

Article Deep Learning-Based Approaches for Classifying Foraminal Stenosis Using Cervical Spine Radiographs

Jiho Park ¹, Jaejun Yang ², Sehan Park ^{2,*} and Jihie Kim ^{1,*}

- ¹ Department of Artificial Intelligence, Dongguk University, Seoul 04620, Republic of Korea
- ² Department of Orthopedic Surgery, Dongguk University Ilsan Hospital, Goyang-si 10326, Republic of Korea
- * Correspondence: birdone86@gmail.com (S.P.); jihie.kim@dgu.edu (J.K.)

Abstract: Various disease detection models, based on deep learning algorithms using medical radiograph images (MRI, CT, and X-ray), have been actively explored in relation to medicine and computer vision. For diseases related to the spine, primarily MRI-based or CT-based studies have been conducted, but most studies were associated with the lumbar spine, not the cervical spine. Foraminal stenosis offers important clues in diagnosing cervical radiculopathy, which is usually detected based on MRI data because it is difficult even for experts to diagnose using only an X-ray examination. However, MRI examinations are expensive, placing a potential burden on patients. Therefore, this paper proposes a novel model for diagnosing foraminal stenosis using only X-ray images. In addition, we propose methods suitable for cervical spine X-ray images to improve the performance of the proposed classification model. First, the proposed model adopts data preprocessing and augmentation methods, including Histogram Equalization, Flip, and Spatial Transformer Networks. Second, we apply fine-tuned transfer learning using a pre-trained ResNet50 with cervical spine X-ray images. Compared to the basic ResNet50 model, the proposed method improves the performance of foraminal stenosis diagnosis by approximately 5.3–6.9%, 5.2–6.5%, 5.4–9.2%, and 0.8–4.3% in Accuracy, F1 score, specificity, and sensitivity, respectively. We expect that the proposed model can contribute towards reducing the cost of expensive examinations by detecting foraminal stenosis using X-ray images only.

Keywords: foraminal stenosis; cervical spine X-ray preprocessing; transfer learning; fine-tuning; Spatial Transformer Network

1. Introduction

In modern medicine, certain disease diagnoses and clinical treatments are based on findings obtained from medical images, such as X-rays, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT). This is also applicable to cervical radiculopathy.

Cervical radiculopathy is often a result of disc herniation or cervical spondylosis, resulting in pain in the neck and arm and nerve paralysis or sensory loss by pressing on nerves in the arm [1]. As a symptom underlying the diagnosis of cervical radiculopathy, foraminal stenosis refers to the narrowing of the foramen between the cervical spine, in which the nerves extending from the cervical spine are compressed, causing pain or decreased sensation and paralysis in the arms. Foraminal stenosis arises as disc degeneration with age causes decreased disc height and foraminal narrowing [1]. Therefore, the presence of foraminal stenosis is important in determining early diagnosis and treatment for cervical radiculopathy. To determine the presence or absence of foraminal stenosis, this paper focuses on deep learning-based approaches to diagnosing foraminal stenosis.

Foraminal stenosis is analyzed by experts based on radiographs, such as X-rays, MRIs, and CTs. Diagnosis is mainly made based on MRI because it has the highest accuracy, and the predictive success rate is about 88% [2]. However, MRI diagnostic tests are a potential burden to patients because they are expensive. In contrast, X-rays are relatively



Citation: Park, J.; Yang, J.; Park, S.; Kim, J. Deep Learning-Based Approaches for Classifying Foraminal Stenosis Using Cervical Spine Radiographs. *Electronics* 2023, 12, 195. https://doi.org/10.3390/ electronics12010195

Academic Editors: Laura Falaschetti and Turchetti Claudio

Received: 25 October 2022 Revised: 15 December 2022 Accepted: 19 December 2022 Published: 31 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). inexpensive, but it is difficult for experts to diagnose foraminal stenosis using only X-rays. However, this does not mean that there are no diagnostic clues on X-rays. According to [1], it can be confirmed even with the naked eye that the foramen is narrowed in the X-ray oblique view. Therefore, this paper aims to propose a novel foraminal stenosis classification model by applying a deep learning algorithm to X-ray images in order to learn features that are difficult to identify with the naked eye automatically and efficiently. It is expected that the proposed model, as an auxiliary tool, will help experts diagnose foraminal stenosis more consistently. Furthermore, it can be expected that the patient will be relieved of the burden of the cost of the examination as the proposed model uses X-ray images rather than MRIs or CTs. In addition, the diagnosis of foraminal stenosis can be automated with the proposed classification model, without requiring much professional expertise.

To date, there have been few cases of foraminal stenosis diagnosis by applying deep learning algorithms to cervical spine X-rays. Most studies using X-ray images were mainly focused on chest X-ray images [3–5] or lumbar spine radiographs [6,7] rather than the cervical spine. Therefore, this paper proposes a classification model to diagnose foraminal stenosis by applying deep learning algorithms to cervical spine X-ray images. In addition, most studies related to diagnosing spinal diseases used MRIs or CTs [6,8,9], which are expensive for patients. This study aims to diagnose foraminal stenosis using X-ray images only, which will be less expensive. It is often difficult to obtain a large amount of data owing to the characteristics of medical data. This paper proposes methods that can substantially increase the accuracy of the model with only a small amount of data by using various image preprocessing and data augmentation methods.

Our contributions can be summarized as follows: (i) we introduce a new technique for classifying foraminal stenosis. (ii) We propose a classification model using cervical spine X-ray images. (iii) We demonstrate that the proposed methods are suitable for a small number of X-ray images.

