

# Automated detection of sacroiliitis on plain radiograph using EfficientDet algorithm in young patients with back pain: a pilot study

**Sung Hyun Noh**

Ajou University Hospital

**Chansik An** (✉ [chansikan@gmail.com](mailto:chansikan@gmail.com))

National Health Insurance Service Ilsan Hospital

**Dain Kim**

Yonsei University

**Seung Hyun Lee**

National Health Insurance Service Ilsan Hospital

**Min-Yung Chang**

National Health Insurance Service Ilsan Hospital

**Hyun Cheol Oh**

National Health Insurance Service Ilsan Hospital

---

## Research Article

**Keywords:** Artificial Intelligence, Deep Learning, Ankylosing Spondylitis, X-Ray

**Posted Date:** March 22nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-272234/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

A computer algorithm that automatically detects sacroiliac joint abnormalities on plain radiograph would help radiologists avoid missing sacroiliitis. This study aimed to develop and validate a deep learning model to detect and diagnose sacroiliitis on plain radiograph in young patients with low back pain.

## Methods

This Institutional Review Board-approved retrospective study included 478 and 468 plain radiographs from 241 and 433 young (< 40 years) patients who complained of low back pain with and without ankylosing spondylitis, respectively. They were randomly split into training and test datasets with a ratio of 8:2. Radiologists reviewed the images and labeled the coordinates of a bounding box and determined the presence or absence of sacroiliitis for each sacroiliac joint. We fine-tuned and optimized the EfficientDet-D4 object detection model pre-trained on the COCO 2107 dataset on the training dataset and validated the final model on the test dataset.

## Results

The mean average precision, an evaluation metric for object detection accuracy, was 0.918 at 0.5 intersection over union. In the diagnosis of sacroiliitis, the area under the curve, sensitivity, specificity, accuracy, and F1-score were 0.932 (95% confidence interval, 0.903–0.961), 96.9% (92.9–99.0), 86.8% (81.5–90.9), 91.1% (87.7–93.7), and 90.2% (85.0–93.9), respectively.

## Conclusions

The EfficientDet, a deep learning-based object detection algorithm, could be used to automatically diagnose sacroiliitis on plain radiograph.

## Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease that mainly involves the axial skeleton and tendon attachment, characterized by the formation of pathological new bones in the cortical region and the loss of the trabecular bone [1]. Diagnosis of AS is typically made when patients have  $\geq 3$  months of back pain, age of onset < 45 years, sacroiliitis on imaging with  $\geq 1$  other radiologic finding, and Human Leukocyte Antigen-B27 positivity with  $\geq 2$  clinical features [2]. Traditionally, and given the high frequency of joint affectation (>90%), the X-ray evidence of sacroiliitis has been an essential criterion for AS since the first diagnostic criteria were established in 1961 [3].

AS typically begins between 20 to 30 years of age with the progressive deformation and stiffness of the spine and sacroiliac joint (SIJ) [4]. If left unmanaged, AS may lead to irreversible structural damage and reduced spinal mobility, leading to an economic burden and productivity loss, since AS afflicts economically active young people [5–7]. Therefore, early diagnosis of AS followed by proper management is crucial.

However, AS diagnosis is often delayed considerably, with an average delay of 5–14 years between the symptom onset and the diagnosis [6]. An important factor contributing to the delayed diagnosis is the difficulty in distinguishing inflammatory back pain due to AS from other highly prevalent types of low back pain in the general population.

Therefore, it is imperative not to overlook sacroiliitis on plain X-ray radiograph, a diagnostic finding of AS, in patients with low back pain even when they are not explicitly suspected of having AS. Unfortunately, however, it is not easy for human radiologists to maintain full attention for an extended time to SIJ abnormalities on every radiograph, especially after they read a lot of CT or MRI images and have encountered hundreds of normal plain radiographs in young people with back pain in daily practice [8].

Therefore, a computer algorithm that automatically detects SIJ abnormalities on plain radiograph would help radiologists avoid missing sacroiliitis. This study aimed to develop and validate a deep learning model to detect and diagnose sacroiliitis on plain radiograph automatically.

