



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/18942>

To cite this version :

Claudio VERGARI, Wafa SKALLI, Laurent GAJNY - A convolutional neural network to detect scoliosis treatment in radiographs - International Journal of Computer Assisted Radiology and Surgery - Vol. 15, p.1069–1074 - 2020

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



A convolutional neural network to detect scoliosis treatment in radiographs

Claudio Vergari¹ · Wafa Skalli¹ · Laurent Gajny¹

Abstract

Purpose The aim of this work is to propose a classification algorithm to automatically detect treatment for scoliosis (brace, implant or no treatment) in postero-anterior radiographs. Such automatic labelling of radiographs could represent a step towards global automatic radiological analysis.

Methods Seven hundred and ninety-six frontal radiographies of adolescents were collected (84 patients wearing a brace, 325 with a spinal implant and 387 reference images with no treatment). The dataset was augmented to a total of 2096 images. A classification model was built, composed by a forward convolutional neural network (CNN) followed by a discriminant analysis; the output was a probability for a given image to contain a brace, a spinal implant or none. The model was validated with a stratified tenfold cross-validation procedure. Performance was estimated by calculating the average accuracy. **Results** 98.3% of the radiographs were correctly classified as either reference, brace or implant, excluding 2.0% unclassified images. 99.7% of brace radiographs were correctly detected, while most of the errors occurred in the reference group (i.e. 2.1% of reference images were wrongly classified).

Conclusion The proposed classification model, the originality of which is the coupling of a CNN with discriminant analysis, can be used to automatically label radiographs for the presence of scoliosis treatment. This information is usually missing from DICOM metadata, so such method could facilitate the use of large databases. Furthermore, the same model architecture could potentially be applied for other radiograph classifications, such as sex and presence of scoliotic deformity.

Keywords Spine deformity · Brace · Implant · Detection · Machine learning

Introduction

Adolescent idiopathic scoliosis (AIS) is a deformity of the spine which can lead, if left untreated, to respiratory or locomotor impairment and, more in general, to a decreased quality of life. It affects 1–4% of the population, and 42% of patients require treatment by brace, which is the most effective non-operative approach in mild to moderate spinal curves [1]. The aim of bracing is to stop the progression of the deformity, and its long-term effect on the deformed spine is related to its immediate correction [2]. Therefore, in-brace radiographs of the patient are often acquired to evaluate such correction on the three-dimensional deformity of the patient. There is still no consensus on which types of braces are more

effective, or more adapted to correct certain topologies of deformities. This results in a vast variety of brace models, all characterized by different correction mechanisms and radiological appearance (Fig. 1).

This lack of consensus is due to the fact that bracing is only effective in slowing down the progression in about 70% of the cases [3]; surgical intervention is necessary for 18–68% of patients, since their deformity progresses further [4]. The aim of surgery is to straighten and de-rotate the spine, while respecting the physiological curves of the spine in the sagittal plane, and this is usually obtained by fusing those vertebral levels more affected by the deformity. Again, radiological examination is necessary to evaluate the result of the surgery, not only considering the fused spinal segment, but also the compensation mechanisms actuated by the patient to find a new balance.

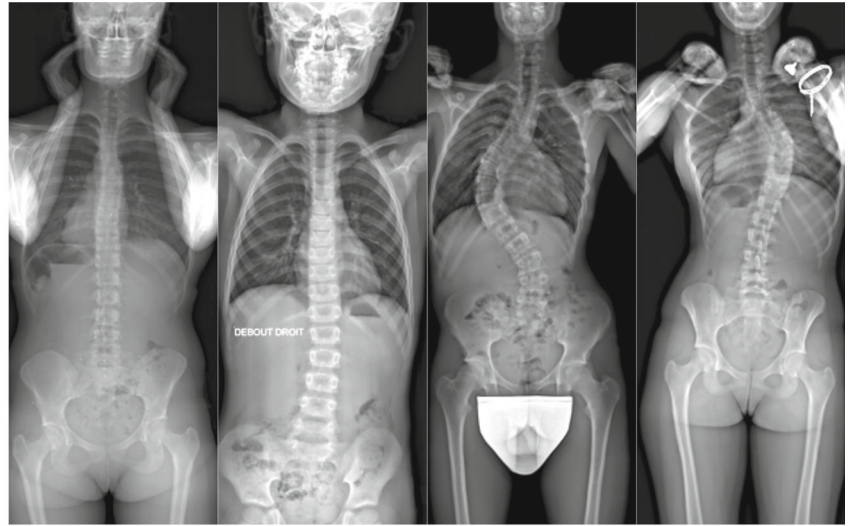
Several imaging analysis methods are being developed for research and clinical use [5]. Machine learning-based methods were described in the literature to detect or seg-

