



Detection and Localization of Spine Disorders from Plain Radiography

İlkay Yıldız Potter¹ · Diana Yeritsyan^{2,3} · Edward K. Rodriguez^{2,3} · Jim S. Wu⁴ · Ara Nazarian^{2,3,5} · Ashkan Vaziri¹

Received: 8 April 2024 / Revised: 16 May 2024 / Accepted: 9 June 2024 / Published online: 27 June 2024

© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2024

Abstract

Spine disorders can cause severe functional limitations, including back pain, decreased pulmonary function, and increased mortality risk. Plain radiography is the first-line imaging modality to diagnose suspected spine disorders. Nevertheless, radiographical appearance is not always sufficient due to highly variable patient and imaging parameters, which can lead to misdiagnosis or delayed diagnosis. Employing an accurate automated detection model can alleviate the workload of clinical experts, thereby reducing human errors, facilitating earlier detection, and improving diagnostic accuracy. To this end, deep learning-based computer-aided diagnosis (CAD) tools have significantly outperformed the accuracy of traditional CAD software. Motivated by these observations, we proposed a deep learning-based approach for end-to-end detection and localization of spine disorders from plain radiographs. In doing so, we took the first steps in employing state-of-the-art transformer networks to differentiate images of multiple spine disorders from healthy counterparts and localize the identified disorders, focusing on vertebral compression fractures (VCF) and spondylolisthesis due to their high prevalence and potential severity. The VCF dataset comprised 337 images, with VCFs collected from 138 subjects and 624 normal images collected from 337 subjects. The spondylolisthesis dataset comprised 413 images, with spondylolisthesis collected from 336 subjects and 782 normal images collected from 413 subjects. Transformer-based models exhibited 0.97 Area Under the Receiver Operating Characteristic Curve (AUC) in VCF detection and 0.95 AUC in spondylolisthesis detection. Further, transformers demonstrated significant performance improvements against existing end-to-end approaches by 4–14% AUC (p -values $< 10^{-13}$) for VCF detection and by 14–20% AUC (p -values $< 10^{-9}$) for spondylolisthesis detection.

Keywords Spine disorder · Plain radiography · Vertebral compression fracture · Spondylolisthesis · Deep learning

Ara Nazarian and Ashkan Vaziri have contributed equally to this work as senior authors.

İlkay Yıldız Potter
ilkay.yildiz@biosensics.com

¹ BioSensics, LLC, 57 Chapel Street, Newton, MA 02458, USA

² Beth Israel Deaconess Medical Center (BIDMC), Carl J. Shapiro Department of Orthopedic Surgery, Harvard Medical School, 330 Brookline Avenue, Stoneman 10, Boston, MA 02215, USA

³ Musculoskeletal Translational Innovation Initiative, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Boston, MA RN123, USA

⁴ Department of Radiology, Massachusetts General Brigham (MGB), Harvard Medical School, 75 Francis Street, Boston, MA 02215, USA

⁵ Department of Orthopaedics Surgery, Yerevan State University, 0025 Yerevan, Armenia

Introduction

Spine disorders affect the mechanical properties and anatomic morphology of the vertebrae, intervertebral discs, facet joints, tendons, ligaments, muscles, spinal cord, and spine nerve roots [1]. These include pain syndromes, disc degeneration, spondylosis, radiculopathy, stenosis, spondylolisthesis, fractures, and tumors [2]. While some spine disorders are common, relevant epidemiology research is in its earlier stages compared to, e.g., cancer and cardiovascular disorders [3]. For example, low back pain has been identified as the most common cause of disability in the USA for patients younger than 45 years old [4], and an estimated \$13.8 billion is spent annually in the USA due to vertebral compression fractures (VCFs) and associated comorbidities [5]. Despite their significance, spine disorders can also be asymptomatic. For instance, intervertebral disc disorders in asymptomatic individuals have a 29–96% prevalence in individuals 20 to 80 years of age [6], and only one-third of

VCFs are symptomatic [7]. This implies that the burden of spine disorders is underestimated, particularly in older ages. Undiagnosed spine disorders can cause significant functional limitations and disabilities such as severe back pain, decreased appetite, poor nutrition, progressive kyphosis, decreased pulmonary function, and increased mortality risk [8, 9]. The elderly population poses a challenge to healthcare systems and clinicians due to multiple comorbidities, such as reduced bone mass density, osteoporosis, decreased mobility, poor balance, and a greater tendency to fall [9].

Plain radiographs are the first-line imaging modality to diagnose subjects with suspected bone disorders, as they are ubiquitous and readily accessible. They provide high-resolution studies with insights into bone geometry, microstructure, and density at a low ionizing radiation dose [10–12]. Nevertheless, radiographical appearance is not always sufficient for final diagnosis due to highly variable patient parameters such as BMI and positioning, image quality, and disease severity [13]. These factors can lead to misdiagnosis or delayed diagnosis [14, 15], e.g., up to 50% for VCFs [16], and thus, increase the risk of complications and potential mortality, as well as time and cost of hospitalization. Therefore, it is of utmost importance to develop an analytical tool to aid in the identification of subjects with spine disorders and alleviate the associated health, societal, and economic burdens.

Employing an accurate automated detection model for spine disorders from radiographs can increase the workflow efficiency of clinical experts and potentially reduce human errors by relieving their workloads [17]. This can facilitate earlier detection of diseases along with improved diagnostic accuracy. Earlier and more accurate detection can further alleviate the need for follow-up imaging, reducing exposure to radiation and contrast agents [18]. To enable accurate detection, deep learning-based computer-aided diagnosis (CAD) tools have significantly outperformed traditional CAD software by particularly decreasing false positive predictions [19], which can potentially reduce radiologist review time by an estimated 17% and mitigate unnecessary burdens on patients. In particular, many studies on spine disorder analysis from radiographs have employed deep learning models and demonstrated high accuracy for potential clinical deployment [20–25]. Nevertheless, these approaches lacked explicit localization of identified spine disorders. Providing the location of identified disorders allows the clinician to visualize and overread automated detection results to confirm the result or decide on further evaluation. Thus, several studies have focused on detecting and localizing spine disorders from radiographs, albeit requiring multiple cascaded models and steps, including vertebrae segmentation, keypoint detection, and vertebral height prediction [26–32]. Developing and evaluating such cascaded approaches is less computationally efficient than end-to-end one-stage detection and localization and potentially requires

manual data cleaning between cascaded models to address error propagation.

A few recent works have proposed deep learning models for end-to-end detection and localization of spine disorders from radiographs via feature pyramid networks [33–35]. Nevertheless, most of these models were only assessed to differentiate types of spine disorders over cases that were *known to* include spine disorders. This precludes applicability in the clinical care environment, where the critical task is to diagnose spine disorders, *i.e.*, differentiate cases with spine disorders from healthy counterparts [36]. In recent years, transformer networks have also become integral to deep learning approaches in medical image analysis [37]. However, they have not yet been evaluated in radiography-based detection of spine disorders.

Motivated by these observations, we employed a deep learning-based approach for end-to-end detection and localization of spine disorders from plain radiographs. Our contributions are as follows:

1. We took the first steps in employing state-of-the-art transformer networks for end-to-end detection and localization of multiple spine disorders from plain radiographs. Our approach can differentiate images of spine disorders from healthy counterparts and localize the identified disorders without requiring multiple cascaded models.
2. We extensively tested the performance of transformer networks over two publicly available datasets comprising spine radiographs, focusing on VCF and spondylolisthesis disorders due to their high prevalence and potential severity (c.f. “Materials and Methods” section). Transformer-based models exhibited 0.97 Area Under the Receiver Operating Characteristic Curve (AUC) in VCF detection and 0.95 AUC in spondylolisthesis detection. Further, transformers demonstrated significant performance improvements against existing end-to-end approaches by 4–14% AUC (p -values $< 10^{-13}$) for VCF detection and by 14–20% AUC (p -values $< 10^{-9}$) for spondylolisthesis detection.

Materials and Methods

VCF Dataset

A VCF is a vertebral body fracture occurring in the vertebra's anterior column (anterior half of the vertebral body) [38]. VCFs are the most common complication of osteoporosis, affecting more than 700,000 Americans annually [39]. They can result in severe back pain, decreased appetite that leads to poor nutrition, and progressive kyphosis, which can lead to decreased pulmonary function [8]. These result

in severe deterioration of health status and a 15% increased risk of mortality [9].

To develop an automated VCF detection model, we employed the publicly available VinDr-SpineXR dataset, comprising plain spine radiographs [40–42]. The Institutional Review Board of Hospital 108 (H108) and Hanoi Medical University Hospital (HMUH) approved the study in Vietnam. Informed consent was waived since the study did not impact the clinical care of these hospitals, and protected health information had been removed from the data. Spine radiographs in DICOM format were retrospectively collected from the Picture Archive and Communication System (PACS) of H108 and HMUH between 2010 and 2020. Low-quality images were excluded. Each included spine image was assigned to one of the three participating radiologists, each with at least 10 years of experience. The radiologists examined each image for thirteen types of spine disorders. Images that did not exhibit these spine disorders were annotated as normal. For images with spine disorders, the location of each disorder was annotated by drawing a bounding box around the corresponding spine region. The dataset was randomly partitioned into stratified training and test sets. Images from 20% of the subjects were held out for testing, and images from the remaining 80% were allocated for training. The aforementioned data collection, inclusion/exclusion, annotation, and partitioning steps were performed by the VinDr-SpineXR dataset providers and are publicly available [41]. The spine segment, such as cervical, thoracic, or lumbar, that each image belongs to was not shared by dataset providers.

For VCF detection, we used all available images annotated with VCFs in the publicly available VinDr-SpineXR dataset and a matching number of subjects with normal images. VCF was named alternatively as “vertebral collapse” in the publicly available dataset, as a collapsed bone

in the spine is called a VCF [43]. The resulting VCF dataset comprised 337 images with VCFs collected from 138 subjects (36 test, 102 training) and 624 normal images collected from 337 subjects (69 test, 268 training). Table 1 shows the image-by-image distribution with respect to age, sex, and VCF presence, as images from the same subject may have been collected at different ages. Agreeing with the published reports on VCF incidence rates [44, 45], the number of subjects with VCFs increases by age for both sexes, with more female subjects than male. Overall, the dataset exhibited diversity over all age groups.

