



End-To-End Computerized Diagnosis of Spondylolisthesis Using Only Lumbar X-rays

Fatih Varçın¹ · Hasan Erbay² · Eyüp Çetin³ · İhsan Çetin⁴ · Turgut Kültür⁵

Received: 24 January 2020 / Revised: 6 October 2020 / Accepted: 18 November 2020 / Published online: 11 January 2021
© Society for Imaging Informatics in Medicine 2021

Abstract

Lumbar spondylolisthesis (LS) is the anterior shift of one of the lower vertebrae about the subjacent vertebrae. There are several symptoms to define LS, and these symptoms are not detected in the early stages of LS. This leads to disease progress further without being identified. Thus, advanced treatment mechanisms are required to implement for diagnosing LS, which is crucial in terms of early diagnosis, rehabilitation, and treatment planning. Herein, a transfer learning-based CNN model is developed that uses only lumbar X-rays. The model was trained with 1922 images, and 187 images were used for validation. Later, the model was tested with 598 images. During training, the model extracts the region of interests (ROIs) via Yolov3, and then the ROIs are split into training and validation sets. Later, the ROIs are fed into the fine-tuned MobileNet CNN to accomplish the training. However, during testing, the images enter the model, and then they are classified as spondylolisthesis or normal. The end-to-end transfer learning-based CNN model reached the test accuracy of 99%, whereas the test sensitivity was 98% and the test specificity 99%. The performance results are encouraging and state that the model can be used in outpatient clinics where any experts are not present.

Keywords Lumbar spondylolisthesis · Convolutional neural networks · Yolo · Transfer learning

Introduction

Lumbar spondylolisthesis, one of the most common spinal diseases among humans, is defined as the anterior shift in one of the lower vertebrae about the subjacent vertebrae. It is identified as the slip progression after the skeletal maturity is connected to disc degeneration at the slip grade. As the biochemical and biomechanical entirety of the disc is lost,

the lumbosacral slip starts to progress and becomes unsteady [1]. It is postulated that the slip progression after the skeletal maturity is in the vast majority of the cases, related to disc degeneration at the slip level. It is recognized as a result of the progression of spondylolysis disease which is often seen at the L5 vertebrae between 85 and 95% and the L4 vertebrae between 5 and 15% [2]. It has been reported that the etiology of the disease and treatment strategy success have had considerable interindividual variability. This fact is conditioned by the clinical variations of the functional spinal unit and deviations in the disease pathogenesis related to mechanical and biochemical processes [3]. Lumbar spondylolisthesis had been classified into six categories in terms of etiology such as dysplastic, isthmic, degenerative, traumatic, pathologic, and iatrogenic [4, 5].

Although there are several symptoms of lumbar spondylolisthesis, most of the time, they are not felt in the early stages. This leads the disease to progress further and, as a result, more demanding treatment is required. Therefore, a reliable and robust automatic diagnosis of spondylolisthesis is important in terms of early diagnosis, rehabilitation, and treatment planning.

✉ Fatih Varçın
fatihvarcin@kku.edu.tr

¹ Department of Computer Engineering, Faculty of Engineering, Kırıkkale University, 71451 Kırıkkale, Turkey

² Department of Computer Engineering, Faculty of Engineering, University of Turkish Aeronautical Association, 06790 Ankara, Turkey

³ Department of Neurosurgery, Faculty of Medicine, Van Yüzüncü Yıl University, 65080 Van, Turkey

⁴ Department of Medical Biochemistry, Faculty of Medicine, Hitit University, 19040 Corum, Turkey

⁵ Department of Physical Medicine and Rehabilitation, Faculty of Medicine, Kırıkkale University, 71450 Kırıkkale, Turkey

Spondylolisthesis is divided into 5 severity stages according to Meyerding Grading System [6]. The system is based on the percentage of slippage of one vertebra relative to the caudal such as grade-1 (0-25%), grade-2 (26-50%), grade-3 (51-75%), grade-4 (76-99%), and grade-5 (99% or more—spondyloptosis). As one expects, the diagnosis of lumbar spondylolisthesis becomes easier as the percentage of slippage increases.

Computerized methods are utilized in many various fields. Aggarwal et al. [7] proposed a method for detecting and recognizing the vehicle license plates. Kumar et al. [8] proposed a model to identify image forgery. Kumar et al. [9] studied to detect skin cancer. The determination of the stage of spondylolisthesis is no exception. It has been done manually for a long time, but there have been some recent studies to automate the process. To the best of our knowledge, the first study with this purpose was presented by Lie et al. in 2016 [10]. Their study consists of two main stages such that automatic spine labeling and critical region detection. In the first stage, a learning-based spine labeling method is used to detect vertebrae by integrating both the image appearance and lumbar spine geometry information. Then, in the second stage, a hierarchical learning method based on the previous studies by Zhan et al. [11, 12] is proposed to determine the spondylolisthesis grade. Another study with the same aim belongs to Cai et al. in 2017 [13]. They developed a technique that is directly capable of spondylolisthesis detection and measurement for lumbar spine images. Their detection method is based on the set of the learning-based detector, while the classification method is k-nearest neighbor with a synthesized image dictionary. Besides detection, there are studies for the treatment such as Liu et al. [14] developed a new type of hollow cement-augmented pedicle screw for the treatment of lumbar spondylolisthesis with osteoporosis.

On the other hand, convolutional neural networks (CNNs) have reached state-of-the-art performances in many diverse fields including medical imaging, natural language processing, and hyperspectral image processing since their first introduction in 2012 [15]. The main reasons behind their success in image analysis tasks are the creation of large-scale annotated datasets like ImageNet [17], improvements in graphics processing unit technologies (GPUs) [16], and the ability to build complicated CNN architectures [18, 19].