✉ Laurent Gajny
laurent.gajny@ensam.eu

¹ Arts et Métiers, Institut de Biomécanique Humaine Georges Charpak, 151 bd de l'Hôpital, 75013 Paris, France

Fig. 1 Example of frontal radiographies of scoliotic patients, patients wearing a brace and patients with a spinal implant. Radiographies could include superimposed text, radiation covers, jewellery or clothes radiopaque features. The * symbol indicates an image where the bottom was padded by mirroring the bottom portion of the image

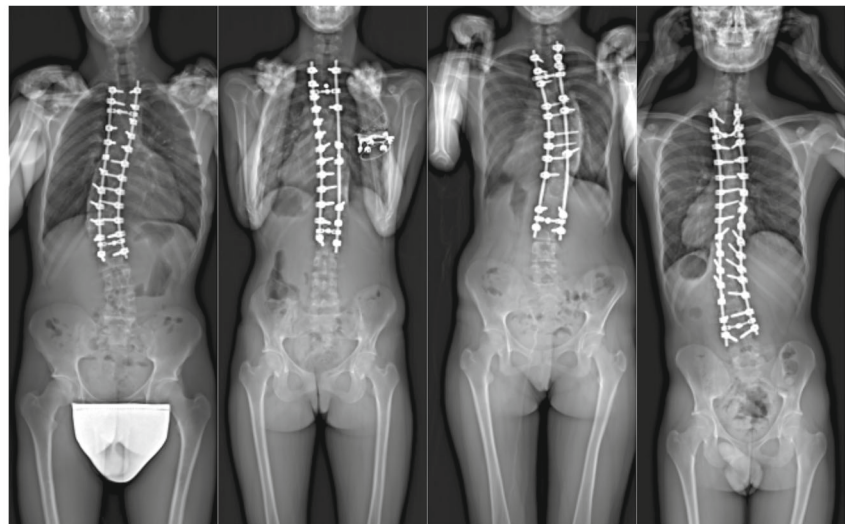
Reference



Brace



Implant



ment anatomical landmarks [2, 3, 6, 7], the full geometry of the spine [8, 9] or specific geometrical parameters, such as Cobb angle [10], lumbar lordosis [11] and sagittal verti-

cal axis [12]. Besides, applications of machine learning are not limited to radiographs but also extend to computerized tomography [13] and magnetic resonance imaging [14, 15].

However, to the best of our knowledge, no work has been done to automatically detect scoliosis treatment (brace, implant or lack thereof) in postero-anterior radiographs. This information is of value either to develop automatic radiographic analysis tools, but it is usually not available in radiographs' DICOM metadata. Therefore, images showing a patient in-brace or after fusion surgery must be manually labelled for further use in image analysis. This can be time-consuming and a potential source of user error which could be limited with automatic labelling through image analysis. In this work, we proposed a classification algorithm based on a CNN to detect scoliosis treatment in postero-anterior radiographs.

Methods

Data

Seven hundred and ninety-six frontal radiographies of adolescents were collected from previous studies. They were extracted from biplanar radiographs obtained thanks to the EOS system (EOS Imaging, Paris, France). These included 84 patients wearing a brace, 325 with a spinal implant and 387 reference images with no treatment; examples are provided in Fig. 1. Patients were scanned while in free-standing position [16]. Images were reduced to a common size of 400×160 pixels: first, the image was resized to reach the target width, then its lower part was cropped when necessary or, if the resulting image was too short, the bottom was padded by mirroring the bottom portion of the image. Finally, image contrast was improved by stretching the greyscale (imposing 1% top and bottom pixel values to be saturated).

The whole dataset was augmented, first by flipping all images. Then, the brace group was further augmented by adding a random vertical translation (between 20 and 60 pixels), enlargement and cropping, and adding noise. Table 1 reports the final dataset size.

Model

The model is composed by a forward CNN, the output of which is further processed by discriminant analysis; the full model is schematized in Fig. 2. The CNN's architecture was inspired by LeNet-5 [17]; the network is composed by three series of convolution, batch normalization and max-pooling layers. The output is fed through a dropout layer [18], a rectified linear unit and then flattened to three values, indicating the input's similarity to each of the three groups. A discriminant analysis was further performed on this output, which allowed assigning three probabilities to the input image, each representing the probability that it belongs to one of the three groups. The image was thus assigned to the group represented

by the highest probability, unless its value was lower than 0.9; in that case, the image was considered as "not classified".

