Dentomaxillofacial Radiology (2022) 51, 20220070

birpublications.org/dmfr

RESEARCH ARTICLE

Convolutional neural network-based automatic cervical vertebral maturation classification method

[1]Haizhen Li, [1]Yanlong Chen, [2,3]Qing Wang, [1]Xu Gong, [4]Yi Lei, [5]Jialiang Tian and [1]Xuemei Gao

[1]Department of Orthodontics, Peking University School and Hospital of Stomatology, 22 Zhongguancun South Avenue, Haidian District, Beijing, P.R. China; [2]Department of Automation, Tsinghua University, No.30 Shuangqing Road, Haidian District, Beijing, P.R. China; [3]Pharmacovigilance Research Center for information technology and Data Science , Cross-strait Tsinghua Research Institute, NO.516 Qishan North Road, Huli District, Xiamen, P.R. China; [4]School of Software Engineering, Faculty of Information Technology, No.100 Pingleyuan, Chaoyang District, Beijing, P.R. China; [5]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, No.10 Xitucheng Road, Haidian District, Beijing, P.R. China

Objectives: This study aimed to develop a fully automated artificial intelligence-aided cervical vertebral maturation (CVM) classification method based on convolutional neural networks (CNNs) to provide an auxiliary diagnosis for orthodontists.
Methods: This study consisted of cephalometric images from patients aged between 5 and 18 years. After grouping them into six cervical stages (CSs) by orthodontists, a data set was constructed for analyzing CVM using CNNs. The data set was divided into training, validation, and test sets in the ratio of 70, 15, and 15%. Four CNN models namely, VGG16, GoogLeNet, DenseNet161, and ResNet152 were selected as the candidate models. After training and validation, the models were evaluated to determine which of them is most suitable for CVM analysis. Heat maps were analyzed for a deeper understanding of what the CNNs had learned.
Results: The final classification accuracy ranking was ResNet152>DenseNet161>GoogLeNet>VGG16, as evaluated on the test set. ResNet152 proved to be the best model among the four models for CVM classification with a weighted κ of 0.826, an average AUC of 0.933 and total accuracy of 67.06%. The F1 score rank for each subgroup was: CS6>CS1>CS4>CS5>CS3>CS2. The area of the third (C3) and fourth (C4) cervical vertebrae were activated when CNNs were assessing the images.
Conclusion: CNN models proved to be a convenient, fast and reliable method for CVM analysis. CNN models have the potential to provide automatic auxiliary diagnostic tools in the future.
Dentomaxillofacial Radiology (2022) **51**, 20220070. doi: 10.1259/dmfr.20220070

Cite this article as: Li H, Chen Y, Wang Q, Gong X, Lei Y, Tian J, et al. Convolutional neural network-based automatic cervical vertebral maturation classification method. *Dentomaxillofac Radiol* (2022) 10.1259/dmfr.20220070.

Keywords: Artificial intelligence; Convolutional neural network; Cervical vertebral maturation analysis

Introduction

Bone age evaluation is very important to orthodontists in many aspects, such as the choice of treatment methods,[1] the decision of treatment timing[2] and the judgment of recurrence trend.[3] Handwrist radiographs were originally used for bone age assessment, and it is considered a reliable and reasonable method.[4] In orthodontics, Lamparski first proposed that bone age can also be evaluated by observing changes in the size and shape of the cervical vertebrae in 1975.[5] Since then, the results of the cervical vertebral maturation (CVM) method have been shown to be consistent with those of the handwrist method (HWM) in many studies, and the former may serve as an alternative to the

Correspondence to: Xuemei Gao, E-mail: xmgao@263.net

Received 23 February 2022; revised 05 May 2022; accepted 10 May 2022; published online 07 June 2022

latter.[6,7] Moreover, HWM requires an additional radiograph in orthodontic practice, thus raising the patients' undue concerns about the extra but quite negligible radiation doses, while the CVM method is performed on lateral cephalometric radiographs and is routinely available for the orthodontic examination which could eliminate the extra radiation exposure.

However, assessing the CVM is a complex and time-consuming process that can only be conducted by experienced specialists, even though there was an improved version of the CVM method proposed by Baccetti based on the second (C2), third (C3), and fourth (C4) cervical vertebrae in 2003.[8] The study by Flavio Uribe indicated that the accuracy of CVM classification is extremely sensitive to the clinical experience of the specialist,[9] so mastery of this method takes a prolonged time. Attempts should be made to develop auxiliary diagnosis methods to reduce the clinical burden.

Artificial intelligence (AI) is a subfield of computer science that aims to create systems that can perform cognitive tasks that are ordinarily performed by humans.[10] In recent years, computer vision technology based on convolutional neural networks (CNNs) has recently made remarkable achievements in the field of medical imaging.[11] Tasks like image classification, object detection and semantic segmentation can be solved by this technology. In dentistry, CNNs are also widely used in many fields,[12,13] such as cephalometric measurements,[14,15] tooth detection,[16] temporomandibular disorder diagnosis[17] and caries detection[18,19] CVM analysis is also an ideal target for CNN technology which takes pictures as input and outputs specific labels. Shin et al have proved that Tanner-Whitehouse 3 (TW3)-based fully automated bone age assessment system using CNN can be effectively utilized for HVM evaluation.[20] Therefore, it can be assumed that the CNN models may be helpful in classifying CVM stages.

Thus, this study aimed to develop a fully automated AI-aided CVM classification method to provide an auxiliary diagnosis to aid orthodontists.

