Practical Applications of Artificial Intelligence in Spine Imaging: A Review

Upasana Upadhyay Bharadwaj, MD[a,1], Cynthia T. Chin, MD[b,1,*], Sharmila Majumdar, PhD[a]

KEYWORDS

• Artificial intelligence • Machine learning • Deep learning • Spine

KEY POINTS

- Artificial intelligence (AI) tools enable automated vertebral, disc, neuroforamen, canal, and facet segmentation.
- AI may enhance spine imaging reconstruction improving image quality, acquisition time, and reduced radiation dose.
- AI may assist characterizing common spine conditions: degeneration; infection; inflammation; trauma; metastatic disease; and deformity.
- AI clinical deployment may assist semiautomated reporting under radiologist supervision to provide more consistent, objective, and efficient reporting.

INTRODUCTION

Recent artificial intelligence (AI) advances may potentially transform all aspects of radiology including acquisition, interpretation, and radiology report generation, enhancing accuracy, efficiency, and clinical utility.[1–6]

AI enables automated vertebral, disc, and canal segmentation even with noise, artifacts, and anatomic variations.[7,8] Models for training large data sets can identify complex features, enhancing diagnosis and classification of fractures, tumors, or degeneration.[9] Specialized algorithms and variational autoencoders may reconstruct images from undersampled/noisy data, resulting in faster acquisition times while preserving diagnostic accuracy.[10,11]

AI capabilities present opportunities for clinical applications in spine imaging: early detection improving patient outcomes, enhanced diagnosis and surgical planning, improved patient experience with faster acquisition times, and assisting radiologists' efficiency with end-to-end tools for automation.[12]

With exponential increased publications on novel imaging AI, assessing a particular approach's practical utility and clinical relevance is challenging. This review presents an AI overview for practical spine imaging and potential clinical adoption.

OVERVIEW OF ARTIFICIAL INTELLIGENCE

AI broadly encompasses techniques enabling computer systems to perform tasks typically requiring human intelligence.[13] AI algorithms enable computers to learn and perceive, through problem-solving, pattern recognition, decision-making, and natural language understanding.

AI can be classified into two main categories: narrow AI and general AI. Narrow AI systems perform specific tasks: image recognition, language translation, or playing chess and cannot

[a] Department of Radiology and Biomedical Imaging, University of California San Francisco, 1700 4th Street, Byers Hall, Suite 203, Room 203D, San Francisco, CA 94158, USA; [b] Department of Radiology and Biomedical Imaging, University of California San Francisco, 505 Parnassus Avenue, Box 0628, San Francisco, CA 94143, USA
[1] CT Chin and UU Bharadwaj contributed equally to the manuscript.
* Corresponding author.
E-mail address: cynthia.t.chin@ucsf.edu

Radiol Clin N Am 62 (2024) 355–370
https://doi.org/10.1016/j.rcl.2023.10.005
0033-8389/24/

generalize beyond their specific expertise.[14] General AI represents autonomous systems that learn and apply knowledge across various domains. General AI, while a goal for future research and development, is largely a theoretic discipline.

Although AI, "machine learning" (ML), and "deep learning" (DL) are sometimes used interchangeably, there are major differences between each of these related terms (**Fig. 1**).

Machine Learning

ML, the foundation of AI, is a subset focusing on algorithms and statistical models enabling computers to automatically learn and make predictions or decisions based on data. ML can be broadly categorized into supervised, unsupervised, and reinforcement learning.[15]

Supervised learning involves training a model on labeled data, the input data and corresponding output or target labels are provided. The model learns to generalize from the training data and make predictions or decisions on new, unseen data. This approach is commonly used for classification, regression, and object detection tasks in medical imaging applications.[16]

Unsupervised learning trains models on unlabeled data, discovering hidden patterns, structures, or relationships. Clustering, dimensionality reduction, and anomaly detection are some common applications of unsupervised learning. In medical imaging, unsupervised learning techniques can aid in tasks such as data exploration, grouping similar images, or identifying outliers.[17,18]

In reinforcement learning, an agent interacts with an environment and improves its performance by receiving feedback (rewards or penalties). This approach, used in robotics and control systems, learns optimal actions through trial and error. It has limited practical use in medical imaging but has potential for future applications such as optimizing imaging protocols or designing personalized treatment plans.[19]

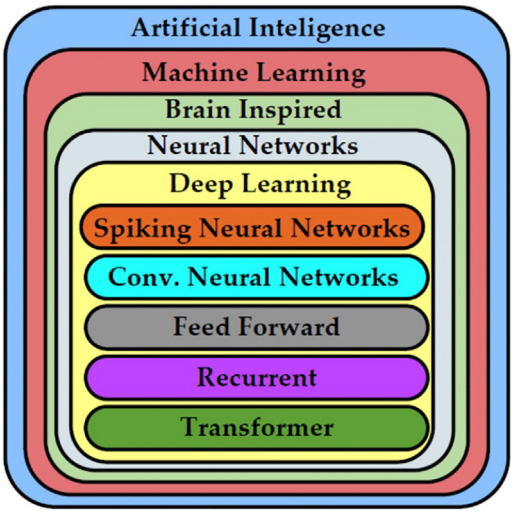

Fig. 1. AI taxonomy with associated learning paradigms. (https://www.researchgate.net/figure/Taxonomy-of-AI-and-its-sub-fields_fig1_352189762.) Khan, F.H.; Pasha, M.A.; Masud, S. Advancements in Microprocessor Architecture for Ubiquitous AI—An Overview on History, Evolution, and Upcoming Challenges in AI Implementation. Micromachines 2021, 12, 665. https://doi.org/10.3390/mi12060665.

Deep Learning

DL, a subset of ML, has recently gained tremendous success.[1,20] It involves training artificial neural networks with multiple layers, allowing the model to learn complex hierarchical representations directly from raw data. Neural networks, mathematical or computer models mimicking the structure and function of biological neural networks (the central nervous system—the brain), consist of a large number of artificial neurons made using a variety of connection techniques: convolutional neural network (CNN), generative adversarial networks (GANs), and recurrent neural networks are examples.[21]

CNN, a widely used DL medical imaging architecture, excels in image reconstruction, segmentation, and classification.[22–25] AI's success in medical imaging is driven by the availability of large, labeled data sets, advancements in computational power, and development of DL models.[26]

APPLICATIONS IN SPINE IMAGING

Image Reconstruction and Acquisition

AI may enhance spine imaging reconstruction, improve image quality, acquisition time, and reduce radiation dose.

AI can reduce MR imaging (MRI) and CT image noise, using DL techniques learning noise patterns in data. DL noise reduction algorithm improved signal-to-noise ratio (SNR) of lumbar spine (LS) MRI scans up to 30%.[27] DL reconstruction algorithms can enhance MR image resolution by learning relationships between different coordinates in the image, significantly improving LS MR image quality.[28,29]

LS CT studies show DL-enhanced images have significantly lower noise compared with the original scan.[30] Reducing radiation dose in LS CT scans is an important direction in image reconstruction from fewer data points with high-quality images acquired at doses up to 72% lower than standard of care (SOC).[31]

Fast MR Acquisition

Reconstruction and noise reduction techniques can also reduce LS MR imaging acquisition times. A rapid LS MR imaging protocol, 3D imaging with DL reconstruction, yielded 54% acquisition time reduction.[32]

Fast MR acquisitions modify conventional imaging protocol parameters, decreasing scan times while maintaining resolution at the cost of increased image noise (reduced SNR). Common strategies shortening acquisition times exploit k-space data redundancy or spatial correlation. Modifications include reducing excitations, raising bandwidth, and increasing parallel imaging factors. These acceleration approaches inherently suffer from reduced SNR, blurring, resulting in insufficient imaging quality. DL-based image denoising methods applied to compromised fast scan data can restore SNR, maintaining image sharpness and SOC quality.[33]

Using a DL reconstruction method to improve SNR and reduce artifacts (commercially available AIR Recon DL, GE Healthcare) on LS MR images, Han and colleagues demonstrated DL reconstruction combined with fast acquisitions has potential for diagnostic image quality noninferior to SOC LS MRIs. The LS MR imaging protocol was 52% faster and able to provide scores noninferior to the standard protocol for apparent SNR, anatomic structure visualization, and diagnostic confidence as evaluated in a blinded fashion by one junior and two senior subspecialty radiologists[34] (**Fig. 2**).