Training was performed by feeding the network batches of 32 images at each iteration. Furthermore, the network was trained in two stages: in the first 25 epochs, only two-thirds of the data were fed to the network, and the full training dataset was used in the following 20 epochs. This significantly reduced overfitting. No improvement was observed after total 45 epochs.

The model was implemented in MATLAB 2017b (The MathWorks, Natick, USA) using the MatConvNet toolbox [19].

Validation

The model's parameters (filters' size and number, layer depths, etc.) were optimized by trial and error in a first phase, in which the dataset was split into a training set (50% of the data), validation (25%) and test set (25%), each containing the same ratio of radiographs from the three groups. Candidate architectures (including the final one) were eventually validated with a stratified tenfold cross-validation procedure, with the group ratios reported in Table 1. The dataset was split into ten sets of similar size, each containing the same ratio of radiographs from the three groups. Patients were assigned to the training or validation set before augmentation, so that augmented radiographs of the same patients could not belong to both groups. Then, the model was run ten times, each time using one of the subsets for testing and the rest of the data for training. Overall performance was estimated by calculating the average accuracy (and 90% confidence intervals).

Results

Table 1 reports the classifier's performance, while Table 2 shows a confusion matrix. Overall, 96.5% of the radiographs were correctly classified as either reference, brace or implant. However, among the remaining images, 2.0% were unclassified; if these are excluded, 98.3% could be considered correctly classified. In particular, 99.7% brace radiographs were correctly classified, while most of the errors were in the reference group (i.e. 2.1% of reference images were wrongly classified).

Discussion

In this work, a classifier was developed to detect orthotic or surgical treatment in radiographs of scoliotic patients. A deep learning framework was adopted to tackle the issue of large variability of braces/instrumentations appearances and similarities with other common objects (buttons, bras,

Table 1 Dataset size and distribution followed by the results of the classifier model

| | Reference | Brace | Implant | Overall |
|---------------------------------|-------------------|------------------|------------------|------------------|
| Original dataset | 387 | 84 | 325 | 796 |
| Dataset after augmentation | 774 | 672 | 650 | 2096 |
| <i>Tenfold cross-validation</i> | | | | |
| Training | 698 | 608 | 586 | 1892 |
| Validation | 76 | 64 | 64 | 204 |
| Correct classification (90% CI) | 97.9% (93.7–97.9) | 97.3 (87.5–97.3) | 99.7 (98.1–99.7) | 98.3 (94.1–98.3) |
| Wrong classification (90% CI) | 2.1% (0–4.2) | 2.7% (0–9.8) | 0.3% (0–1.6) | 1.7 (0–4.2) |
| Not classified (90% CI) | 4.1% (0.7–12.6) | 1.7 (0–5.8) | 0.2 (0–0.8) | 2.0 (0–6.1) |

Correct and wrong accuracies are given as percentages of the total classified values, while not classified are percentages of the total dataset

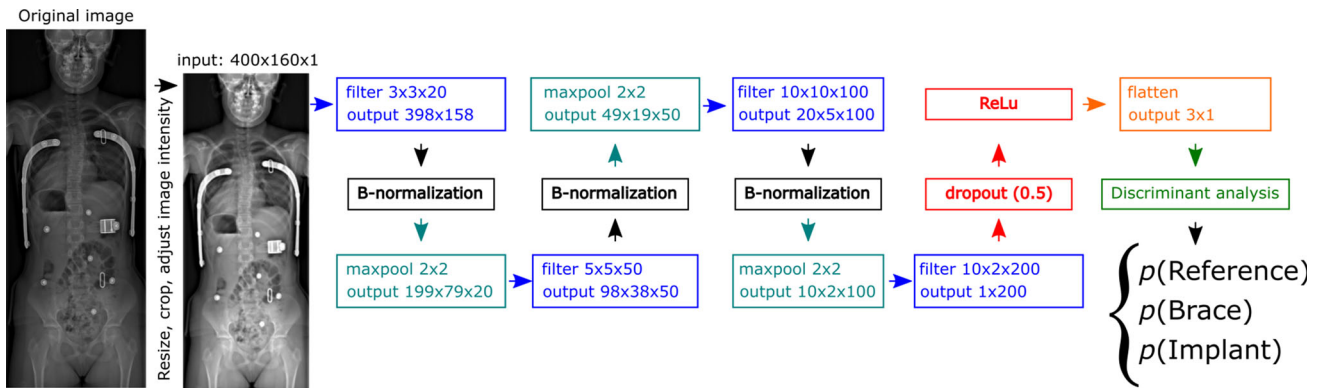


Fig. 2 Architecture of the model developed. The model takes a 400×160 pixel grayscale image, which is processed through a convolutional neural network alternating convolutional, batch normalization and max-pooling layers. The networks outputs through a dropout layer and a rectified linear unit, and the result is further processed by discriminant analysis. This yields a probability for the image to belong to any of the three group; if the highest probability is lower than 0.9, the image is considered as “not classified”