## Methods

### Subjects

Our Institutional Review Board approved this retrospective study and waived the requirement to obtain informed consent. By searching our electronic medical records, we identified 232 adult patients <40 years old who were diagnosed with ankylosis spondylitis at a tertiary hospital from February 2002 to March 2020 and underwent lower spine or abdominopelvic X-ray imaging. As a control group, we selected by matching age and gender patients who complained of lower back pain but were not diagnosed with ankylosis spondylitis, approximately 1.5 times the number of AS patients. Thus, the final study cohort had similar numbers of radiographs in the normal and abnormal groups. A board-certified abdominal radiologist (C.A. with 10 years of experience in reading plain X-ray) reviewed the radiographs and excluded 502 images because of the low image quality (n = 32) or the lack of the radiologic findings of sacroiliitis in patients with ankylosing spondylitis (n = 470). The final study subjects were randomly split into training and test sets with a ratio of 8:2, while preserving the proportion of positive cases and preventing the same patient's images from being assigned to different sets (Fig. 1).

### Data

## Image preprocessing

The X-ray images were downloaded in the Digital Imaging and Communications in Medicine format following de-identification and converted to the Joint Photographic Experts Group format. The converted images were first resized along the shorter dimension to 1,024 pixels and then resized along the longer dimension to 1,024 pixels such that the original aspect ratio was maintained.

## Image annotation

Our goal was to automatically detect both SIJs on a plain radiograph and classify them as either normal or abnormal (i.e., the absence or presence of sacroiliitis). Therefore, two types of ground truth were required: the coordinates of a bounding box surrounding a SIJ and the presence or absence of sacroiliitis findings. First, the bounding boxes were drawn manually by the radiologist (C.A.) and a machine learning researcher (D.K.) in consensus using ImageJ software [9]. Second, regarding the presence or absence of sacroiliitis findings in each SIJ, two board-certified musculoskeletal radiologists (M.Y.C. and S.H.L, with 7 and 8 years of experience, respectively) reviewed the X-ray images and graded in consensus the severity according to the New York criteria: grade 0, normal; grade 1, suspicious (some blurring of the joint margins); grade 2, minimal sclerosis with some erosion; grade 3, definite sclerosis or severe erosion with joint space widening; grade 4, complete ankylosis [10]. They were unaware of the patients' diagnosis and demographic information during the image review session. For the ground-truth, grades 0/1 and grades 2/3/4 were considered the absence and presence of sacroiliitis, respectively.

## Deep Learning

### Training

We used Python 3 with TensorFlow 2 Object Detection application programming interface (API) [11]. Among available models, we used EfficientDet-D4 1024 x 1024, pre-trained on COCO 2017 dataset. EfficientDet is currently a state-of-the-art architecture for object detection. It was built upon its predecessor, EfficientNet, and incorporates a novel bi-directional feature network and new scaling rules [12, 13].

We fine-tuned the pre-trained EfficientDet-D4 model on our training dataset with a batch size of 4. To avoid overfitting, we performed random data augmentation for training: horizontal flip with a probability of 0.5, brightness adjustment with a scale range of 0.9–1.1, contrast adjustment with a scale range of 0.9–1.1, and rescaling with a scale range of 0.8–1.2. The scaled image was then either cropped or padded to maintain its original size. A loss function was the weighted sigmoid focal function for classification loss and the smooth L1 function for localization loss. We used the Adam optimizer with learning rate warm-up, a heuristic method that increases the learning rate linearly over the warm-up period [14]. We optimized the learning rate parameters through 5-fold cross-validation and random grid search.

The warm-up learning rate, learning rate base, warm-up steps used for our final model were  $1.0000001e-05$ ,  $0.00019999998$ , and  $2,500$ , respectively, which means that the initial learning rate of  $1.0000001e-05$  was increased by  $0.00019999998$  over  $2,500$  steps. For other hyperparameters, we used the default hyperparameter configuration provided by TensorFlow Object Detection API. The optimal number of iterations was  $70,000$  (i.e.,  $17,500$  epochs); it took approximately five hours for two Titan RTX graphic processing units to train our model.