A prospective, randomized, multicenter study assessed DL enhancement to preserve perceived spine MR imaging quality despite 40% scan time reduction. Three experienced neuroradiologists rated perceived superiority of 61 spine MR images for SNR; spatial resolution; imaging artifacts; cord delineation; cord/cerebrospinal fluid (CSF) contrast; disc pathology; bone lesions; and facet/ligamentous pathology. The readers assessed image consistency for anatomy and pathology and found overall diagnostic quality of DL-enhanced MR images statistically equivalent or subjectively better (perceived benefits in SNR and artifact reduction) than SOC across all assessed features, suggesting potential for clinical practice utility.[35]

Synthetic Artificial Intelligence

Synthesizing new images from available images is an active MR imaging research area. DL reconstruction can create synthetic images from existing data sets. Virtually generated MR imaging may make the physical acquisition of particular sequences no longer necessary. GANs can generate synthetic images from different MR contrasts as input. A GAN (generator network and discriminator network) learns to synthesize realistic images (generator) and distinguish real from fake (synthesized) images (discriminator). During the learning process, these networks compete against each other, resulting in the generator network progressively learning to synthesize images with more and more realistic appearances.[36]

T2-weighted (T2W) fat sat (FS) spine sequences are important, requiring significant scan time. Using GAN-generated T2WFS images from conventional T1W and non-FS T2W images, Schlaeger and colleagues compared synthetic T2WFS images to their true counterparts for image quality, FS quality, and diagnostic agreement.

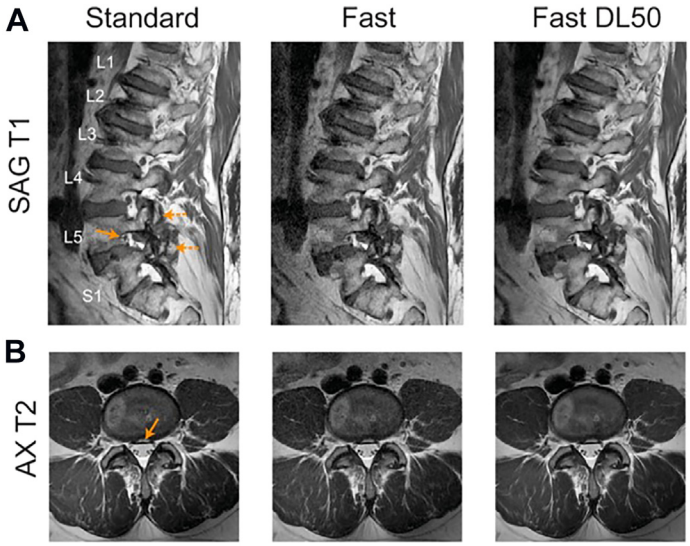

Fig. 2. Comparison of standard versus fast versus fast DL50 imaging on sag T1 axial T2 sequences. (*A*) L5–S1 transitional anatomy, moderate NF stenosis (solid *arrow*), severe L4–L5 facet arthropathy (dashed *arrows*). (*B*) Central annular fissure (solid *arrow*). These features are well delineated on fast DL50 images.[34]

Apparent signal- and contrast-to-noise ratios measured in true and synthetic T2WFS sequences by two neuroradiologists were not significantly different. Subjective image quality was graded higher for synthetic T2WFS (P=.023). In a Turing test, synthetic and true T2WFS could not be distinguished from each other. The inter-method agreement between synthetic and original protocol ranged from substantial to almost perfect agreement for six evaluated spine pathologies. Overall scan time was reduced approximately 40% compared with conventional spine examinations.[37]

A multicenter, multireader study evaluated synthetically created short tau inversion recovery (STIR) spine MR images compared with acquired STIR. A digital imaging and communications in medicine (DICOM)-based DL application generated a synthetically created STIR series from sagittal T1 and T2 images. Three neuroradiologists, one musculoskeletal (MSK) radiologist, and one general radiologist rated STIR quality and classified disease pathology; assessed presence/absence of findings typically evaluated with STIR in trauma. The radiologists evaluated either acquired STIR or synthetically created STIR in a blinded and randomized fashion with a 1-month washout period. The interchangeability of acquired and synthetically created STIR was assessed using a noninferiority threshold of 10%.

For classification, there was a decrease in interreader agreement expected by randomly introducing synthetically created STIR of 3.23%. For trauma, there was an overall increase in interreader agreement by 11.9%. The lower bound of confidence for both exceeded the noninferiority threshold, indicating interchangeability of synthetically created with acquired STIR. Results showed higher image quality scores for synthetic STIR over acquired STIR (P<.0001). The investigators concluded synthetic STIR spine MR images were diagnostically interchangeable with acquired STIR while providing significantly higher image quality, suggesting routine clinical practice potential. The investigators also avoided the use of GANs, which can be prone to introducing structures in synthesized images that are not present in the source images.[38]

CT and MR imaging are complimentary, routinely obtained for evaluation and surgical planning in spine patients. Roberts and colleagues developed a DL algorithm producing 3D LS CT images from MR imaging data using a supervised 3D cycle-GAN model, thus the potential to reduce patient radiation.[36]

They evaluated the accuracy of synthetic LS CTs by comparing 24 clinically relevant measurements on 20 matched synthetic CTs and true CTs by four clinical evaluators (neurosurgeons and radiologists). The outcome measured was the mean difference in measurements performed by the group of evaluators between real CT and synthetic CTs. Measurements in the sagittal plane had a10% relative error and pedicle measurements in the axial plane were considerably less accurate (relative error up to 34%). The investigators concluded that computer-generated synthetic CTs demonstrated a high level of accuracy when performed in-plane to original MR images used for synthesis. Measurements performed on axial reconstructed images were less accurate, attributable to the images being synthesized from nonvolumetric sagittal T1W MR images.[36]

LOCALIZATION OF SPINAL STRUCTURES

Localization of spinal structures on imaging is essential for accurate diagnosis and treatment.[7,8,39,40] They are used to enhance the performance of end-to-end DL systems and also as intermediate visualization tools that can assist diagnosis. The finest form of localization is segmentation, structural delineation from images. Similarity of various vertebrae, curvature, hardware artifacts, and transitional vertebrae are some challenges developing and implementing AI solutions for automatic labeling.[41,42]

Before DL, segmented structures were traditionally processed with thresholding (classifying pixels as foreground or background based on intensity, texture features). More sophisticated techniques such as contour modeling and watershed transform group pixels iteratively based on image gradients improved performance.[43]

DL, a powerful tool for segmentation, can identify different structures from large data sets of labeled images, without necessitating manual feature engineering[8] (**Fig. 3**).

DL models require relatively large, labeled data sets, because as more labeled images become available, the models tend to be more accurate.[44] New DL architectures specifically for spine segmentation tend to be more efficient and accurate than standard DL architectures. Transfer learning, a technique allowing DL models to be trained on one task and subsequently applied to another task, has been effective for spine segmentation. Models trained on labeled images of other structures have shown significant performance improvements to spine segmentation.[45]

The U-Net architecture with skip connections has been the de facto model to precisely localize spinal structures using the encoder–decoder architecture.[46] Several improvements, such as DeepLab that uses dilated convolutions and

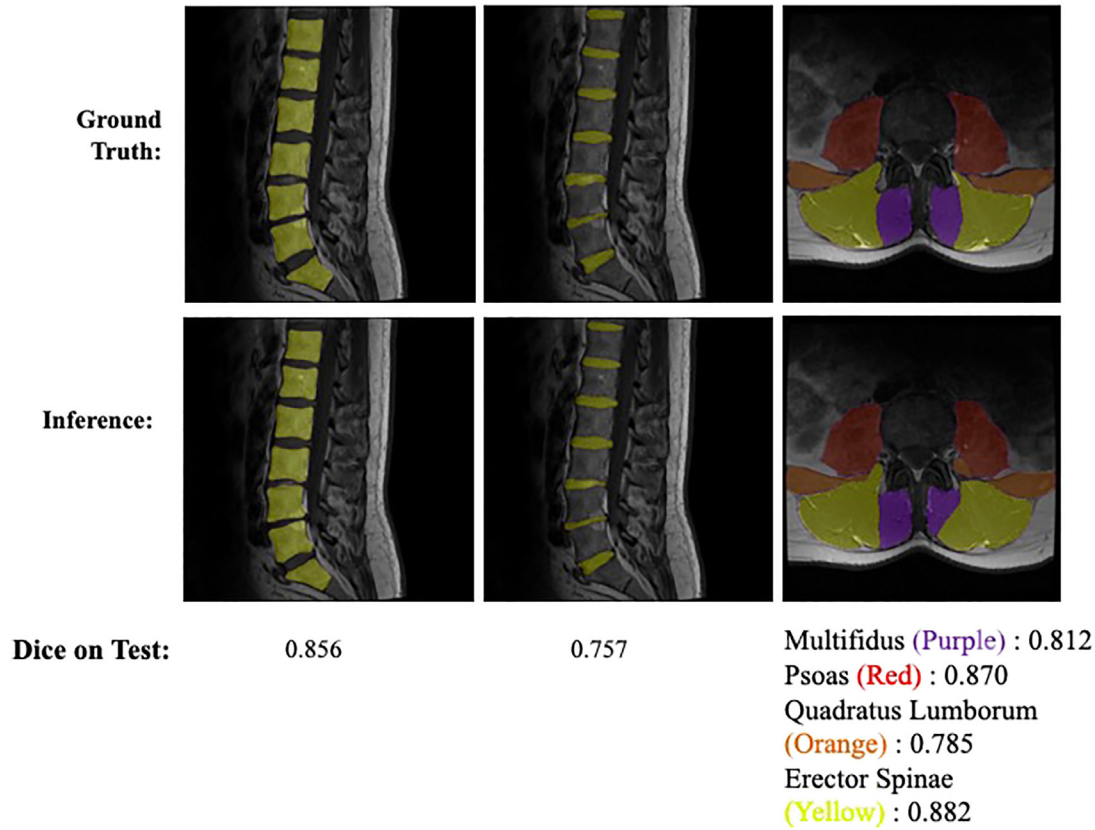

Fig. 3. Vertebral, disc, and paraspinal muscle segmentation results (first, second, and third columns, respectively).[8]