Methods and materials

The present research was a retrospective study. All data were processed to eliminate personally identifiable information in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the ethics committee of Peking University Hospital of Stomatology and the approval number is PKUSSIRB-202054025.

Data collection

A total of 6079 cephalometric images were collected from Peking University Hospital of Stomatology between 2017 and 2020. Only those samples that fulfilled the following criteria were included in the study: the patient medical records were complete, cephalometric images were qualified, and the patients were less than 18 years old. The exclusion criteria were as follows: the presence of congenital syndromes, metabolic diseases, diseases affecting growth and development, and history of special drugs. The patients were 2576 males and 3503 females, aged from 5 to 18 years. The age and sex distribution of the patients are shown in Figure 1.

The cephalometric images were routinely taken in three scanners of the same type, Veraviewepocs 2D (J Morita Corp, Kyoto, Japan), with the following parameters: scanning time 4.9 s, tube current 5–10 mA; tube voltage 90 kV. All the images were stored and read in JPG format.

Data processing

The rectangular region consisting of C2, C3, and C4 was manually cropped from the cephalograms to remove the influence of other characteristics. To facilitate the transmission of pictures to the CNN model, the pictures were resized to 224 × 224 pixels while keeping the picture scale unchanged.

The total data were divided into the training, validation, and test data sets, respectively in the ratio of 70, 15, and 15% (Table 1). The training data set was used to train the CNN models. The validation data set was used to tune the hyperparameters of CNNs. The test set was used to evaluate the performance of the model. The flowchart of the experimental process of this study is shown in Figure 2.

Data labelling

To label the data, two experienced orthodontists examined the images manually and grouped them into six stages: cervical vertebral 1 (CS1), cervical vertebral2 (CS2), cervical vertebral 3 (CS3), cervical vertebral 4 (CS4), cervical vertebral 5 (CS5) and cervical vertebral 6 (CS6), with reference to the user guide delivered by McNamara.[21] For images whose classification the two orthodontists did not agree on, a third orthodontist with extensive experience was consulted to come to a final decision.

Convolutional neural network selection

Due to the lack of experience in applying CNN to CVM analysis, four classical CNNs were selected as candidates for our study:VGG16,[22] GoogLeNet,[23] DenseNet161,[24] and ResNet152.[25] The high performance of these CNNs was validated in ImageNet Large Scale Visual Recognition Competition.[26] The above four models had been trained in the present study and their performance on the test set was compared to identify the most suitable model for the CVM analysis.

Data augmentation

To prevent overfitting, the following data augmentation techniques were applied in this study: random translation within a quarter of the width of the image, random rotation in the range of 15° clockwise and 15° counter-clockwise, and adaptive histogram equalization. In each epoch of training, 50% of the training data set was randomly selected for data augmentation.

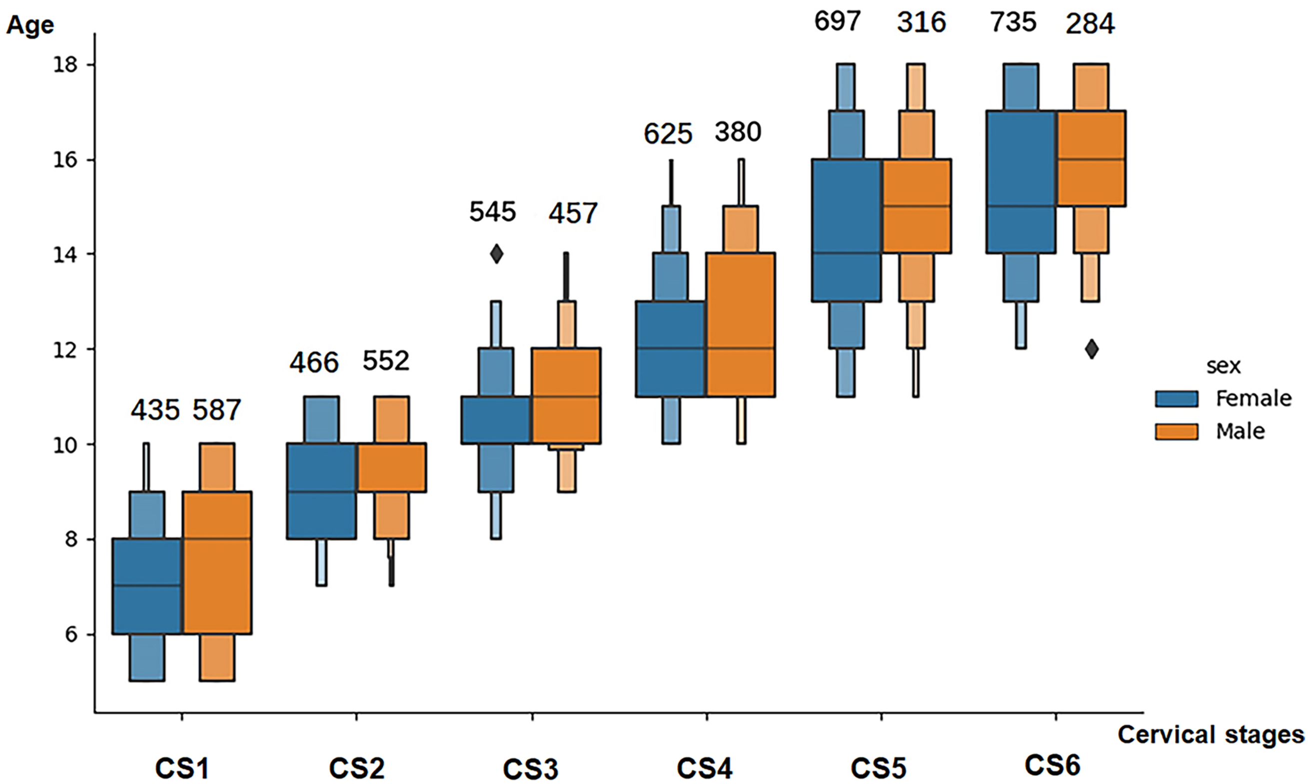

Figure 1 Age, sex, and cervical stages distribution of the enrolling patients.