## Evaluation

We fit the final model on the entire training dataset and validated the trained model on the test dataset. We had the model make  $\sim 100$  inferences on the bounding box coordinate and whether sacroiliitis is present with its confidence score and chose one bounding box with the highest score for each SIJ. Inferences with a score lower than  $0.5$  were discarded.

In the qualitative analysis, the radiologist (C.A.) determined whether a drawn bounding box contains a SIJ correctly with an appropriate size. In the quantitative analysis, we used a mean average precision (mAP) of Intersection over Union (IoU) as an evaluation metric. IoU measures the overlap between a ground-truth bounding box and a predicted box, and mAP is an average of maximum precision scores across all recall values. We can predefine a threshold to determine whether a prediction is correct. For example, an mAP at  $0.5$  IoU is the score when at least a  $50\%$  overlap with the ground truth bounding box is considered a correct prediction. In addition, we calculated the sensitivity, specificity, accuracy, precision, NPV, and F1-score in the diagnosis of sacroiliitis.

## Results

### Subjects

The final study subjects were  $946$  lower spine or abdominopelvic plain radiographs from  $674$  patients with lower back pain:  $478$  images from  $241$  AS patients and  $468$  images from  $433$  patients without AS (Fig. 1). The mean ( $\pm$ standard deviation) age was  $33.3$  ( $\pm 5.2$ ) years. The male:female ratio was  $2.9:1$  (Table 1).

### Accuracy in detecting sacroiliac joint

Most of the bounding boxes ( $96.3\%$ ,  $366/380$ ) were determined to be drawn correctly and contain SIJs sufficiently with appropriate box sizes in the qualitative analysis (Fig. 2). In ten patients, either left or right bounding box was mismarked. In two patients, both bounding boxes were incorrectly drawn. These incorrect bounding boxes were considered wrong diagnoses when evaluating the diagnostic performance (Fig. 3). In the quantitative analysis, the model achieved an mAP at  $0.5$  IoU of  $0.918$ .

# Diagnostic Performance

## Per joint performance

In the test set, the sensitivity, specificity, and accuracy of our model in diagnosis of SIJ ankylosis were 96.9% (95% confidence interval [CI], 92.9–99.0), 86.8% (95% CI, 81.5–90.9), and 91.1% (95% CI, 87.7–93.7), respectively. Precision, NPV, and F1-score were 84.3% (95% CI, 78.3–89.2), 97.4% (95% CI, 94.1–99.2), and 90.2% (95% CI, 85.0–93.9). AUC was 0.932 (95% CI, 0.903–0.961) (Table 2; Figs 2 and 3).

## Per image performance

In diagnosis of AS per image, sensitivity, specificity, and accuracy were 94.6% (95% CI, 87.8–98.2), 79.6% (95% CI, 70.3–87.1), and 91.3% (95% CI, 87.7–93.7), respectively, in the test set. Precision, NPV, and F1-score were 81.3% (95% CI, 72.6–88.2), 94.0% (95% CI, 86.5–98.0), and 87.4% (95% CI, 79.5–92.9). AUC was 0.876 (95% CI, 0.831–0.922) (Table 2).

## Discussion

Our study results suggest that a deep learning model using the EfficientDet algorithm may be used for automatic detection of sacroiliitis on plain radiograph, which is prone to be overlooked, especially in young patients. Our trained model detected sacroiliac joints successfully most of the time (>95%) and diagnosed the presence of abnormality with an accuracy over 90%. With further development and validation, we hope to use the deep learning model for computer-aided diagnosis of sacroiliitis on plain radiograph and reduce the frequency of delayed AS diagnosis.

In this study, we used a two-step approach based on object detection instead of the conventional deep learning model for image classification. An object detection model first detects an object and then classifies it as one of the pre-defined classes. In contrast, the conventional deep learning model reads and classifies a whole input image. The object detection approach's primary advantage in this case is efficiency; it can achieve comparable performance with less data than the conventional model. In typical medical imaging, only a small portion of the vast image data is relevant to diagnosing a specific disease. Thus, the model requires a large amount of data simply to learn which area to focus on. However, if we use our domain knowledge and let the model know where to look at (e.g., SIJ in this study), the model can perform a task more efficiently because it does not need to learn which part of the image has crucial information.