Table 2 Confusion matrix

| Detection | Ground truth | | |
|----------------|--------------|-------------|-------------|
| | Reference | Brace | Implant |
| Reference | 730 (94.3%) | 18 (2.6%) | 0 |
| Brace | 15 (1.9%) | 643 (95.7%) | 2 (0.3%) |
| Implant | 0 | 0 | 647 (99.5%) |
| Not classified | 29 (3.7%) | 11 (1.6%) | 1 (0.2%) |

Correct and wrong detections are given as percentages of the total classified values, while not classified are percentages of the total dataset

metallic objects) that can challenge classical machine learning approaches. The model was composed by a 13-layer deep CNN, the output of which was further processed by discriminant analysis. Preliminary tests showed that increasing the network complexity (more layers, larger filters, etc.) only lead to overfitting, while decreasing its complexity did not reach good accuracy. On the other hand, complexity was increased relative to LeNet-5 by adding a set of layers (convolution + normalization + max pooling) to obtain a slower

image size reduction. The additional discriminant analysis allowed improving the separation between the classes in the results space, yielding a global better accuracy. Indeed, with linear discriminant analysis, accuracy was higher for all groups, and overall accuracy improved from 96.9 to 98.3%.

2.0% of the dataset was flagged as “unclassified”; these radiographs should be checked manually for proper classification. Nevertheless, 1.7% wrong classifications remained, most of them concentrated in the reference group. One-third of these images (7 out of 19) contained foreign objects, such as superimposed text or radiopaque markers (Fig. 3), while the other did not show any peculiarity. It is possible that removing such images might have improved the results, but the model would have lost in generality. Instead, the model can be considered robust to the presence of small image annotations, jewellery, etc., within the limits reported in Table 1.

The main limitation of this study is the large data augmentation that was performed, due to difficulties in acquiring large annotated datasets of patient radiographs. In particular, the brace group was augmented $8 \times$; this could have led to an overestimation of the classification accuracy (which