Strategy and hyperparameter set of training

The strategy of transfer learning and fine-tuning techniques were used in this study. The pre-trained models on the ImageNet data set of the four candidate CNNs were downloaded from Pytorch Zoo (https://pytorch.org/). After initializing, the networks with the pre-trained models, the output layer of the model was changed into six outputs so that it could correspond to the six stages of the CVM method. Because the characteristics of the medical image are quite different from those of the general database, the parameters of the whole CNN layers were fine-tuned. The networks were trained using a Stochastic Gradient Descent (SGD) optimizer for 100 epochs with a mini-batch size of 64. Hyperparameters of the CNNs were adjusted manually according to model performance on the validation set in order to maximize the classification capabilities of each model. After testing various times, the individualized hyperparameter combination was set as follows: initial learning rate = 0.001; momentum = 0.09; weight decay = 0.01. Our training was performed on the server of the computing platform of ******* university, with NVDIA Tesla P100 graphic processing unit.

Table 1 Numbers of patients assigned to training, validation, and test sets of the six CVM subgroups

	CS1	*CS2*	*CS3*	*CS4*	*CS5*	*CS6*	*Total*
Train set	715	712	701	703	709	713	4253
Validation set	153	153	150	151	152	153	912
Test set	154	153	151	151	152	153	914
Total	1022	1018	1002	1005	1013	1019	6079

CVM, cervical vertebral maturation.

Model testing and evaluation metrics

For each of the four CNNs, the model with the lowest loss on the validation set was selected to verify its final classification performance on the test set. To evaluate the classification performance of the CNN models, the following indicators were used to test the model performance: accuracy, precision, recall rates, F1 score and confusion matrix. The receiver operating characteristic (ROC) curves were also plotted and the area under the curve (AUC) was calculated.

For a deeper understanding of the learned model, a heat map was generated using class activation mapping (CAM).[27] This map visually highlights the regions which are mostly informative in distinguishing the CVM classification.

Statistical analysis

The weighted κ statistic was used to verify the consistency of the model's output with researcher annotation results. The classification results of the four CNNs were compared using McNemar's χ^2 analysis. SPSS software (Windows v. 27.0; SPSS Inc. Chicago, IL) was used to perform data analysis.