Since this was a pilot study to test a deep learning model's feasibility and performance on a small dataset, there is still room for improvement in several aspects. First, we need more data to improve the robustness and performance, especially unusual or rare case images such as overlying projection of sacrum and ilium at SIJ (Fig. 3a), pediatric patients (Fig. 4a), or underexposed images (Fig. 4b). Second,

we need to extend our model to detect spine abnormality on X-ray as well because spine involvement is a key finding to AS, with the radiological hallmark being the growth of bony spurs (i.e., syndesmophytes) and florid spondylitis [15]. Third, the proportion of patients with sacroiliitis was unrealistically high in our study population. In Korea, the prevalence of AS in 2015 was 52.3 per 100,000 people [16], although it is likely higher in people with lower back pain. Therefore, the sensitivity and precision presented in this study are highly likely overly optimistic. Lastly, the dataset used in this study was from a single institution; it is unsure whether the model can be generalized well to independent external datasets. Therefore, now that we confirmed a deep learning model's potential, we plan to conduct a further study with larger and more heterogeneous datasets from multiple institutions with a realistic prevalence of AS.

## **Conclusions**

In conclusion, we developed and validated the deep learning-based model to detect SIJs and determine whether they are sclerosed or ankylosed on plain X-ray radiograph in young patients with low back pain.

## **Declarations**

## **Ethics approval and consent to participate**

The Institutional Review Board of National Health Insurance Service Ilsan Hospital (NHIMC 2020-06-005) approved this Health Insurance Portability and Accountability Act-compliant retrospective study and waived the informed consent. All methods were performed in accordance with relevant guidelines and regulations.

## **Consent for publication**

Not applicable.

## **Availability of data and materials**

Due to the institutional policy, data can only be made available to researchers who subject to a non-disclosure agreement, upon reasonable request.

## **Competing interests**

The authors declare that they have no conflict of interest.

## **Funding**

Not applicable.

## Authors' contributions

SHN, CA, HCO conceived the study. SHN and DK obtained and extracted data. CA and DK cleaned data. SHL and MYC reviewed images. CA and DK performed deep learning. CA performed statistical analysis. SHN and CA wrote the paper. All authors have taken due care to ensure the integrity of this work, and all authors read and approved the final manuscript. CA and HCO were in charge of the overall direction.

## Abbreviations

AS, ankylosing spondylitis; SIJ, sacroiliac joint; CT, computed tomography; MRI, magnetic resonance imaging; API, application programming interface; mAP, mean average precision; IoU, Intersection over Union, NPV, negative predictive value, CI, confidence interval; AUC, area under the receiver operating characteristics curve.

## References

1. Braun J, Sieper J. Ankylosing spondylitis. *Lancet*. 2007;369:1379–90.
2. Rudwaleit M. New approaches to diagnosis and classification of axial and peripheral spondyloarthritis. *Curr Opin Rheumatol*. 2010;22:375–80.
3. Kellgren JH. Diagnostic criteria for population studies. *B Rheum Dis*. 1962;13:291–2.
4. Dagfinrud H, Kjekshus I, Mowinckel P, Hagen KB, Kvien TK. Impact of functional impairment in ankylosing spondylitis: impairment, activity limitation, and participation restrictions. *J Rheumatology*. 2005;32:516–23.
5. Palla I, Trieste L, Tani C, Talarico R, Cortesi PA, Mosca M, et al. A systematic literature review of the economic impact of ankylosing spondylitis. *Clin Exp Rheumatol*. 2012;30 4 Suppl 73:S136-41.
6. Yi E, Ahuja A, Rajput T, George AT, Park Y. Clinical, Economic, and Humanistic Burden Associated With Delayed Diagnosis of Axial Spondyloarthritis: A Systematic Review. *Rheumatology Ther*. 2020;7:65–87.
7. Seo MR, Baek HL, Yoon HH, Ryu HJ, Choi H-J, Baek HJ, et al. Delayed diagnosis is linked to worse outcomes and unfavourable treatment responses in patients with axial spondyloarthritis. *Clin Rheumatol*. 2015;34:1397–405.
8. Bruno MA, Walker EA, Abujudeh HH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *Radiographics*. 2015;35:1668–76.
9. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9:671–5.
10. Geijer M, Göthlin GG, Göthlin JH. The validity of the new york radiological grading criteria in diagnosing sacroiliitis by computed tomography. *Acta Radiol*. 2009;50:664–73.