atrous spatial pyramid pooling, have become popular given their overall performance and ability to capture multiscale contextual information for accurate spine segmentation.[47] A recent study using U-Net achieved a dice score of 0.85 for segmenting discs on spine MR images.[7,8,48,49] Using DeepLab achieved a Jaccard index of 0.82 for vertebral CT segmentation. Three-dimensional (3D) CNNs have shown superior performance in segmenting complex spine structures, achieving dice scores over 0.90 for various structures.[7,8]

CLASSIFYING AND DIAGNOSING SPINE PATHOLOGIES

Degenerative Spine

Given the increasing prevalence of spinal degenerative disease, there is great potential for AI-assisted ML MR imaging interpretation to streamline care for these patients. In providing rapid automated analysis of MR imaging scans, ML technology may also assist in reducing radiologist workloads. ML models can generate quantitative parameters from imaging data, which are time-consuming for a radiologist to produce.

Spinal Stenosis

Clinically, spinal stenosis (SS) diagnosis relies on subjective evaluation, and spine MRIs are essential for accurate evaluation. DL is promising in automatically learning representative imaging features to perform classifications. Initially focused on automating vertebral numbering and disk classification, DL more recently has assessed automated SS grading. Studies have looked at individual ordinal/multiclass (normal, mild, moderate, severe) and binary/dichotomous stenosis grading (stenosis vs no stenosis; normal–mild vs moderate–severe; normal-mild-moderate vs severe).

Lu and colleagues developed a DL model (Deep Spine) to grade central canal (CC) and neural foramen (NF) stenosis using axial and sagittal T2W images. A large data set of LS MRIs was used but relied on natural language processing labels from existing radiology reports. Average class accuracy

(normal, mild, moderate, or severe stenoses) was 70.6% CC and 67.1% NF stenosis.[49]

SpineNet,[42] a multitask architecture, developed automated classification of several spinal conditions, including CC stenosis on sagittal and axial T2W images.[50] SpineNet achieved agreement of 65.7% for ordinal gradings (normal, mild, moderate, or severe) compared with an expert radiologist and 94% (κ = 0.75) for binary/dichotomous grading of normal-mild-moderate versus severe stenosis compared with an experienced orthopedic surgeon.[51]

Lateral Recess Stenosis

Hallinan and colleagues showed comparable agreement with subspecialist radiologists for classifying CC and lateral recess stenosis, with slightly lower agreement for NF stenosis on LS MR imaging. Dichotomous classification (normal/mild vs moderate/severe) showed good agreement for both radiologists and DL model: κ values 0.98, 0.98, and 0.96 CC; 0.92, 0.95, and 0.92 lateral recesses; and 0.94, 0.95, and 0.89 NF ($P < .001$) for an MSK radiologist, subspecialist radiologists 1 and 2, (31, 5 and 9 year experience, respectively). The DL model also showed good agreement for dichotomous classification of NF stenosis (κ = 0.89; $P < .001$), which was slightly reduced compared with subspecialist radiologists (κ = 0.94, 0.95; $P < .001$). These results compare favorably with the average DL model from the SpineNet study, which had κ values 0.82, 0.96 for ordinal and dichotomous classification of CC stenosis, respectively ($P < .001$).[40]

Facet Arthropathy

Bharadwaj and colleagues proposed a two-staged learning system that automatically evaluated T2W axial LS MR images classifying CC and NF stenosis and for the first time facet arthropathy (Fig. 4). The first stage—localization of anatomic regions—was performant with excellent volumetric Dice scores (well above 0.90) for the dural sac and disk, and with no additional training or fine-tuning, localization of the foramen and facet was also favorable. In the second stage, the interpretable approach to multiclass grading (normal, mild, moderate, severe) of CC stenosis was in line with pairwise agreements between three radiologists (two senior subspecialty radiologists each with greater than 20 year experience; one junior radiologist with 3 year experience) and significantly outperformed a black-box CNN. Models also showed accurate binary classification (normal/mild vs moderate/severe) of both NF stenosis and facet arthropathy with area under the receiver operating characteristic (AUROC) curves 0.92 and 0.93, respectively (Fig. 5).

Facet arthropathy, prevalent in 15% to 45% of patients presenting with chronic low back pain, is a very important underdiagnosed etiology. In clinical practice, MR imaging is used to identify disc herniation, NF stenosis, and CC stenosis, whereas facet arthropathy is often omitted as a descriptor in radiology MR imaging reports. This approach targets a more comprehensive evaluation of LS MR imaging and assessment of features associated with back pain.[7]

Modic Changes

Modic changes (MCs), endplate–adjacent marrow signal abnormalities representing sequela of structural and inflammatory changes, are hypothesized to be potentially associated with pain. MC type 1: edema or fibrovascular changes (hypointense T1W, hyperintense T2W images); MC type 2: fatty marrow (hyperintense T1W, iso-hyperintense T2WFS, and non-FS T2W sequences, respectively); MC type 3: sclerotic (hypointense on both sequences). The semiquantitative nature of MC classification is highly susceptible to variability in

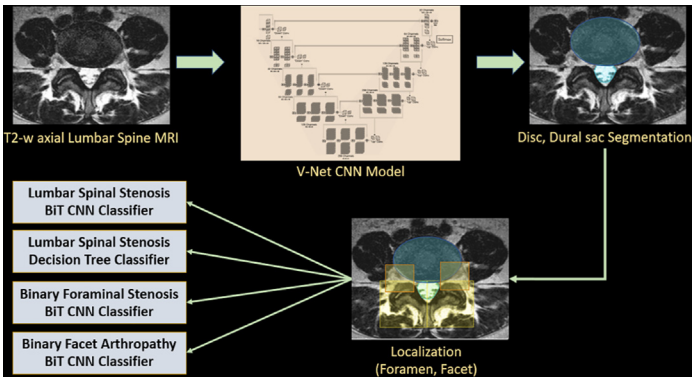

Fig. 4. Overview of DL-pipeline. Axial T2-weighted slices passed into V-Net segmentation models obtain masks for disc and dural sac (DDS). Geometric rules based on DDS localize foramen and facet bounding boxes and passes into its corresponding classifier: Big Transfer (BiT) CNN classifies lumbar SS, foraminal stenosis, and facet arthropathy. Lumbar SS interpretable classification (decision tree) relies on additional quantitative metrics extracted from the DDS segmentations.[7]

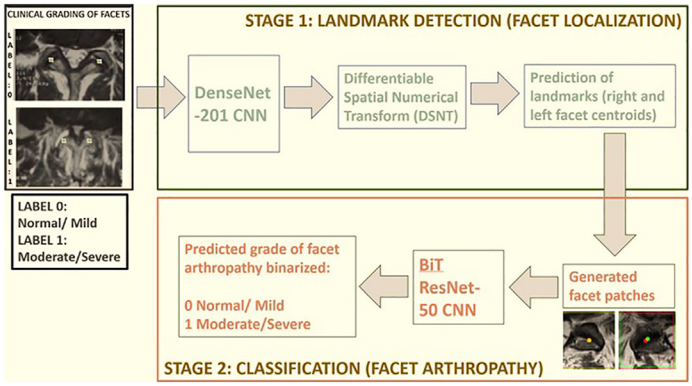

Fig. 5. Second stage: landmark detection (facet localization) and binary classification (facet arthropathy). Landmark coordinate regression system uses a DSNT (differentiable spatial-to-numerical transform) layer on the top of a DenseNet–CNN backbone architecture to predict two coordinates per slice, corresponding to right and left facet. Crop patches of facets (36 mm × 36 mm bounding boxes) center around predicted facet landmarks. Green, ground truth; Red, model prediction. Binary classification model (BiT) with a ResNet-50 backbone architecture classifies the patches from first stage as normal/mild versus moderate/severe arthropathy.[7].

non-standardized imaging.[52] Wang and colleagues extracted morphologic and signal intensity-based metrics from contours of MCs, reporting improved inter- and intra-rater agreement compared with unassisted MC classification.[53] The need for labor-intensive manual demarcation of MCs is a significant limitation.

