscopy, 2025; [2] Karnuta et al, J Arthroplasty, 2021; [3] Steele et al, J Arthroplasty, 2023; [4] Dutt et al, Radiol Artif Intell, 2022; [5] Selby et al, J Am Acad Orthop Surg, 2012; [6] Murtagh et al, Radiology, 2011; [7] Sivakumar et al, PeerJ Comput Sci, 2024; [8] Yi et al, Eur Spine J, 2023;[9] Yeh et al, Sci Rep. 2021; [10] Lee et al, BMC Neurol, 2022

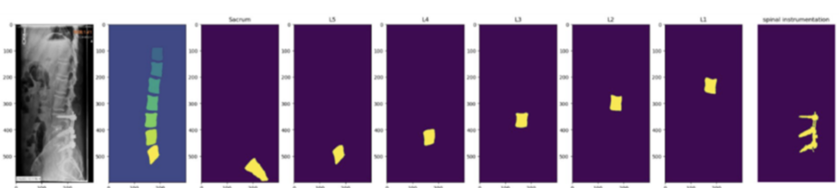

Figure 1. Watershed segmentation algorithm successfully identified vertebrae levels

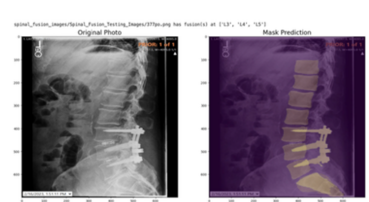

Figure 2. Masking prediction and spinal fusion levels prediction in test set