First, to detect foraminal stenosis, the proposed model needs to focus on the foramen. In an original X-ray image, the cervical spine oblique view contains not only the foramen but also other bone parts such as teeth and skull, so we cropped the input image only to the Region of Interest (ROI). In order to crop the desired section of the image, we applied YOLOv5 [10] to learn the ROI, as described in Section 3.1.

Second, as the Convolutional Neural Network (CNN)-based model tends to be sensitive to the input, to emphasize the foramen part, we applied Histogram Equalization, which is one of the most popular methods for X-ray images [11,12]. Histogram equalization makes the image clearer because the contrast between the bone part and non-bone part is emphasized. As shown in Section 3.2, such image preprocessing can help CNN-based approaches [13] learn X-ray classification models more effectively.

Third, in the case of the oblique view of the cervical spine X-ray used in this study, the labels of the left view and the right view may be different even for the same patient. Therefore, the left and right X-ray images are learned separately. As the amount of data used in this paper is limited, we perform data augmentation by using flipped images of the left view for training the right view, and vice versa. As a result, we can double the number of images for generating the model, as described in Section 3.3.

Fourth, as the CNN-based model tends to be sensitive to input, by applying the Spatial Transformer Network (STN), the slope of the cervical spine, which is different for each person, is aligned into a similar slope. In Section 3.4, we show the effectiveness of STN in increasing the accuracy of the model for diagnosing foraminal stenosis.

Finally, in Section 3.5, this paper proposes a novel foraminal stenosis classification model based on ResNet50 [14], which utilizes cervical spine X-rays by effectively processing low amounts of medical data. Transfer learning [15] was performed using a pre-trained model to utilize the small amount of data, and fine-tuning was applied to reflect the characteristics of the medical data domain to the parameters of the pre-trained model.

Reference	Task	Method	Modality	Metric
Jamaludin et al. 2017 [16]	Spinal stenosis diagnosis	3D CNN	12,018 MRI images of 2009 subjects	Accuracy (Acc)
Won et al., 2020 [17]	Spinal stenosis diagnosis	R-CNN, RPN, ResNet50, VGG	12,018 MRI images	Acc, F1 score
Dong et al., 2018 [18]	Chest organ segmentation	ResNet18	Chest X-ray images	IoU
Saiz et al., 2020 [19]	Lung detection	VGG16, Fast R-CNN	987 Chest X-ray images	Acc, Specificity, Sensitivity
Brunese et al., 2020 [20]	Covid-19 classification	Variant VGG16, Grad-CAM	6523 Chest X-ray images	Acc, F-measure, Specificity, Sensitivity
Al-Kafri et al., 2019 [21]	Spinal stenosis diagnosis	SegNet, DeepLab, RefineNet, VGG16	48,345 MRI images of 515 subjects	IoU, Acc. BF-score
Fan et al., 2020 [22]	Spinal stenosis diagnosis, semantic segmenatation	3D U-Net	1681 CT images of 31 subjects	DC
Gaonkar et al., 2019 [23]	Spinal stenosis diagnosis, Disc segmenatation	Deep U-Net	MRI images of 1755 subjects	Dice score, Hausdorff distance, and average surface distance
Bharati et al., 2021 [24]	Covid-19 classification	CO-ResNet: ResNet101, ResNet50, ResNet152	5935 Chest X-ray images	Acc, AUC, F1-score, Precision, Recall, Sensitivity
Nayak et al., 2021 [25]	Covid-19 detection	ResNet34, ResNet50, GoogleNet, VGG16, AlexNet, MobileNetV2, InceptionV3, SqueezeNet	406 Chest X-ray images	Acc, AUC, F1-score, Precision, Specificity, Sensitivity
Chen et al., 2022 [26]	Scoliosis diagnosis	Faster R-CNN, ResNet50, LBP, SVM	3600 Spine X-ray images	AUC, Precision, Specificity, Sensitivity

Table 1. Overview of studies using deep learning for medical images, especially X-ray images or spine data.

2. Related Work

With the advent of CNN, the image classification field has developed rapidly. Starting with AlexNet [27], the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) error decreased remarkably. Models that have performed well in the ILSVRC, such as VGGNet (VGG) [28] and ResNet [14], are still widely used as pre-trained models. These models are used not only for datasets similar to ImageNet but also for medical data. As presented in Table 1, the VGG model has been widely used in COVID-19-related models that use chest X-ray images after the outbreak of COVID-19 [19,20,24,25]. In addition, the ResNet model has been used upon various types of medical images [17,29,30] including X-ray images [18,24–26].

According to Table 1 and Qu et al. [31], a paper on current development and prospects of deep learning in spine images, most of the spine-related studies [16,17,21–23] so far have proposed models using MRI or CT images. Furthermore, these studies considered the lumbar spine, not the cervical spine. Among these studies, VGG and ResNet were used in Won et al. [17]. Other spinal-related studies [21–23] were segmentation-related studies, not classification model-related studies. Therefore, a U-Net based model was used in these studies [21–23]. SpineGEM [6] is similar to this study in that the proposed model classifies spine diseases based on the VGG-M model. However, SpineGEM used MRI images, not X-ray images, to classify diseases, and they classified the diseases of the lumbar spine, not the cervical spine. Another previous study related to spine diseases [8] detects foraminal stenosis in the same manner as in this study but uses MRI to detect foraminal stenosis of the lumbar spine, not the cervical spine. In this previous study [9], an MRI image of the lumbar spine was used, and the accuracy of the lumbar spine disc state classification was increased to 87% by applying the ROI method and fine-tuning several pre-trained