Gao and colleagues used DL-based models to automatically map MCs. Overall, these results demonstrate substantial agreement of the detection model with radiologist-annotated grading and a novel Modic mapping technique providing grading assistance. MCs are often transitional (27.2% regarded as mixed, comprising characteristics of multiple Modic types).[54] Capturing this granularity of mixed MCs is challenging for the human eye. A voxel-wise MC segmentation method was therefore implemented due to its key capability of visualizing the heterogeneity of mixed MCs. In addition, the segmentation methodology offers higher degree of supervision, retaining context of the neighboring tissue and improving label specificity. Further works using this approach can unravel attributes of progressive or transitional MCs that may interact with pain, as heterogeneous tissues are often correlated with degeneration[9] (**Fig. 6**).

In the MC detection component, the distribution of predicted MCs across the LS was predominantly in the L4–S1 range (74.4%), matching well with the radiologist annotations (78.8%) and past work (75.5%).[54] MC voxel-wise classification yielded high predictive value of MC types 1 and 2, the groups most important to classify due to their prevalence and strong association of MC 1 with nonspecific low back pain (LBP).[55] The additional utility of the model predictions improved agreement of the junior radiologist (3 year experience) grading the MCs with the two senior (each >25 year experience) subspecialty radiologists ($\Delta\kappa = +0.06$ and $\Delta\kappa = +0.03$ reader 1 and reader 2, respectively).[9]

Spine Infection and Inflammation

Modic changes versus pyogenic spondylitis

Spinal infection diagnosis is difficult due to nonspecific clinical and laboratory findings. Noninfectious conditions can mimic the imaging findings of spondylodiscitis and it can be difficult to distinguish between degenerative and infectious endplate abnormalities.

MC type 1 can mimic infection on MR imaging. Differences between inflammatory, degenerative, and infectious pathologies have a significant impact on prognosis. Many imaging techniques, including conventional plain radiography, CT, MR imaging, and radionuclide studies, have been used to diagnose spinal infections.

A retrospective study evaluated the performance of a CNN to differentiate pyogenic spondylitis from MC on MR imaging. Fifty MRIs, each from pyogenic spondylitis and MC patients, were reviewed, comparing the performance of the CNN to four clinicians: a radiologist, spine surgeon, and two orthopedic surgeons (17, 20, 5–6 year experience, respectively).

The CNN-based AUROC curve from the T1 weighted image (T1W), T2W, and STIR images was 0.95, 0.94, and 0.95, respectively. The accuracy of the CNN was significantly greater than that of the four clinicians on T1W and STIR ($P < .05$), and better than a radiologist and one orthopedic surgeon on the T2W ($P < .05$). The sensitivity was significantly better than the four clinicians on T1W and STIR ($P < .05$) and better than a radiologist and one orthopedic surgeon on the T2W ($P < .05$). The specificity was significantly better than one orthopedic surgeon on T1W and T2W ($P < .05$) and better than both orthopedic

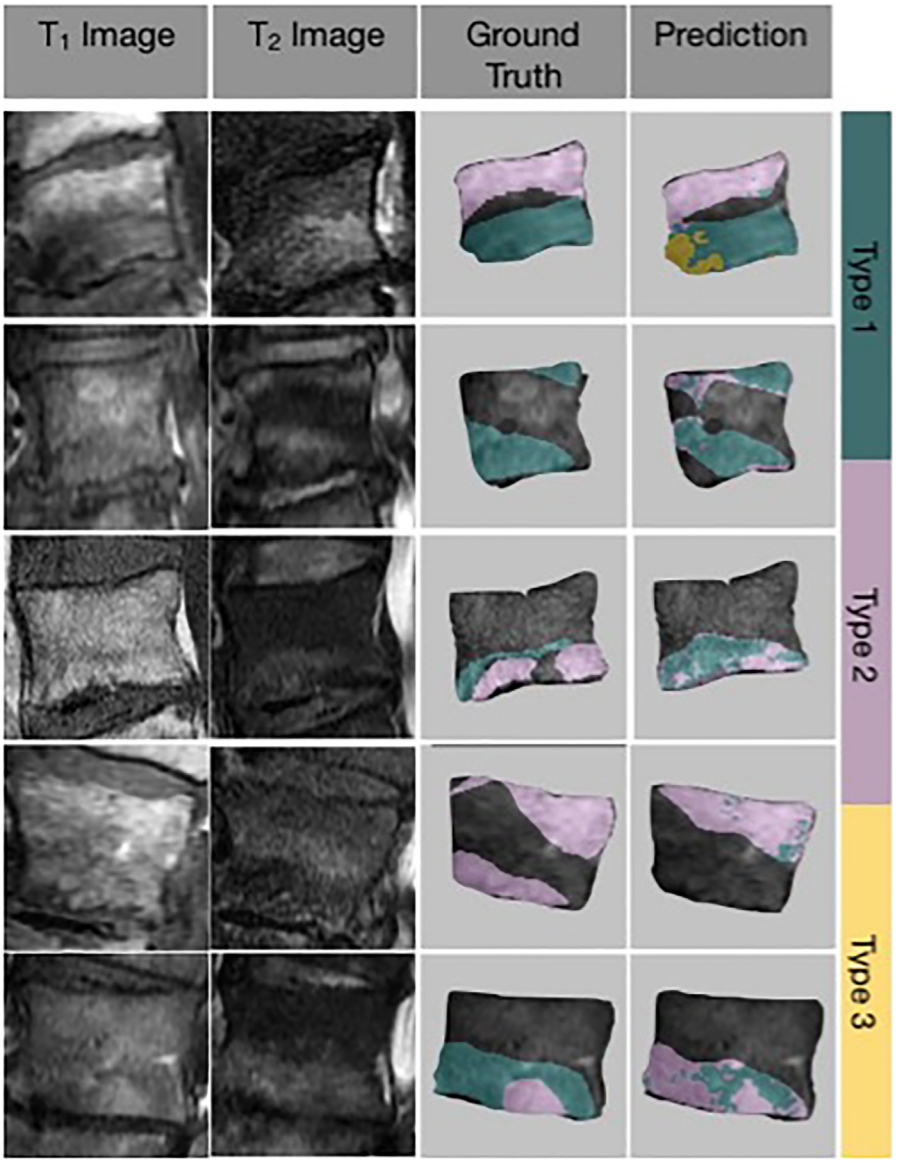

Fig. 6. Modic maps. Representative examples of inputs (T_1 and T_2 images), radiologist-annotated ground truth segmentations, predicted Modic maps. Model is advantageous for visualizing heterogeneity and transitional pathology: top row: model detects MC type 3 in anterior inferior endplate; second row: small MC type1 region in anterior superior endplate, unmarked by the radiologist and annotated by the model.[9]

surgeons on STIR ($P < .05$). The investigators concluded that the CNN model was able to differentiate between MCs and pyogenic spondylitis and the model performance was comparable to, or better than, that of the radiologist, spine surgeon, and two orthopedic surgeons.[56]

Multiple sclerosis

The spinal cord (SC) is frequently affected by demyelinating lesions and atrophy in multiple sclerosis (MS) patients. Spinal cord and lesion segmentation may provide diagnosis, prognosis, and longitudinal monitoring in MS. Automating segmentation may assist in decreasing inter-rater variability and increase efficiency of large-throughput analysis pipelines. Precise outlining of lesions is challenging due to their heterogeneity, and reliable segmentation across multisite SC data is difficult due to variability related to acquisition parameters and image artifacts.

Gros and colleagues created an original automated SC and MS lesion segmentation method based on two CNNs (Spinal Cord Toolbox [SCT], open-source readily available). When compared

against manual segmentation, the investigators demonstrated this CNN-based approach showed a median Dice of 95% versus 88% compared with PropSeg ($P < .05$), a state-of-the-art SC segmentation method. On MS data, the framework provided a Dice of 60%, a relative volume difference of 15%, and a lesion-wise detection sensitivity and precision of 83% and 77%, respectively. Their SC segmentation results outperformed a state-of-the-art method on a multisite and highly heterogeneous clinical data set. Lesion segmentation results were generally within range of manual segmentations, although the false-positive rate (FP) warrants further investigations.[57]

Spine Trauma

Fracture

AI has been studied for detecting both osteoporotic and traumatic spinal fractures. The clinical management of these patients requires timely and accurate interpretation of volumetric imaging. Automated image analysis has the potential to streamline care of patient with spinal fractures.