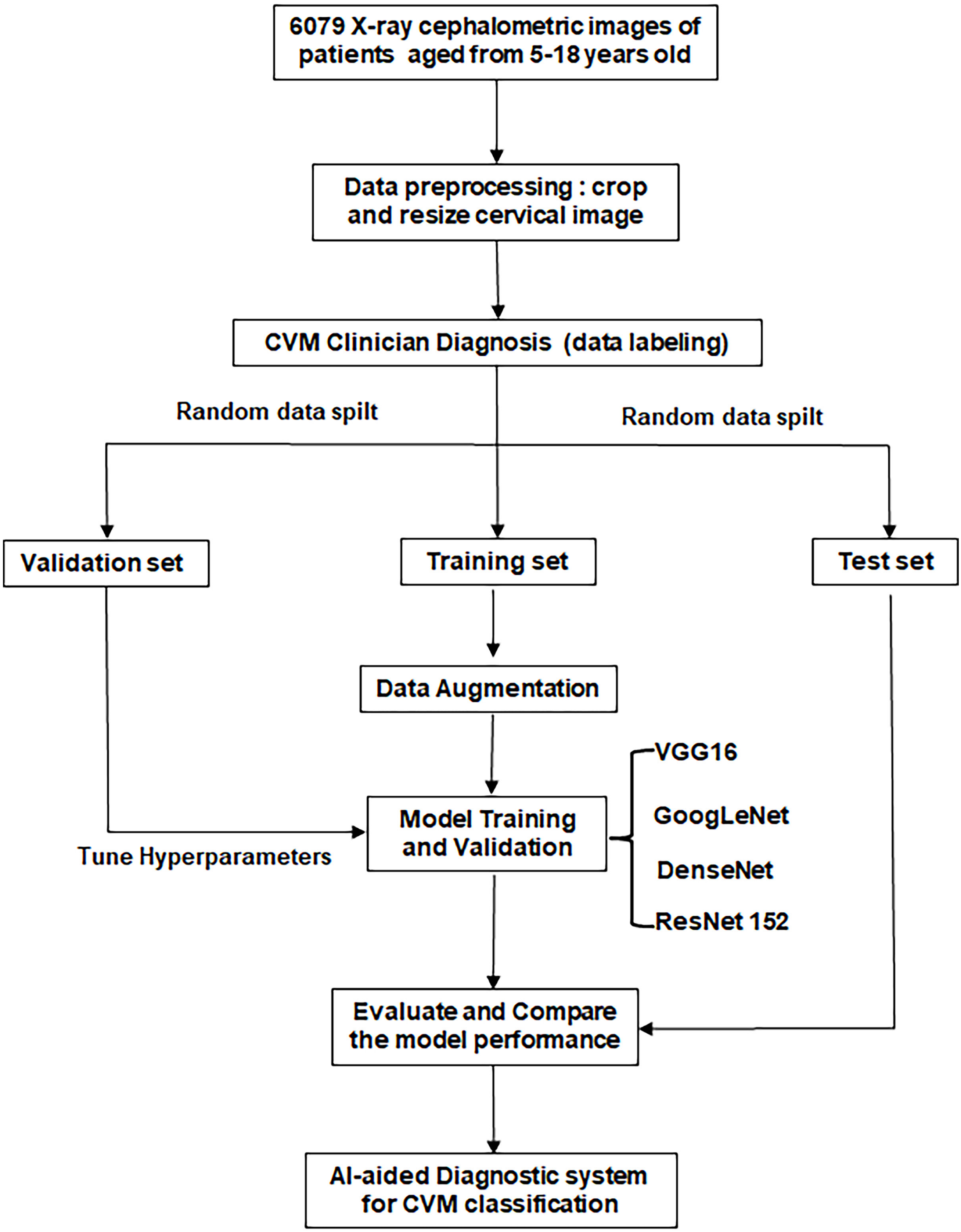

Figure 2 Flowchart of the experimental process. CVM, cervical vertebral maturation

Table 2 The size, max train accuracy and max validation accuracy of the four candidate models

	Model size (MB)	*Max train accuracy*	*Max validation accuracy*
VGG16	512	67.34%	63.04%
GoogLeNet	21.5	64.90%	63.37%
DenseNet161	102	68.04%	66.22%
ResNet152	222	68.51%	67.32%

CNN, convolutional neural network.
Model size: Total parameter size in the CNN model; Max train accuracy: The biggest accuracy on the train set among all the epochs. Max validation accuracy: The biggest accuracy on the validation set among all the epochs.

Results

The model size and training results of the CNN models are shown in Table 2. The accuracy and loss values for each epoch during training are displayed in Figure 3. Although the VGG16 model had the most parameters, it had a worse training effect. Even though there were only 21.5MB parameters in GoogLeNet, the training results were quite good. The training results of ResNet152 and DenseNet161 were very close to each other and better than those of VGG16 and GoogLeNet.

Table 3 Classification performance of the four candidate CNNs on the test set

	Weighted κ	*Average AUC*	*Total accuracy*	*Inference time*
VGG16	0.796***	0.920	61.15%	0.26 s
GoogLeNet	0.811***	0.926	64.11%	0.083 s
DenseNet161	0.818***	0.928	64.22%	0.32 s
ResNet152	0.826***	0.933	67.06%	0.32 s

AUC, area under the curve; CNN, convolutional neural network.
Weighted κ: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

The results of weighted κ, average AUC and total accuracy which were used to compare the final classification performance of the four CNNs are given in Table 3. The weighted κ statistic showed that the classification results of the four CNNs were in good consistency with the researcher annotation results