models. Most of the X-ray-related studies [18–20,24,25] in Table 1 are studies using chest X-rays and ResNet-based models. In [32], the study used spine X-ray images to detect scoliosis. They performed experiments using a variety of models, i.e., ResNet34, ResNet50, GoogleNet, VGG16, AlexNet, MobileNetV2, InceptionV3, and SqueezeNet, to obtain the scoliosis classification model. The best-performing model of [32] is a ResNet-based model. In addition, the ResNet50 model used as the base has parameters trained on the ImageNet dataset. However, there is a difference between the ImageNet dataset and the X-ray dataset used in this study. In a study [33] that performed gender detection using cervical spine X-rays, the model was trained by fine-tuning the pre-trained model. This study fine-tuned the ResNet50 model to fit the model parameters to the X-ray dataset. Therefore, this paper proposes a novel ResNet50-based model with transfer learning and fine-tuning using a pre-trained model's parameters, as there is not much data to learn.

YOLOv5 [10] shows high-accuracy performance in object detection, and many studies [33–35] apply YOLOv5 for object detection. Consequently, this paper employed YOLOv5 to crop the ROI part from cervical spine X-ray images to remove the unnecessary parts for learning the proposed model.

The study related to X-ray image preprocessing [11] suggests that the Histogram Equalization-applied dataset's accuracy was 2% higher than the non-Histogram-Equalization-applied dataset's accuracy. For this reason, this paper suggests Histogram Equalization as a preprocessing method to improve the performance of the proposed model.

Thus, transfer learning, fine-tuning, and ROI methods were applied using the pretrained model to increase the accuracy of our proposed model.

3. Methods

Owing to the characteristics of medical data, it is difficult to secure a large amount of labeled data. Human experts have to label medical data directly, and consent must be obtained for the use of medical data. In addition, there were few prior studies and no open dataset for the cervical spine oblique view considered in this study. Therefore, to optimize the performance of the foraminal stenosis classification model using a small amount of data, various data preprocessing techniques and data augmentation techniques are applied. The overview of the proposed model is shown in Figure 1. The methods applied in this paper are as follows.



Figure 1. The overview of the proposed model based on ResNet50.

3.1. YOLOv5-Based Region of Interest Detection in X-ray Images

The original oblique view of cervical spine X-ray images includes not only the cervical spine and foramina, which are critical for foraminal stenosis diagnosis, but also additional information, such as teeth, skull, clavicle, and an alphabet indicating left or right view, as shown in Figure 2a,b. Therefore, parts that are not required for foraminal stenosis detection and parts that may be mislearned by the model should be excluded from learning. Therefore, we applied ROI crop to pre-process the input data of the model in this paper. Object detection methods can be applied to detect ROI in the images, i.e., the area of the cervical spine and foramina in the case of cervical spine X-ray images. We used the YOLOv5 [10] model well-known for object detection to obtain ROI-cropped images, as illustrated in Figure 2c,d. In the cervical spine X-ray image, only the cervical spine part was annotated as a bounding box by a clinician. A total of 100 X-ray images were annotated by a clinician and trained the YOLO model using these annotated data. ROI cropping of the other X-ray images (without annotation) was performed by the trained YOLO model. Therefore, the classification model does not intensively learn features other than the cervical spine and foramina. As a result, we separately pre-trained the YOLOv5 model on cervical spine X-ray images and obtained the trained model with the mAP score of 0.97. The pre-trained YOLOv5 model is applied to all of the raw data and detects the ROI part. The detected ROI part was cropped. The output data of this step, i.e., the ROI crop image, become the input data of the next step, Histogram Equalization, of the proposed model.



Figure 2. Results of ROI crop using YOLOv5. (**a**) right original oblique view, (**b**) left original oblique view, (**c**) right ROI cropped oblique view, and (**d**) left ROI cropped oblique view.

3.2. Contrast Improvement in X-ray Images

Then, next data preprocessing method is Histogram Equalization [36], which is often applied to X-ray image preprocessing [11]. Histogram Equalization (HE) is an image preprocessing method using a histogram to adjust the contrast of the image to make the image clear. HE is useful in image data consisting of a narrow range of grayscale values, such as X-ray data. Equalization, i.e., transforming the histogram of the image to the entire grayscale range, is performed to enhance the contrast of the image. The HE implementation is as follows.

First, the number of pixel values g is in the image using the histogram function h defined by

$$h(g) = N_g. \tag{1}$$

Second, the histogram value is obtained when w is the width and h is the height of the image using the function p defined by

$$p(g) = \frac{h(g)}{w \times h}.$$
(2)

Then, using the function *cumulative distribution function* defined by

$$cdf(g) = \sum_{0 \le i \le g} p(i), \tag{3}$$

the accumulated histogram value is obtained.

Finally, the result value of histogram equalization y is obtained by applying the following operation:

$$y = round(cdf(g) \times L_{max}), \tag{4}$$

where *L* is the total number of grayscale levels in them image.

After HE, the bone area becomes brighter in the X-ray image, while the dark area other than the bone area becomes darker so that the contrast becomes more distinct, as shown in Figure 3. As a result, the foramen, an important feature to detect foraminal stenosis, becomes further clarified. In this study, the input X-ray images are first preprocessed to obtain the ROI-cropped images, followed by histogram equalization.



