Original Article

Deep learning models to classify skeletal growth phase on 3D radiographs

Nazila Ameli[1], Manuel Oscar Lagravere[1], Hollis Lai[1]

[1]Department of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton Clinical Health Academy, University of Alberta, Edmonton, Alberta, Canada.

*Corresponding author:

Nazila Ameli,
Department of Dentistry,
Faculty of Medicine and
Dentistry, University of Alberta,
Edmonton Clinical Health
Academy, University of Alberta,
Edmonton, Alberta, Canada.

nazila@ualberta.ca
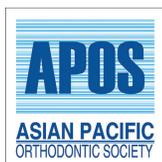
ABSTRACT

Objectives: Cervical vertebral maturation (CVM) is widely used to evaluate growth potential in orthodontics. This study aims to develop an artificial intelligence (AI) algorithm that automatically predicts the CVM stages in terms of growth phases using cone-beam computed tomography images.

Material and Methods: A total of 30,016 slices were obtained from 56 patients with an age range of 7–16 years. After cropping the region of interest, a convolutional neural network (CNN) was built to classify the slices based on the presence of a good vision of vertebrae. The output was used to train another model capable of categorizing the slices into phases of growth, which were defined as Phase I (prepubertal), Phase II (circumpubertal), and Phase III (postpubertal). After training the model, 88 new images were used to evaluate the performance of the model using multi-class classification metrics.

Results: The average classification accuracy of the first and second CNN-based deep learning models was 96.06% and 95.79%, respectively. The multi-class classification metrics also showed an overall accuracy of 84% for predicting the growth phase in unseen data. Moreover, Phase I ranked the highest accuracy in terms of F1-score (87%), followed by Phase II (83%) and Phase III (80%).

Conclusion: Our proposed models could automatically detect the C2–C4 vertebrae and accurately classify slices into three growth phases without the need for annotating the shape and configuration of vertebrae. This will result in the development of a fully automatic and less complex system with reasonable performance.

Keywords: Cervical vertebral maturation, Skeletal age, Deep learning, Convolutional neural network

INTRODUCTION

In medicine and dentistry, understanding growth and development is crucial for diagnosis and treatment.[1,2] Bone age provides more accurate maturation insights than chronological age.[3] In orthodontics, treatment timing is vital for selecting appliances and influencing jaw growth.[1,4] Hand-wrist radiographs, the gold standard for skeletal age determination, offer simplicity and minimal radiation exposure but are criticized for time consumption, expertise demand, and inter/intra-rater variability.[5,6]

Evaluating cervical vertebral maturation (CVM)-introduced by Baccetti *et al.* using the morphological changes in the C2, C3, and C4 vertebral bodies[4]-can be performed on the lateral cephalometric radiographs.[7] Cephalometry is crucial in orthodontics for diagnosis, planning, and growth assessment.[8,9] Thus, in orthodontics, an obvious advantage of CVM evaluation is the

prevention of additional exposure to radiation by eliminating the need for a hand-wrist radiograph.[4]

According to this evidence, the CVM stages 1 and 2 have been referred to as prepubertal; stage 3 has been referred to as circumpubertal; and stages 4, 5, and 6 have been defined as postpubertal.[10] Some studies have reported that this technique is inherently subjective and influenced by the practitioner's experience.[11] Moreover, some authors believe that due to the high level of radiographic noise and intrinsic limitations of 2D lateral cephalograms that affect the magnification and image accuracy, the estimation of bone age using CVM may be difficult for practitioners lacking adequate knowledge and experience.[4,11]

Based on the limitations listed above and the fact that accurate image analysis plays a crucial role in achieving a successful orthodontic outcome, automatizing the task will provide time saving, efficiency, accuracy, and repeatability in orthodontic treatment planning and assist clinicians in alleviating their enormous workload.[4]

Machine learning (ML) employs algorithms to predict outcomes based on inherent statistical patterns in data.[12,13] Deep learning (DL) involves network architectures with multiple hidden layers, which is particularly effective for analyzing complex data like images.[12,14] Convolutional neural networks (CNNs) have revolutionized the direct interpretation, recognition, and classification of medical images, with a focus on cephalometric radiograph analysis and landmark auto-identification; however, skeletal age assessment from lateral cephalograms is an emerging area of study.[14-16]

Cone-beam computed tomography (CBCT) is gaining popularity in orthodontics, offering a three-dimensional (3D) evaluation of hard and soft tissues with advantages such as reduced radiation, clearer images, precision, and cost-effectiveness compared to conventional computed tomography scans.[5,17-19] Given the importance of CVM classification in clinical applications is to determine the optimum timing for growth modification treatments, and as there is no data available regarding the performance of CNN models to estimate the CVM on 3D radiographs, the objective of this study is to demonstrate the application of CNN in dental imaging for classifying phases of growth that works in a fully automatic manner without the need for annotating the images.