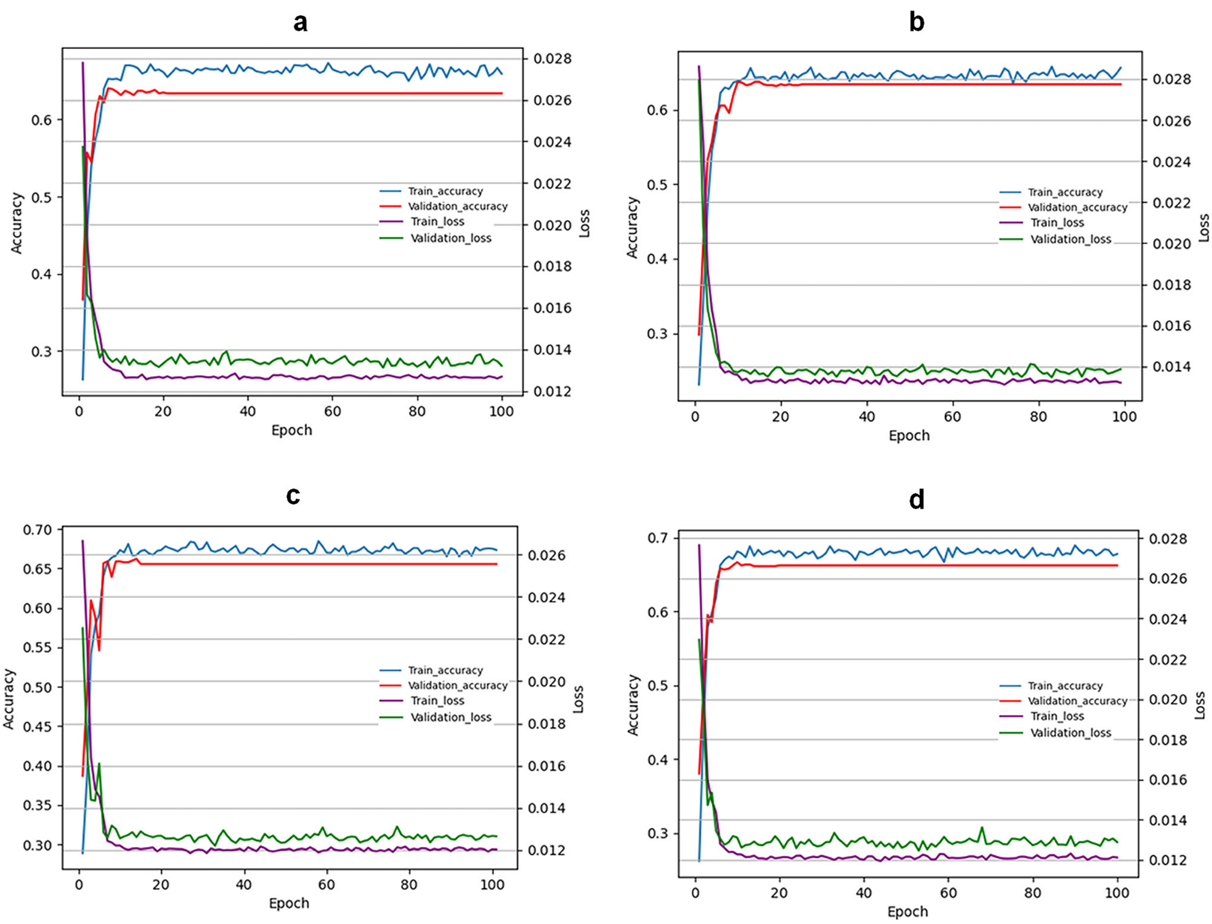

Figure 3 Training results for the four candidate CNN models. (a) Training results of VGG16 (b) training results of GoogLeNet (c) training results of DesneNet161 (d) training results of ResNet152. The horizontal axis of the graph represents the training epochs. The left vertical axis of the graph represents the classification accuracy and the right vertical axis of the graph represents the loss value. CNN, convolutional neural network.

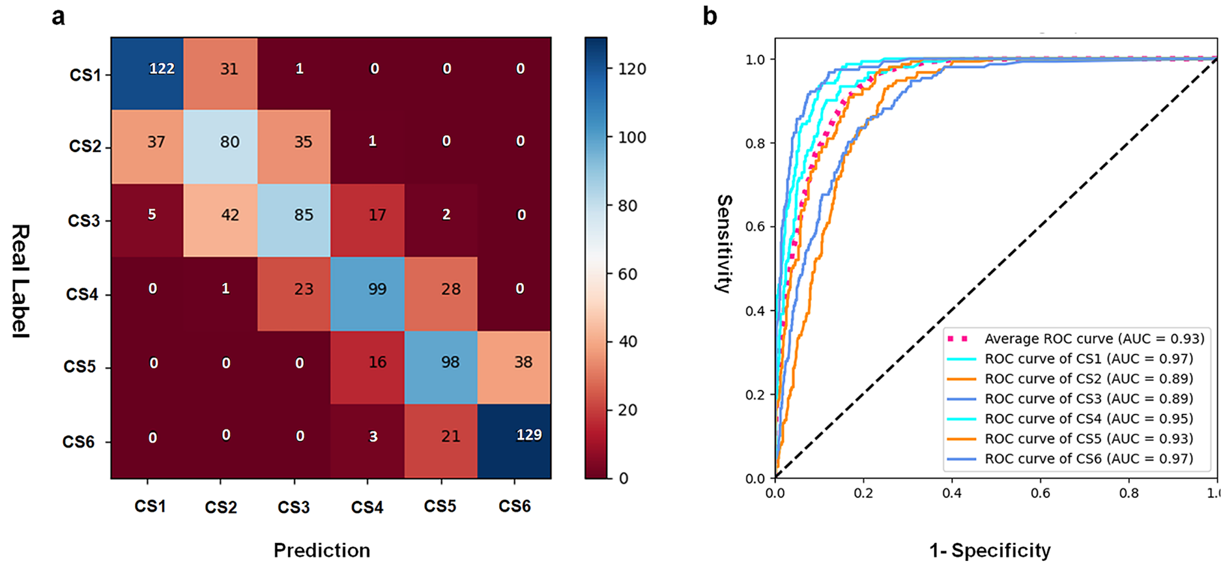

Figure 4 (a) Confusion matrix obtained using ResNet152 model on test set. (b) ROC curves of ResNet152 model for the cervical stages. ROC, receiver operating characteristic.

(all the weighted κ >0.7, p < 0.001). The ranking of classification accuracy, average AUC and weighted κ were as follows: ResNet152>DenseNet161>GoogLeNet>VGG16. ResNet152 proved to be the best among the four models (McNemar's chi-square test: ResNet152 *vs* DenseNet161, p = 0.024; ResNet152 *vs* GoogLeNet, p = 0.026; ResNet152 *vs* VGG16, p < 0.001), with a weighted κ of 0.826, average AUC of 0.933, and total accuracy of 67.06%. The ranking of the inference speed for the four models was as follows: GoogleNet>VGG16>DenseNet161=ResNet152.

The classification results of ResNet152 on the test set are shown as a confusion matrix in Figure 4a. The confusion matrix showed a good classification result. By observing the color and number distribution of each cell, it can be intuitively found that the data were mainly distributed on the diagonal cell or around the diagonal cell. The model always tends to achieve the correct classification; even when the final classification is incorrect, the model tends to a neighboring rather than distant category from the correct one. The precision, recall rates, and F1 score of each cervical stage were calculated according to the confusion matrix, and the results are shown in Table 4. CS2 and CS3 phases had the lowest precision, recall rates and F1 score, CS4 and CS5 phases were in the middle rank, while CS1 and CS6 phases had the highest precision, recall rates and F1 score. The ROC curves for each cervical stage are shown in Figure 4b and the AUC was also calculated.

Figure 5 shows the heat maps created using the CAM. The method activated the significant areas which influenced the diagnostic results in the inference process of the CNN models. It shows the features that the trained model learned to perform the CVM analysis. In this study, the activation area of the neural network was on the areas of C3 and C4. The CAM was highly consistent with the regions on which orthodontists performed the CVM classification judgments.