Figure 3. Results of Histogram Equalization. (**a**) right ROI cropped oblique view, (**b**) left ROI cropped oblique view, (**c**) right ROI cropped oblique view with HE applied, and (**d**) left ROI cropped oblique view with HE applied.

3.3. X-ray Data Augmentation for Foraminal Stenosis

Finally, the Flip data augmentation method is applied. As the amount of labeled data is insufficient, the Flip method was used to double the amount of data, as presented in Table 2. In the case of the dataset used in this study (the left and right cervical spine oblique view X-ray images), the direction and the angle of X-ray imaging are different. Therefore, each view of the images has its own label. The data used in this study are X-rays of cervical vertebrae on the left and X-rays of cervical vertebrae on the right. Therefore, there are two X-ray images per patient, each labeled. We consider the characteristics of the data and use left-and-right flip as a data augmentation method to leverage the labels of the original images. This is because the data cannot be combined if you proceed with the up-and-down flip. Therefore, as shown in Figure 1, the dataset was doubled by flipping the output data of the HE in opposite directions and combining them with the opposite dataset. Then, the data doubled could be used as input data for model training as shown in Figure 1. The original labels are preserved. The configuration of the flipped dataset is listed in the last row of Table 2.

3.4. Attention-Based Spatial Transformation

Spatial Transformer Network (STN) [37] is a generalization of differentiable attention that can be applied to any spatial transformation. STN allows the neural network to learn what spatial transformations to perform on the input image in order to enhance the geometric invariance. STN is a network that transforms a distorted object image into a straight shape by performing an affine transform. It is composed of three parts: Localization network, Grid generator, and Sampler. First, the localization network receives the feature map of the input image as input and outputs *theta*. *theta* is a parameter for affine transform. Thereafter, the grid generator samples the pixel values of the output image using *theta*. Finally, the sampler generates the final transformed output image using the value sampled by the grid generator and the input image.

As CNNs [13,27] have sensitive output fluctuations for transformed inputs, such as rotation, scaling, and STN can be a very useful mechanism to overcome these. In addition, one of the advantages of STN is that it can be easily connected to the existing model with very little modification. Therefore, this study applies STN to the proposed model to overcome the difference across multiple X-ray images. In this study, as there are two types of spatial input data, i.e., in the left-slope and right-slope direction, the STN learns the spatial features of the image by dividing the X-ray image into left–right sides. The results of the STN are illustrated in Figure 4, and the average loss and accuracy of the STN are 0.0218 and 99%, respectively. The STN is trained using the cervical spine X-ray image, and then the weights are frozen and added as a layer to pre-process the data right before the proposed model training as shown in Figure 1.



Figure 4. Comparative results before and after STN. Dataset images: before STN; Transformed image: after STN.

3.5. Transfer Learning

The proposed model is based on ResNet50 [14]. The ResNet model uses the Residual Block (BottleNeck Architecture) to solve the vanishing gradient problem that occurs when backpropagating in deep learning. This structure has a shortcut structure in which an input value is added to an output value as it is, so it is possible to solve the vanishing gradient problem in which an input value is forgotten as the layer deepens in a model with a deep structure. In addition, the ResNet model is one of the models that is still widely used as it has a simple structure.

The parameters of YOLOv5 and STN in the proposed model are frozen as shown in Figure 1. Those parameters were obtained through separate experiments described in Sections 3.1 and 3.4. The trainable part of the proposed model consists of the pre-trained ResNet50 model and the Fully Connected (FC) layer, which is attached to suit foraminal stenosis classification. The parameters of the pre-trained ResNet50 are initialized using the ImageNet [38] dataset, which is a domain different from our cervical spine X-ray dataset. Therefore, the proposed model applied transfer learning [15] and fine-tuned the parameters of the pre-trained ResNet50 and FC layer using the cervical spine X-ray images to improve the performance.

In the final part of the proposed model, the cervical spine X-ray input image is classified as abnormal if it has foraminal stenosis, and normal if it does not have foraminal stenosis. As there are two classification classes in the proposed model, this study uses a binary cross entropy function as a loss function.

4. Experiments Results

4.1. Dataset

The data used in this study comprises cervical spine X-ray images of 798 patients provided by Dongguk University Ilsan Hospital. For each of the 798 patients, two X-ray images (a left oblique view and a right oblique view) were provided per person. The oblique view was captured by adjusting the angle so that the foramen was clearly visible when X-rays of the cervical spine were taken. Even in the same foramen, the foraminal stenosis results of the left oblique view and the right oblique view were different, so each was labeled separately. If foraminal stenosis was present, the image was labeled as abnormal and if not, it was labeled as normal. For accurate labeling, labeling for X-ray images was based on MRI examinations. For each data on the left and right, 60% was used for the training dataset, and 20% each for the test and validation dataset. An overview of the detailed dataset is presented in Table 2. The original size of input cervical spine X-ray images 1180×2012 was resized to 512×512 . We downscale the size of input data as 512×512 , as the size of the ROI crop data were not the same but slightly different after the YOLO crop step. Additionally, when the scale was larger than 512×512 , the results were similar while the training time increased significantly. On the other hand, the performance was poor when the size of the input data was 256×256 . Therefore, we unified the size of input data to 512×512 .

Table 2. Configuration of the cervical spine X-ray image dataset. The first and second row of the table present the configuration of the raw data before applying Flip method. The last row of the table is the configuration of the result of the Flip: flipped left view added to right view and flipped right view added to left view.