11. Google. TensorFlow 2 Object Detection Application Programming Interface. Available via [https://github.com/tensorflow/models/tree/master/research/object\\_detection](https://github.com/tensorflow/models/tree/master/research/object_detection). Accessed 20 Nov 2020.
12. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Arxiv. 2019;1905.11946.
13. Tan M, Pang R, Le QV. EfficientDet: Scalable and Efficient Object Detection. Arxiv. 2019;1911.09070.
14. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. Arxiv. 2014;1412.6980
15. Østergaard M, Lambert RGW. Imaging in ankylosing spondylitis. Ther Adv Musculoskelet Dis. 2012;4:301–11.
16. Park JS, Hong JY, Park YS, Han K, Suh SW. Trends in the prevalence and incidence of ankylosing spondylitis in South Korea, 2010–2015 and estimated differences according to income status. Sci Rep. 2018;8:7694.

## Table

**Table 1.** Patient characteristics

	No AS	AS	Total
No. of patients	433	241	674
Age, mean ( $\pm$ SD)	32.3 ( $\pm$ 5.3)	34.1 ( $\pm$ 5.1)	33.3 ( $\pm$ 5.2)
Sex, Male:Female	323:110	179:62	502:172
No. of X-rays	468	478	946
Radiologic grade*			
Grade 0	67 (15.5%)	0 (0%)	67 (10.0%)
Grade 1	366 (84.5%)	0 (0%)	366 (57.4%)
Grade 2	0 (0%)	211 (44.1%)	211 (31.3%)
Grade 3	0 (0%)	106 (22.2%)	106 (15.7%)
Grade 4	0 (0%)	161 (33.7%)	161 (23.9%)

AS, ankylosing spondylitis; SD, standard deviation

\*the worse grade of the left and right sacroiliac joints

**Table 2.** Diagnostic performance of the trained EfficientDet model in the test dataset

TP/FN/FP/TN	AUC	Sensitivity	Specificity	Accuracy	Precision	NPV	F1-score
Per joint performance							
156/29/5/190	0.932 (0.903– 0.961)	96.9% (92.9– 99.0)	86.8% (81.5– 90.9)	91.1% (87.7– 93.7)	84.3% (78.3–89.2)	97.4% (94.1– 99.2)	90.2% (85.0– 93.9)
Per image performance							
87/20/5/78	0.876 (0.831– 0.922)	94.6% (87.8– 98.2)	79.6% (70.3– 87.1)	91.3% (87.7– 93.7)	81.3% (72.6– 88.2),	94.0% (86.5– 98.0)	87.4% (79.5– 92.9)

Values in parentheses are 95% confidence intervals. TP, true positive; FN, false negative; FP, false positive; TN, true negative; NPV, negative predictive value