Figure 6 shows the process of inference using the ResNet152 model. The output of the neural network is the confidence that the image belongs to each category. The index corresponding to the highest confidence is the predicted label of the CNN model.

Discussion

In orthodontics, bone age is more frequently used than chronological age for growth prediction because the latter is easily influenced by gender, genetic characteristics, environmental factors, etc. This can also be proved in this research by the fact that the chronological age and bone age could overlap as is shown in Figure 1. HWM has been proved to be a mature method in determining bone age, but it is complained by the patients as exposed to extra radiation. CVM is also recommended

Table 4 Precision, recall rates and F1 score of ResNet152 on test set for each CVM subgroup

	Precision	*Recall*	*F1 score*	*Sample size of each subgroup*
CS1	0.74	0.79	0.77	154
CS2	0.52	0.52	0.52	153
CS3	0.59	0.56	0.58	151
CS4	0.73	0.66	0.69	151
CS5	0.66	0.64	0.65	152
CS6	0.77	0.84	0.81	153

Precision = TP/TP +FP; Recall = TP/TP +FN; F1 score = 2*Precision * Recall/Precison + Recall. TP is true CVM, cervical vertebral maturation; Precison, Recall and F1 score are calculated as follows with confusion matrix in Figure 3.positive, FP is false positive, FN is false negative, and TN is true negative.

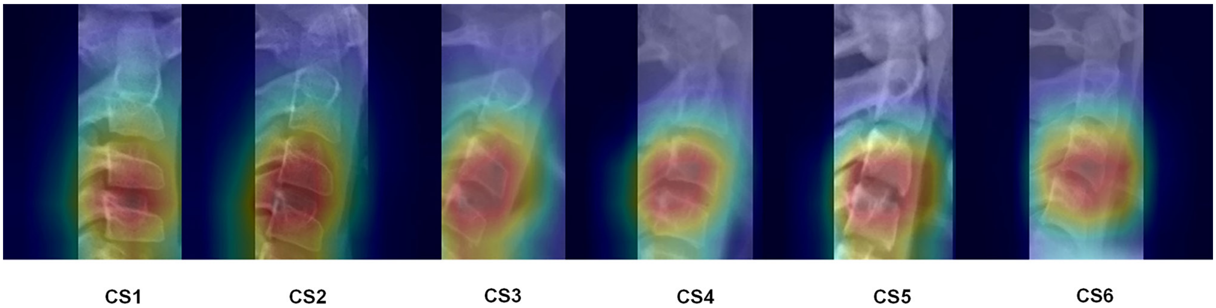

Figure 5 Representative class activation map of the correct classification in each cervical stage.

in orthodontics with its relative facility, despite it is not as precise as HWM. The CVM method is not easy to master because it relies heavily on clinical experience.[28] The present study tested an automatic classification method based on CNN to assist orthodontics in the decision-making process for CVM analysis. To find a suitable CNN model for CVM analysis, four CNN models were trained using the transfer learning method. Upon comparing the performance of the four models on the test set, ResNet152 proved to be the best model for our data set. The model exhibited a weighted κ of 0.826 and an average AUC of 0.933 compared to the researcher's diagnosis. Besides, the time for ResNet152 model to analyze a single cephalogram is only about 0.32 s. CNN model proved to be an effective and fast method for CVM stage classification, demonstrating its potential as an AI-aided diagnostic tool.

For better application of CNNs algorithm in our research, four CNNs with different structures were compared for CVM classification. The ranking of the classification accuracy for CVM classification was as follows: ResNet152>DenseNet161>GoogLeNet>VGG16. By reviewing the literature, it could be seen that the structure of CNNs was updated and improved from VGG16 to ResNet152.[22–25] Correspondingly, the classification accuracy of these CNNs on our data set is getting better and better. This reminds us that the progress of the computer algorithm can improve the accuracy of medical diagnosis.

As a representative algorithm of deep learning, the CNNs can effectively learn relevant features from a large number of images by convolution calculation and backpropagation algorithms,[29] avoiding the process of manual feature extraction. It is more efficient and simpler compared to traditional machine learning algorithms in medical image analysis.[11] Previous studies have used other machine learning algorithms for CVM analysis. For example, Pisa proposed a method for CVM assessment by Naïve Bayes algorithm with 188 lateral cephalograms as early as 2012[30] and Amasya developed an artificial neural network model for CVM analysis with 647 lateral cephalograms in 2020.[31,32] However, due to the limitations of the selected algorithm, they only achieved semi-automated functions, meaning that all these methods required orthodontists to extract the features manually by depicting anatomical landmarks of the cervical vertebrae, whereas the CNN algorithm in the present study requires little involvement of orthodontists. CVM analysis based on CNNs is more intelligent and could reduce the workload for orthodontists.