	Label	Training (60%)	Validation (20%)	Testing (20%)	Total
left obligue view	abnormal	298	100	100	498
len oblique view	normal	180	60	60	300
right obligue view	abnormal	313	105	105	523
light oblique view	normal	165	55	55	275
left / right flip applied	abnormal	611	205	205	1021
ien/ rigin inp applied	normal	345	115	115	575

4.2. Preliminary Experiments

Before proceeding with the experiments, some preliminary investigations were required to obtain the necessary setup information for the experiment. When the batch size was increased to 8, the accuracy was 75.0%, which is lower than when the batch size was set to 4. In the case of the epoch size, the accuracy was approximately 74% when the epoch size was set to 20, and 75.93% when set to 30, which was almost the same as when it was set to 25. Therefore, the experiments set a batch size of 4 and trained for 25 epochs using a Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9 and an initial learning rate of 0.001. The learning rate was decreased by 0.1 every 7 epochs. Furthermore, the experiments use binary cross-entropy and applied additionally calculated weights to overcome the class imbalance.

A preliminary experiment was conducted to determine a suitable pre-trained model for the cervical spine X-ray image dataset among the pre-trained models used in previous studies about various medical data, especially X-ray images. This experiment used only left oblique view X-ray images. Additionally, this experiment is for the comparison of pre-trained models and choosing the suitable baseline model for X-ray image data. Therefore, only YOLOv5 cropping process was used and the other proposed method in this paper, i.e., HE, Flip, and STN were not used. The experiment results are summarized in Table 3. As the ResNet50 [14] pre-trained model was slightly superior to VGG16 [28] as presented in the Table 3, ResNet50 was selected as the pre-trained model in this study.

	Accuracy	F1 Score	Specificity	Sensitivity
ResNet50	70.62	60.5	77.0	60.0
VGG16	70.0	59.82	78.0	58.33
VGG19	66.87	49.52	81.0	43.33

Table 3. Result of various pre-trained models trained with the left oblique view.

As this paper proposes the model based on the pre-trained ResNet50 model trained by ImageNet, a preliminary experiment was conducted to check whether it is effective to fine-tune the parameters using the cervical spine X-ray data considered. The results of the experiment are summarized in Table 4. As predicted, the fine-tuned model performed better than the model using the frozen pre-trained model parameters. Therefore, fine-tuning the pre-trained ResNet50 model is recommended.

	Fine-Tuning	Accuracy	F1 Score	Specificity	Sensitivity
left oblique view	×	70.62	60.5	77.0	60.0
left oblique view	1	71.25	60.95	78.0	59.93
right oblique view	X	70.0	58.62	74.28	61.81
right oblique view	1	73.12	69.81	86.66	67.27

Table 4. Result of fine-tuning in left and right oblique view, respectively.

The ablation study results confirm that the proposed performance enhancement methods, i.e., HE, Flip, and STN, have a positive effect on the performance of the proposed model, as can be observed by comparing the data in Tables 4 and 5. First, agreeing with the previous studies using X-ray data, an ablation experiment was performed to confirm whether it is effective to apply Histogram Equalization. The result of Histogram Equalization is shown in Figure 3, and the results of training the model using the data to which HE is applied are presented in Table 5. According to Table 5, the performance was better in most metrics than in non-histogram equalization applied data, as this study predicted. In the case of the right oblique view, the accuracy increased by approximately 3%, and in the case of the left oblique view, all metrics increased except specificity. In the left oblique view, the accuracy increased by approximately 1%, the F1 score increased by approximately 3%, and sensitivity increased by approximately 5%.

	HE	Flip	STN	Accuracy	F1 Score	Specificity	Sensitivity
	X	×	×	73.12	69.81	86.66	67.27
	1	×	×	76.74	63.38	85.89	59.26
right oblique view	X	1	×	72.81	61.33	78.04	60.86
	1	×	1	73.12	51.68	89.52	41.81
	X	×	1	76.25	65.45	81.9	65.45
	1	1	1	76.93	65.15	83.41	62.6
	X	×	×	71.25	60.95	78.0	59.93
left oblique view	1	×	×	72.84	63.68	78.0	64.06
	X	1	×	71.8	60.86	85.89	59.26
	1	×	1	70.0	57.89	79.0	55.0
	X	×	1	73.12	61.94	82.0	58.33
	1	1	1	75.93	65.77	82.43	64.34

Table 5. Ablation experiment results from applying HE, Flip, and STN, respectively, for test dataset. The model is based on the fine-tuned ResNet50.

Second, an ablation experiment was performed to check whether the performance of the model improved when the amount of data doubled by applying Flip, a preprocessing technique that considered the characteristics of the cervical spine oblique view X-ray dataset. The data configuration as a consequence of applying the Flip method is presented in the last row of Table 2. The results of training the proposed model for left and right oblique views are presented in Table 5. According to Table 5, even though the F1 Score and sensitivity values slightly declined by approximately 0.1% and about 0.7%, respectively, the accuracy and specificity values improved about 0.6% and 7.8%, respectively, for each in the case of the left view, and the difference between the accuracy of the left and right views was reduced. However, all metrics are decreased in the case of the right view. In order to verify the effectiveness of the flip, we compared the result of the experiment: HE and STN were applied but without Flip (fourth row of each view on the Table 5), with the result of the experiment with all the proposed methods (sixth row of each view on the Table 5). According to the flip ablation experiment, when the flip was applied, the metrics were better than when it was not applied, except for the specificity of the right oblique view. The flip method doubled the number of data, and as a result, the model learned more data, showing performance improvement in all metrics except for the specificity of the right oblique view. Therefore, this study suggests that flip is an effective method to increase the amount of the cervical spine X-ray data.