Medical image analysis is also one of the most frequently used areas of CNNs. Studies in this field often use magnetic resonance imaging (MRI) scans, computed tomography (CT) images, ultrasound images, and X-ray images for different medical tasks, for example, male pelvic organ segmentation [20] and detection and

segmentation of pulmonary nodules [21] in CT dataset, hippocampus analysis in MR images [22], and white matter hyperintensities segmentation in MRI scans [23]. In addition, CNNs are used for diagnosing lung cancer in endobronchial ultrasound images [24], automated detection and classification of thyroid nodules in ultrasound images [25]. There are a lot of studies that used X-ray images such as automated detection and localization of pneumonia on chest X-ray images [26], detection of rheumatoid arthritis on hand radiographs [27], and coronary artery segmentation in X-ray angiograms [28]. Moreover, Goyal et al. [29] developed a CNN-based mobile application to detect skin cancer using a public dataset.

After the introduction of CNNs, different architectures based on CNNs have been developed to detect diverse target problems such as object detection algorithms, for example, R-CNN [30], Fast-RCNN [31], Faster-RCNN [32], and Yolo [33] are the most well-known object detection algorithms and used for many different tasks [34–37]. Among these, Yolo is the fastest object recognition algorithm because it uses a single artificial neural network. It also scans the entire image and makes more accurate predictions.

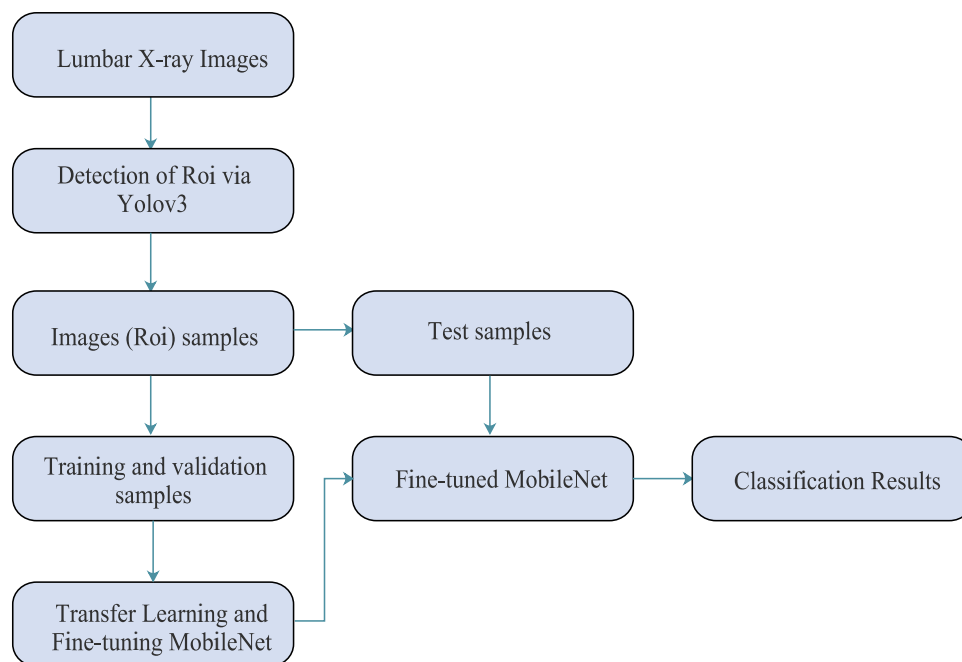
Herein, an end-to-end computerized method to diagnose spondylolisthesis using only lumbar X-ray images is proposed to minimize possible human (manual) errors during both detection and treatment procedures by helping specialists as a second reader. The main steps of the method are presented in Fig. 1. It starts with Yolov3 [38], the last version of Yolo models [33, 39], which is responsible for extracting the region of interest (ROI) in the images. Then, the ROIs enter the fine-tune MobileNet convolutional neural network to complete the diagnosis procedure by classifying as spondylolisthesis or normal.

Materials and Methods

Image Acquisition

Ethical approval certificate for the study was obtained from the Non-interventional Clinical Researches Ethics Board in Hatay Mustafa Kemal University with certificate date September 26, 2019, and record number 04. Some criteria were taken into account while including a patient into the dataset. The patients currently under treatment in the neurosurgery clinic, suffering from degenerative spondylolisthesis, between the ages of 40-70 from both sex groups were included in the dataset. However, the patients suffering from low back pain other than degenerative spondylolisthesis diagnosed with spinal disorder other than degenerative spondylolisthesis and aged out of the range 40-70 were excluded from the study.

Fig. 1 The process of the diagnosis of spondylolisthesis based on Yolov3 and fine-tuned MobileNet



Data Augmentation

As mentioned above, the performance of CNN-based methods depends highly on the amount of data; otherwise, training CNN using a relatively small amount of data might result in convergence and/or over-fitting problems. Most real-world problems usually do not have enough labeled data; especially, it is difficult to get sufficient data in the medicine where specialist annotations are expensive and lesions are scarce. These problems can be overcome by data augmentation techniques and the performance of CNNs is improved. See [40] for details. In this study, a multi-stage offline image augmentation procedure, called augmentor [41], is employed for each image. The augmentor is a software package and typically used in the artificial generation of image data for machine learning problems. It allows us to create an augmentation pipeline, which chains the images together with the operations such as rotate, crop, zoom, etc. The operations are applied stochastically, where the parameters supplied to each of these operations are chosen at random fashion, within a prespecified range. In this way, each time a different image is derived from an image that is passed through the pipeline.

The dataset 1 used in the study contains a total of 2746 lumbar X-ray images after the augmentation procedure. Among them, 600 lumbar X-ray images collected from the Department of Neurosurgery in Hatay Dörtyol State Hospital between January 1, 2016, and August 30, 2019, in which 300 cases were labeled as spondylolisthesis by two specialists and the rest 300 cases labeled as normal.

The sizes of images in the dataset were varying between 900×800 and 950×860 pixels.

The augmentor, in this study, includes zoom, flip, and brightness operations. Table 1 shows the augmented data and its splitting. Observe that the original dataset grew after the augmentation process.

Here is the right place to mention that the training and validation data are used during the training of the model, but the test data are used to measure the trained model's capability of detecting spondylolisthesis.

Yolo

ROI is a portion of an image for a specific purpose, and extracting the ROI is considered as an object detection task. In this study, ROI was the portion of the image containing L4, L5, and S1, and each ROI automatically detected through Yolov3, the latest version of Yolo.