## Figures

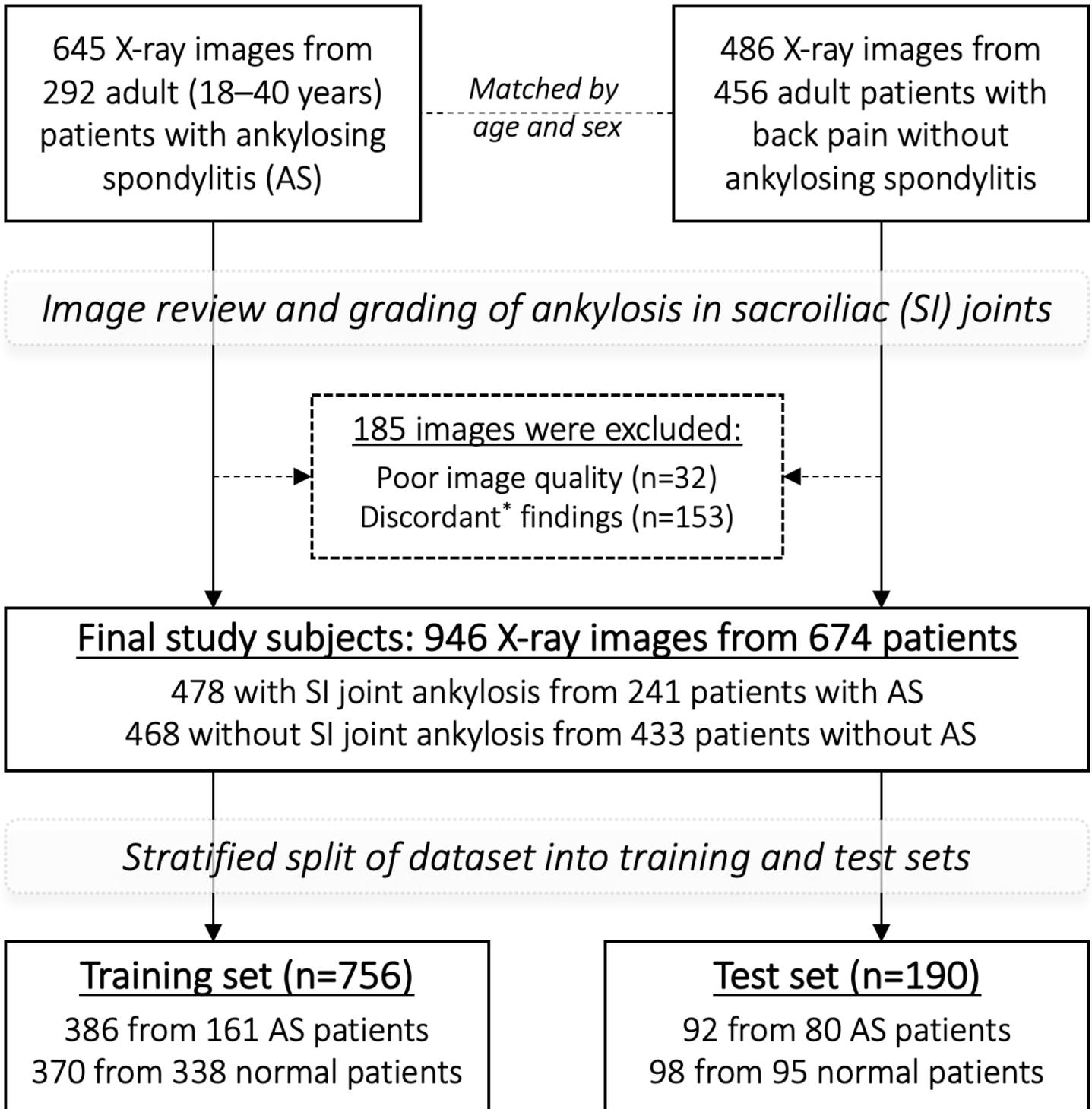