MATERIAL AND METHODS

This study was approved by the Health Research Ethics Board-Pro00118171. All patients aged between 7 and 16 years without congenital or acquired malformation of the cervical vertebrae, who underwent CBCT (120 kVp, 5 mA, and 4 s) sagittal views of craniofacial structures between 2013 and 2020 were included in the study. CBCTs were obtained from a database where they were taken for aid in diagnosis and treatment planning for orthodontic patients.

All collected images were kept in DICOM format, so they were all transformed into portable network graphics (PNG) images using the ITK-SNAP software (726 × 644 pixels). Obtained images were preprocessed by resizing and enhancement techniques. The sagittal views, which consisted of 536 slices for each patient, were classified by two orthodontists (A. S. and N. A.) with more than 6 years of experience. In the case of any conflicts, a third orthodontist (S. F.) evaluated the slices to determine the class of CVM. CVM was classified into six stages according to the methodology from the previous studies.[4] Then, slices were grouped into three growth phases (I, II, and III) by combining the CS1 and 2 as Phase I, CS 3 as Phase II, and CS4, 5, and 6 as Phase III. Then, the slices were exported into Google Colaboratory. First, regions of interest (ROI), which included the C2–C4 vertebrae, were cropped from the original slices for CVM classification. The cropping was done using the coordinates of the lower right quarter of every slice where these vertebrae were present. The result was a collection of 536 slices for each patient (a total of 30,016 slices).

To fully automate analysis without labeling target structures, two classification models were developed using a 3D lateral cephalogram. The first model used resized and cropped ROI from the original image as input to classify C2–C4 vertebrae views. Operating on fixed-sized images (344 × 350 pixels), it determined the presence or absence of the preferred view. The output, containing slices with preferred views, fed into the second CNN model, predicting the three growth phases. For training the first CNN model, 638 slices were utilized. About 20% (127 slices) were designated for validation, and the rest were employed for training. Using the Keras library, a CNN classification model was constructed to distinguish between preferred and non-preferred vertebrae views. The model, organized in a "Sequential" container, started with a convolutional layer featuring 32 filters, a (3, 3) kernel size, and a specified input shape. Non-linearity was introduced through the "ReLU" activation function, followed by a 2 × 2 max-pooling layer to reduce spatial dimensions and computational complexity. Subsequently, a "Flatten" layer converted 2D feature maps into a 1D vector, leading to a fully connected layer with 64 units and a ReLU activation function. The final dense layer, utilizing the "Sigmoid" activation function, produced probability scores for each class. The model was compiled with "categorical cross-entropy" loss and "adam" optimizer. The output comprised 1705 slices, with 88 slices reserved for testing the second model, representing growth Phases I, II, and III.

To train the second CNN model, 1617 slices were randomly split into training (1294 or 80%) and validation (323 or 20%)

datasets. To avoid data leakage, all preprocessing steps were independently applied to training and validation datasets. Moreover, to address overfitting, dropout layers were incorporated in both CNN models and early stopping was implemented by monitoring validation loss.

The second model replicated the first's architecture but differed by removing "dropout" in the third hidden layer, adjusting epochs to 25, and utilizing "sparse categorical cross-entropy" with "softmax" activation. Epoch selection involved a grid search for optimal hyperparameters. Evaluation involved testing with 88 unseen slices that were not used for training the model, using multi-class metrics after model training.

The consistency between the two raters for classifying the growth phase was assessed using Cohen's Kappa, a measure of inter-rater reliability (IRR). The IRR was measured using Python and the "sklearn.metrics" library.