Finally, a third ablation experiment on the effectiveness of applying STN was performed. According to Table 5, accuracy, F1 Score, and specificity are improved by approximately 2%, 1%, and 4%, respectively, in the case of the left view, whereas sensitivity decreased by approximately 1%. However, all metrics except accuracy are decreased in the case of the right view. The accuracy of the right view was improved by approximately 3%, but F1 Score, specificity, and sensitivity decreased by approximately 4%, 5%, and 2%, respectively.

4.3. Experiment Results and Evaluation

The results of the proposed model after applying HE, Flip, and STN are presented in Table 6. All the values of the first row of each view, i.e., the results of basic ResNet50 model, are from Table 4. The final accuracy of the proposed model was 76.93% in the right oblique view and 75.93% in the left oblique view. For the right oblique view, the proposed model was ahead of other basic, ResNext, and WideResNet in all metrics, but Res2Net101 had higher F1 score and sensitivity values than the proposed model, as illustrated in Table 6.

However, in the case of the F1 score value, there was a difference of approximately 0.1 from the proposed model, and the accuracy of the proposed model was approximately 3% higher than that of Res2Net101. Therefore, the proposed model was finally selected. For the left oblique view, the proposed model was finally selected because the proposed model showed better performance than other comparable models for all metrics as illustrated in Table 6. The accuracy of the right oblique view was improved by approximately 6.9%, while the accuracy of the left oblique view was improved by approximately 5.3% compared to the basic ResNet50 model. In the case of the F1 Score, the right oblique view was improved by approximately 6.5%, and the left oblique view was improved by approximately 5.2%. The specificity of the right oblique view was improved by approximately 9.2%, while the specificity of the left oblique view was improved by approximately 5.4%. Finally, the values of the sensitivity of the right oblique view and the left oblique view were improved by approximately 0.8% and 4.3%, respectively. Therefore, all metrics show performance improvement compared to the basic ResNet50 model. As a result, the difference between the left and right widens for all metrics after applying the fine-tuning was improved after applying the proposed methods, especially Flip. According to the confusion matrix, it seemed difficult for the proposed model to classify the normal class. Therefore, the metrics could be improved by focusing on learning the difference between abnormal class and normal class by applying contrastive learning.

Table 6. Comparison of the result values of the ResNet50 model that applied all the methods, i.e., finetuning, ROI cropping, HE, Flip, and STN, proposed in this study (proposed model), three latest ResNet models (ResNext50, WideResNet50, Res2Net101) that applied all the proposed methods, and no method applied ResNet50 model (basic ResNet50).

	Model	Accuracy	F1 Score	Specificity	Sensitivity
	basic ResNet50	70.0	58.62	74.28	61.81
	ResNext50	74.37	62.38	82.92	59.13
right oblique view	WideResNet50	73.43	64.43	77.07	66.95
	Res2Net101	73.75	65.28	76.58	68.69
	proposed model	76.93	65.15	83.41	62.6
	basic ResNet50	70.62	60.5	77.0	60.0
1 4 11.	ResNext50	68.43	56.65	74.63	57.39
left oblique view	WideResNet50	69.37	57.01	76.58	56.52
	Res2Net101	74.37	63.71	80.97	62.60
	proposed model	75.93	65.77	82.43	64.34

Furthermore, the Receiver Operating Characteristic (ROC) curve and Area under the ROC Curve (AUC) value of the best-performing model are shown in Figure 5. The ROC curve shows the classification performance at all classification thresholds. A ROC curve plots True Positive Rate (TPR) vs. False Positive Rate (FPR) at various classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing FP and TP, as shown in Figure 5. AUC measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1). The range of the AUC is 0.0 to 1.0, and the model whose predictions are 100% correct has an AUC of 1.0. Therefore, the closer the AUC value is to 1.0, the better the model. The AUC score of the best-performing model is 0.83 as shown in Figure 5.



Figure 5. The ROC curve and AUC. (**a**) ROC curve and AUC value of the right oblique view, (**b**) ROC curve and AUC value of the left oblique view.

5. Discussion

This paper proposes a novel model to classify the presence or absence of foraminal stenosis, a diagnostic component of cervical radiculopathy, using X-ray images. It also suggests effective methods for preprocessing and augmentation to overcome the challenges arising from the limited number of X-ray images available for training. The accuracy of the best-performing model is approximately 77%. In addition, fine-tuning and transfer learning are suitable when the pre-trained model is used in distinctive domains such as medical datasets. As a preprocessing method, we demonstrate that HE and STN are the most effective methods for X-ray images, as summarized in Table 5. HE increases the contrast between bone and non-bone parts in X-ray images, so the performance of the model is improved. STN learns the spatial features of the slope of the cervical spine and makes the slope, which varies for each patient, align to reduce the geometric invariance of the input dataset for the CNN-based model and improve the performance of the model. We also suggest that Flip is a suitable method to overcome the lack of cervical spine X-ray data in this study. Flip is a specialized method for cervical spine X-ray data to augment considering the characteristics of the data as shown in Table 5. The proposed model can be a help as a reference to the clinical judgment process for cervical root disease, including determining whether to perform MRI or diagnostic root block. In addition, we expect that the proposed model can contribute to reducing the cost of expensive examinations by detecting foraminal stenosis using X-ray images only rather than MRIs or CTs. It is expected that the proposed model, as an auxiliary tool, will help experts diagnose foraminal stenosis more consistently. While foraminal stenosis is difficult to diagnose by a physician only based on oblique radiograph images, a deep learning model could have detected features not easily recognized by human eyes. Recent clinical studies [39] have also suggested that deep learning models could give feedback to physicians regarding radiograph interpretation and that clinicians could learn from deep learning models. The proposed model can be expected to be further utilized later by increasing accuracy, such as automating the diagnosis of foraminal stenosis. However, a major limitation of this study is the limited amount of the labeled cervical spine X-ray data. Therefore, it is necessary to conduct further research in order to overcome this limitation and improve the performance of the proposed model by applying a more effective attention module [40] in the future. While we applied Flip and STN to double the amount of data and to align the slope of the cervical spine, future research can conduct data augmentation using other methods to further augment the dataset by considering various angles of the cervical spine. Afterward, we plan to compare which method is more suitable for the cervical spine X-ray image dataset from the perspective of improving the performance of the foraminal stenosis classification model. Furthermore, the self-supervised learning method [32,41–44] can be applied to unlabeled data to increase the amount of the data. We also plan to apply contrastive learning [45] to