Osteoporotic Tomita and colleagues compared a CNN model detecting osteoporotic vertebral fractures (VFs) to the radiologists' diagnoses from 1432 CT scan radiology reports. The CNN achieved an accuracy of 89.2% and F1 score of 90.8% on the test/validation data set (F1 score: ML metric measuring model accuracy). Although the F1 score for the model matched the performance of practicing radiologists on the test set, the radiologists' diagnoses were more precise and had higher specificity relative to the model. The investigators proposed the model could be used to assist and improve the diagnosis of osteoporotic VFs in clinical settings by pre-screening routine CT examinations and flagging suspicious cases before review by radiologists.[58]

Bush and colleagues created an automated ML system to detect and classify osteoporotic VFs according to Genant standards and to measure bone density of thoracic and lumbar vertebral bodies on CT using 3D CT images. They retrospectively analyzed 210 thoracic and lumbar vertebrae with VFs that were electronically marked and classified by a radiologist.

The sensitivity for detection/localization was 95.7% with a FP of 0.29 per patient. Sensitivity was 98.7% and specificity was 77.3% at case-based ROC curve analysis. Accuracy for classification by Genant type (anterior, middle, or posterior height loss) was 0.95 with weighted $\kappa = 0.90$. Accuracy for categorization by Genant height loss grade was 0.68, with a weighted $\kappa = 0.59$. The average bone attenuation for T12–L4 vertebrae was 146 HU$\pm$29 (standard deviation) in patients versus 173 HU$\pm$42 in controls ($P < .001$).[59]

Traumatic thoracic lumbar Murata and colleagues used a CNN on anteroposterior and lateral thoracolumbar radiographs of 300 patients to detect VFs with accuracy 86% and sensitivity 84.7%. They compared model performance with orthopedic surgeons and residents and determined it was non-inferior to orthopedic surgeons and had superior sensitivity compared with orthopedic residents. The model did not reveal which vertebra was fractured.[60]

Burns and colleagues also designed an automated system for retrospective detection/localization of traumatic thoracic and lumbar VFs on CT in 104 patients compared with the localizations and classifications marked by a radiologist according to Denis column involvement.

Testing set sensitivity for the detection and localization of fractures within each vertebra was 0.81, with an FP of 2.7. The most common cause of FPs was nutrient foramina (39%).

Origins of false-negative (FN) findings (misses) and FP fracture line detections were decided in a consensus review by two board-certified fellowship-trained radiologists (19 years of experience). Most of FNs were fracture lines paralleling in close proximity to vertebral end plates and to degenerative joint disease.

The investigators concluded their fully automated computer system detected and anatomically localized thoracic and lumbar VFs on CT with a high sensitivity and a low FP rate.[61]

Traumatic cervical Cervical spine injury can be associated with high morbidity and mortality. Multidetector CT has emerged as a critical SOC imaging technique to evaluate cervical spine trauma for rapid diagnosis and intervention.

An FDA-approved CNN by Aidoc (www. aidoc.com) to detect cervical VFs on CT was used to analyze 665 examinations. Ground truth was established by retrospective visualization of VFs on CT using all available CTs, MRIs and CNN output information. The finalized cervical spine CT reports were simultaneously independently reviewed by two fellowship-trained neuroradiologists.

The CNN sensitivity (79%) was lower than the radiologists (93%) and CNN accuracy of 92% compared with 96% for the radiologists. Time from image acquisition to CNN analysis was shorter than time from image acquisition to radiologist report finalization emphasizing the value of the CNN in worklist prioritization. CNN false-negative examinations demonstrated that the

locations of CNN misses closely matched those of radiologists, similar by level and location. Fractures the CNN missed included severe fracture–dislocation, distractions, distal spinous processes, and lower cervical spine where fine bony detail was obscured by CT beam attenuation. There were a few instances in which the CNN detected a fracture that the radiologist missed, underscoring the ability of the CNN to function as a useful complementary tool in fracture detection that radiologists would review before report finalization.[62]

Spinal Cord Injury

The degree of abnormal T2 signal on sagittal and axial MR imaging has strongly correlated with diagnostic and prognostic value after SC injury (SCI). The intramedullary lesion length (IMLL) and the Brain and Spinal Injury Center (BASIC) score are two metrics based on single two-dimensional (2D) images that are approximations for true 3D size and distribution of injured SC.

The SCT[63] is an open-source anatomic atlas allowing quantitative volumetric injury analysis. Using a semiautomated image processing pipeline incorporating many tools available freely as part of the open-source SCT, volumetric injury measures were extracted from preoperative MR images in 47 patients who underwent 3T MR imaging within 24 hours of SCI. This customized image analysis and processing pipeline integrated three different novel 2D-CNN architectures for both whole SC and traumatic lesion segmentation. Segmentation results from CNNs were compared with each other and with standard manual segmentation as well as with two current state-of-the-art SC segmentation algorithms (PropSeg; DeepSeg). Compared with manual labeling, the average test set Dice coefficient for the BASIC segmentation model was 0.93 for SC segmentation versus 0.80 for PropSeg and 0.90 for DeepSeg (both SCT components).

These segmented volumetric measures predicted lower extremity motor function at discharge from the hospital more strongly than standard 2D radiographic parameters including BASIC score and IMLL. The investigators concluded that the volume of the T2 lesion after SCI was a more accurate imaging biomarker of injury severity than conventional 2D MR imaging measures of injury[64] (**Fig. 7**).

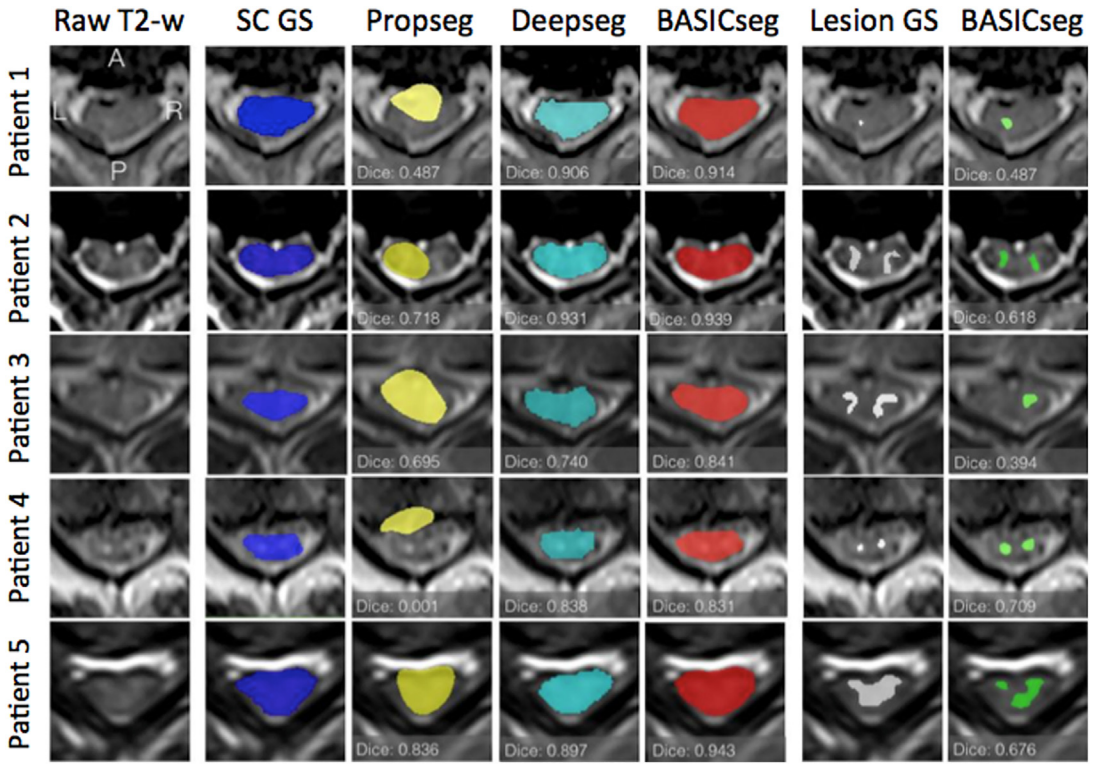

Fig. 7. Whole SC and lesion SC segmentation performance: BASICseg algorithms in five SCI patients at representative axial slices compared with manual ground truth segmentation (GS), PropSeg and DeepSeg models: column 1: conventional axial T2; columns 2 to 5: whole SC segmentation; columns 6 to 7: cord lesion segmentation. (*Courtesy* of JF Talbott, San Francisco, CA.[64])

Spine oncology

Radiomics-based feature analysis and CNNs are two popular ML imaging techniques used in oncology. Radiomics-based techniques extract first and higher order statistical features from radiological images. Interpretability of radiomic features has been a major limitation of radiomics as the features are not standardized and there is difficulty relating them to the underlying biology of the tissue of interest. As the features are limited to the knowledge of the radiologist or clinician, the accuracy of the developed algorithm could be reduced.[65,66]

Differentiating Benign Versus Metastatic Vertebral Fractures

Differentiating benign VFs from pathologic/malignant VFs is critically important for treatment decisions. A novel Two-Stream Compare and Contrast Network (TSCCN: three-class classification: normal, benign, and malignant VFs)[67] model was tested on 239 VFs on median sagittal T1W, T2WFS, and a combination T1W/T2WFS sequences.