Spondylolisthesis Dataset

More than 3 million people are diagnosed yearly with spondylolisthesis, in which a vertebra slips out of its position relative to the adjacent vertebra [46]. Similar to VCF, spondylolisthesis progresses with aging due to wear and tear on the spine, with up to 30% prevalence among the elderly [47]. If not treated, spondylolisthesis can cause severe back pain, numbness in the extremities, sensory loss by nerve root compression, and spinal deformities [48].

We combined two publicly available datasets for spondylolisthesis to develop an automated spondylolisthesis detection model. First, we employed the VinDr-SpineXR dataset, using all available images annotated with spondylolisthesis in the publicly available dataset and a matching number of subjects with normal images. All data collection, inclusion/exclusion, annotation, and partitioning steps were performed by the VinDr-SpineXR dataset providers and are publicly available [41]. This portion of the spondylolisthesis dataset comprised 285 images with spondylolisthesis collected from 272 subjects (52 test, 220 training) and 654 normal images collected from 349 subjects (69 test, 280 training).

Table 1 Distribution of images with respect to age, sex, and VCF presence for the VCF dataset. The numbers of the subjects (*n*) in each group is presented below each subgroup

| Age | Female | | Male | | Unknown (<i>n</i> = 1) | Sum (<i>n</i> = 1) |
|---------|--------------------------|-----------------------------|-------------------------|----------------------------|----------------------------|------------------------|
| | VCF (<i>n</i> = 104) | Normal (<i>n</i> = 286) | VCF (<i>n</i> = 33) | Normal (<i>n</i> = 50) | | |
| | | | | | | |
| < 18 | 4 | 39 | 1 | 5 | 2 | 0 |
| 18–30 | 0 | 31 | 0 | 4 | 0 | 35 |
| 31–40 | 0 | 63 | 1 | 3 | 0 | 67 |
| 41–50 | 3 | 57 | 2 | 8 | 0 | 70 |
| 51–60 | 15 | 18 | 8 | 3 | 0 | 44 |
| 61–70 | 13 | 5 | 3 | 2 | 0 | 23 |
| 71–80 | 11 | 2 | 2 | 0 | 0 | 15 |
| 81–90 | 31 | 0 | 4 | 0 | 0 | 35 |
| Unknown | 177 | 314 | 60 | 68 | 0 | 2 |
| Sum | 254 | 529 | 81 | 93 | 2 | 961 |

We also incorporated the BUU-LSPINE sample dataset comprising frontal and lateral view plain radiographs of the lumbar spine [49]. BUU-LSPINE was developed by cooperation between the Burapha University (BUU), Thailand, and the Korea Institute of Oriental Medicine (KIOM). The Institutional Review Board of the BUU approved the corresponding study, and informed consent was waived. Lumbar spine radiographs in DICOM format were retrospectively collected from the PACS databases of BUU and KIOM. Exclusion criteria included images that were too low in quality, spines with embedded medical devices, not showing all lumbar vertebrae, and imaging views other than frontal and lateral. Too low quality was defined as being challenging to distinguish the vertebrae [50], which can preclude analysis for both humans and CAD tools. All included images underwent an anonymization process to eliminate protected health information. Each included vertebra was annotated by marking the coordinates of its four corners. Each pair of consecutive lumbar vertebrae was then reviewed by radiologists and orthopedics doctors of KIOM and BUU for spondylolisthesis. Each vertebrae pair was classified as positive or negative based on the presence of spondylolisthesis between the two vertebral bodies. The aforementioned data collection, inclusion/exclusion, and annotation steps were performed by the BUU-LSPINE dataset providers and are publicly available [49].

We used all available images that were annotated with spondylolisthesis in the publicly available BUU-LSPINE dataset, as well as a matching number of subjects with normal images. This portion of the spondylolisthesis dataset comprised 128 images with spondylolisthesis collected from 64 subjects and 128 normal images collected from 64 subjects. For consistency with the VCF dataset and following Klinwichit et al. [50], we mapped the vertebrae corner annotations to bounding boxes. In particular, we fit a bounding

box to tightly contain each vertebrae pair annotated as positive for spondylolisthesis presence based on the corresponding publicly available corner annotations. Note that we did not perform new annotations and only reformatted publicly available annotations to uniformize formats across datasets. We randomly partitioned the included subjects from BUU-LSPINE into stratified training and test sets. Images from 20% of the subjects were held out for testing, and images from the remaining 80% were used for training.

The training (test) sets from VinDr-SpineXR and BUU-LSPINE were merged to create the final training (test) set of the spondylolisthesis dataset. The final spondylolisthesis dataset comprised 413 images with spondylolisthesis collected from 336 subjects (65 test, 271 training) and 782 normal images collected from 413 subjects (82 test, 331 training). Table 2 shows the image-by-image distribution with respect to age, sex, and spondylolisthesis presence. Agreeing with the published reports on spondylolisthesis incidence rates, the number of subjects with spondylolisthesis increases by age for both sexes, with more female subjects than male [47]. Overall, the dataset exhibited diversity over all age groups.

Overall, the VCF dataset included all available images annotated with VCFs in the publicly-available VinDr-SpineXR dataset, along with a matching number of subjects with normal images, while the spondylolisthesis dataset combined all available images annotated with spondylolisthesis from the two publicly-available VinDr-SpineXR and BUU-LSPINE datasets along with a matching number of subjects with normal images. Thus, both VCF and spondylolisthesis datasets included embedded medical devices such as fixation apparatus, different spine segments of cervical, thoracic, and lumbar, and different imaging views such as frontal, lateral, and oblique, since images from VinDr-SpineXR participated in both datasets. Accordingly, overall

Table 2 Distribution of images with respect to age, sex, and spondylolisthesis presence for the spondylolisthesis dataset. The numbers of the subjects (n) in each group is presented below each subgroup

| Age | Female | | Male | | Unknown | | Sum | |
|------------|-------------------|------------|------------|------------|-------------------|----------|-------------|--|
| | Spondylolisthesis | | Normal | | Spondylolisthesis | | | |
| | (n = 246) | (n = 324) | (n = 87) | (n = 87) | (n = 3) | (n = 2) | | |
| <18 | 8 | 49 | 3 | 6 | 1 | 0 | 67 | |
| 18–30 | 0 | 58 | 0 | 11 | 0 | 0 | 69 | |
| 31–40 | 6 | 74 | 1 | 9 | 0 | 0 | 90 | |
| 41–50 | 19 | 61 | 5 | 12 | 0 | 0 | 97 | |
| 51–60 | 20 | 34 | 8 | 13 | 0 | 0 | 75 | |
| 61–70 | 38 | 16 | 16 | 8 | 0 | 0 | 78 | |
| 71–80 | 27 | 11 | 13 | 4 | 0 | 0 | 55 | |
| 81–90 | 10 | 1 | 7 | 2 | 0 | 0 | 20 | |
| Unknown | 172 | 343 | 57 | 68 | 2 | 2 | 644 | |
| Sum | 300 | 647 | 110 | 133 | 3 | 2 | 1195 | |

exclusion criteria were the same in both, which were low-quality images defined as challenging to distinguish the vertebrae [50].

Proposed Spine Disorder Detection Approach

We employed a deep learning-based approach for end-to-end detection and localization of spine disorders from plain radiographs following the recent influx and success of deep learning in analyzing various clinical conditions from medical imaging [37, 51–53]. In doing so, we employed state-of-the-art transformer networks to differentiate images of multiple spine disorders from normal images and localize the identified disorders. Figure 1 summarizes our approach from image preprocessing to inference. We describe each step in our approach below, followed by the instantiation of specific transformer networks we employed in the “Transformer Networks” section. Then, in the “Feature Pyramid Networks” section, we describe the end-to-end deep learning models used in recent literature and were compared with transformers in our experiments. Each model described in the “Transformer Networks” and “Feature Pyramid Networks” sections was implemented using the same training and inference steps described in this section. To do so, we used the Python programming language and MMDetection, an open-source object detection toolbox based on the PyTorch library.

We prepared each spine radiograph via contrast-limited adaptive histogram equalization (CLAHE), a typical radiography-based bone disease analysis technique to reduce noise and improve image quality [30, 35, 54–56]. Each image was then normalized to the range 0 to 1 via min–max normalization, following the standards in deep learning literature for medical imaging to aid training stability [57, 58].

Each deep learning model received a plain radiograph of the spine and made the following predictions: (i) rectangular bounding boxes circumscribing candidate spine disorder regions and (ii) a confidence score in the range of 0 to 1 associated with each detected box. The confidence score governed the likelihood of spine disorder existence

and was thresholded in post-processing stages to detect a spine disorder. In the clinical care environment, a CAD tool is expected to predict the presence of each indicated clinical condition, as the presence of one condition does not imply the lack of another. Thus, we trained and evaluated separate deep-learning models for VCF and spondylolisthesis detection [59, 60].

To accelerate training, we employed transfer learning to initialize the parameters of each deep learning model, with weights pre-trained on a benchmark object detection dataset named COCO [61]. Pre-training weights for each model were obtained from the publicly available weights provided by MMDetection. Following initialization, each deep learning model for each dataset was trained over the pairs of training images and corresponding ground-truth spine disorder bounding boxes for 100 epochs via stochastic gradient descent with a momentum factor of 0.9 and batch size of 1 [62]. The learning rate was divided by ten after 50 and 75 epochs to aid training convergence [63]. To aid performance generalization, training images were augmented by horizontal flipping and resizing, where image height was fixed at 1333 and width was varied between 512 and 800 by increments of 32. Moreover, initialized parameters of the first quarter of layers were not fine-tuned, while all trained parameters were regularized via weight decay with regularization level 10^{-4} [64]. We also explain our hyperparameter selections in detail in Supplementary Table S1.