The neural network is often described as a black box because it does not provide any feedback as to why and how it performs its predictions. CAM technology is one of the approaches for investigating the features neural networks learn during the training process.[27,33] It highlights the discriminative regions used by CNN to identify the special categories. In the present study, the area of C3 and C4 was activated when CVM was assessing the images. CNN utilized features similar to those used by orthodontists, but some tiny features were ignored, such as the concavity in the inferior border of the second vertebra.[8]

In this study, it is believed that there were two special points in applying CNN models to medical data. The first one is transfer learning, which is widely used in the medical imaging AI field. Transforming the experience

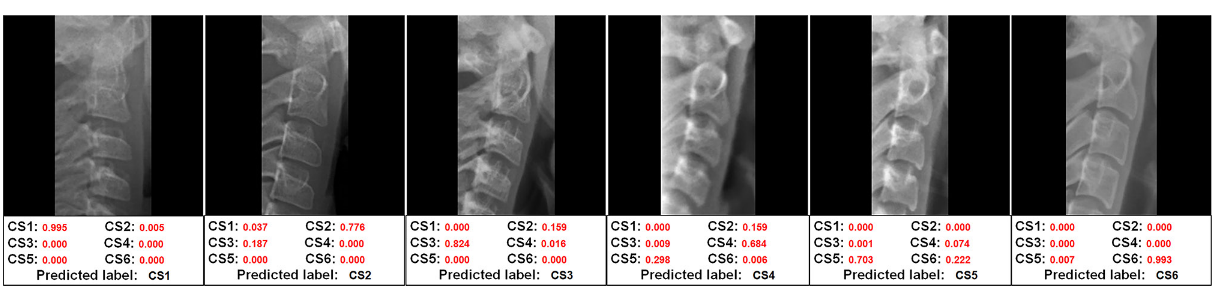

Figure 6 Representative demo of inference process using the CNNs. The red text is the confidence of the ResNet152 model prediction. CNNs, convolutional neural networks

learned from general pictures into medical imaging is necessary at present because it can help the CNN model to converge faster even if there is not enough data.[34] It is very suitable for medical data which are usually relatively small and difficult to collect. The second point is data augmentation. As a common method to prevent overfitting in deep learning, data augmentation can help the CNN model extract more information from the training data set, thereby narrowing the gap between the training set and the validation set.[35] Notably, selecting personalized data augmentation methods is important for medical data sets because general data augmentation methods may add unnecessary variance to highly structured medical data sets.[36]

The limitation of this study is that the trained model only achieved an accuracy of 67.06% on the test set, and the F1 scores of CS2 and CS3 were only about 0.5, which suggests that the algorithm needs further improvement. Three reasons are thought to be responsible for the limitation, and the first one is the quality and quantity of the dataset. Even though the sample size of this study was much larger than those of previous studies, it was still relatively small for deep learning. Besides, the quality of data labelling is severely affected by the subjective nature of CVM analysis.[28] Establishing a standard database based on multicentres and more experts is a feasible solution to improve the model performance in the future. The second reason is that the CVM staging is sometimes imprecise due to the gradual changes in size and shape of the three vertebral bodies that occur over time, especially for the maturational stages that last a shorter length of time such as C2 and C3. The third reason lies in the CNN algorithm. The CAM technology shows that the CNN algorithm used in this study could not find some special features related to cervical stages, such as the concavity in the inferior border of the second vertebra. Creating a customized CNN model based on the existing network structure may be the solution to this problem. This needs future systematic work to develop a better CNN model for CVM analysis, which is beyond the scope of this work.

Conclusion

The present study demonstrated the effectiveness of applying CNN in CVM classification as an automatic AI-aided diagnostic tool that can assist orthodontists. CNN models could provide a convenient, fast and reliable CVM diagnosis in clinical practice. Future work should focus on creating a customized network structure and constructing a standard database that is still needed to enhance model performance.

Acknowledgment

We would thank the support of High-performance Computing Platform of Peking University.

Funding

This research was supported by Beijing Municipal Natural Science Foundation[L192068].

Ethics approval

The study has been approved by the ethics committee of Peking University Stomatology School and Hospital (PKUSSIRB-202054025)

REFERENCES

1. Lee K-M, Chung D-H, Lee J-W, Lee S-M. Cervical vertebrae maturation, dentoalveolar, head postural and respiratory parameters in predicting the stable outcome of face-mask treatment. *Eur J Orthod* 2015; **37**: 251–56. https://doi.org/10.1093/ejo/cju035
2. Baccetti T, Franchi L, Toth LR, McNamara JA. Treatment timing for twin-block therapy. *Am J Orthod Dentofacial Orthop* 2000; **118**: 159–70. https://doi.org/10.1067/mod.2000.105571
3. Melrose C, Millett DT. Toward a perspective on orthodontic retention? *Am J Orthod Dentofacial Orthop* 1998; **113**: 507–14. https://doi.org/10.1016/s0889-5406(98)70261-6
4. Flores-Mir C, Nebbe B, Major PW. Use of skeletal maturation based on hand-wrist radiographic analysis as a predictor of facial growth: a systematic review. *Angle Orthod* 2004; **74**: 118–24. https://doi.org/10.1043/0003-3219(2004)074<0118:UOSMBO>2.0.CO;2
5. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae. *American Journal of Orthodontics* 1975; **67**: 458–59. https://doi.org/10.1016/0002-9416(75)90038-X
6. Szemraj A, Wojtaszek-Słomińska A, Racka-Pilszak B. Is the cervical vertebral maturation (CVM) method effective enough to replace the hand-wrist maturation (HWM) method in determining skeletal maturation?-A systematic review. *Eur J Radiol* 2018; **102**: S0720-048X(18)30095-0: 125–28: . https://doi.org/10.1016/j.ejrad.2018.03.012
7. San Román P, Palma JC, Oteo MD, Nevado E. Skeletal maturation determined by cervical vertebrae development. *Eur J Orthod* 2002; **24**: 303–11. https://doi.org/10.1093/ejo/24.3.303
8. Baccetti T, Franchi L, McNamara JA. An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth. *Angle Orthod* 2002; **72**: 316–23. https://doi.org/10.1043/0003-3219(2002)072<0316:AIVOTC>2.0.CO;2
9. Khajah A, Tadinada A, Allareddy V, Kuo CL, Nanda R, Uribe F. Influence of type of radiograph and levels of experience and training on reproducibility of the cervical vertebral maturation method. *Am J Orthod Dentofacial Orthop* 2020; **157**: S0889-5406(19)30831-5: 228–39: . https://doi.org/10.1016/j.ajodo.2019.03.025
10. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 2020; **99**: 769–74. https://doi.org/10.1177/0022034520915714
11. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; **42**: S1361-8415(17)30113-5: 60–88: . https://doi.org/10.1016/j.media.2017.07.005