Three radiologists (11, 15, and 8 year experience) assessed the same MR images twice at different times within 1 month and were blinded to patient history, treatment details for fair comparison with the TSCCN.

The model achieved average sensitivity, specificity, and accuracy of 92.6%, 96.3%, and 95.2%, respectively. The sensitivity of the model overlapped or was less than the sensitivity of the most experienced radiologist; the accuracy and specificity of the model was higher than achieved by the radiologists. The investigators concluded that the model had the potential to enhance VF diagnostic accuracy, sensitivity, and specificity.[68]

Deformity

ML has also analyzed spinal deformity. Spinal parameters are time-consuming to manually annotate and inter-rater reliability can vary. Research has focused on automated quantitative spinal parameters localizing various landmarks: endplates, hip joints, S1 angle, T4–T12 kyphosis, L1–L5 lordosis, Cobb angle, pelvic incidence, sacral slope, and pelvic tilt. Although model-generated parameters can show good correlation to radiologist-derived values, the standard errors of the estimated parameters can range from 2.7 for pelvic tilt to 11.5 for L1–5 lordosis.[69,70]

Landmark annotations are not feasible within the clinical workflow as they require significant user input and are vulnerable to user error. Severe spinal deformity, overlapping soft tissues, lead shields, body habitus, osteoporosis, transitional anatomy, and variable skeletal maturities are several factors reducing visibility and reliable manual identification. Iriondo and colleagues developed a new method for automatic extraction of vertebral midline from biplanar radiographs and 3D spine shape models. The developed landmark extraction algorithms demonstrated robust performance across the tested data sets, are fully automatic, and may integrate into the clinical workflow, allowing temporal evaluation of deformity progression (**Figs. 8** and **9**).

A surprising finding was the range in actual sagittal and coronal imbalance among images linked to radiology reports stating “no imbalance” with true measurements of approximately $\pm$2.5 cm of coronal imbalance and $\pm$5 cm sagittal imbalance. This has important implications, when no exact measurement was provided, qualitative descriptions of spinal alignment were subjective.[71]

CLINICAL IMPLEMENTATION

Many AI models discussed can be integrated into automated reading workflow. Reporting individual spinal CC and NF stenoses, disc, and facet degeneration is time-consuming. Clinically deploying DL models could assist in semiautomated reporting under radiologist supervision to provide more consistent, objective, and efficient reporting.

A published DL algorithm, Spine AI, automatically classifies LS CC, lateral recess, and NF stenoses on LS MRIs.[72] Eight radiologists (2–13 year experience), retrospectively reviewed studies with and without DL assistance with a 1-month washout were compared with test data labeled by an external MSK radiologist (32 year experience) as standard. Interpretation time reduced 62% to 74% for DL-assisted radiologists: mean 124-274 to 47-71 seconds ($P < .001$) with greatest time savings for in-training radiologists: mean 274 (unassisted) to 71 seconds (assisted) ($P < .001$).

DL-assisted radiologists had superior or equivalent interobserver agreement for all stenosis gradings compared with unassisted radiologists. DL-assisted general and in-training radiologists improved their interobserver agreement for four-class NF stenosis, $\kappa = 0.71$ and 0.70 (DL) versus 0.39 without DL, respectively (both $P < .001$).[73]

DL assistance can streamline report generation, which involves image review and a separate text input into a reporting module. DL assistance can detect regions of interest (ROIs), grade stenosis, and automatically generate a sentence directly into the reporting module. The radiologist can change and control the DL-assisted predictions before a report is generated, important for safety and patient preference.[74] Strategic “one-click” solutions integrated within the normal radiologist workflow will be prerequisite for successful implementation.[75]

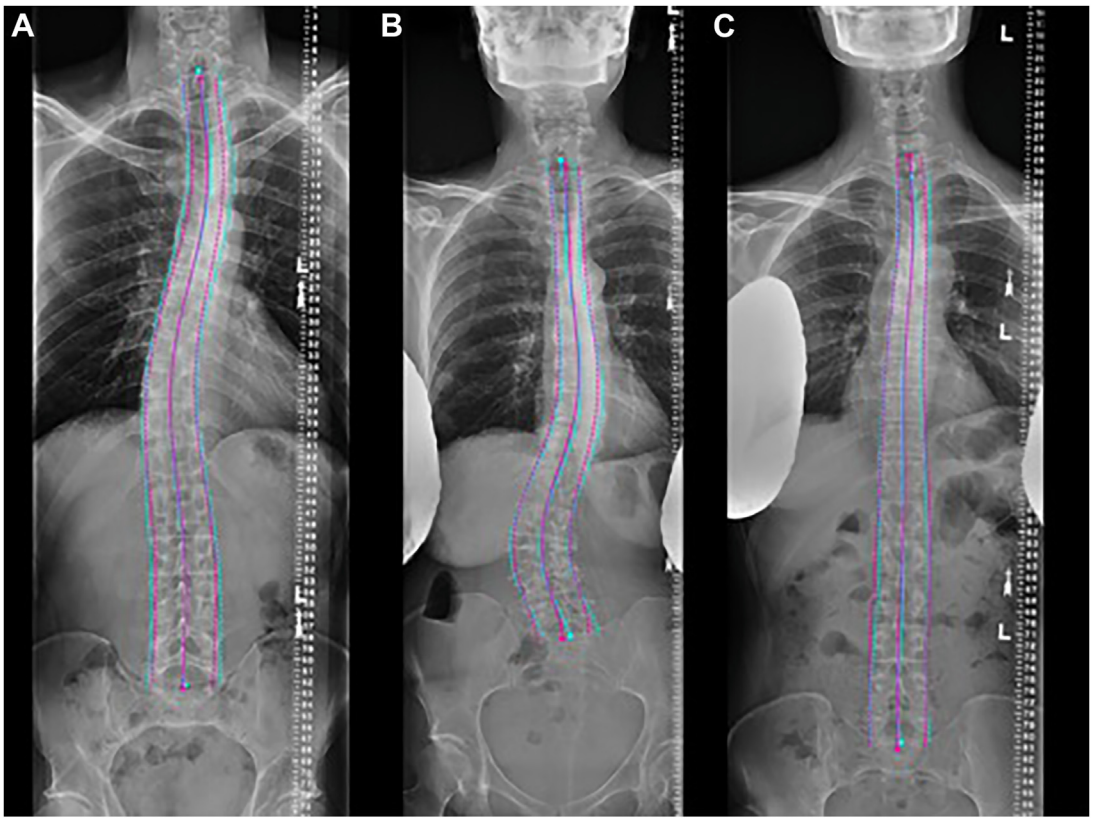

Fig. 8. Frontal spinal radiographs show accurate spine midline identification: model predicted curves (magenta) closely correlate with ground truth curves (blue). Paired dashed lines (—): spine contour curves; Center solid lines (—): midline curves.

CHALLENGES AND FUTURE DIRECTIONS

Most of the AI spine imaging studies are retrospective, single-center with small sample sizes. Randomized, controlled, multicenter studies will be required to validate applications, facilitating clinical practice.[66,73] Future work will test pooled, publicly available data sets from collaborating institutions by multidisciplinary teams, comparing results with published work.[71] Increasing high-quality competitions (ie, 2022 RSNA cervical spine AI fracture challenge) may further research toward novel AI spine algorithms.

Required trained human input limits practical clinical implementation. Manual radiologist image labeling, most accurate for model training, is labor-intensive. Many models also require trained human post-processing (placing ROIs for segmentation).[40] Hardware and innovative data processing developing robust ML models with human-level performance will be needed.[75]

Cultural challenges are also barriers. ML tools requiring large, tens of thousands, annotated data sets face systemic data privacy concerns (storage, transmission, usage) subject to strict regulations despite standard data anonymization.[70]

Another cultural challenge, AI perceived as a "black box," relates to the opaque nature of DL systems. We see the input and output; however, the system's code or logic producing the output is not inherently transparent. Explainable AI is a branch attempting to make the methodology transparent to users.