We applied each trained model for each dataset on each image in the corresponding test set to record bounding box detections and their confidence scores. We represented each image with the detection corresponding to the maximum confidence score to be thresholded for spine disorder detection. We determined the detection threshold for each dataset and each model as the score that maximized the geometric mean of sensitivity and specificity [65]. In the clinical care environment aided by this binary prediction, experts are expected to review the positive-flagged images and decide on the presence of spine disorder. Thus, the focus of our approach was not to miss positive images; instead, it focused on one high-confidence detection for each image.

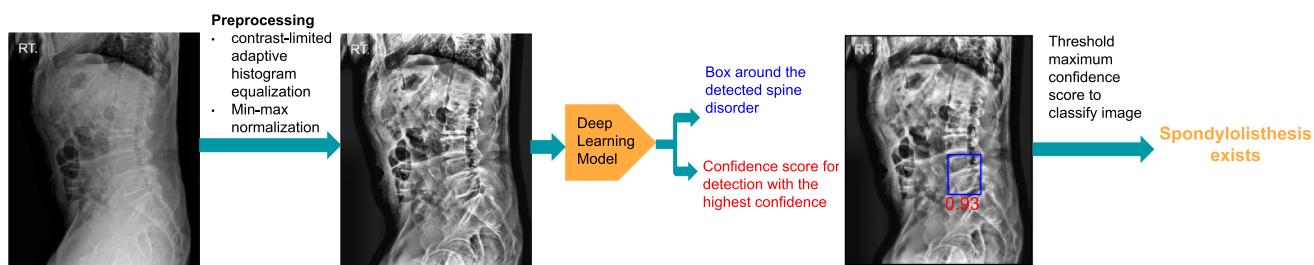


Fig. 1 Summary of the proposed deep learning-based approach for end-to-end detection and localization of various spine disorders from plain radiographs

Transformer Networks

In this study, we employed state-of-the-art transformer networks for end-to-end detection and localization of spine disorders from plain radiographs. Transformer networks involve attention mechanisms that learn weighting coefficients capturing long-range dependencies over sequences, such as pixels in images or time points in signals [66]. Particularly for object detection, transformers were used to forego hand-designed components requiring prior knowledge, such as spatial anchors or non-maximal suppression [67]. In particular, a detection transformer (DETR) [67] comprises a sequence of a convolutional network (such as ResNet-50 [68]) and a transformer encoder to extract latent image features, followed by a transformer decoder to predict object bounding boxes.

We compared four extensions of DETR that constituted the state-of-the-art in detecting common objects in recent years. Conditional DETR [69] and DAB-DETR [70] learn latent features to capture spatial positions and contents of objects separately. Deformable DETR accelerates training by applying attention to learned sampling points around a reference pixel rather than all neighboring pixels [71]. DINO replaces the convolutional feature extractor with another transformer network and improves training by learning from perturbed ground-truths and combinations of positional and content-based features [72].

In our experiments, transformer network-based models (DINO and Deformable DETR in particular) demonstrated significant performance improvements against existing end-to-end approaches to radiography-based detection of spine disorders, which are discussed below. We list the loss function used to train each transformer in the Supplementary Table S2. We further provide details on the architectures of best performing transformers DINO and Deformable DETR in Supplementary Figs S1 and S2, respectively.

Feature Pyramid Networks

Existing end-to-end approaches to radiography-based detection of spine disorders are instances of Feature Pyramid Networks (FPNs). An FPN receives a 2D image of any size and begins with extracting a hierarchy of features at multiple scales via a base neural network [73]. The base network comprises a sequence of stages, each containing convolutional and residual layers. The activation output of each stage's last residual layer is part of the feature pyramid. Base network features are complemented by upsampling to extract higher-resolution features, merged with lower-resolution base network features of the same size to form a multi-scale feature pyramid. For localization, each pixel location at each feature pyramid level is mapped back to the original input scale, and coordinates of bounding boxes circumscribing

object locations are predicted via a head network comprising fully-connected or convolutional layers.

We compared six FPNs against transformer networks based on the recent works on radiography-based detection of spine disorders via deep learning [33–35]. Faster R-CNN [74] and Cascade R-CNN [75] involve region proposal networks to predict bounding box locations relative to pre-defined anchor boxes. AC-Faster R-CNN is an extension of Faster R-CNN that combines fully connected and convolutional layers in the head network [35]. RetinaNet incorporates focal loss to combat the imbalance between background vs. object locations [76]. FCOS does not require anchor boxes and directly predicts bounding boxes and confidence scores for each pixel location on feature pyramids [77]. RepPoints is also anchor-free and extracts object features from learned boundary points useful for detection and localization [78]. All FPNs were implemented with ResNeXt-101 as their base neural network [79]. We list the loss function used to train each of these networks in Supplementary Table S2.

Performance Metrics

Spine disorder detection performance was assessed via several clinically relevant evaluation metrics, each taking values in the range 0–1. The Area Under the Receiver Operating Characteristic Curve (AUC) was computed using confidence scores before thresholding. After thresholding for binary classification of each image as positive or negative for spine disorder, sensitivity, specificity, accuracy, and positive and negative predictive values were computed.

The benchmark Intersection Over Union (IOU) metric was used to assess spine disorder localization performance, governed by the overlap percentage between a ground-truth spine disorder box and the corresponding detected box [80]. IOU was computed over the true positive images, as these were the only images with both ground-truth and detected boxes after thresholding for spine disorder detection and reported in the range from 0 to 1 for consistency with other metrics.

We reported each metric and its 95% confidence interval [81]. To assess the significance when comparing two metrics, we reported *p*-values for the two-sided Mann–Whitney nonparametric test [82].

Results

Spine Disorder Detection

Tables 3 and 4 display the detection rates for VCF and spondylolisthesis across all compared models. A transformer model outperformed others in both categories, achieving

Table 3 Comparison of DINO with the highest VCF detection performance to other models on the VCF dataset. Detection metrics include Area Under the Receiver Operating Characteristic Curve (AUC), sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). Localization is assessed by Intersection over Union (IOU). The row with the highest value for each metric is written in bold. Below each metric, the 95% confidence interval (CI) is written. For each model except DINO, *p*-values (for the two-sided Mann–Whitney nonparametric test) comparing the metrics to those of DINO are also written

| Model | | AUC | NPV | Specificity | PPV | Sensitivity | Accuracy | IOU |
|------------------|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DINO | Metric | 0.97 | 0.98 | 0.97 | 0.94 | 0.96 | 0.96 | 0.49 |
| Faster R-CNN | CI (+/-) | 0.02 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.08 |
| | Metric | 0.93 | 0.94 | 0.95 | 0.91 | 0.88 | 0.93 | 0.53 |
| RetinaNet | CI (+/-) | 0.03 | 0.02 | 0.02 | 0.02 | 0.03 | 0.02 | 0.08 |
| | <i>p</i> -value | 10^{-21} | 10^{-29} | 10^{-13} | 10^{-15} | 10^{-31} | 10^{-19} | 10^{-4} |
| Cascade R-CNN | Metric | 0.9 | 0.94 | 0.89 | 0.82 | 0.9 | 0.89 | 0.51 |
| | CI (+/-) | 0.03 | 0.02 | 0.03 | 0.03 | 0.02 | 0.03 | 0.08 |
| FCOS | <i>p</i> -value | 10^{-28} | 10^{-27} | 10^{-35} | 10^{-36} | 10^{-27} | 10^{-32} | 10^{-2} |
| | Metric | 0.9 | 0.92 | 0.94 | 0.89 | 0.86 | 0.91 | 0.5 |
| RepPoints | CI (+/-) | 0.03 | 0.02 | 0.02 | 0.03 | 0.03 | 0.02 | 0.08 |
| | <i>p</i> -value | 10^{-31} | 10^{-33} | 10^{-21} | 10^{-23} | 10^{-34} | 10^{-22} | 0.2 |
| AC-Faster R-CNN | Metric | 0.83 | 0.81 | 0.88 | 0.74 | 0.62 | 0.79 | 0.65 |
| | CI (+/-) | 0.04 | 0.03 | 0.03 | 0.04 | 0.04 | 0.03 | 0.07 |
| Deformable DETR | <i>p</i> -value | 10^{-36} | 10^{-36} | 10^{-34} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-26} |
| | Metric | 0.96 | 0.98 | 0.94 | 0.9 | 0.96 | 0.95 | 0.41 |
| Conditional DETR | CI (+/-) | 0.02 | 0.01 | 0.02 | 0.02 | 0.01 | 0.02 | 0.08 |
| | <i>p</i> -value | 10^{-28} | 10^{-34} | 10^{-32} | 10^{-35} | 10^{-34} | 10^{-33} | 0.09 |
| DAB DETR | Metric | 0.96 | 0.98 | 0.94 | 0.9 | 0.96 | 0.95 | 0.41 |
| | CI (+/-) | 0.02 | 0.01 | 0.02 | 0.02 | 0.01 | 0.02 | 0.08 |
| DINO | <i>p</i> -value | 0.06 | 0.70 | 10^{-16} | 10^{-19} | 0.3 | 10^{-4} | 10^{-4} |
| | Metric | 0.88 | 0.95 | 0.73 | 0.65 | 0.93 | 0.8 | 0.49 |
| Deformable DETR | CI (+/-) | 0.03 | 0.02 | 0.04 | 0.04 | 0.02 | 0.03 | 0.08 |
| | <i>p</i> -value | 10^{-33} | 10^{-24} | 10^{-36} | 10^{-36} | 10^{-17} | 10^{-36} | 0.9 |
| DINO | Metric | 0.78 | 0.84 | 0.73 | 0.6 | 0.75 | 0.74 | 0.44 |
| | CI (+/-) | 0.05 | 0.03 | 0.04 | 0.04 | 0.04 | 0.04 | 0.08 |
| | <i>p</i> -value | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-36} | 0.02 |

0.97 specificity and 0.96 sensitivity for identifying VCF and 0.87 specificity and 0.94 sensitivity for identifying spondylolisthesis. In doing so, transformer models significantly outperformed all FPNs. For VCF detection, DINO outperformed all FPNs by 4–14% AUC, 4–17% NPV, 2–9% specificity, 5–20% PPV, 6–34% sensitivity, and 5–17% accuracy (*p*-values $< 10^{-13}$). For spondylolisthesis detection, Deformable DETR outperformed all FPNs by 14–20% AUC, 10–13% NPV, 8–20% specificity, 11–22% PPV, 13–20% sensitivity and 10–17% accuracy (*p*-values $< 10^{-9}$). As transformer models are often computationally complex, we further evaluated the inference time of DINO, which had the lowest inference time among the tested transformers. DINO took 420 ms on average to analyze an image using a Quadro RTX 600 GPU and 12 s on average to analyze an image using an Intel Xeon Bronze CPU in a desktop computer. These results confirm the state-of-the-art performance of transformer models in detection and localization tasks based on image data, including spine disorders based on plain radiographs.