Statistical analysis

Classification accuracy measures were used to evaluate outcomes from the validation image set. In ML, different evaluation metrics are applied according to the type of problem. Accuracy, precision, recall, and the F1-score are used for classification tasks. As this study was based on a classification task, the evaluation criteria of accuracy, precision, recall, and the F1-score were used to evaluate the classification performance of the proposed model. A confusion matrix was used to calculate these values.[20] The confusion matrix has true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values. The equations for accuracy, recall, precision, and the F1-score, which are performance evaluation metrics, are provided below:

Accuracy = (TP + TN)/TP +TN + FP + FN

Recall = TP/(TP + FN)

Precision = TP/(TP + FP)

F1-score = 2 × (precision × recall)/(precision + recall)

RESULTS

[Table 1] summarizes the descriptive characteristics of the images and growth phases included in the study. CBCT images belonging to 56 patients (consisting of 536 slices per patient) were first categorized into three growth phases by two orthodontists with a strong IRR of 89.3%. [Table 2] demonstrates the performance of the first CNN model to predict preferred versus non-preferred views of C2–C4 vertebrae on a new set of images. The training and validation accuracies were found to be 91.78% and 88.19%, respectively. According to the table, all slices of new images, including a

Table 1: Descriptive information of the included images.

Growth phase	Number of patients	Age	Number of slices
	n (%)	(Mean±SD)	*n* (%)
I	18 (32)	8 years and 9 month±1 year and 5 months	536 (31.4)
II	15 (27)	11 years±9 months	527 (49)
III	23 (41)	13 years and 7 months±1 year and 3 months	642 (37.6)

SD: Standard deviation

Table 2: Model performance of detecting ROI on the test dataset.

	Predicted ROI[a]	
	Not preferred	**Preferred**
Actual (true) ROI		
Not preferred	103	72
Preferred	0	41

[a]ROI: Region of interest

good vision of vertebrae for classification (n = 41), could be predicted correctly.

[Table 3] demonstrates the multi-class classification metrics applied to the validation dataset and a group of 88 images as a new unseen dataset. The overall accuracy on this set of new slices was found to be 84%. The average classification accuracy of our CNN-based DL model was 98.92% and 95.79% on the training and validation datasets, respectively.

DISCUSSION

In this study, CNN models were designed to classify images according to the presence or absence of the ROI and then into three phases of growth. The annotating step was skipped in the proposed model, which resulted in a more time-efficient image pre-processing. To fully automate the process of CVM classification, a recent study by Atici *et al.*[21] was conducted. They proposed an innovative, custom-designed deep CNN to detect and classify the CVM stages. A layer of tunable directional filters was applied to fully automate the procedure, and they achieved a validation accuracy of 84.63% in CVM stage classification using 1018 cephalometric images from 56 patients. They stated that this level of accuracy was higher compared to other DL models investigated. Our proposed fully automated model was successful in determining the growth phase of patients using the CVM staging with a validation accuracy of 95.79%, which is higher compared to Atici *et al.* findings.[21] This can be due to the higher resolution and accuracy of the input images in our study, which enhances the training accuracy of the model.

Table 3: Model performance on validation and test datasets for categorizing slices into three growth phases.

Growth phase	Test data				Validation data			
	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
I	0.77	1.00	0.87	0.84	0.97	0.97	0.97	0.96
II	1.00	0.71	0.83		0.94	0.93	0.93	
III	0.83	0.77	0.80		0.96	0.97	0.96	

F1: F-measure.

Depending on the task to be performed, various architectures of CNN models have been proposed so far. For instance, Makaremi *et al.* utilized a semi-automatic CNN-based model to assess the maturation of cervical vertebrae; however, it needed manual segmentation of the region of interest.[22] Since then, many novel methods of image segmentation based on fully convolutional network (FCN) have been utilized for medical image analysis.[23,24] In a study conducted by Seo *et al.*, the performance of six CNN-based DL models was evaluated and compared for CVM analysis on conventional 2D cephalometric images. Inception-ResNet-v2 demonstrated the highest classification accuracy due to its capability of focusing on all three vertebrae compared to other DL models. They stated that most studied DL techniques classify CVM by focusing on a specific area of the cervical vertebrae. Thus, they suggested that the application of high-quality input data and better-performing CNN architectures that are capable of segmenting images will help in creating models with higher performance.[25]

