External Application of Deep Convolutional Neural Networks Trained on Radiographs to Cross-Sectional Imaging in the Spine: Applications for Semantic Labeling

Tae Kyung Kim, BA, Paul H Yi, MD, Jinchui Wei, BSE, Ji Won Shin, BSE, Tae Soo Kim, MSE, Gregory Hager, PhD, Jan Fritz, MD, Haris Sair, MD

Johns Hopkins Radiology Artificial Intelligence Lab (RAIL) & Johns Hopkins Malone Center for Engineering in Healthcare

Background.
The purpose of this study was to develop and test the performance of deep convolutional neural networks (DCNNs) for automated classification of 1) cervical spine (C-spine) radiographs (XRs) by view and 2) lateral cervical (C), thoracic (T), and lumbar (L) spine XRs by anatomic region. A secondary purpose was to test the ability of the 2nd DCNN to identify analogous CT and MRI images.

Methods.
We used 2 datasets: 1) 150 AP, lateral, and odontoid view C-spine XRs (50 each); and 2) 150 lateral XRs of the C, T, and L-spine (50 each). For each DCNN, XRs were split into training (70%), validation (10%), and test (20%) datasets. Training & validation datasets were augmented 22x using standard preprocessing techniques. The ResNet-18 DCNN pretrained on ImageNet was trained and validated using augmented images for classification of 1) C-spine XR view and 2) lateral spine XR anatomic region. Receiver operating characteristic (ROC) curves with area under the curve (AUC) were used to evaluate test performance. The 2nd DCNN was then tested on 2 sets of 45 midline sagittal images each of CT and MRI (15 each anatomic region). DCNN development & testing was performed on a 2.5 GHz Intel Haswell 12-core dual socket with 128 GB RAM and 2 NVIDIA K80 GPUs. DCNN AUCs were statistically compared.

Results.
DCNNs trained for C-spine XR view achieved AUC of 1.0 for all 3 views (p=1). DCNNs trained for lateral spine XR anatomic region achieved AUC of 1.0 for C and L-spine and 0.99 for T spine (p=0.3). In external CT & MRI test sets, the DCNNs performed best for L-spine (AUC of 0.76 [CT] & 0.87 [MRI]) and C-spine (AUC of 0.62 [CT] & 0.71 [MRI]); in contrast, T-spine had AUC of 0.3 [CT] and 0.42 [MRI]) [p<0.01 for all]. During testing, the DCNNs classified images at a rate of 33 images/second.

Conclusion.
DCNNs can accurately classify C-spine XRs by view and lateral spine XRs by anatomic region at superhuman speeds. We have also demonstrated the diagnostic utility of DCNNs trained using XRs for spine anatomic region localization towards CT and MRI. Similar DCNNs may improve PACS workflow through automated XR labeling and identification of relevant comparison examinations, particularly in the setting of mislabeled studies. The proof-of-concept from our work can be applied to other joints and radiographic views to create an all-encompassing semantic labeling DCNN.