In the clinical care environment, a false negative prediction by a CAD tool can have severe consequences, including delayed diagnosis, unrecognized disorders with risk of mortality, and increased time and cost of hospitalization [14, 15, 83]. Thus, in Fig. 2, we visually analyzed false negative predictions by transformer models with the highest detection performances (DINO for VCF and Deformable DETR for spondylolisthesis). DINO misclassified only three (out of 69, 0.96 sensitivity) test images exhibiting VCF, as shown in Fig. 2a–c. All of the false negative images captured the cervical spine, which was underrepresented in the dataset due to being the least common spine region for fractures compared to the thoracic and lumbar spine [84]. Figure 2a shows a cervical spine with fixation apparatus that was annotated as positive for VCF. Post-fixation follow-up imaging has been observed to exhibit 88–90% of the calculated height of an injured vertebra [85, 86], which accordingly still exhibits height loss as per the clinical definition of VCF [87]. To the best of our knowledge, the remaining post-fixation

Table 4 Comparison of Deformable DETR with the highest spondylolisthesis detection performance to other models on the Spondylolisthesis dataset. Detection metrics include Area Under the Receiver Operating Characteristic Curve (AUC), sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). Localization is assessed by Intersection over Union (IOU). The row with the highest value for each metric is written in bold. Below each metric, the 95% confidence interval (CI) is written. For each model except Deformable DETR, *p*-values (for the two-sided Mann–Whitney nonparametric test) comparing the metrics to those of Deformable DETR are also written

| Model | AUC | NPV | Specificity | PPV | Sensitivity | Accuracy | IOU | |
|------------------|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|
| Deformable DETR | Metric | 0.95 | 0.96 | 0.87 | 0.81 | 0.94 | 0.89 | 0.44 |
| Faster R-CNN | CI (+/−) | 0.02 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.08 |
| | Metric | 0.81 | 0.86 | 0.79 | 0.7 | 0.78 | 0.79 | 0.48 |
| | CI (+/−) | 0.03 | 0.02 | 0.02 | 0.02 | 0.03 | 0.02 | 0.08 |
| RetinaNet | <i>p</i> -value | 10^{-20} | 10^{-28} | 10^{-9} | 10^{-13} | 10^{-34} | 10^{-20} | 10^{-3} |
| | Metric | 0.78 | 0.83 | 0.77 | 0.66 | 0.74 | 0.76 | 0.5 |
| | CI (+/−) | 0.03 | 0.02 | 0.03 | 0.03 | 0.02 | 0.03 | 0.08 |
| Cascade R-CNN | <i>p</i> -value | 10^{-29} | 10^{-26} | 10^{-35} | 10^{-36} | 10^{-29} | 10^{-31} | 0.5 |
| | Metric | 0.75 | 0.84 | 0.68 | 0.59 | 0.78 | 0.72 | 0.49 |
| | CI(+/−) | 0.03 | 0.02 | 0.02 | 0.03 | 0.03 | 0.02 | 0.08 |
| FCOS | <i>p</i> -value | 10^{-31} | 10^{-32} | 10^{-16} | 10^{-25} | 10^{-35} | 10^{-27} | 0.5 |
| | Metric | 0.78 | 0.84 | 0.69 | 0.6 | 0.78 | 0.72 | 0.47 |
| | CI (+/−) | 0.04 | 0.03 | 0.03 | 0.04 | 0.04 | 0.03 | 0.07 |
| RepPoints | <i>p</i> -value | 10^{-36} | 10^{-36} | 10^{-35} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-24} |
| | Metric | 0.79 | 0.85 | 0.67 | 0.6 | 0.81 | 0.72 | 0.47 |
| | CI (+/−) | 0.03 | 0.03 | 0.02 | 0.03 | 0.04 | 0.03 | 0.08 |
| AC-Faster R-CNN | <i>p</i> -value | 10^{-31} | 10^{-36} | 10^{-28} | 10^{-34} | 10^{-36} | 10^{-35} | 10^{-12} |
| | Metric | 0.81 | 0.84 | 0.76 | 0.66 | 0.76 | 0.76 | 0.5 |
| | CI (+/−) | 0.03 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.08 |
| Conditional DETR | <i>p</i> -value | 10^{-26} | 10^{-32} | 10^{-35} | 10^{-35} | 10^{-35} | 10^{-33} | 0.02 |
| | Metric | 0.8 | 0.92 | 0.63 | 0.6 | 0.91 | 0.74 | 0.46 |
| | CI (+/−) | 0.02 | 0.01 | 0.02 | 0.02 | 0.01 | 0.02 | 0.08 |
| DAB DETR | <i>p</i> -value | 10^{-4} | 0.9 | 10^{-19} | 10^{-18} | 0.6 | 10^{-6} | 10^{-11} |
| | Metric | 0.78 | 0.82 | 0.68 | 0.58 | 0.75 | 0.7 | 0.41 |
| | CI (+/−) | 0.03 | 0.02 | 0.04 | 0.04 | 0.02 | 0.03 | 0.08 |
| DINO | <i>p</i> -value | 10^{-33} | 10^{-22} | 10^{-36} | 10^{-36} | 10^{-15} | 10^{-36} | 0.5 |
| | Metric | 0.94 | 0.94 | 0.86 | 0.8 | 0.9 | 0.88 | 0.46 |
| | CI (+/−) | 0.05 | 0.03 | 0.04 | 0.04 | 0.04 | 0.04 | 0.08 |
| | <i>p</i> -value | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-36} | 10^{-5} |

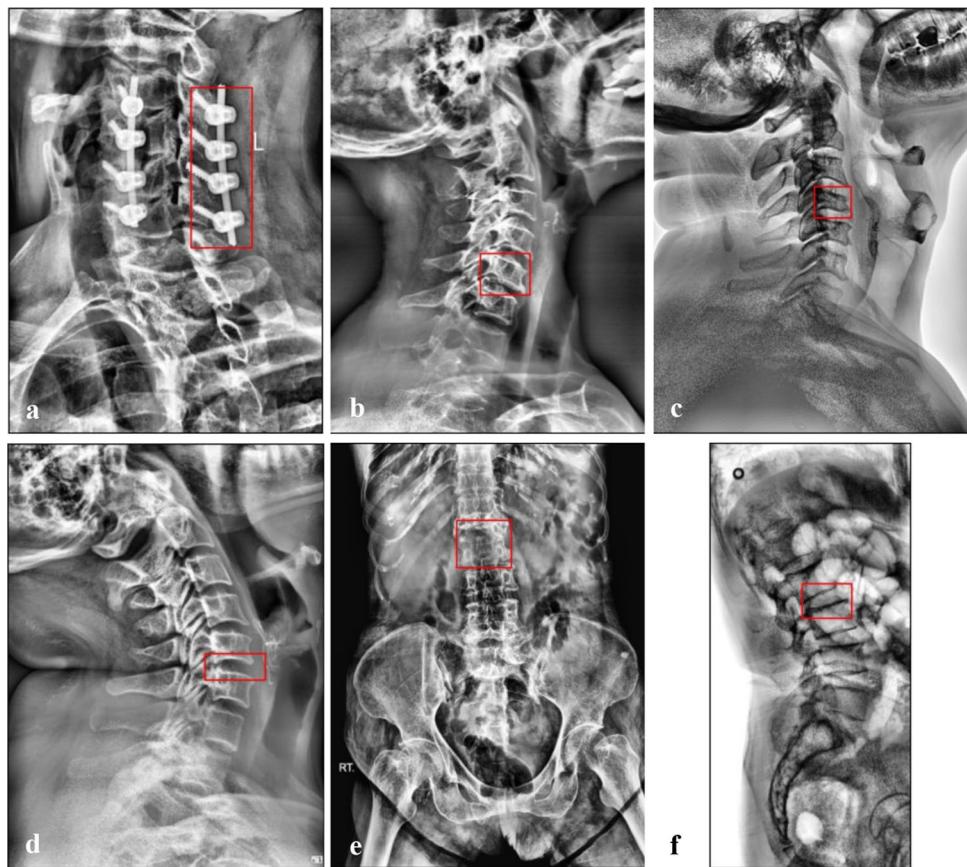
vertebral height loss was the reason of positive annotation for vertebrae with fixation apparatus in Fig. 2a.

Deformable DETR misclassified only six (out of 93, 0.94 sensitivity) test images exhibiting spondylolisthesis, with examples in Fig. 2d–f. Two of the six false negative predictions belonged to images of the cervical spine (*e.g.*, Fig. 2d), which was even more underrepresented in the spondylolisthesis dataset as the BUU-LSPINE subset contained only lumbar spine images [49]. When evaluated over only lumbar spine images from the BUU-LSPINE subset, Deformable DETR attained the same high accuracy of 0.89 as the full spondylolisthesis dataset. We further observed that the remaining four false negative images were due to image quality, such as Fig. 2e, which was more zoomed out of the spine than Figs. 3, 4, and Fig. 2f with lower brightness than Fig. 2a–e.

Spine Disorder Localization

Tables 3 and 4 also show VCF and spondylolisthesis localization performances of all compared models with respect to IOU. While FPN models tended to outperform transformer models with respect to IOU, none of the compared models attained a high IOU. The main cause of this observation is our design choice of focusing on one high-confidence detection per image, which reduced the IOU for all models, as an image can exhibit disorders in multiple locations. As our primary end goal for clinical applications is to correctly classify each image as positive or negative based on spine disorder presence, this design choice provided sufficient information for spine disorder detection with high performance, attaining 0.97 specificity and 0.96 sensitivity for VCF detection and 0.87 specificity and 0.94 sensitivity for spondylolisthesis detection.