Yolo family is one of the most powerful and fastest state-of-art deep learning object detection algorithms. It is a real-time object detection system (45 frames per second) and based

Table 1 The distribution of normal and lumbar spondylolisthesis cases in training, validating, and testing groups of dataset 1

Cases	Training	Validation	Test
Normal	972	101	334
Lumbar spondylolisthesis	950	86	303
Total	1922	187	637

on single CNN which is called DarkNet. DarkNet can directly predict the confidence target box and the class at the same time from the whole image. This makes Yolo more successful than its competitors in generalizing the representation of objects. Working principle of Yolo can be explained roughly as follows:

1. The input image is divided into a grid of size $S \times S$. If the center of the target falls into a grid cell, the grid cell is going to detect the object. Each grid cell predicts B bounding boxes and the confidence of these detection boxes. The total number of bounding boxes for each input image is $S \times S \times B$. The confidence score can also be explained as an indication of whether there are any objects in the bounding box.
2. For a number of C classes, there $C + 5$ values are predicted for each bounding box. The predicted values are the coordinates x, y of the midpoint of the bounding box, the width and height w, h of the bounding box, the confidence score Conf calculated by the formula 1, and C many class probabilities.

$$\text{Conf} = \text{Pr}(\text{Object}) \times \text{IOU}_{\text{predicted}}^{\text{ground truth}} \quad (1)$$

The intersection over union (IOU) is calculated by the predicted and the ground truth bounding boxes. If the grid cell doesn't contain any object, the confidence score must be zero, otherwise, is equal to IOU [33].

3. In addition to previous steps, conditional class probabilities $\text{Pr}(\text{Class}_i | \text{Object})$ are predicted for each bounding box.
4. The class-specific confidence scores for each bounding box are calculated by multiplying the

confidence predictions and the conditional class probabilities at the test time of Yolo (see formula 2). Finally, an $S \times S \times B \times [x, y, w, h, \text{Conf}, C]$ tensor is obtained for each the input image.

$$\begin{aligned} & \text{Pr}(\text{Class}_i | \text{Object}) \times \text{Pr}(\text{Object}) \times \text{IOU}_{\text{predicted}}^{\text{ground truth}} \\ & = \text{Pr}(\text{Class}_i) \times \text{IOU}_{\text{predicted}}^{\text{ground truth}} \end{aligned} \quad (2)$$

As mentioned above, this study utilizes Yolov3 for detecting the L4, L5, and S1 as illustrated in Fig 2. All images in the original dataset were resized to the 400×600 resolution before fed into Yolov3. Yolov3 is an improved version of Yolo in terms of a few additional methods that used in the downsampling of feature maps and predicting confidence scores of the bounding boxes and class confidence probabilities. Furthermore, Yolov3 uses a more strong feature extractor network than DarkNet which is called DarkNet-53 [38].

Transfer Learning and Fine-Tuning

Training of CNNs using medical images can be generally done in three different ways. The first one is to start with a scratch network and train it. It requires a well-annotated and well-distributed large-scale dataset. However, building such a large-scale medical image dataset is very troublesome because of the cost of the expert annotation and difficulty of data acquisition. In addition, starting to train CNN from scratch is a time-consuming and costly process in terms of memory and computing resources. Apart from these, successfully developing a deep CNN architecture from scratch is very

Fig. 2 Example of detecting L4, L5, and S1 from a lumbar X-ray image using Yolov3

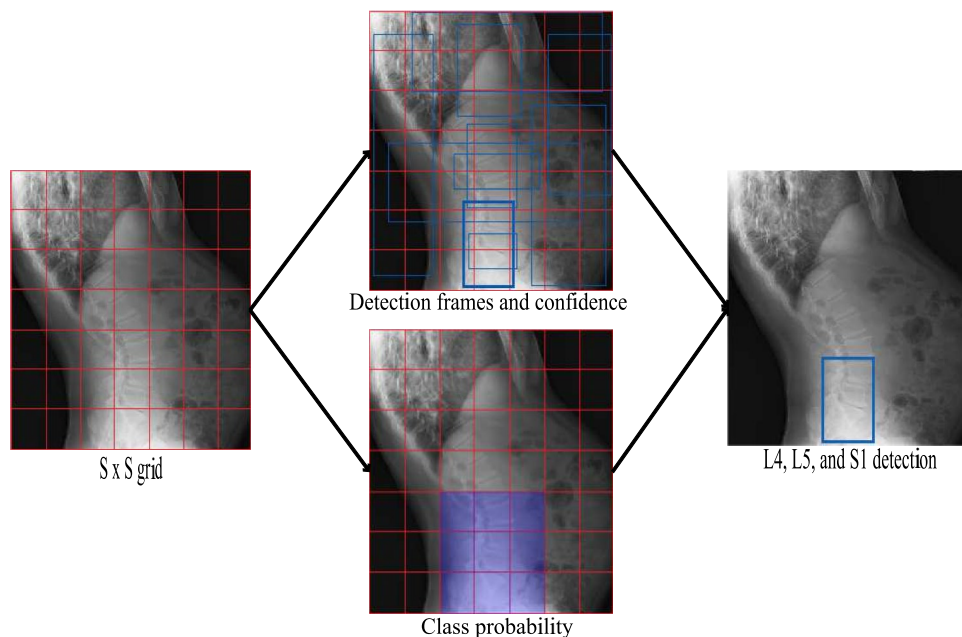


Table 2 Hyperparameters and frozen weights

	MobileNet
Optimizer	Adam
Momentum	0.900
Initial Learn Rate	$1.0000e - 03$
Validation Patience	10
Maximum Epochs	25
Mini Batch Size	32
Shuffle	Every Epoch
Frozen weights (layers)	1:16

laborious and requires profound expertise due to technical details [18]. The second technique is to utilize the “off-the-shelf” CNN features as input to a machine learning classifier. Early layers of the CNNs retain simple features while later layers include more complex features of the images. So, the “off-the-shelf” CNN features are high-level convolution features obtained before fully connected layers [42, 43]. The third and promising alternative approach is to employ a CNN that has been trained on a large dataset on a related task, in short, to use transfer learning. It relies on the idea of utilizing the knowledge acquired for one task to solve related ones, as humans do. However, the pre-trained model must be modified to fit the needs of the new task. First, the original classifier is replaced by the new classifier that fits the new task’s purpose. Then, it must be fine-tuned. Since the features at lower layers end up with general features, the strategy followed is to keep the problem-independent lower layers frozen, just fine-tune the problem-dependent high layers. Here, the weights (except weights in the last fully connected layer) are transferred to a different network that has the same architecture as the pre-trained network. This process can be considered as a starting point for a new model. Then, the last fully connected layer is replaced with a new fully connected layer that contains as many nodes as the number of classes in the target dataset. In