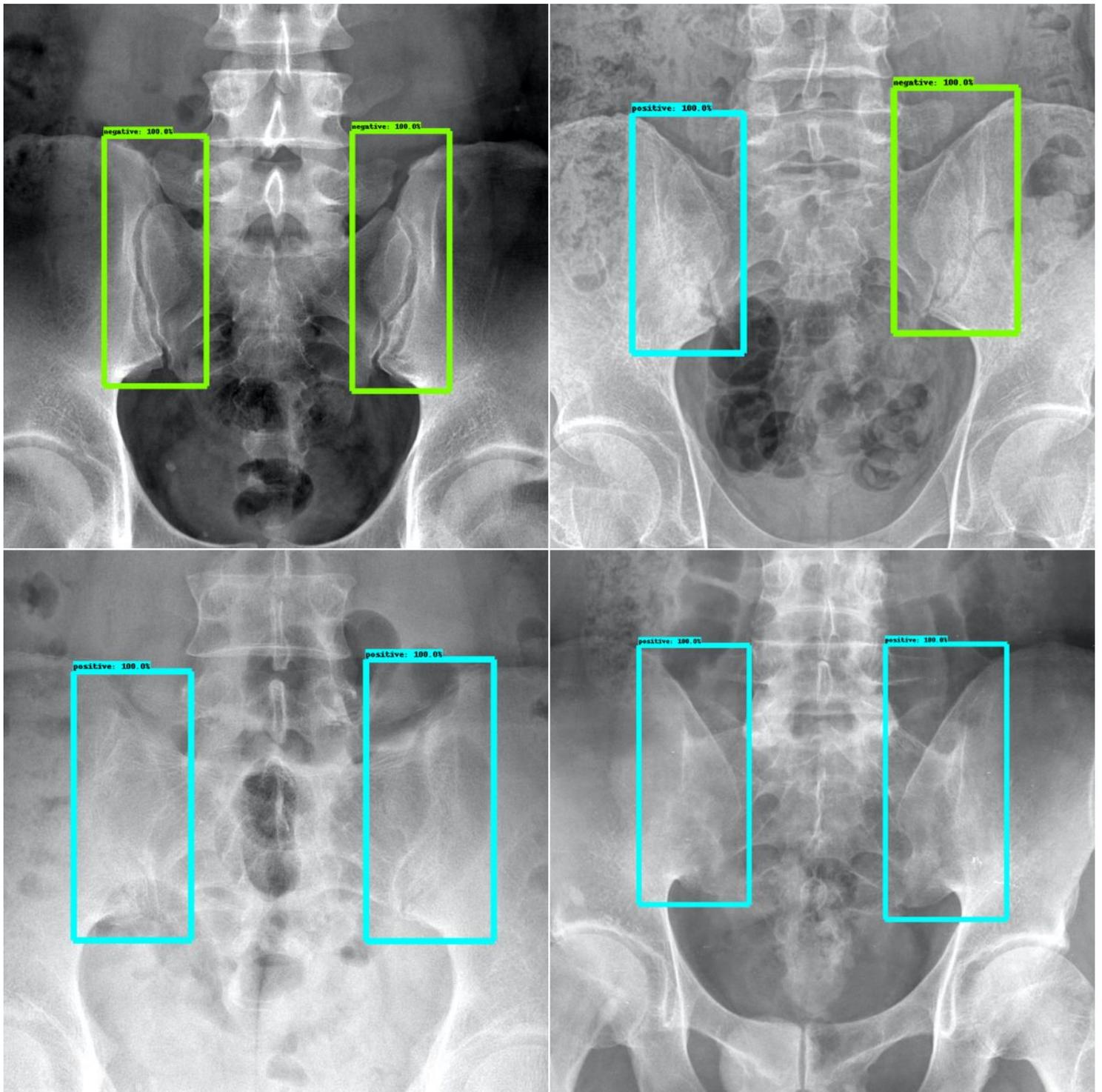