Fig. 2 Ground truths corresponding to false negative predictions by DINO with the highest detection performance for the VCF dataset (first row) and Deformable DETR with the highest detection performance for the spondylolisthesis dataset (second row)



Figures 3 and 4 further visually compare localization examples from transformer models with the highest detection performances (DINO for VCF and Deformable DETR for spondylolisthesis) to FPN models with the highest localization performances (FCOS for VCF and AC-Faster R-CNN for spondylolisthesis). For the VCF dataset, DINO could not localize the ground-truth box correctly in only two true positive test images (Fig. 3b, c), compared to FCOS with mislocalization on one true positive test image (Fig. 3a). Moreover, in one of the two images, DINO exhibited mislocalization, FCOS could not correctly classify the image as positive for VCF presence (Fig. 3b). Both transformer and FPN models made localization mistakes on true positive test images for the spondylolisthesis dataset, as exemplified in Fig. 4a–c. In doing so, Deformable DETR had only 6% less IOU than AC-Faster R-CNN while improving spine disorder detection performance by 11% specificity and 18% sensitivity. Overall, we observed similar localization performances from transformer and FPN models, including images with challenging quality, such as Fig. 4a, with correct localization by Deformable DETR. IOU was primarily affected by localizing only one ground-truth box per image with the highest confidence (Figs. 3d, 4d) or shape differences between ground-truth and predicted boxes (Fig. 4b). Crucially, while maintaining similar localization performances to FPNs,

transformer models significantly outperformed FPN models in detecting spine disorders, as discussed above.

Discussion

We proposed a deep learning-based approach for end-to-end detection and localization of spine disorders from plain radiography images, focusing on VCF and spondylolisthesis due to their high prevalence and potential severity [8, 9, 40, 42, 43]. The VCF dataset comprised 337 images with VCFs collected from 138 subjects (104 females, 33 males, 1 with unknown sex) and 624 normal images collected from 337 subjects (286 females, 50 males, 1 with unknown sex). The spondylolisthesis dataset comprised 413 images with spondylolisthesis collected from 336 subjects (246 females, 87 males, 3 with unknown sex) and 782 normal images collected from 413 subjects (324 females, 87 males, 2 with unknown sex).

Automated detection via deep learning has successful applications in many domains involving image data, such as face detection [88], pedestrian detection [89], license plate [90], and traffic sign [91] detection in autonomous driving, as well as event and object detection in daily living [92, 93]. Transformer models have shown increasing

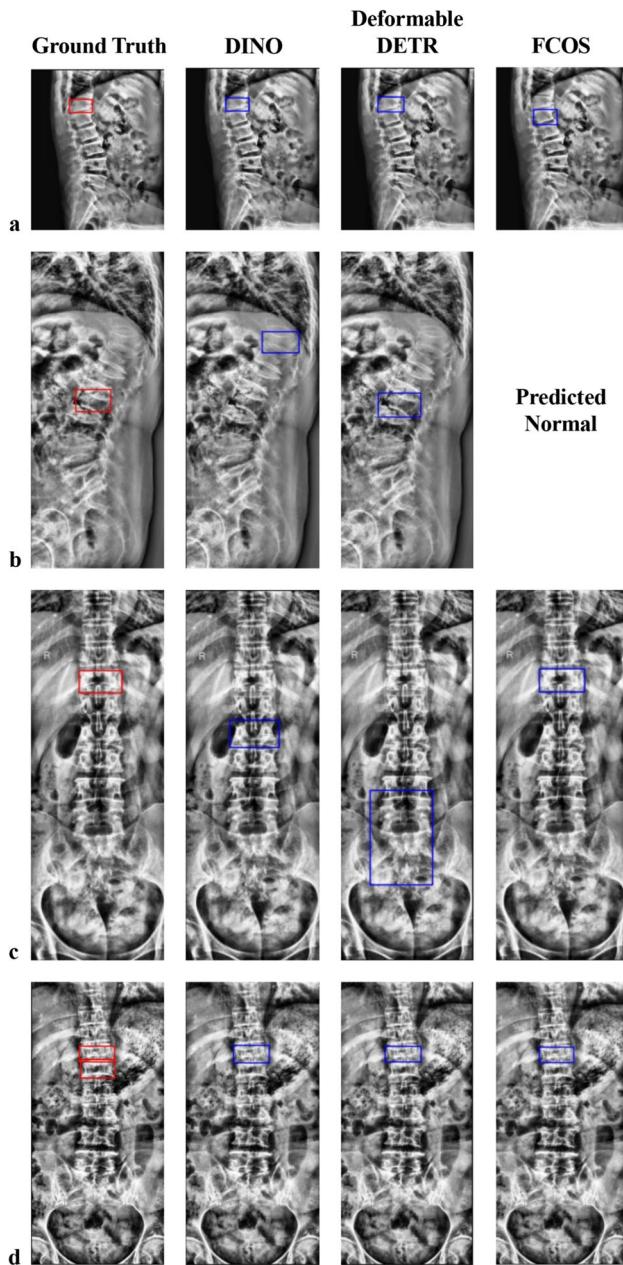


Fig. 3 VCF localization examples. **a** Correct localization by DINO and Deformable DETR. **b** Correct localization by Deformable DETR. **c** Correct localization by FCOS. **d** Correct localization by all three models

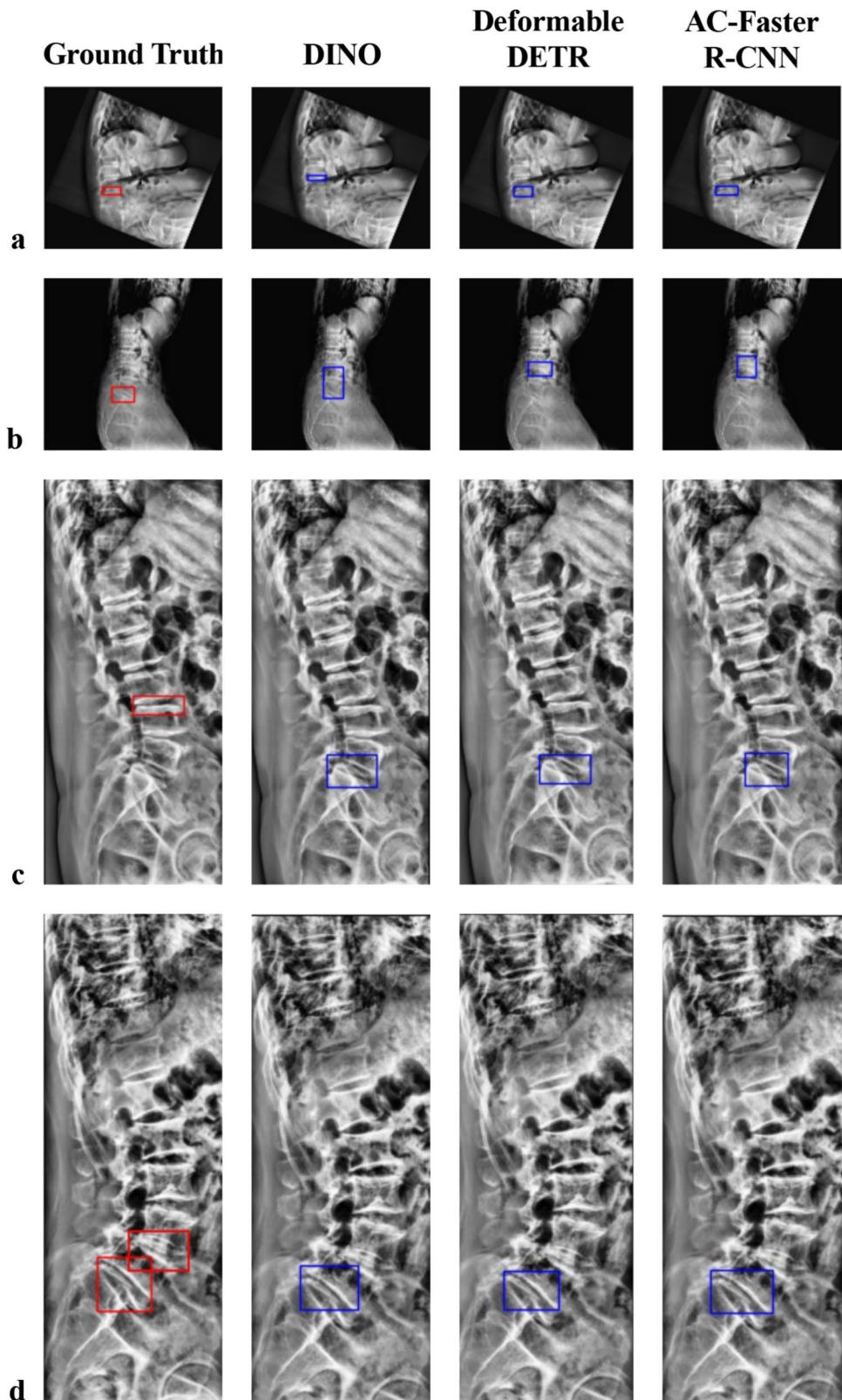
success in recent image analysis literature, including image recognition [94], segmentation [95], super-resolution [96], generation [97], and visual question answering [98]. Particularly for detecting objects of interests from images, DETR transformer [67] (c.f. “Transformer Networks” section) was proposed to remove the dependence on hand-crafted operations such as region proposals in FPN-based models (c.f. “Feature Pyramid Networks” section) and showed comparable performance to a popular instance of FPNs (Faster

R-CNN). DETR has been successfully employed in detection tasks for medical images, such as lymph node detection from MRI with 0.92 sensitivity [99] and peripheral blood leukocyte detection with 0.96 average precision [100], and further gone through technical improvements in terms of feature extraction quality and training efficiency [69–72]. Motivated by these observations, this study employed transformer models based on improvements of DETR for a detection task in medical imaging.