Table 3 The distribution of normal and lumbar spondylolisthesis cases in training, validating, and testing groups of dataset 2

Cases	Training	Validation	Test
Normal	972	102	318
Lumbar spondylolisthesis	950	85	280
Total	1922	187	598

addition, processes such as changing some parameters (e.g., learning rate, momentum) and freezing weights of the first few layers may be applied for better training performance. All changes applied after copying of weights are called “fine-tuning” and have been used successfully in many studies in different tasks [44].

Fine-Tuning MobileNet

The initial MobileNet with 30 layers was presented by Howard et al. [45] in 2017 and trained on ImageNet. The parameters of MobileNet were transferred, and MobileNet was fine-tuned through training images in the dataset 2 (Table 3). The last layer of MobileNet was removed, and the first 16 convolution layers were frozen due to having high saturation. (It was decided which layers should be frozen after a few experiments.) First, a global average pooling layer was added and we also used a batch normalization and a dropout ratio of 0.7 to speed up training and improve success. We then used a dense layer ($1 \times 1 \times 512$) with ReLU and a dropout ratio of 0.5. Finally, a new softmax classifier with two outputs (normal and spondylolisthesis) employed as the classifier of fine-tuned MobileNet. Additionally, fine-tuning the model was done using the Adam optimization algorithm, categorical cross-entropy, batch size of 32, and an initial learning rate of 0.001 for 25 epochs. Table 2 shows the hyperparameters along with frozen weights. Fine-tuned MobileNet is shown in Fig. 3.

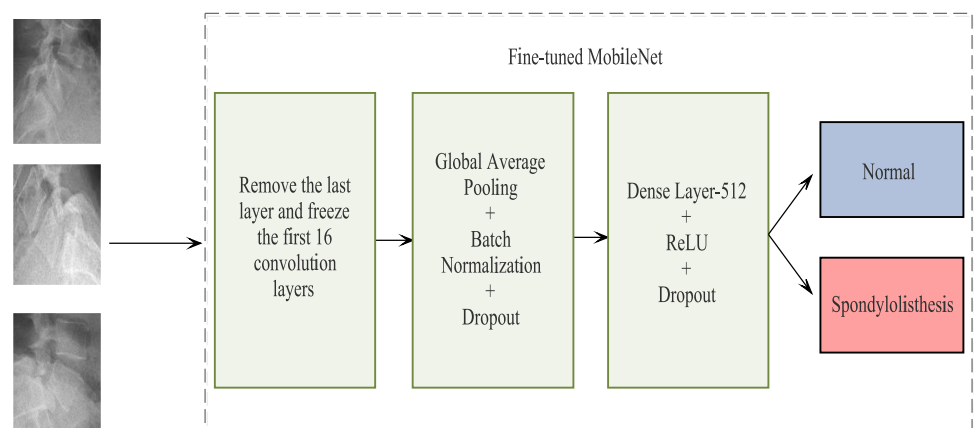
Fig. 3 Fine-tuned MobileNet

Table 4 Performance statistics

Measure	Value
Accuracy	0.99
Sensitivity	0.98
Specificity	0.99
AUC	0.99
Precision	0.99
F1 Score	0.98

Experimental Results

Performance evaluation of the proposed diagnosis pipeline was carried out in two stages. In the first step, the performance of the re-trained Yolov3 model was evaluated using the Intersection over Union (IoU) metric. If the IoU rate of the region detected by Yolov3 was greater than 70%, the detection would be considered successful. The IoU rate of applying Yolov3 to the testing images (a region was detected in 598 images) in dataset 1 is 96%. The number of images on which Yolo failed to detect was 39, and these images weren't included in calculating the IoU rate. Figure 8 shows a few examples of detected and undetected images.

In the second step, the classification performance of fine-tuned MobileNet was analyzed on dataset 2 in terms of

different metrics such as accuracy, sensitivity, specificity, area under curve (AUC), etc. (Table 4). These metrics were obtained through the confusion matrix and receiver operating characteristic (ROC) curve shown in Fig. 4 and 7, respectively. In addition, plots of accuracy and loss over 25 epochs for the fine-tuned MobileNet are illustrated in Figs. 5 and 6.

Discussion

Herein, an automatic computerized method to diagnose spondylolisthesis using lumbar X-ray images is proposed to reduce possible human errors during both detection and training procedures by helping specialists as a second reader.

It is perceived as a result of the progression of spondylolysis disease which occurs at the L5 vertebrae between 85 and 95% and the L4 vertebrae between 5 and 15%. Lumbar spondylolisthesis was classified into five categories in terms of etiology by Wiltse et al. in 1976 [4], later, the categories were updated as dysplastic, isthmic, degenerative, traumatic, pathologic, and iatrogenic in 1995. There are several symptoms to identify lumbar spondylolisthesis, and most of the time, these symptoms are not felt in the early stages. This leads to disease progress further; thus, more challenging