Medical accountability is another challenge. Would the clinician using the ML system or its manufacturer be responsible for an AI-generated misdiagnosis? This also affects marketing approval and cost of novel AI tools, requiring deeper testing and verification relative to other technologies and longer time to market.[76]

Potential biases are concerning with unintended biases resulting from scarce data sets (rare pathologies, ethnicities) and privileged easier-accessed data sets.[77] Efforts toward governing AI aim to build robust public trust.[78] The European Union's General Data Protection Regulation expanded patient rights considerably with an explicit opt-in policy regarding data processing permission. Open

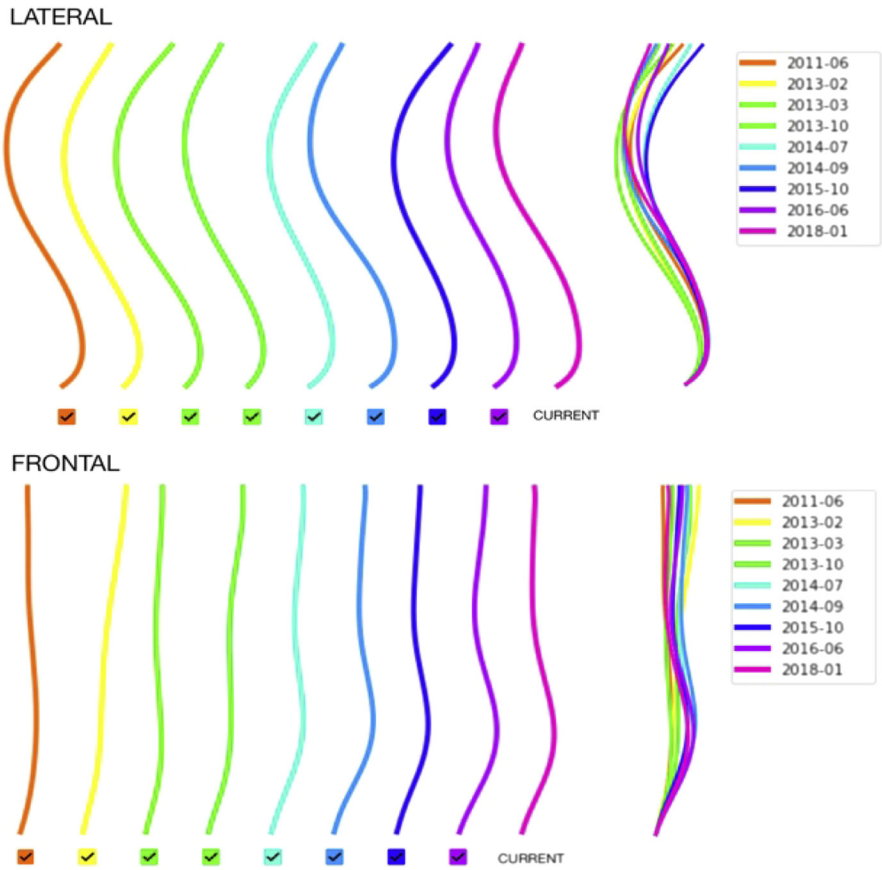

Fig. 9. Sagittal and coronal views of spine midlines in a 12-year-old boy monitored over 6.5 years allow overall comparisons demonstrating progressive levoscoliotic deformity over time.[71]

access to ML models and training data may improve public trust, accountability, and prediction bias. As models improve and clinicians more heavily rely on automated diagnoses, continued multidisciplinary collaboration involving radiologists, clinicians, engineers, data scientists, ethicists, and policy makers is mandatory globally.[76,75]

SUMMARY

AI potentially increases efficiency, reducing time-consuming tasks and assisting radiologists in specific diagnoses with a goal to provide comprehensive automated image analysis. Characterizing lesions and identifying lesions that might be missed are of great benefit, allowing earlier diagnoses, reporting, and treatment.[75] Combining clinical data with image analysis may permit improved and more personalized treatment decision-making. Although challenges remain, AI technology continues to rapidly progress with great potential to improve patient care and outcomes.[70]

CLINICS CARE POINTS

- Artificial intelligence (AI) models can assist in diagnosing spine pathologies such as degenerative disease, tumor, infection, fracture, with increasing sensitivity and specificity when used in parallel with radiologists' input and supervision.
- AI models integrated into automated reading workflow can significantly decrease interpretation time. This is most evident for radiologists in training.
- AI models can improve SNR and reduce artifacts inherent in rapid spine imaging protocols resulting in diagnostic quality imaging with 40-50% reduction in image acquisition time.
- AI models may assist in 'pre-screening' routine imaging examinations and flagging suspicious cases before review by radiologists. This allows added value through worklist prioritization.

- Clinically deploying AI models may assist in semiautomated reporting under radiologist supervision to provide more consistent and objective reporting.

DISCLOSURE

C.T. Chin research support UCSF Edward A. Dickson Professorship Award.

REFERENCES

1. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021;4(1):65.
2. Cheng PM, Montagnon E, Yamashita R, et al. Deep Learning: An Update for Radiologists. Radiographics 2021;41(5):1427–45.
3. Do S, Song KD, Chung JW. Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. Korean J Radiol 2020;21(1):33–41.
4. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(8):500–10.
5. Kelly BS, Judge C, Bollard SM, et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). Eur Radiol 2022;32(11):7998–8007.
6. Montagnon E, Cerny M, Cadrin-Chenevert A, et al. Deep learning workflow in radiology: a primer. Insights Imaging 2020;11(1):22.
7. Bharadwaj UU, Christine M, Li S, et al. Deep learning for automated, interpretable classification of lumbar spinal stenosis and facet arthropathy from axial MRI. Eur Radiol 2023;33(5):3435–43.
8. Hess M, Allaire B, Gao KT. Deep Learning for Multi-Tissue Segmentation and Fully Automatic Personalized Biomechanical Models from BACPAC Clinical Lumbar Spine MRI. Pain Med 2022. https://doi.org/10.1093/pm/pnac142.
9. Gao KT, Tibrewala R, Hess M, et al. Automatic detection and voxel-wise mapping of lumbar spine Modic changes with deep learning. JOR Spine 2022;5(2):e1204.
10. Almansour H, Herrmann J, Gassenmaier S, et al. Deep Learning Reconstruction for Accelerated Spine MRI: Prospective Analysis of Interchangeability. Radiology 2023;306(3):e212922.
11. Schlaeger S, Drummer K, Husseini ME, et al. Implementation of GAN-Based, Synthetic T2-Weighted Fat Saturated Images in the Routine Radiological Workflow Improves Spinal Pathology Detection. Diagnostics 2023;13(5).
12. Fasterholdt I, Naghavi-Behzad M, Rasmussen BSB, et al. Value assessment of artificial intelligence in medical imaging: a scoping review. BMC Med Imag 2022;22(1):187.
13. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med 2022;28(1):31–8.
14. Shevlin H, Vold K, Crosby M, et al. The limits of machine intelligence: Despite progress in machine intelligence, artificial general intelligence is still a major challenge. EMBO Rep 2019;20(10):e49177.
15. Erickson BJ, Korfiatis P, Akkus Z, et al. Machine Learning for Medical Imaging. Radiographics 2017;37(2):505–15.
16. Aljuaid A, Anwar M. Survey of Supervised Learning for Medical Image Processing. SN Comput Sci 2022;3(4):292.
17. Raza K, Singh NK. A Tour of Unsupervised Deep Learning for Medical Image Analysis. Curr Med Imaging 2021;17(9):1059–77.
18. Rana M, Bhushan M. Machine learning and deep learning approach for medical image analysis: diagnosis to detection. Multimed Tools Appl 2022;1–39.
19. Hu M, Zhang J, Matkovic L, et al. Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions. J Appl Clin Med Phys 2023;24(2):e13898.
20. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 2017;19:221–48.
21. Cui Y, Zhu J, Duan Z, et al. Artificial Intelligence in Spinal Imaging: Current Status and Future Directions. Int J Environ Res Publ Health 2022;19(18).
22. Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 2018;9(4):611–29.
23. Soffer S, Ben-Cohen A, Shimon O, et al. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. Radiology 2019;290(3):590–606.
24. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Trans Med Imag 2016;35(5):1299–312.
25. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evol Intell 2022;15(1):1–22.
26. Kiryati N, Landau Y. Dataset Growth in Medical Image Analysis Research. J Imaging 2021;7(8).
27. Kashiwagi N, Tanaka H, Yamashita Y, et al. Applicability of deep learning-based reconstruction trained by brain and knee 3T MRI to lumbar 1.5T MRI. Acta Radiol Open 2021;10(6). 20584601211023939.
28. Sun S, Tan ET, Mintz DN, et al. Evaluation of deep learning reconstructed high-resolution 3D lumbar spine MRI. Eur Radiol 2022;32(9):6167–77.
29. Zochowski KC, Tan ET, Argentieri EC, et al. Improvement of peripheral nerve visualization using a deep