A significant body of research has used deep learning models to identify or classify spine disorders from radiographs, albeit lacking explicit localization of identified spine disorders [20–25]. Friawan et al. [20] compared fourteen convolutional neural network models to classify spine radiographs based on the presence of scoliosis, spondylolisthesis, or neither. Dong et al. [21] manually extracted vertebral bodies from spine radiographs and classified each vertebra based on VCF presence using GoogleNet [101]. Naguib et al. [22] compared AlexNet [102] and GoogleNet to classify cervical spine radiographs based on the presence of a dislocation, a fracture, or neither. Xu et al. [23] used ResNet-18 [68] to classify lumbar spine radiographs based on the presence of VCFs. Saravagi et al. [24] compared VGG16 [103] and InceptionV3 [104] to classify lumbar spine radiographs based on the presence of spondylolisthesis; Varçın et al. [25] tackled the same task by comparing AlexNet and GoogleNet. Nevertheless, these approaches lacked explicit localization of identified spine disorders. Instead, we focused on end-to-end detection and localization of spine disorders, allowing the clinician to visualize and overread automated detection results to confirm or decide on further evaluation.

Several studies have detected and localized spine disorders from radiographs via machine learning and deep learning, albeit requiring multiple cascaded models [26–32]. Trinh et al. [28] segmented and extracted each vertebra on lumbar spine radiographs via a customized network called LumbarNet, detected corner points of each extracted vertebra, and detected the presence of spondylolisthesis based on the corner features of each adjacent vertebrae pair. Seo et al. [30] segmented each vertebra on spine radiographs using a recurrent residual U-Net [105], extracted each sequence of three vertebral bodies, and predicted vertebral compression ratio for each vertebrae triplet by comparing ResNet50 [68] and DenseNet121 [106]. Kim et al. [31] incorporated dilated convolutions into the recurrent residual U-Net for vertebrae segmentation and predicted the vertebral compression ratio based on vertebrae heights extracted from segmentations. In addition to using cascaded models, incorporating vertebrae segmentation requires predicting each pixel in a radiograph, which is more computationally expensive than drawing boxes around localized vertebrae as in our approach (Figs. 3, 4). Varçın et al. [28] instead automatically extracted the lumbosacral region in spine radiographs and classified

Fig. 4 Spondylolisthesis localization examples. **a** Correct localization by Deformable DETR and AC-Faster R-CNN. **b** Correct localization by DINO. **c** Mislocalization by all three models. **d** Correct localization by all three models



the extracted image based on spondylolisthesis presence via MobileNet [107]. Nguyen et al. [29] identified the corner points of each vertebra in lumbar spine radiographs using

VGG [103]. Then, they used the corner points to predict the severity of lumbar spondylolisthesis based on their relative distances. Klinwichit et al. [50] localized each vertebra

from spine radiographs by comparing several object detection models, extracted the corner points of the localized vertebra, and detected the presence of spondylolisthesis based on the corner features of each adjacent vertebrae pair using a Support Vector Machine (SVM) [108]. Such cascaded approaches are less computationally efficient than end-to-end one-stage detection and localization and potentially require manual data cleaning between cascaded models to address error propagation. Instead, our approach performs end-to-end detection and localization of spine disorders via one deep-learning model based on transformer networks.

Closer to our work, a few recent studies have performed end-to-end detection and localization of spine disorders from radiographs [33–35]. Thanh et al. [34] compared Faster R-CNN [74], RetinaNet [76], and FCOS [77] to detect osteophytes, spondylolisthesis, disc space narrowing, VCF, stenosis, and surgical implants from spine radiographs; Zhong et al. [35] proposed an extension of Faster R-CNN that combines fully-connected and convolutional layers for detecting the same abnormalities. Nevertheless, these models were only assessed to differentiate types of spine disorders over cases that were *known to* include spine disorders. This precludes applicability in the clinical care environment, where the critical task is to differentiate cases with spine disorders from healthy counterparts [36]. Nguyen et al. [40] and Zhang et al. [33] compared Faster R-CNN, RetinaNet, EfficientDet [109], and Sparse R-CNN [110] to detect multiple spine disorders from radiographs, this time including normal cases.

Overall, we took the first steps in employing state-of-the-art transformer networks to differentiate images of multiple spine disorders from normal images and localize the identified disorders in an end-to-end manner. Our approach significantly outperformed six FPNs based on recent works [33–35] by 4–14% AUC (p -values $< 10^{-13}$) for VCF detection and by 14–20% AUC (p -values $< 10^{-9}$) for spondylolisthesis detection. In doing so, DINO transformer took only 12 s on average to analyze an image on a desktop computer without GPU acceleration, providing a promising direction for adoption in clinical settings without high-performance computing resources. An accurate CAD tool such as ours can be integrated with hospital cloud services to enable automatic, opportunistic screening of spine disorders by investigating each plain radiograph of the spine to (i) flag the presence of each disorder within the intended use, and (ii) draw bounding boxes to localize the identified disorders. Providing the location of identified disorders facilitates the interpretability of the proposed model based on complex transformer networks by allowing the clinician to visualize and overread automated detection results towards confirming the result or deciding on further evaluation. As radiographs are ubiquitous, this approach can aid in low-resource settings with limited access to trained experts by alleviating the burdens associated with misdiagnoses and delayed diagnoses, such as increased risk of complications and potential mortality [16].

Moreover, in high-resource settings, opportunistic screening can facilitate early identification and ease the workflow of clinicians and radiologists by saving time and resources [18].

Our study has some limitations due to the imbalance of sexes in the datasets we analyze: 104 females vs. 33 males with VCF and 246 females vs. 87 males with spondylolisthesis. While this imbalance agrees with the literature [111, 112], collecting more images from male subjects to augment our dataset will improve performance generalization. Moreover, our study aimed to advance CAD tools for analyzing spine disorders from radiographs, focusing on VCF and spondylolisthesis due to their high prevalence and potential severity. In doing so, we focused on providing a first-line of aid for clinicians by accurately flagging cases with VCFs and spondylolisthesis, to be reviewed by experts for determining subsequent treatment approaches. Conducting a Multiple Readers and Multiple Cases (MRMC) study by comparing detection accuracies of human readers with and without the aid of the proposed approach would strengthen our conclusions towards adoption in the clinical-care environment. Expanding our approach for clinical settings by including more spine disorders such as osteophytes [113] and stenosis [114] is an open research direction. Further assessing the severity of a detected disorder is more informative for treatment decisions. We plan to tackle this in future work by employing other predictive models tailored for multi-class classification and regression tasks rather than the highly accurate binary predictions we focused on.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10278-024-01175-x>.

Author Contributions All listed authors have participated actively in the entire study. I.Y.P. performed data analysis and wrote the first draft of the manuscript. All authors commented on previous versions and participated in and approved the final submission.

Funding Research reported in this publication was supported in part by BioSensics and in part by the National Institute On Aging of the National Institutes of Health under Award Number R44AG084327. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability The datasets analyzed in this study are publicly available at: www.physionet.org/content/vindr-spinexr/1.0.0/ and services.informatics.buu.ac.th/spine/.

Declarations

Ethics Approval and Informed Consent The institutional review boards of data providers at Hospital 108, Hanoi Medical University Hospital, Burapha University, and the Korea Institute of Oriental Medicine approved the datasets analyzed in this study. Written informed consent was waived.

Competing Interests I.Y.P., D.Y., E.R., and J.W. declare that they have no financial interests. A.V. and A.N. received the research grant as principal investigators. A.N. is also a consultant with BioSensics, LLC, on an unrelated project.