12. Heo M-S, Kim J-E, Hwang J-J, Han S-S, Kim J-S, Yi W-J, et al. Artificial intelligence in oral and maxillofacial radiology: what is currently possible? *Dentomaxillofac Radiol* 2021; **50**(3): 20200375. https://doi.org/10.1259/dmfr.20200375
13. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofac Radiol* 2020; **49**(1): 20190107. https://doi.org/10.1259/dmfr.20190107
14. Yoon H-J, Kim D-R, Gwon E, Kim N, Baek S-H, Ahn H-W, et al. Fully automated identification of cephalometric landmarks for upper airway assessment using cascaded convolutional neural networks. *Eur J Orthod* 2022; **44**: 66–77. https://doi.org/10.1093/ejo/cjab054
15. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Artificial intelligence in orthodontics : evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. *J Orofac Orthop* 2020; **81**: 52–68. https://doi.org/10.1007/s00056-019-00203-8
16. Umer F, Habib S, Adnan N. Application of deep learning in teeth identification tasks on panoramic radiographs. *Dentomaxillofac Radiol* 2022; **2022**: 20210504. https://doi.org/10.1259/dmfr.20210504
17. Diniz de Lima E, Souza Paulino JA, Lira de Farias Freitas AP, Viana Ferreira JE, Barbosa J da S, Bezerra Silva DF, et al. Artificial intelligence and infrared thermography as auxiliary tools in the diagnosis of temporomandibular disorder. *Dentomaxillofac Radiol* 2022; **51**(2): 20210318. https://doi.org/10.1259/dmfr.20210318
18. Schwendicke F, Rossi JG, Göstemeyer G, Elhennawy K, Cantu AG, Gaudin R, et al. Cost-effectiveness of artificial intelligence for proximal caries detection. *J Dent Res* 2021; **100**: 369–76. https://doi.org/10.1177/0022034520972335
19. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018; **77**: S0300-5712(18)30225-2: 106–11: . https://doi.org/10.1016/j.jdent.2018.07.015
20. Shin N-Y, Lee B-D, Kang J-H, Kim H-R, Oh DH, Lee BI, et al. Evaluation of the clinical efficacy of a TW3-based fully automated bone age assessment system using deep neural networks. *Imaging Sci Dent* 2020; **50**: 237–43. https://doi.org/10.5624/isd.2020.50.3.237
21. McNamara JA, Franchi L. The cervical vertebral maturation method: A user's guide. *Angle Orthod* 2018; **88**: 133–43. https://doi.org/10.2319/111517-787.1
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014.
23. Szegedy C, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Anguelov D, ed. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Boston, MA, USA. ; June 2015. https://doi.org/10.1109/CVPR.2015.7298594
24. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: Weinberger KQ, ed. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI. ; July 2017. https://doi.org/10.1109/CVPR.2017.243
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Sun J, ed. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA. ; June 2016. https://doi.org/10.1109/CVPR.2016.90
26. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015; **115**: 211–52. https://doi.org/10.1007/s11263-015-0816-y
27. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. In: Torralba A, ed. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA. ; June 2016. https://doi.org/10.1109/CVPR.2016.319
28. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Cervical vertebrae maturation method morphologic criteria: poor reproducibility. *Am J Orthod Dentofacial Orthop* 2011; **140**: 182–88. https://doi.org/10.1016/j.ajodo.2011.04.013
29. Leonard J, Kramer MA. Improvement of the backpropagation algorithm for training neural networks. *Computers & Chemical Engineering* 1990; **14**: 337–41. https://doi.org/10.1016/0098-1354(90)87070-6
30. Baptista RS, Quaglio CL, Mourad LMEH, Hummel AD, Caetano CAC, Ortolani CLF, et al. A semi-automated method for bone age assessment using cervical vertebral maturation. *Angle Orthod* 2012; **82**: 658–62. https://doi.org/10.2319/070111-425.1
31. Amasya H, Cesur E, Yıldırım D, Orhan K. Validation of cervical vertebral maturation stages: artificial intelligence vs human observer visual analysis. *Am J Orthod Dentofacial Orthop* 2020; **158**: S0889-5406(20)30602-8: e173–79: . https://doi.org/10.1016/j.ajodo.2020.08.014
32. Amasya H, Yildirim D, Aydogan T, Kemaloglu N, Orhan K. Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofac Radiol* 2020; **49**(5): 20190441. https://doi.org/10.1259/dmfr.20190441
33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: Batra D, ed. 2017 IEEE International Conference on Computer Vision (ICCV); Venice. ; October 2017. https://doi.org/10.1109/ICCV.2017.74
34. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 2016; **35**: 1299–1312. https://doi.org/10.1109/TMI.2016.2535302
35. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data* 2019; **6**(1). https://doi.org/10.1186/s40537-019-0197-0
36. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J. Automated skeletal classification with lateral cephalometry based on artificial intelligence. *J Dent Res* 2020; **99**: 249–56. https://doi.org/10.1177/0022034520901715