Fig. 4 Confusion matrix

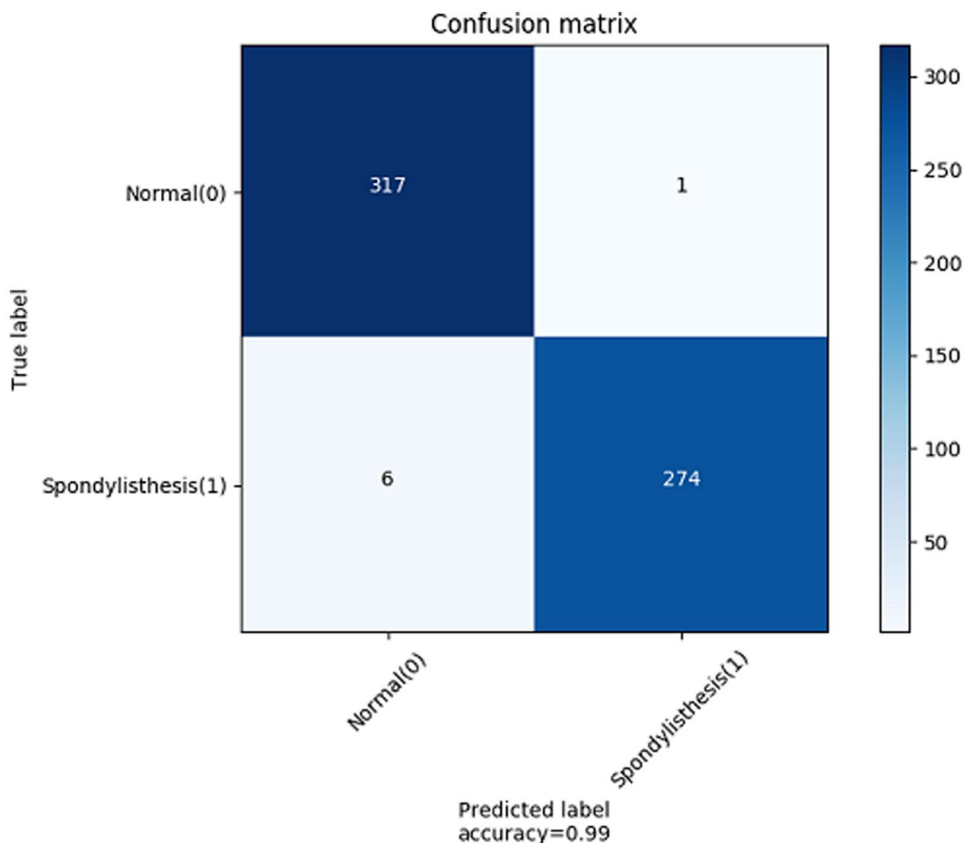
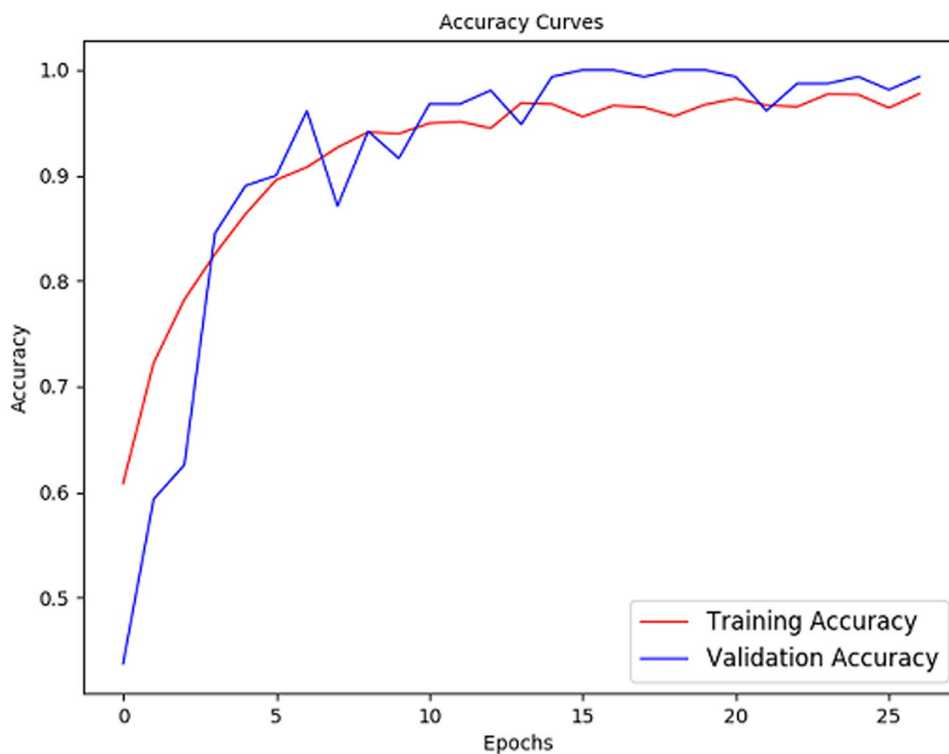


Fig. 5 Accuracy of the fine-tuned MobileNet



treatment is required. Therefore, a reliable and robust automatic method for diagnosing spondylolisthesis is important in terms of early diagnosis, rehabilitation, and treatment planning.

A diagnosis pipeline based on deep learning for lumbar spondylolisthesis is proposed in this paper. Firstly, Yolov3 is used to detect ROIs in lumbar X-rays and then fine-tuned MobileNet classifies X-rays as spondylolisthesis or normal.