References

1. Elfering A, Mannion AF. Epidemiology and risk factors of spinal disorders. In: Boos N, Aebi M, editors. Spinal disorders: fundamentals of diagnosis and treatment. Berlin (DE): Springer; 2008.
2. Alshami, A.M., 2015. Prevalence of spinal disorders and their relationships with age and gender. *Saudi medical journal*, 36(6), p.725.
3. Manek NJ, MacGregor AJ. Epidemiology of back disorders: prevalence, risk factors, and prognosis. *Curr Opin Rheumatol*. 2005;17:134–140.
4. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet*. 1999;354:581–585.
5. Wong CC, McGirt MJ. Vertebral compression fractures: a review of current management and multimodal therapy. *J Multidiscip Healthc*. 2013;6:205–14. Epub 2013/07/03. <https://doi.org/10.2147/jmdh.S31659>. PubMed PMID: 23818797; PMCID: PMC3693826.
6. Brinjikji, W., Luetmer, P.H., Comstock, B., Bresnahan, B.W., Chen, L.E., Deyo, R.A., Halabi, S., Turner, J.A., Avins, A.L., James, K. and Wald, J.T., 2015. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. *American journal of neuroradiology*, 36(4), pp.811-816.
7. McCarthy J, Davis A. Diagnosis and Management of Vertebral Compression Fractures. *Am Fam Physician*. 2016;94(1):44–50. Epub 2016/07/09. PubMed PMID: 27386723.
8. Alexandru, D. and So, W., 2012. Evaluation and management of vertebral compression fractures. *The Permanente Journal*, 16(4), p.46.
9. Fehlings, M.G., Tetreault, L., Nater, A., Choma, T., Harrop, J., Mroz, T., Santaguida, C. and Smith, J.S., 2015. The aging of the global population: the changing epidemiology of disease and spinal disorders. *Neurosurgery*, 77, pp.S1-S5.
10. Priolo, F., Cerase, A.: The current role of radiography in the assessment of skeletal tumors and tumor-like lesions. *European Journal of Radiology* 27, S77–S85 (1998).
11. Tang, C., Aggarwal, R.: Imaging for musculoskeletal problems. *InnovAiT* 6(11), 735–738 (2013).
12. Santiago, F.R., Ramos-Bossini, A.J.L., Wang, Y.X.J. and Zúñiga, D.L., 2020. The role of radiography in the study of spinal disorders. *Quantitative imaging in medicine and surgery*, 10(12), p.2322.
13. Lenchik L, Rogers LF, Delmas PD, Genant HK. Diagnosis of osteoporotic vertebral fractures: importance of recognition and description by radiologists. *AJR Am J Roentgenol*. 2004;183(4):949–58. Epub 2004/09/24. <https://doi.org/10.2214/ajr.183.4.1830949>. PubMed PMID: 15385286.
14. Pinto, A., Beritto, D., Russo, A., Ricciutello, F., Caruso, M., Belfiore, M.P., Papapietro, V.R., Carotti, M., Pinto, F., Giovagnoni, A., et al.: Traumatic fractures in adults: Missed diagnosis on plain radiographs in the emergency department. *Acta Bio Medica: Atenei Parmensis* 89(1), 111 (2018).
15. Bruno, M.A., Walker, E.A. and AbuJudeh, H.H., 2015. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *Radiographics*, 35(6), pp.1668–1676.
16. Gehlbach SH, Bigelow C, Heimisdottir M, May S, Walker M, et al. Recognition of vertebral fracture in a clinical setting. *Osteoporos Int*. 2000;11(7):577–82.
17. Trockel, M.T., Menon, N.K., Rowe, S.G., Stewart, M.T., Smith, R., Lu, M., Kim, P.K., Quinn, M.A., Lawrence, E., Marchalik, D. and Farley, H., 2020. Assessment of physician sleep and wellness, burnout, and clinically significant medical errors. *JAMA network open*, 3(12), pp.e2028111-e2028111.
18. Van Leeuwen, K.G., de Rooij, M., Schalekamp, S., van Ginneken, B. and Rutten, M.J., 2022. How does artificial intelligence in radiology improve efficiency and health outcomes?. *Pediatric Radiology*, pp.1–7.
19. Mayo, R.C.; Kent, D.; Sen, L.C.; Kapoor, M.; Leung, J.W.T.; Watanabe, A.T. Reduction of False-Positive Markings on Mammograms: A Retrospective Comparison Study Using an Artificial Intelligence-Based CAD. *J. Digit. Imaging* 2019, 32, 618–624.
20. Fraiwan, M., Audat, Z., Fraiwan, L. and Manasreh, T., 2022. Using deep transfer learning to detect scoliosis and spondylolisthesis from X-ray images. *Plos one*, 17(5), p.e0267851.
21. Dong, Q., Luo, G., Lane, N.E., Lui, L.Y., Marshall, L.M., Kado, D.M., Cawthon, P., Perry, J., Johnston, S.K., Haynor, D. and Jarvik, J.G., 2022. Deep learning classification of spinal osteoporotic compression fractures on radiographs using an adaptation of the genant semiquantitative criteria. *Academic radiology*, 29(12), pp.1819–1832.
22. Naguib, S.M., Hamza, H.M., Hosny, K.M., Saleh, M.K. and Kassem, M.A., 2023. Classification of cervical spine fracture and dislocation using refined pre-trained deep model and saliency map. *Diagnostics*, 13(7), p.1273.
23. Xu, F., Xiong, Y., Ye, G., Liang, Y., Guo, W., Deng, Q., Liang, Z. and Zeng, X., 2023. Deep learning-based artificial intelligence model for classification of vertebral compression fractures: A multicenter diagnostic study. *Frontiers in Endocrinology*, 14, p.1025749.
24. Saravagi, D., Agrawal, S., Saravagi, M., Chatterjee, J.M. and Agarwal, M., 2022. Diagnosis of lumbar spondylolisthesis using optimized pretrained CNN models. *Computational Intelligence and Neuroscience*, 2022.
25. Varçın, F., Erbay, H., Çetin, E., Çetin, İ. and Kültür, T., 2019, September. Diagnosis of lumbar spondylolisthesis via convolutional neural networks. In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1–4). IEEE.
26. Varçın, F., Erbay, H., Çetin, E., Çetin, İ. and Kültür, T., 2021. End-to-end computerized diagnosis of spondylolisthesis using only lumbar X-rays. *Journal of Digital Imaging*, 34, pp.85–95.
27. Kim, K.C., Cho, H.C., Jang, T.J., Choi, J.M. and Seo, J.K., 2021. Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation. *Computer Methods and Programs in Biomedicine*, 200, p.105833.
28. Trinh, G.M., Shao, H.C., Hsieh, K.L.C., Lee, C.Y., Liu, H.W., Lai, C.W., Chou, S.Y., Tsai, P.I., Chen, K.J., Chang, F.C. and Wu, M.H., 2022. Detection of lumbar spondylolisthesis from X-ray images using deep learning network. *Journal of Clinical Medicine*, 11(18), p.5450.
29. Nguyen, T.P., Chae, D.S., Park, S.J., Kang, K.Y. and Yoon, J., 2021. Deep learning system for Meyerding classification and segmental motion measurement in diagnosis of lumbar spondylolisthesis. *Biomedical Signal Processing and Control*, 65, p.102371.
30. Seo, J.W., Lim, S.H., Jeong, J.G., Kim, Y.J., Kim, K.G. and Jeon, J.Y., 2021. A deep learning algorithm for automated measurement of vertebral body compression from X-ray images. *Scientific Reports*, 11(1), p.13732.
31. Kim, D.H., Jeong, J.G., Kim, Y.J., Kim, K.G. and Jeon, J.Y., 2021. Automated vertebral segmentation and measurement of vertebral compression ratio based on deep learning in X-ray images. *Journal of digital imaging*, 34, pp.853–861.
32. Ribeiro, E.A., Nogueira-Barbosa, M.H., Rangayyan, R.M. and Azevedo-Marques, P.M.D., 2012, October. Detection of vertebral compression fractures in lateral lumbar X-ray images. In *XXIII Congresso Brasileiro em Engenharia Biomédica (CBEB)* (pp. 1–4).
33. Zhang, J., Lin, H., Wang, H., Xue, M., Fang, Y., Liu, S., Huo, T., Zhou, H., Yang, J., Xie, Y. and Xie, M., 2023. Deep learning system assisted detection and localization of lumbar spondylolisthesis. *Frontiers in Bioengineering and Biotechnology*, 11.
34. Thanh, B.P.N. and Nguyen, P., 2023, October. Comparative study of object detection models for abnormality detection on spinal

- X-ray images. In 2023 International Conference on Multimedia Analysis and Pattern Recognition (MAPR) (pp. 1–5). IEEE.
35. Zhong, B., Yi, J. and Jin, Z., 2023. AC-Faster R-CNN: an improved detection architecture with high precision and sensitivity for abnormality in spine x-ray images. *Physics in Medicine & Biology*, 68(19), p.195021.
 36. Kim, G.U., Chang, M.C., Kim, T.U. and Lee, G.W., 2020. Diagnostic modality in spine disease: a review. *Asian spine journal*, 14(6), p.910.
 37. Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S. and Fu, H., 2023. Transformers in medical imaging: A survey. *Medical Image Analysis*, p.102802.
 38. Donnally IC, DiPompeo CM, Varacallo M. Vertebral Compression Fractures. StatPearls. Treasure Island (FL): StatPearls Publishing. Copyright © 2021, StatPearls Publishing LLC.; 2021.
 39. Riggs BL, Melton LJ. The worldwide problem of osteoporosis: insights afforded by epidemiology. *Bone*. 1995;17(5 suppl):S05S-S11S.
 40. Nguyen, H.T., Pham, H.H., Nguyen, N.T., Nguyen, H.Q., Huynh, T.Q., Dao, M. and Vu, V., 2021. VinDr-SpineXR: A deep learning framework for spinal lesions detection and classification from radiographs. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24 (pp. 291–301). Springer International Publishing.
 41. Pham, H. H., Nguyen Trung, H., & Nguyen, H. Q. (2021). VinDr-SpineXR: A large annotated medical image dataset for spinal lesions detection and classification from radiographs (version 1.0.0). PhysioNet. <https://doi.org/10.13026/q45h-5h59>.
 42. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. 101 (23), pp. e215–e220.
 43. University of Maryland Medical System. (2003). A Patient's Guide to Lumbar Compression Fracture. Retrieved March 29, 2024, from <https://www.umms.org/ummc/health-services/orthopedics/services/spine/patient-guides/lumbar-compression-fractures>
 44. Thibault *et al.*, “Volume of lytic vertebral body metastatic disease quantified using computed tomography-based image segmentation predicts fracture risk after spine stereotactic body radiation therapy,” *International Journal of Radiation Oncology & Biology & Physics*, vol. 97, no. 1, pp. 75–81, 2017.
 45. B. Garg, V. Dixit, S. Batra, R. Malhotra, and A. Sharan, “Non-surgical management of acute osteoporotic vertebral compression fracture: a review,” *Journal of clinical orthopaedics and trauma*, vol. 8, no. 2, pp. 131–138, 2017.
 46. Baker, S. (2019) spondylolisthesis: Another Good Reason To Take Better Care Of Your Back. <https://www.keranews.org/health-science-tech/2019-08-05/spondylolisthesis-another-good-reason-to-take-better-care-of-your-back> (Accessed: February 23, 2024).
 47. Karsy, M., Chan, A.K., Mummaneni, P.V., Virk, M.S., Bydon, M., Glassman, S.D., Foley, K.T., Potts, E.A., Shaffrey, C.I., Shaffrey, M.E. and Coric, D., 2020. Outcomes and complications with age in spondylolisthesis: an evaluation of the elderly from the Quality Outcomes Database. *Spine*, 45(14), pp.1000–1008.
 48. spondylolisthesis symptoms & treatment. Mount Sinai Health System. (n.d.). <https://www.mountsinai.org/locations/spine-hospital/conditions/spondylolisthesis>
 49. BUU Datasets. (2023). BUU Spine Dataset. Retrieved March 29, 2024, from <https://services.informatics.buu.ac.th/spine/>
 50. Klinwichit, P., Yookwan, W., Limchareon, S., Chinnasarn, K., Jang, J.S. and Onuean, A., 2023. BUU-LSPINE: A Thai open lumbar spine dataset for spondylolisthesis detection. *Applied Sciences*, 13(15), p.8646.
 51. Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H.J. and Kim, N., 2019. Deep learning in medical imaging. *Neurospine*, 16(4), p.657.
 52. Yang, R. and Yu, Y., 2021. Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. *Frontiers in oncology*, 11, p.638182.
 53. Liu, X., Gao, K., Liu, B., Pan, C., Liang, K., Yan, L., Ma, J., He, F., Zhang, S., Pan, S. and Yu, Y., 2021. Advances in deep learning-based medical image analysis. *Health Data Science*, 2021.
 54. Karanam, S. R., Srinivas, Y., & Chakravarty, S. (2022). A systematic approach to diagnosis and categorization of bone fractures in X-Ray imagery. *International Journal of Healthcare Management*, 1–12.
 55. Lu, S., Wang, S., & Wang, G. (2022). Automated universal fractures detection in X-ray images based on deep learning approach. *Multimedia Tools and Applications*, 1–17.
 56. Mall, P. K., Singh, P. K., & Yadav, D. (2019, December). GLCM based feature extraction and medical x-ray image classification using machine learning techniques. In *2019 IEEE Conference on Information and Communication Technology* (pp. 1–6).
 57. Kibriya, H., Amin, R., Alshehri, A.H., Masood, M., Alshamrani, S.S. and Alshehri, A., 2022. A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers. *Computational Intelligence and Neuroscience*, 2022.
 58. Reddy, G.T., Bhattacharya, S., Ramakrishnan, S.S., Chowdhary, C.L., Hakak, S., Kaluri, R. and Reddy, M.P.K., 2020, February. An ensemble based machine learning model for diabetic retinopathy classification. In 2020 international conference on emerging trends in information technology and engineering (ic-ETITE) (pp. 1–6). IEEE.
 59. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T. and Riegler, M., 2017, June. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 164–169).
 60. Joshi, R.C., Singh, D., Tiwari, V. and Dutta, M.K., 2022. An efficient deep neural network based abnormality detection and multi-class breast tumor classification. *Multimedia Tools and Applications*, 81(10), pp.13691–13711.
 61. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13 (pp. 740–755). Springer International Publishing.
 62. Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint [arXiv:1609.04747](https://arxiv.org/abs/1609.04747).
 63. You, K., Long, M., Wang, J. and Jordan, M.I., 2019. How does learning rate decay help modern neural networks? arXiv preprint [arXiv:1908.01878](https://arxiv.org/abs/1908.01878).
 64. Krogh A, Hertz J. A simple weight decay can improve generalization. Advances in neural information processing systems. Proceedings of the 4th International Conference on Neural Information Processing Systems, 950–957 (1991).
 65. Fawcett T. An introduction to ROC analysis. *Pattern recognition letters* 2006;27(8):861–874.
 66. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
 67. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S., 2020, August. End-to-end object detection with transformers. In European conference on computer vision(pp. 213–229). Cham: Springer International Publishing.
 68. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE*