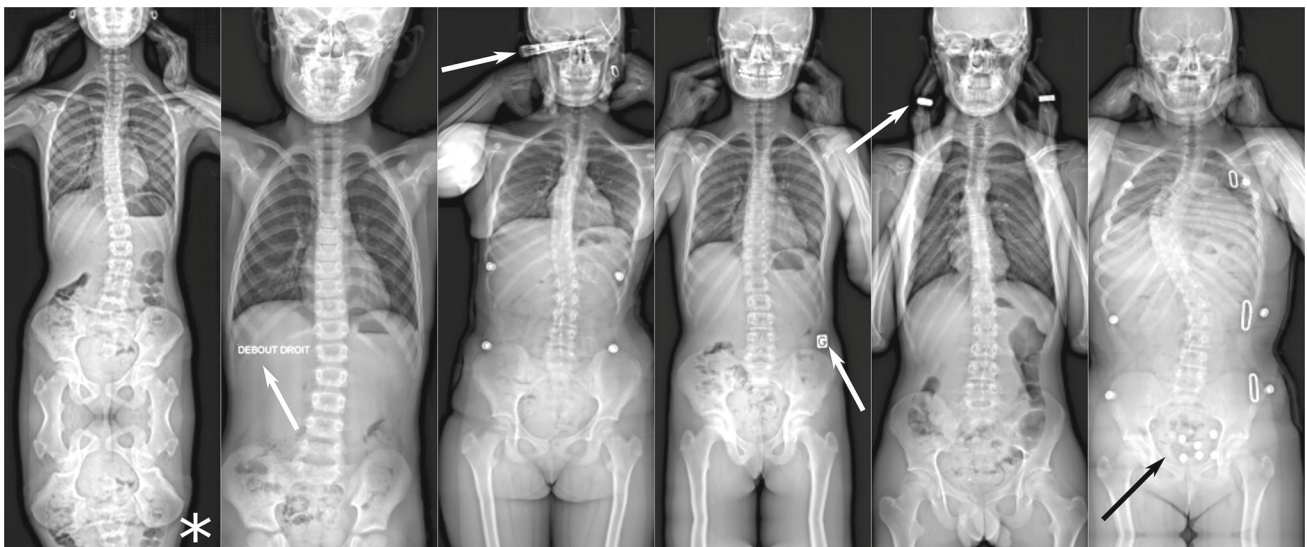


Fig. 3 Examples of incorrectly classified radiographs. Seven out of nineteen of the radiographs that were not correctly classified contained foreign objects, such as superimposed text or radiopaque buttons or jewellery (arrows). One of these images (*) also needed a very large portion of padding

was 99.7%), although the tenfold cross-validation method should reduce such overestimation. Image flipping is sometimes not performed on medical images since it could result in non-physiological anatomical features; however, this dataset contained both antero-posterior and postero-anterior radiographs, so horizontal image flipping generated images which could be expected in clinical routine. The same applies for the other data augmentation methods, which generate realistic radiographs: translation, enlarging, noise addition. The latter holds a particular interest since the current tendency is to reduce the radiation dose when imaging young scoliotic patients [20], resulting in higher noise content.

All images were acquired with the patient in free-standing position [16]. This is the most adopted position in radiological examination of scoliosis because the arms do not hide the spine neither in frontal radiographs nor in lateral ones, and the ribcage remains fairly visible in both views. Indeed, for a three-dimensional assessment of scoliosis, biplanar radiographs are acquired when possible, since they allow 3D reconstruction of the bony structures [21, 22]. It is possible that feeding radiographs with patients in different position (such as in lateral bending [23]) could alter the results. Nevertheless, the sample can be considered heterogeneous since the radiographs were acquired from five different hospitals in a time span of about 8 years.

In conclusion, the proposed classification model could help automatic labelling radiographs for scoliosis treatment. The same model architecture could potentially be applied for other radiograph classifications, such as sex and presence of scoliotic deformity. This work opens the door to automatic medical images pre-labelling, and it is a fundamental first

step towards a global automatic radiologic analysis framework.

Acknowledgements The authors are grateful to the ParisTech BiomecAM chair program on subject-specific musculoskeletal modelling (with the support of ParisTech and Yves Cotrel Foundations, Société Générale, Proteor and Covea).

Compliance with ethical standards

Conflict of interest Wafa Skalli holds patents related to the EOS system and associated 3D reconstruction methods, with no personal financial benefit (royalties rewarded for research and education). The other authors declare that they have no conflict of interest.

Ethical standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent For this type of study, formal consent is not required.

References

1. Cheng JC, Castelein RM, Chu WC, Danielsson AJ, Dobbs MB, Grivas TB, Gurnett CA, Luk KD, Moreau A, Newton PO, Stokes IA, Weinstein SL, Burwell RG (2015) Adolescent idiopathic scoliosis. *Nat Rev Dis Primers* 1:15030. <https://doi.org/10.1038/nrdp.2015.30>
2. Negrini S, Donzelli S, Aulisa AG, Czaprowski D, Schreiber S, de Mauroy JC, Diers H, Grivas TB, Knott P, Kotwicki T, Lebel A, Marti C, Maruyama T, O'Brien J, Price N, Parent E, Rigo M, Romano M, Stikeleather L, Wynne J, Zaina F (2018) 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. *Scoliosis Spinal Disord* 13:3. <https://doi.org/10.1186/s13013-017-0145-8>