Fig. 6 Loss of the fine-tuned MobileNet

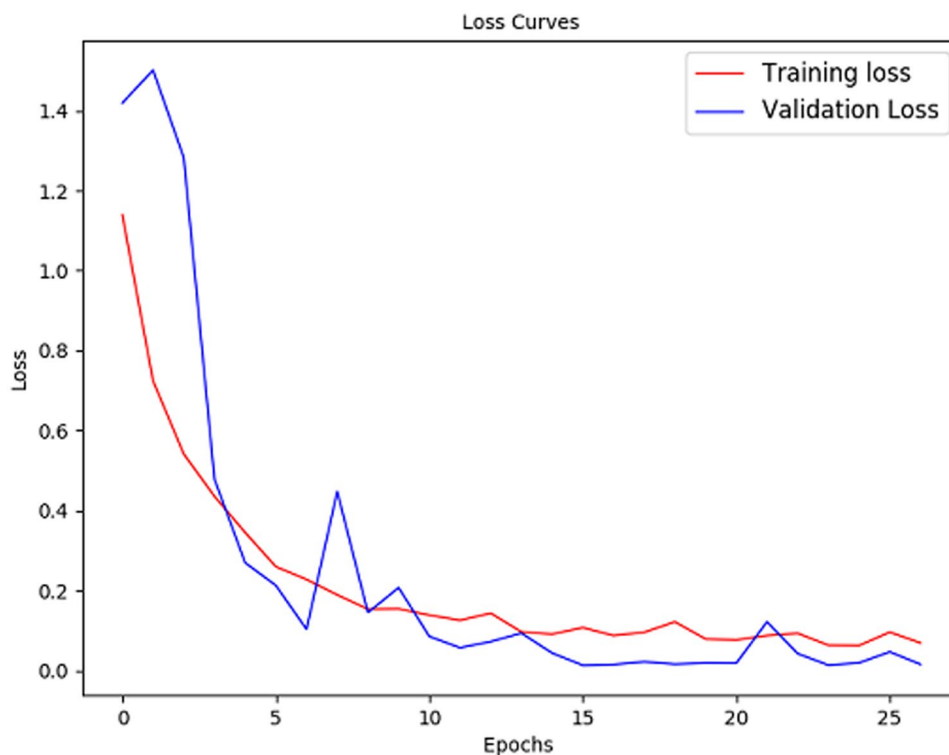


Fig. 7 The ROC curve of classifying test samples from dataset

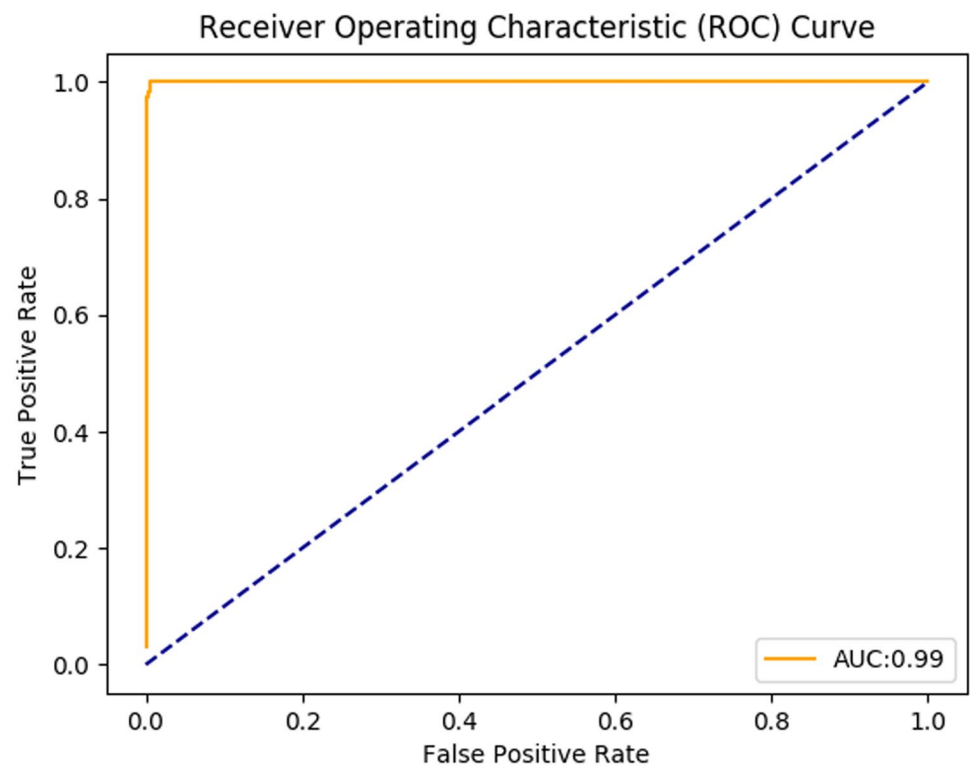
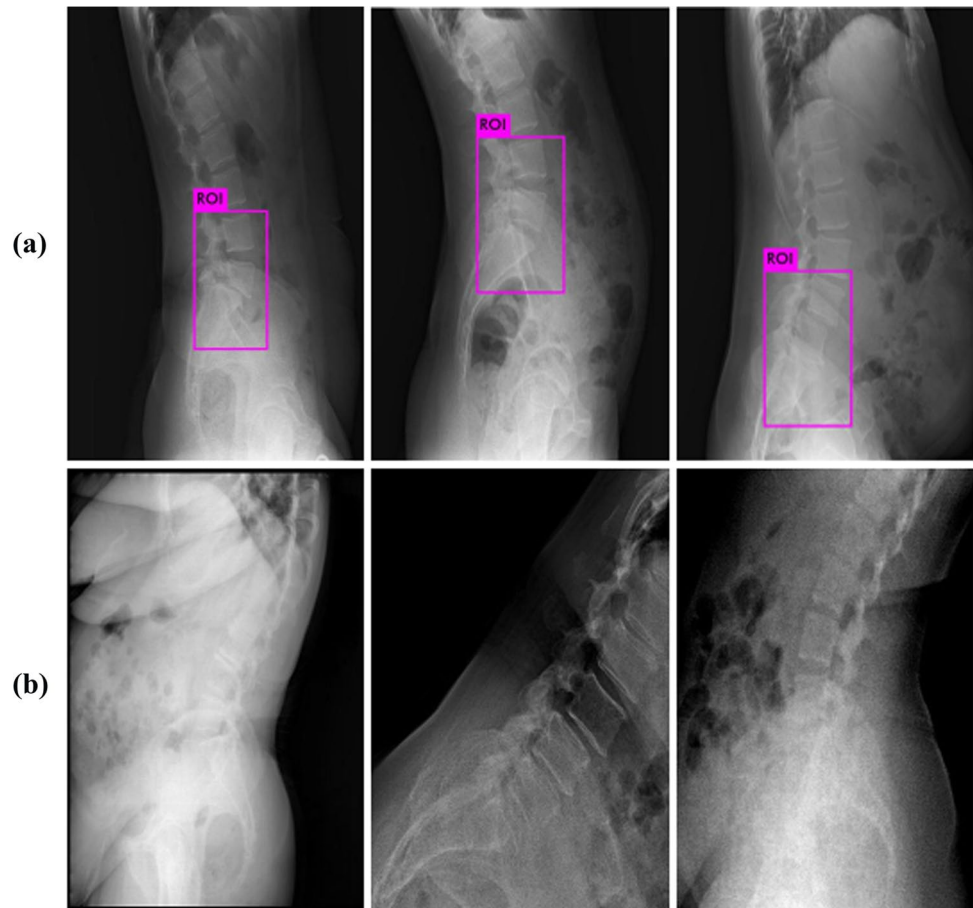