Our study used CBCT slices of the vertebrae to determine the skeletal age of the patients. CBCT accuracy and reliability in several aspects of dentistry, such as assessment of tumor lesions, orthognathic surgery planning, and implant placement, have been reported.[26] There is universal agreement that CBCT images are more accurate compared to 2D cephalometrics for craniofacial studies.[27,28] This can be an explanation for the higher amount of accuracy our model achieved. A recent systematic review by Rossini *et al.*[29] also showed that 3D cephalometric analysis outperforms the conventional 2D cephalometrics in terms of accuracy and reproducibility.[17] However, the amount of radiation exposure, which is higher in comparison to a 2D cephalogram, is the biggest controversy about its use in dental imaging.[30] It is suggested that CBCT images can be a valid and useful tool for the assessment of skeletal age using CVM, although they should not be used solely for that purpose.[31] CBCT imaging for CVM analysis is particularly beneficial in patients with craniofacial fractures, cleft lip/palate deformities, temporomandibular joint concerns, or obstructive sleep apnea. Despite increased radiation exposure, the clinical benefits make CBCT a valuable tool for these specific patients.[32-34]

Our model accuracy in predicting a group of unseen images was greater than <80%, with the highest performance at Phase I (F1-score:87%), which is consistent with the previous studies. According to the literature, CVM stages are sometimes difficult to differentiate according to the continuous nature of morphological changes in cervical vertebrae.[35] Thus, the CS 1 and CS 6 stages are easier to identify. Our model performed well in predicting the CS3 (phase II) with an F1 score of 85%. This was in contrast with a study conducted by Zhou *et al.*[36] who reported an F1-score of 31% for diagnosing the pubertal spurt on the cephalometric radiograph. As the authors mentioned, this could be due to their insufficient training set of CS3 for growth spurt is short and difficult to find in clinical practice.

In contrast to previous studies, we only classified patients according to the three growth phases. However, according to the main clinical application of CVM staging, which is to determine the growth potential of the patients, our classification method can be justified in terms of orthodontic treatment planning and correction of jaw discrepancies.

CONCLUSION

Our proposed model could automatically detect C2–C4 required for CVM staging and accurately classify images into three growth phases without the need for annotating the shape and configuration of vertebrae. This will result in the development of a fully automatic and less complex system with reasonable performance. Classical methods are time-consuming and prone to inter- and intra-rater variability; thus, using methods that automate this process will be of value.