improve the metrics of the proposed model in order to strengthen the classification of the normal class.

Author Contributions: Conceptualization, J.P., S.P. and J.K.; Methodology, J.P., S.P. and J.K.; Software, J.P.; Validation, J.P., S.P. and J.K.; Formal analysis, J.P., J.Y., S.P. and J.K.; Investigation, J.P., S.P. and J.K.; Resources, J.P., S.P. and J.K.; Data curation, J.P., J.Y., S.P. and J.K.; Writing—original draft, J.P.; Writing—review & editing, J.P., S.P. and J.K.; Visualization, J.P.; Supervision, S.P. and J.K.; Project administration, S.P. and J.K.; Funding acquisition, S.P. and J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01789) (50%), and under the High-Potential Individuals Global Training Program (RS-2022-00155054) (50%), supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Institutional Review Board Statement: Informed consent was waived due to the retrospective nature of the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of DUIH 2022-04-025-001.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: We thank Hyeryung Jang, from the Department of Artificial Intelligence, Dongguk University and Hyungjin Chang, from the School of Computer Science, University of Birmingham for their helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kang, K.C.; Lee, H.S.; Lee, J.H. Cervical radiculopathy focus on characteristics and differential diagnosis. *Asian Spine J.* 2020, 14, 921. [CrossRef] [PubMed]
- Brown, B.M.; Schwartz, R.H.; Frank, E.; Blank, N.K. Preoperative evaluation of cervical radiculopathy and myelopathy by surface-coil MR imaging. *Am. J. Neuroradiol.* 1988, 9, 859–866. [CrossRef] [PubMed]
- Luo, L.; Chen, H.; Zhou, Y.; Lin, H.; Heng, P.A. Oxnet: Deep omni-supervised thoracic disease detection from chest X-rays. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 537–548.
- Bozorgtabar, B.; Mahapatra, D.; Vray, G.; Thiran, J.P. Salad: Self-supervised aggregation learning for anomaly detection on X-rays. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 468–478.
- Haghighi, F.; Hosseinzadeh Taher, M.R.; Zhou, Z.; Gotway, M.B.; Liang, J. Learning semantics-enriched representation via selfdiscovery, self-classification, and self-restoration. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 137–147.
- Kuang, X.; Cheung, J.P.Y.; Ding, X.; Zhang, T. SpineGEM: A Hybrid-Supervised Model Generation Strategy Enabling Accurate Spine Disease Classification with a Small Training Dataset. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 145–154.
- Sekuboyina, A.; Valentinitsch, A.; Kirschke, J.S.; Menze, B.H. A localisation-segmentation approach for multi-label annotation of lumbar vertebrae using deep nets. *arXiv* 2017, arXiv:1703.04347.
- Hallinan, J.T.P.D.; Zhu, L.; Yang, K.; Makmur, A.; Algazwi, D.A.R.; Thian, Y.L.; Lau, S.; Choo, Y.S.; Eide, S.E.; Yap, Q.V.; et al. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology* 2021, 300, 130–138. [CrossRef]
- Al-kubaisi, A.; Khamiss, N.N. A Transfer Learning Approach for Lumbar Spine Disc State Classification. *Electronics* 2021, 11, 85. [CrossRef]
- Jocher, G.; Stoken, A.; Borovec, J.; Chaurasia, A.; Changyu, L.; Laughing, A.; Hogan, A.; Hajek, J.; Diaconu, L.; Marc, Y.; et al. ultralytics/yolov5: V5. 0-YOLOv5-P6 1280 Models AWS Supervise. ly and YouTube Integrations. *Zenodo* 2021. Available online: https://zenodo.org/record/4679653#.Y6qxsRVByHs (accessed on 1 October).
- 11. Giełczyk, A.; Marciniak, A.; Tarczewska, M.; Lutowski, Z. Pre-processing methods in chest X-ray image classification. *PLoS ONE* **2022**, *17*, e0265949. [CrossRef]
- 12. Caseneuve, G.; Valova, I.; LeBlanc, N.; Thibodeau, M. Chest X-Ray Image Preprocessing for Disease Classification. *Procedia Comput. Sci.* **2021**, 192, 658–665. [CrossRef]
- 13. O'Shea, K.; Nash, R. An introduction to convolutional neural networks. arXiv 2015, arXiv:1511.08458.