Fig. 8 Detected and undetected images



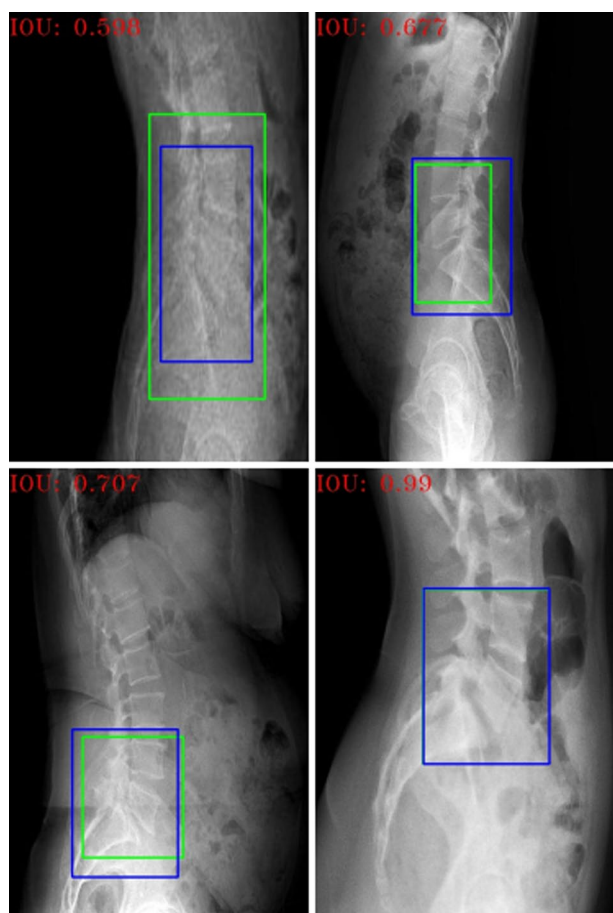


Fig. 9 IOU values of different detections (blue box: ground truth, green box: predicted)

Deep learning-based methods need a huge amount of image dataset for different tasks such as classification, segmentation, and detection [46]. However, it is not easy to build a medical dataset because of the difficulties in obtaining large numbers of quality images. Therefore, we applied a pipeline including different data augmentation techniques to the images in the local dataset and increased the number of images as much as possible. Then, the regions including the L4, L5 and S1 on the images in the dataset 1 were determined using the blue colored boxes by an expert.

When evaluating the performance of Yolov3, detections with an IoU rate of less than 70% were considered unsuccessful. However, as shown in Fig. 9, it is actually very successful in detections with an IoU rate of less than 70%. These detections include the desired vertebrae, but the predicted boxes are larger or smaller than the ground truths. Given that two experts can draw ground truths of different sizes for the same picture, the threshold rate (70%) can be further reduced.

The ROIs detected by Yolov3 are used to re-training and fine-tuning MobileNet. MobileNet is particularly an efficient model for mobile and embedded vision applications. It is based on depth-wise and point-wise convolutions and remarkable reduces the computational model parameters. The most important reason for using MobileNet in this study is the low number of parameters that the model needs to be trained. In this way, although the number of images in the dataset is small compared to large-scale datasets such as ImageNet, it is sufficient for proper convergence of the parameters in the network.

This study aimed to detect whether the patient had spondylolisthesis using lumbar X-ray images. Our study is important because it provides an encouraging start to computerized diagnosis systems that can assist doctors regarding lumbar pathologies. Moreover, with the existence of such a system, early detection of lumbar pathologies becomes possible, as a result, conservative treatment can be applied without any surgical treatment. This serves as the patient's life comfort as well as prevents the patient from possible complications of surgery. Furthermore, the cumulative of similar studies leads to the starting point of clinical usage.

Conclusion

In summary, we proposed a pipeline utilizing two deep learning methods for diagnosis of spondylolisthesis in X-ray spin images. It mainly consists of two main stages: detecting the region of interest containing L4, L5 and S1 via Yolov3 model and classifying obtained sub-images thanks to the pre-trained and fine-tuned MobileNet. The performance of the proposed system was evaluated in two steps using our local dataset. While the performance of Yolov3 was evaluated according to the IOU metric, MobileNet's performance was analyzed according to known deep learning metrics. According to the findings obtained from this study, our method is very encouraging and it can be used in outpatient clinics where no experts are available.

References

1. Floman Y: Progression of lumbosacral isthmic spondylolisthesis in adults. *Spine* 25(3):342–347,2000
2. Gagnet P, Kern K, Andrews K, Elgafy H, Ebraheim N, et al: Spondylolysis and spondylolisthesis: A review of the literature. *J Orthop* 15(2):404–407,2018
3. Sutovsky J, Sutovska M, Kocmalova M, Kazimierova I, Pappova L, Benco M, Grendar M, Bredvold HH, Miklusica J, Franova S, et al: Degenerative lumbar spondylolisthesis. Biochemical aspects and evaluation of stabilization surgery extent in terms of adjacent segment disease theory. *World Neurosurg* 121:554–565,2019
4. Wiltse LL, Newman PH, Macnab I, et al : Classification of spondyloysis and spondylolisthesis. *Clinical Orthopaedics and Related Research* (1976-2007) 117:23–29,1976