- conference on computer vision and pattern recognition (pp. 770–778).
69. Depu Meng, Xiaokang Chen, Zejia Fan, Gang Zeng, Houqiang Li, Yuhui Yuan, Lei Sun, and Jingdong Wang. Conditional detr for fast training convergence. arXiv preprint [arXiv:2108.06152](https://arxiv.org/abs/2108.06152), 2021.
 70. Shilong Liu, Feng Li, Hao Zhang, Xiao Yang, Xianbiao Qi, Hang Su, Jun Zhu, and Lei Zhang. DAB-DETR: Dynamic anchor boxes are better queries for DETR. arXiv preprint [arXiv:2201.12329](https://arxiv.org/abs/2201.12329), 2022.
 71. Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. De-formable detr: Deformable transformers for end-to-end object detection. In ICLR 2021: The Ninth International Conference on Learning Representations, 2021.
 72. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L.M. and Shum, H.Y., 2022. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. arXiv preprint [arXiv:2203.03605](https://arxiv.org/abs/2203.03605).
 73. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117–2125).
 74. Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.
 75. Cai, Z. and Vasconcelos, N., 2019. Cascade R-CNN: High quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence, 43(5), pp.1483–1498.
 76. Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P., 2017. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980–2988).
 77. Tian, Z., Shen, C., Chen, H. and He, T., 2019. FCOS: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9627–9636).
 78. Yang, Z., Liu, S., Hu, H., Wang, L. and Lin, S., 2019. Reppoints: Point set representation for object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9657–9666).
 79. Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K., 2017. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492–1500).
 80. Redmon, J., and Farhadi, A., “YOLO9000: better, faster, stronger.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271. 2017.
 81. Hanley, J.A. and McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), pp.29–36.
 82. McKnight, P.E. and Najab, J., 2010. Mann-Whitney U Test. The Corsini encyclopedia of psychology, pp.1–1.
 83. Taylor, J.A., Clopton, P., Bosch, E., Miller, K.A. and Marcelis, S., 1995. Interpretation of Abnormal Lumbosacra Spine Radiographs: A Test Comparing Students, Clinicians, Radiology Residents, and Radiologists in Medicine and Chiropractic. Spine, 20(10), pp.1147–1153.
 84. Leucht, P., Fischer, K., Muhr, G. and Mueller, E.J., 2009. Epidemiology of traumatic spine fractures. Injury, 40(2), pp.166–172.
 85. Li, Q., Liu, Y., Chu, Z., Chen, J. and Chen, M., 2011. Treatment of thoracolumbar fractures with transpedicular intervertebral bone graft and pedicle screws fixation in injured vertebrae. Chinese Journal of Reparative and Reconstructive Surgery, 25(8), pp.956–959.
 86. Olerud, S. and Sjöström, L., 1988. Transpedicular fixation of thoracolumbar vertebral fractures. Clinical Orthopaedics and Related Research®, 227, pp.44–51.
 87. Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 1993;8(9):1137–1148.
 88. Taigman Y, Yang M, Ranzato MA, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 1701–1708.
 89. A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey,” *Neurocomputing*, vol. 300, pp. 17–33, Jul. 2018.
 90. P. Shivakumara, D. Tang, M. Asadzadehkaljahi, T. Lu, U. Pal, and M. H. Anisi, “CNN-RNN based method for license plate recognition,” *CAAI Trans. Intell. Technol.*, vol. 3, no. 3, pp. 169–175, Sep. 2018.
 91. J. Li and Z. Wang, “Real-time traffic sign recognition based on efficient CNNs in the wild,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 975–984, Mar. 2019.
 92. W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, “Deep joint rain detection and removal from a single image,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 1357–1366.
 93. V. Allken, N. O. Handegard, S. Rosen, T. Schreyeck, T. Mahiout, and K. Malde, “Fish species identification using a convolutional neural network trained on synthetic data,” *ICES J. Mar. Sci.*, vol. 76, no. 1, pp. 342–349, Jan. 2019.
 94. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929), 2020.
 95. L. Ye, M. Rochan, Z. Liu, and Y. Wang, “Cross-modal selfattention network for referring image segmentation,” in CVPR, 2019.
 96. F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, “Learning texture transformer network for image super-resolution,” in CVPR, 2020.
 97. H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, “Pre-trained image processing transformer,” arXiv preprint [arXiv:2012.00364](https://arxiv.org/abs/2012.00364), 2020.
 98. H. Tan and M. Bansal, “LXMERT: Learning cross-modality encoder representations from transformers,” in EMNLP-IJCNLP, 2019.
 99. Mathai, T.S., Lee, S., Elton, D.C., Shen, T.C., Peng, Y., Lu, Z. and Summers, R.M., 2022, April. Lymph node detection in T2 MRI with transformers. In Medical Imaging 2022: Computer-Aided Diagnosis (Vol. 12033, pp. 869–873). SPIE.
 100. Leng, B., Wang, C., Leng, M., Ge, M. and Dong, W., 2023. Deep learning detection network for peripheral blood leukocytes based on improved detection transformer. *Biomedical Signal Processing and Control*, 82, p.104518.
 101. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1–9).
 102. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
 103. Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
 104. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818–2826).
 105. Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention—MICCAI

- 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18 (pp. 234–241). Springer International Publishing.
106. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700–4708).
 107. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861).
 108. Bishop, C.M., 2006. Pattern recognition and machine learning. Springer google schola, 2, pp.5-43.
 109. Tan, M., Pang, R. and Le, Q.V., 2020. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781–10790).
 110. Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., Tomizuka, M., Li, L., Yuan, Z., Wang, C. and Luo, P., 2021. Sparse r-cnn: End-to-end object detection with learnable proposals. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14454–14463).
 111. Felsenberg D, Silman AJ, Lunt M, Armbrecht G, Ismail AA, Finn JD, Cockerill WC, Banzer D, Benevolenskaya LI, Bhalla A, Bruges Armas J, Cannata JB, Cooper C, Dequeker J, Eastell R, Felsch B, Gowin W, Havelka S, Hoszowski K, Jajic I, Janott J, Johnell O, Kanis JA, Kragl G, Lopes Vaz A, Lorenc R, Lyritis G, Masaryk P, Matthis C, Miazgowski T, Parisi G, Pols HA, Poor G, Raspe HH, Reid DM, Reisinger W, Schedit-Nave C, Stepan JJ, Todd CJ, Weber K, Woolf AD, Yershova OB, Reeve J, Ó'Neill TW. Incidence of vertebral fracture in Europe: results from the European Prospective Osteoporosis Study (EPOS). *J Bone Miner Res.* 2002;17(4):716–24. Epub 2002/03/29. <https://doi.org/10.1359/jbmr.2002.17.4.716>. PubMed PMID: 11918229.
 112. Wang, Y.X.J., Káplár, Z., Deng, M. and Leung, J.C., 2017. Lumbar degenerative spondylolisthesis epidemiology: a systematic review with a focus on gender-specific and age-specific prevalence. *Journal of orthopaedic translation*, 11, pp.39–52.
 113. Brown DE, Neumann RD (2004) Orthopedic secrets. In: Osteoarthritis, Chap I, 3rd edn. Elsevier, Philadelphia, pp 1–3
 114. Katz, J.N., Zimmerman, Z.E., Mass, H. and Makhni, M.C., 2022. Diagnosis and management of lumbar spinal stenosis: a review. *Jama*, 327(17), pp.1688–1699.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.