- 14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- 15. Pan, S.J.; Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 2009, 22, 1345–1359. [CrossRef]
- Fairbank, J.; Jamaludin, A.; Lootus, M.; Kadir, T.; Zisserman, A.; Urban, J.; Battié, M.; McCall, I.; The Genodisc Consortium. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRI's) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur. Spine J.* 2017, 26, 1374–1383. [CrossRef]
- 17. Won, D.; Lee, H.J.; Lee, S.J.; Park, S.H. Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. *Spine* **2020**, *45*, 804–812. [CrossRef] [PubMed]
- Dong, N.; Kampffmeyer, M.; Liang, X.; Wang, Z.; Dai, W.; Xing, E. Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 544–552.
- 19. Saiz, F.; Barandiaran, I. COVID-19 detection in chest X-ray images using a deep learning approach. *Int. J. Interact. Multimed. Artif. Intell.* 2020, *in press.* [CrossRef]
- 20. Brunese, L.; Mercaldo, F.; Reginelli, A.; Santone, A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Comput. Methods Programs Biomed.* **2020**, *196*, 105608. [CrossRef] [PubMed]
- Al-Kafri, A.S.; Sudirman, S.; Hussain, A.; Al-Jumeily, D.; Natalia, F.; Meidia, H.; Afriliana, N.; Al-Rashdan, W.; Bashtawi, M.; Al-Jumaily, M. Boundary delineation of MRI images for lumbar spinal stenosis detection through semantic segmentation using deep neural networks. *IEEE Access* 2019, *7*, 43487–43501. [CrossRef]
- Fan, G.; Liu, H.; Wang, D.; Feng, C.; Li, Y.; Yin, B.; Zhou, Z.; Gu, X.; Zhang, H.; Lu, Y.; et al. Deep learning-based lumbosacral reconstruction for difficulty prediction of percutaneous endoscopic transforaminal discectomy at L5/S1 level: A retrospective cohort study. *Int. J. Surg.* 2020, *82*, 162–169. [CrossRef]
- Gaonkar, B.; Villaroman, D.; Beckett, J.; Ahn, C.; Attiah, M.; Babayan, D.; Villablanca, J.; Salamon, N.; Bui, A.; Macyszyn, L. Quantitative analysis of spinal canal areas in the lumbar spine: An imaging informatics and machine learning study. *Am. J. Neuroradiol.* 2019, 40, 1586–1591. [CrossRef]
- Bharati, S.; Podder, P.; Mondal, M.; Prasath, V. CO-ResNet: Optimized ResNet model for COVID-19 diagnosis from X-ray images. Int. J. Hybrid Intell. Syst. 2021, 1–15.. [CrossRef]
- 25. Nayak, S.R.; Nayak, D.R.; Sinha, U.; Arora, V.; Pachori, R.B. Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomed. Signal Process. Control* **2021**, *64*, 102365. [CrossRef]
- Chen, P.; Zhou, Z.; Yu, H.; Chen, K.; Yang, Y. Computerized-Assisted Scoliosis Diagnosis Based on Faster R-CNN and ResNet for the Classification of Spine X-Ray Images. *Comput. Math. Methods Med.* 2022, 2022, 3796202. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 2017, 60, 84–90. [CrossRef]
- 28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
- 29. Liao, H.; Luo, J. A deep multi-task learning approach to skin lesion classification. arXiv 2018, arXiv:1812.03527.
- 30. Chae, J.; Zimmermann, R.; Kim, D.; Kim, J. Attentive Transfer Learning via Self-supervised Learning for Cervical Dysplasia Diagnosis. J. Inf. Process. Syst. 2021, 17, 453–461.
- Qu, B.; Cao, J.; Qian, C.; Wu, J.; Lin, J.; Wang, L.; Ou-Yang, L.; Chen, Y.; Yan, L.; Hong, Q.; et al. Current development and prospects of deep learning in spine image analysis: A literature review. *Quant. Imaging Med. Surg.* 2022, 12, 3454–3479. [CrossRef] [PubMed]
- 32. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13-18 July 2020; pp. 1597–1607.
- Xue, Z.; Rajaraman, S.; Long, R.; Antani, S.; Thoma, G. Gender detection from spine X-ray images using deep learning. In Proceedings of the 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), Karlstad, Sweden, 18–21 June 2018; pp. 54–58.
- Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 2778–2788.
- 35. Kasper-Eulaers, M.; Hahn, N.; Berger, S.; Sebulonsen, T.; Myrland, Ø.; Kummervold, P.E. Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5. *Algorithms* **2021**, *14*, 114. [CrossRef]
- Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* 1987, 39, 355–368. [CrossRef]
- Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* 2015, 28. [CrossRef]
- Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
- 39. Kim, T.; Kim, Y.G.; Park, S.; Lee, J.K.; Lee, C.H.; Hyun, S.J.; Kim, C.H.; Kim, K.J.; Chung, C.K. Diagnostic triage in patients with central lumbar spinal stenosis using a deep learning system of radiographs. *J. Neurosurgery: Spine* **2022**, *1*, 1–8. [CrossRef]

- 40. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
- Noroozi, M.; Favaro, P. Unsupervised learning of visual representations by solving jigsaw puzzles. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 69–84.
- He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.
- Grill, J.B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. Bootstrap your own latent-a new approach to self-supervised learning. *Adv. Neural Inf. Process. Syst.* 2020, 33, 21271–21284.
- 44. Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; Joulin, A. Unsupervised learning of visual features by contrasting cluster assignments. *Adv. Neural Inf. Process. Syst.* 2020, 33, 9912–9924.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. *Adv. Neural Inf. Process. Syst.* 2020, 33, 18661–18673.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.