3. Weinstein SL, Dolan LA, Wright JG, Dobbs MB (2013) Effects of bracing in adolescents with idiopathic scoliosis. *N Engl J Med* 369:1512–1521. <https://doi.org/10.1056/NEJMoa1307337>
4. Bunnell WP (1986) The natural history of idiopathic scoliosis before skeletal maturity. *Spine (Phila Pa 1976)* 11:773–776
5. Galbusera F, Casaroli G, Bassani T (2019) Artificial intelligence and machine learning in spine research. *Jor Spine* 2:e1044. <https://doi.org/10.1002/jsp2.1044>
6. Bakhous C, Aubert B, Vazquez C, Cresson T, Parent S, Guise J De (2018) Automatic pedicles detection using convolutional neural network in a 3D spine reconstruction from biplanar radiographs. In: *Proc SPIE*
7. Ebrahimi S, Gajny L, Skalli W, Angelini E (2019) Vertebral corners detection on sagittal X-rays based on shape modelling, random forest classifiers and dedicated visual features. *Comput Methods Biomech Biomed Eng Imaging Vis* 7:132–144. <https://doi.org/10.1080/21681163.2018.1463174>
8. Aubert B, Vazquez C, Cresson T, Parent S, de Guise JA (2019) Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting. *IEEE Trans Med Imaging* 38:2796–2806. <https://doi.org/10.1109/TMI.2019.2914400>
9. Yang Z, Skalli W, Vergari C, Angelini ED, Gajny L (2019) Automated spinal midline delineation on biplanar x-rays using mask R-CNN - VipIMAGE. In: Natal Jorge RM (ed) Tavares JMRS. Springer International Publishing, Cham, pp 307–316
10. Zhang J, Li H, Lv L, Zhang Y (2017) Computer-aided Cobb measurement based on automatic detection of vertebral slopes using deep neural network. *Int J Biomed Imaging* 2017:9083916. <https://doi.org/10.1155/2017/9083916>
11. Cho BH, Kaji D, Cheung ZB, Ye IB, Tang R, Ahn A, Carrillo O, Schwartz JT, Valliani AA, Oermann EK, Arvind V, Ranti D, Sun L, Kim JS, Cho SK (2019) Automated measurement of lumbar lordosis on radiographs using machine learning and computer vision. *Glob Spine J*. <https://doi.org/10.1177/2192568219868190>
12. Weng C-H, Wang C-L, Huang Y-J, Yeh Y-C, Fu C-J, Yeh C-Y, Tsai T-T (2019) Artificial intelligence for automatic measurement of sagittal vertical axis using ResUNet framework. *J Clin Med* 8(11):1826
13. Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C (2009) Automated model-based vertebra detection, identification, and segmentation in CT images. *Med Image Anal* 13:471–482. <https://doi.org/10.1016/j.media.2009.02.004>
14. Zheng G, Chu C, Belavý DL, Ibragimov B, Korez R, Vrtovec T, Hutt H, Everson R, Meakin J, Andrade IL, Glocker B, Chen H, Dou Q, Heng P-A, Wang C, Forsberg D, Neubert A, Fripp J, Urschler M, Stern D, Wimmer M, Novikov A, Cheng H, Armbrrecht G, Felsenberg D, Li S (2017) Evaluation and comparison of 3D intervertebral disc localization and segmentation methods for 3D T2 MR data: a grand challenge. *Med Image Anal* 35:327–344. <https://doi.org/10.1016/j.media.2016.08.005>
15. Jamaludin A, Lootus M, Kadir T, Zisserman A, Urban J, Battié MC, Fairbank J, McCall I (2017) ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J* 26:1374–1383. <https://doi.org/10.1007/s00586-017-4956-3>
16. Faro FD, Marks MC, Pawelek J, Newton PO (2004) Evaluation of a functional position for lateral radiograph acquisition in adolescent idiopathic scoliosis. *Spine (Phila Pa 1976)* 29:2284–2289
17. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2324. <https://doi.org/10.1109/5.726791>
18. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15:1929–1958
19. Vedaldi A, Lenc K (2015) MatConvNet—convolutional neural networks for MATLAB. In: *Proceeding of the ACM Int Conf on Multimedia*
20. Pedersen PH, Vergari C, Alzakri A, Vialle R, Skalli W (2019) A reduced micro-dose protocol for 3D reconstruction of the spine in children with scoliosis: results of a phantom-based and clinically validated study using stereo-radiography. *Eur Radiol* 29:1874–1881. <https://doi.org/10.1007/s00330-018-5749-8>
21. Dubouset J, Ilharreborde B, Le Huec J-C (2014) Use of EOS imaging for the assessment of scoliosis deformities: application to postoperative 3D quantitative analysis of the trunk. *Eur Spine J* 23:397–405. <https://doi.org/10.1007/s00586-014-3334-7>
22. Vergari C, Aubert B, Lallemand-Dudek P, Haen TX, Skalli W (2019) A novel method of anatomical landmark selection for rib cage 3D reconstruction from biplanar radiography. *Comput Methods Biomech Biomed Eng Imaging Vis*. <https://doi.org/10.1080/21681163.2018.1537860>
23. Hirsch C, Ilharreborde B, Mazda K (2016) Flexibility analysis in adolescent idiopathic scoliosis on side-bending images using the EOS imaging system. *Orthop Traumatol Surg Res* 102:495–500. <https://doi.org/10.1016/j.otsr.2016.01.021>