REFERENCES

1. Korde SJ, Daigavane P, Shrivastav S. Skeletal maturity indicators-review article. Int J Sci Res 2015;6:361-70.
2. Baldin CC, Kitt M, Costa AL, Yasuda CL, Cendes F, Nahás-Scocate AC. Evaluation of the skeletal maturation of cervical vertebrae with magnetic resonance imaging: A pilot study. Braz J Oral Sci 2017;16:e17060.
3. Dzemidzic V, Sokic E, Tiro A, Nakas E. Computer based assessment of cervical vertebral maturation stages using digital lateral Cephalograms. Acta Inform Med 2015;23:364-8.
4. Baccetti T, Franchi L, McNamara JA. The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in Dentofacial orthopedics. Semin Orthod 2005;11:119-29.
5. Cericato GO, Bittencourt MA, Paranhos LR. Validity of the assessment method of skeletal maturation by cervical vertebrae: A systematic review and meta-analysis. Dentomaxillofac Radiol 2015;44:20140270.
6. Bunch PM, Altes TA, McIlhenny J, Patrie J, Gaskin CM. Skeletal development of the hand and wrist: Digital bone age companion-a suitable alternative to the Greulich and Pyle atlas for bone age assessment? Skelet Radiol 2017;46:785-93.
7. 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;49:20190441.
8. Freitas LM, Freitas KM, Pinzan A, Janson G, Freitas MR. A comparison of skeletal, dentoalveolar and soft tissue characteristics in white and black Brazilian subjects. J Appl Oral Sci 2010;18:135-42.
9. Souza KR, Oltramari-Navarro PV, Navarro RL, Conti AC, Almeida MR. Reliability of a method to conduct upper airway analysis in cone-beam computed tomography. Braz Oral Res 2013;27:48-54.
10. Sohrabi A, Babay Ahari S, Moslemzadeh H, Rafighi A, Aghazadeh Z. The reliability of clinical decisions based on the cervical vertebrae maturation staging method. Eur J Orthod 2016;38:8-12.
11. 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;140:182-8.
12. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: Chances and challenges. J Dent Res 2020;99:769-74.
13. Kim EG, Oh IS, So JE, Kang J, Le VN, Tak MK, *et al.* Estimating cervical vertebral maturation with a lateral Cephalogram using the convolutional neural network. J Clin Med 2021;10:5400.
14. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221-48.
15. Huang M, Huang S, Zhang Y, Bhatti U. Medical image segmentation using deep learning with feature enhancement. IET Image Process 2020;14:3324-32.
16. Kok H, Acilar AM, Izgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. Prog Orthod 2019;20:41.
17. Gribel BF, Gribel MN, Frazäo DC, McNamara JA Jr., Manzi FR. Accuracy and reliability of craniometric measurements on lateral cephalometry and 3D measurements on CBCT scans. Angle Orthod 2011;81:26-35.
18. Grauer D, Cevidanes LS, Proffit WR. Working with DICOM craniofacial images. Am J Orthod Dentofacial Orthop 2009;136:460-70.
19. Alqerban A, Willems G, Bernaerts C, Vangastel J, Politis C, Jacobs R. Orthodontic treatment planning for impacted maxillary canines using conventional records versus 3D CBCT. Eur J Orthod 2014;36:698-707.
20. Ertaş K, Pence I, Cesmeli MS, Ay ZY. Determination of the stage and grade of periodontitis according to the current classification of periodontal and peri-implant diseases and conditions (2018) using machine learning algorithms. J Periodontal Implant Sci 2023;53:38-53.
21. Atici SF, Ansari R, Allareddy V, Suhaym O, Cetin AE, Elnagar MH. Fully automated determination of the cervical vertebrae maturation stages using deep learning with directional filters. PLoS One 2022;17:e0269198.
22. Makaremi M, Lacaule C, Mohammad-Djafari A. Deep learning and artificial intelligence for the determination of the cervical vertebrae maturation degree from lateral radiography. Entropy 2019;21:1222.
23. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multiscale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers Med Image Anal 2018;2:21-45.
24. Fourcade A, Khonsari RH. Deep learning in medical image analysis: A third eye for doctors. J Stomatol Oral Maxillofac Surg 2019;120:279-88.
25. Seo H, Hwang J, Jeong T, Shin J. Comparison of deep learning models for cervical vertebral maturation stage classification on lateral cephalometric radiographs. J Clin Med 2021;10:3591.
26. Pinsky HM, Dyda S, Pinsky RW, Misch KA, Sarment DP. Accuracy of three-dimensional measurements using cone beam CT. Dentomaxillofac Radiol 2006;35:410-6.
27. Korbmacher H, Kahl-Nieke B, Schollchen M, Heiland M. Value of two cone-beam computed tomography systems from an orthodontic point of view. J Orofac Orthop 2007;68:278-89.
28. Ludlow JB, Gubler M, Cevidanes L, Mol A. Precision of cephalometric landmark identification: cone-beam computed tomography vs. conventional cephalometric views. Am J Orthod Dentofacial Orthop 2009;136:312.e1-10.
29. Rossini G, Cavallini C, Cassetta M, Barbato E. 3D cephalometric analysis obtained from computed tomography. Review of the literature. Ann Stomatol (Roma). 2011;2:31-9.
30. Valizadeh S, Tavakkoli MA, Karimi Vasigh H, Azizi Z, Zarrabian T. Evaluation of cone beam computed tomography (CBCT) system: Comparison with intraoral periapical radiography in proximal caries detection. J Dent Res Dent Clin Dent Prospects 2012;6:1-5.
31. Bonfim MA, Costa AL, Fuziy A, Ximenez ME, Cotrim-Ferreira FA, Ferreira-Santos RI. Cervical vertebrae maturation index estimates on cone beam CT: 3D reconstructions vs sagittal sections. Dentomaxillofac Radiol 2016;45:20150162.
32. Kapila SD, Nervina JM. CBCT in orthodontics: Assessment of treatment outcomes and indications for its use. Dentomaxillofac Radiol 2015;44:20140282.

33. Li G. Patient radiation dose and protection from cone-beam computed tomography. Imaging Sci Dent 2013;43:63-9.
34. Colceriu-Şimon IM, Băciuţ M, Ştiufiuc RI, Aghiorghiesei A, Ţărmure V, Lenghel M, *et al.* Clinical indications and radiation doses of cone beam computed tomography in orthodontics. Med Pharm Rep 2019;92:346-51.
35. McNamara JA Jr., Franchi L. The cervical vertebral maturation method: A user's guide. Angle Orthod 2018;88:133-43.
36. Zhou J, Zhou H, Pu L, Gao Y, Tang Z, Yang Y, *et al.* Development of an artificial intelligence system for the automatic evaluation of cervical vertebral maturation status. Diagnostics (Basel) 2021;11:2200.