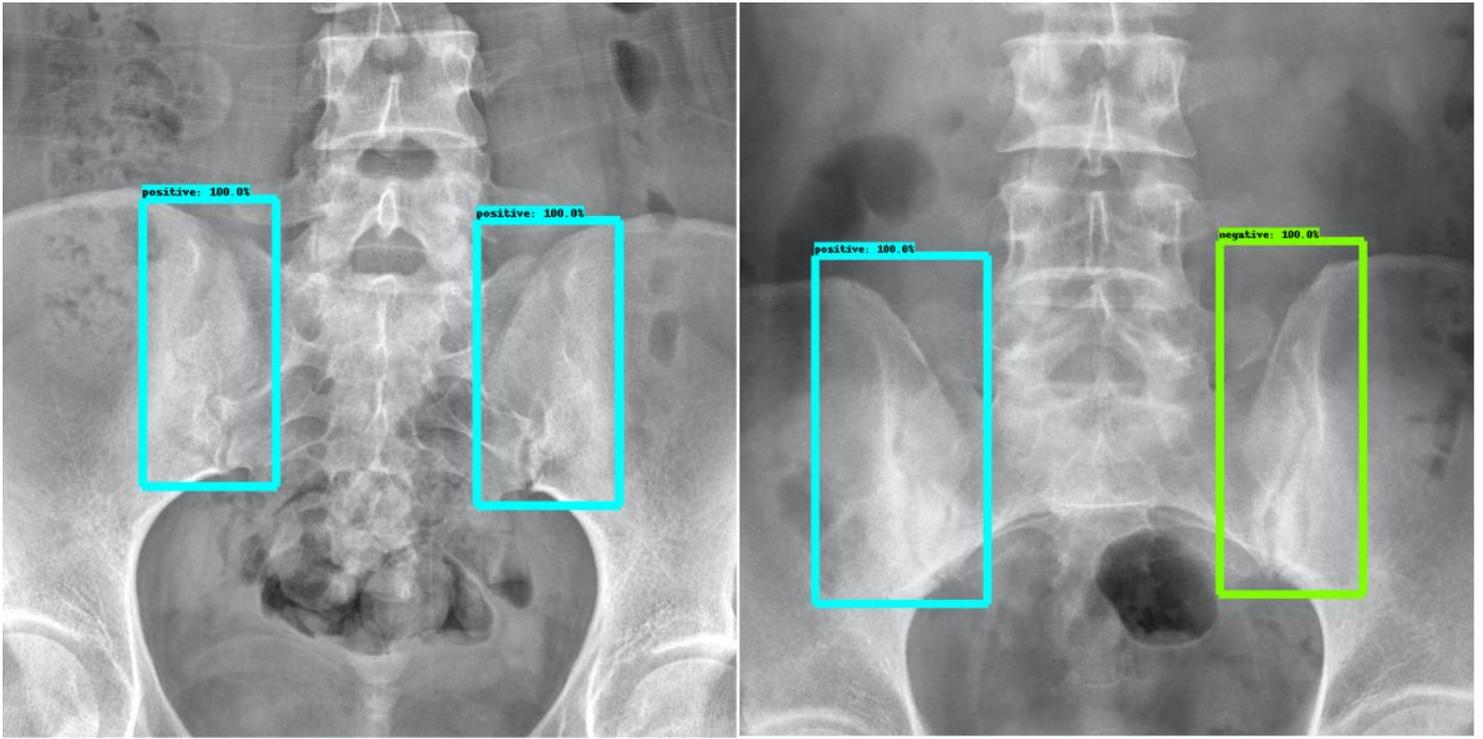
Figure 1

## Study flow



**Figure 2**

Example cases of correct classification by the EfficientDet model. (a) 31-year-old man without sacroiliitis. The model correctly detected and classified both sacroiliac joints (SIJs) as normal; (b) 24-year-old man with sacroiliitis (grade 3 for right SIJ and grade 1 for left SIJ); (c) 39-year-old man with sacroiliitis (grade 3 for both SIJs); (d) 32-year-old man with sacroiliitis (grade 4 for both SIJs).



**Figure 3**

Example cases of wrong classification by the EfficientDet model. (a) 21-year-old woman without sacroiliitis. The model detected both joints but incorrectly classified them as abnormal possibly because of the sacroiliac joint (SIJ)'s atypical appearance due to the overlying projection of sacrum and ilium. (b) 29-year-old man with sacroiliitis (grade 3 for right SIJ and grade 2 for left SIJ). The algorithm incorrectly classified the left SIJ as normal.



## Figure 4

Example cases of detection failure by the EfficientDet model. (a) 13-year-old girl without sacroiliitis. The model failed to detect both sacroiliac joints (SIJs). (b) 39-year-old man with complete ankylosis in both SIJs. However, the model failed to detect both SIJs possibly because of the decreased contrast in this underexposed image.