5. Lasanianos NG, Kanakaris NK, Giannoudis PV, et al: Trauma and orthopaedic classifications: a comprehensive overview. Springer 2014
6. H.W. MEYERDING: Low backache and sciatic pain associated with spondylolisthesis and protruded intervertebral disc: incidence, significance, and treatment. *JBJS* 23(2), 461–470,1941
7. Aggarwal A, Rani A, Kumar M, et al: A robust method to authenticate car license plates using segmentation and roi based approach. *Smart and Sustainable Built Environment*, 2019
8. Kumar M, Srivastava S, Uddin N: Forgery detection using multiple light sources for synthetic images. *Aust J Forensic Sci*, 51(3):243–250,2019
9. Kumar M, Alshehri M, AlGhamdi R, Sharma P, Deep V: A de-ann inspired skin cancer detection approach using fuzzy c-means clustering. *Mob Netw Appl* 25:1319–1329,2020
10. Liao S, Zhan Y, Dong Z, Yan R, Gong L, Zhou XS, Salganicoff M, Fei J, et al: Automatic lumbar spondylolisthesis measurement in ct images. *IEEE Trans Med Imaging* 35(7):1658–1669,2016
11. Zhan Y, Dewan M, Harder M, Krishnan A, Zhou XS, et al: Robust automatic knee mr slice positioning through redundant and hierarchical anatomy detection. *IEEE Trans Med Imaging* 30(12):2087–2100,2011
12. Zhan Y, Dewan M, Harder M, Zhou XS, et al: Robust mr spine detection using hierarchical learning and local articulated model. In *International conference on medical image computing and computer-assisted intervention*, Springer, 2012, pp 141–148.
13. Cai Y, Leung S, Warrington J, Pandey S, Shmuilovich O, Li S, et al: Direct spondylolisthesis identification and measurement in mr/ct using detectors trained by articulated parameterized spine model. In *Medical Imaging 2017: Image Processing*, volume 10133, page 1013319. *International Society for Optics and Photonics*, 2017
14. Liu Y-Y, Xiao J, Yin X, Liu M-Y, Zhao J-H, Liu P, Dai F, et al: Clinical efficacy of bone cement-injectable cannulated pedicle screw short segment fixation for lumbar spondylolisthesis with osteoporosis. *Sci Rep* 10(1):1–9,2020
15. Zhao G, Liu G, Fang L, Tu B, Ghamisi P, et al: Multiple convolutional layers fusion framework for hyperspectral image classification. *Neurocomputing* 339:149–160,2019
16. LeCun Y, Bengio Y, Hinton G, et al: Deep learning. *Nature*, 521(7553):436–444,2015
17. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L, et al: In 2009 IEEE conference on computer vision and pattern recognition 2009, pp 248–255
18. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J, et al: Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans Med Imaging* 35(5):1299–1312,2016
19. Krizhevsky A, Sutskever I, Hinton GE, et al: Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*, 2012, pp 1097–1105
20. Wang S, He K, Nie K, Zhou S, Gao Y, Shen D, et al: Ct male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. *Med Image Anal* 54:168–178,2019
21. Huang X, Sun W, Tseng T-LB, Li C, Qian W, et al: Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic ct scans using deep convolutional neural networks. *Comput Med Imaging Graph* 74:25–36,2019
22. Li F, Liu M, Alzheimer's Disease Neuroimaging Initiative, et al: A hybrid convolutional and recurrent neural network for hippocampus analysis in alzheimer's disease. *J Neurosci Methods* 323:108–118,2019
23. Li H, Jiang G, Zhang J, Wang R, Wang Z, Zheng W-S, Menze B, et al: Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images. *NeuroImage* 183:650–665,2018
24. Chen C-H, Lee Y-W, Huang Y-S, Lan W-R, Chang R-F, Tu C-Y, Chen C-Y, Liao W-C, et al: Computer-aided diagnosis of endobronchial ultrasound images using convolutional neural network. *Comput Meth Prog Bio* 177:175–182,2019
25. Liu T, Guo Q, Lian C, Ren X, Liang S, Yu J, Niu L, Sun W, Shen D, et al: Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks. *Med Image Anal* 58:101555,2019
26. Hu G, Yang X, Zhang Y, Wan M, et al: Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems* 2019, p 100353
27. Üreten K, Erbay H, Maraş HH, et al: Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network. *Clinical rheumatology* 39(4):969–974,2020
28. Fan J, Yang J, Wang Y, Yang S, Ai D, Huang Y, Song H, Hao A, Wang Y, et al: Multichannel fully convolutional network for coronary artery segmentation in x-ray angiograms. *IEEE Access* 6:44635–44643,2018
29. Goyal V, Singh G, Tiwari O, Punia S, Kumar M, et al: Intelligent skin cancer detection mobile application using convolution neural network. *Advanced Research in Dynamical and Control Systems (JARCDS, IASR)* 11(7(SI)):253–259,2019
30. Girshick R, Donahue J, Darrell T, Malik J, et al: Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 2014, pp 580–587
31. Girshick R. Fast r-cnn: In *Proceedings of the IEEE international conference on computer vision* 2015, pp 1440–1448
32. Ren S, He K, Girshick R, Sun J, et al: Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans Patt Anal Mach Intell* 39(6):1137–1149,2016
33. Redmon J, Divvala S, Girshick R, Farhadi A, et al: You only look once: Unified, real-time object detection. *arXiv preprint arXiv:1506.02640*, 2015
34. Cai Z, Vasconcelos N: In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp 6154–6162
35. Li J, Liang X, Shen S, Xu T, Feng J, Yan S, et al: Scale-aware fast r-cnn for pedestrian detection. *IEEE Trans Multimedia* 20(4):985–996,2017
36. Jiang H, Learned-Miller E: In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp 650–657
37. Lan W, Dang J, Wang Y, Wang S, et al: In *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2018, pp 1547–1551
38. Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018
39. Redmon J, Farhadi A. Yolo9000: Better, faster, stronger. *arXiv preprint arXiv:1612.08242*, 2017
40. Shorten C, Khoshgoftaar TM: A survey on image data augmentation for deep learning. *J Big Data* 6(1):60,2019
41. Bloice MD, Roth PM, Holzinger A, et al: Biomedical image augmentation using augmentor. *Bioinformatics* 35(21):4522–4524,2019
42. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M, et al: Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging* 30(4):477–486,2017
43. Nguyen K, Fookes C, Ross A, Sridharan S, et al: Iris recognition with off-the-shelf cnn features: A deep learning perspective. *IEEE Access* 6:18848–18855,2017

44. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM, et al: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35(5):1285–1298,2016
45. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H, et al: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861), 2017
46. Soekhoe D, Der Putten PV, Plaat A, et al: On the impact of data set size in transfer learning using deep neural networks. In *International Symposium on Intelligent Data Analysis*, Springer, 2016, pp 50–60

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