learning-based MR reconstruction algorithm. Magn Reson Imaging 2022;85:186–92.
30. Yeoh H, Hong SH, Ahn C, et al. Deep Learning Algorithm for Simultaneous Noise Reduction and Edge Sharpening in Low-Dose CT Images: A Pilot Study Using Lumbar Spine CT. Korean J Radiol 2021; 22(11):1850–7.
31. Greffier J, Frandon J, Durand Q, et al. Contribution of an artificial intelligence deep-learning reconstruction algorithm for dose optimization in lumbar spine CT examination: A phantom study. Diagn Interv Imaging 2023;104(2):76–83.
32. Chazen JL, Tan ET, Fiore J, et al. Rapid lumbar MRI protocol using 3D imaging and deep learning reconstruction. Skeletal Radiol 2023;52(7):1331–8.
33. Zaharchuk G, Gong E, Wintermark M, et al. Deep Learning in Neuroradiology. AJNR Am J Neuroradiol 2018;39(10):1776–84.
34. Han M, Bahroos E, Hess ME, et al. Technology and Tool Development for BACPAC: Qualitative and Quantitative Analysis of Accelerated Lumbar Spine MRI with Deep-Learning Based Image Reconstruction at 3T. Pain Med 2023;24(Suppl 1):S149–59.
35. Bash S, Johnson B, Gibbs W, et al. Deep Learning Image Processing Enables 40% Faster Spinal MR Scans Which Match or Exceed Quality of Standard of Care : A Prospective Multicenter Multireader Study. Clin Neuroradiol 2022;32(1):197–203.
36. Roberts M, Hinton G, Wells AJ, et al. Imaging evaluation of a proposed 3D generative model for MRI to CT translation in the lumbar spine. Spine J 2023. https://doi.org/10.1016/j.spinee.2023.06.399.
37. Schlaeger S, Drummer K, El Husseini M, et al. Synthetic T2-weighted fat sat based on a generative adversarial network shows potential for scan time reduction in spine imaging in a multicenter test dataset. Eur Radiol 2023;33(8):5882–93.
38. Tanenbaum LN, Bash SC, Zaharchuk G, et al. Deep Learning-Generated Synthetic MR Imaging STIR Spine Images Are Superior in Image Quality and Diagnostically Equivalent to Conventional STIR: A Multicenter, Multireader Trial. AJNR Am J Neuroradiol 2023;44(8):987–93.
39. Cina A, Bassani T, Panico M, et al. 2-step deep learning model for landmarks localization in spine radiographs. Sci Rep 2021;11(1):9482.
40. Hallinan J, Zhu L, Yang K, et al. Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI. Radiology 2021;300(1):130–8.
41. Chen Y, Gao Y, Li K, et al. Vertebrae Identification and Localization Utilizing Fully Convolutional Networks and a Hidden Markov Model. IEEE Trans Med Imag 2020;39(2):387–99.
42. Jamaludin A, Kadir T, Zisserman A. SpineNet: Automated classification and evidence visualization in spinal MRIs. Med Image Anal 2017;41:63–73.
43. Ghosh S, Chaudhary V. Supervised methods for detection and segmentation of tissues in clinical lumbar MRI. Comput Med Imag Graph 2014;38(7): 639–49.
44. Liebl H, Schinz D, Sekuboyina A, et al. A computed tomography vertebral segmentation dataset with anatomical variations and multi-vendor scanner data. Sci Data 2021;8(1):284.
45. Xuan J, Ke B, Ma W, et al. Spinal disease diagnosis assistant based on MRI images using deep transfer learning methods. Front Public Health 2023;11: 1044525.
46. Ronneberger O, Fischer P, Broz T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. 2015.
47. Chen L-C, Papandreou G, Kokkinow I, Murphy K. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. arXiv 2017.
48. Wang S, Jiang Z, Yang H, et al. Automatic Segmentation of Lumbar Spine MRI Images Based on Improved Attention U-Net. Comput Intell Neurosci 2022;2022:4259471.
49. Lu J-T, Pedemonte S, Bizzo B, et al. Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. Proceedings of Machine Learning Research 2018;85:16.
50. Jamaludin A, Lootus M, Kadir T, et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. Eur Spine J 2017;26(5):1374–83.
51. Ishimoto Y, Jamaludin A, Cooper C, et al. Could automated machine-learned MRI grading aid epidemiological studies of lumbar spinal stenosis? Validation within the Wakayama spine study. BMC Muscoskel Disord 2020;21(1):158.
52. Fields AJ, Battie MC, Herzog RJ, et al. Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): recommendations from the ISSLS Degenerative Spinal Phenotypes Group. Eur Spine J 2019;28(10):2266–74.
53. Wang Y, Videman T, Niemelainen R, et al. Quantitative measures of modic changes in lumbar spine magnetic resonance imaging: intra- and inter-rater reliability. Spine 2011;36(15):1236–43.
54. Xu L, Chu B, Feng Y, et al. Modic changes in lumbar spine: prevalence and distribution patterns of end plate oedema and end plate sclerosis. Br J Radiol 2016;89(1060):20150650.
55. Jensen TS, Karppinen J, Sorensen JS, et al. Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. Eur Spine J 2008;17(11):1407–22.

56. Mukaihata T, Maki S, Eguchi Y, et al. Differentiating Magnetic Resonance Images of Pyogenic Spondylitis and Spinal Modic Change Using a Convolutional Neural Network. Spine (Phila Pa 1976) 2023;48(4):288–94.
57. Gros C, De Leener B, Badji A, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. Neuroimage 2019;184:901–15.
58. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Comput Biol Med 2018;98:8–15.
59. Burns JE, Yao J, Summers RM. Vertebral Body Compression Fractures and Bone Density: Automated Detection and Classification on CT Images. Radiology 2017;284(3):788–97.
60. Murata K, Endo K, Aihara T, et al. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. Sci Rep 2020;10(1):20031.
61. Burns JE, Yao J, Munoz H, et al. Automated Detection, Localization, and Classification of Traumatic Vertebral Body Fractures in the Thoracic and Lumbar Spine at CT. Radiology 2016;278(1):64–73.
62. Small JE, Osler P, Paul AB, et al. CT Cervical Spine Fracture Detection Using a Convolutional Neural Network. AJNR Am J Neuroradiol 2021;42(7):1341–7.
63. De Leener B, Levy S, Dupont SM, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 2017;145(Pt A):24–43.
64. McCoy DB, Dupont SM, Gros C, et al. Convolutional Neural Network-Based Automated Segmentation of the Spinal Cord and Contusion Injury: Deep Learning Biomarker Correlates of Motor Impairment in Acute Spinal Cord Injury. AJNR Am J Neuroradiol 2019;40(4):737–44.
65. Parekh VS, Jacobs MA. Deep learning and radiomics in precision medicine. Expert Rev Precis Med Drug Dev 2019;4(2):59–72.
66. Ong W, Zhu L, Zhang W, et al. Application of Artificial Intelligence Methods for Imaging of Spinal Metastasis. Cancers 2022;14(16).
67. Feng S, Liu B, Zhang Y, et al. Two-Stream Compare and Contrast Network for Vertebral Compression Fracture Diagnosis. IEEE Trans Med Imag 2021;40(9):2496–506.
68. Liu B, Jin Y, Feng S, et al. Benign vs malignant vertebral compression fractures with MRI: a comparison between automatic deep learning network and radiologist's assessment. Eur Radiol 2023;33(7):5060–8.
69. Galbusera F, Niemeyer F, Wilke HJ, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. Eur Spine J 2019;28(5):951–60.
70. Merali ZA, Colak E, Wilson JR. Applications of Machine Learning to Imaging of Spinal Disorders: Current Status and Future Directions. Global Spine J 2021;11(1_suppl):23S–9S.
71. Iriondo C, Mehany S, Shah R, et al. Institution-wide shape analysis of 3D spinal curvature and global alignment parameters. J Orthop Res 2022;40(8):1896–908.
72. Ooi BC, Tan KL, Wang S, et al. SINGA: A Distributed Deep Learning Platform. Mm'15: Proceedings of the 2015 Acm Multimedia Conference. 2015:685-688.
73. Lim DSW, Makmur A, Zhu L, et al. Improved Productivity Using Deep Learning-assisted Reporting for Lumbar Spine MRI. Radiology 2022;305(1):160–6.
74. Young AT, Amara D, Bhattacharya A, et al. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021;3(9):e599–611.
75. Martin-Noguerol T, Onate Miranda M, Amrhein TJ, et al. The role of Artificial intelligence in the assessment of the spine and spinal cord. Eur J Radiol 2023;161:110726.
76. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. JOR Spine 2019;2(1):e1044.
77. Pesapane F, Volonte C, Codari M, et al. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging 2018;9(5):745–53.
78. Winfield AFT, Jirotka M. Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philos Trans A Math Phys Eng Sci 2018;376(2133).