

Article

Development of an Artificial Intelligence System for the Automatic Evaluation of Cervical Vertebral Maturation Status

Jing Zhou ¹, Hong Zhou ¹, Lingling Pu ¹, Yanzi Gao ¹, Ziwei Tang ¹, Yi Yang ¹, Meng You ², Zheng Yang ³, Wenli Lai ¹ and Hu Long ^{1,*}

¹ State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, China; nkuzhoujing@163.com (J.Z.); zh17736839697@163.com (H.Z.); dangdapll@163.com (L.P.); yanmei100@163.com (Y.G.); tziwei2016@163.com (Z.T.); yy18583233717@163.com (Y.Y.); wenlilai@scu.edu.cn (W.L.)

² State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Department of Oral Radiology, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, China; youmeng@scu.edu.cn

³ State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Department of General Dentistry, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, China; hxkqyx@163.com

* Correspondence: hlong@scu.edu.cn



Citation: Zhou, J.; Zhou, H.; Pu, L.; Gao, Y.; Tang, Z.; Yang, Y.; You, M.; Yang, Z.; Lai, W.; Long, H. Development of an Artificial Intelligence System for the Automatic Evaluation of Cervical Vertebral Maturation Status. *Diagnostics* **2021**, *11*, 2200. <https://doi.org/10.3390/diagnostics11122200>

Academic Editor: Keun Ho Ryu

Received: 21 October 2021

Accepted: 21 November 2021

Published: 25 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Cervical vertebral maturation (CVM) is widely used to evaluate growth potential in the field of orthodontics. This study is aimed to develop an artificial intelligence (AI) system to automatically determine the CVM status and evaluate the AI performance. Methods: A total of 1080 cephalometric radiographs, with the age of patients ranging from 6 to 22 years old, were included in the dataset (980 in training dataset and 100 in testing dataset). Two reference points and thirteen anatomical points were labelled and the cervical vertebral maturation staging (CS) was assessed by human examiners as gold standard. A convolutional neural network (CNN) model was built to train on 980 images and to test on 100 images. Statistical analysis was conducted to detect labelling differences between AI and human examiners, AI performance was also evaluated. Results: The mean labelling error between human examiners was 0.48 ± 0.12 mm. The mean labelling error between AI and human examiners was 0.36 ± 0.09 mm. In general, the agreement between AI results and the gold standard was good, with the intraclass correlation coefficient (ICC) value being up to 98%. Moreover, the accuracy of CVM staging was 71%. In terms of F1 score, CS6 stage (85%) ranked the highest accuracy. Conclusions: In this study, AI showed a good agreement with human examiners, being a useful and reliable tool in assessing the cervical vertebral maturation.

Keywords: artificial intelligence; cervical vertebral maturation; skeletal age; deep learning; convolutional neural network; orthodontics

1. Introduction

Dental malocclusion, with a prevalence of 20–83% among both adolescents and adults, manifests as misaligned teeth resulting in poor masticatory function and esthetic problems [1–4]. Of particular, among adolescents, early interventions could eliminate or intercept the development of malocclusion [5], e.g., mandibular advancement therapy for adolescents with mandibular retrusion [6]. To ensure the success of early interventions, meticulous and correct assessment of growth potential and the timing of growth spurt is very important. Traditionally, skeletal age was used to assess the growth potential, among which hand-wrist bone age and cervical vertebral maturation (CVM) staging were widely used by dental practitioners [7,8]. Since hand-wrist radiographs require additional radiographic examinations, both orthodontists and patients are reluctant to use this method in

orthodontic practice. Fortunately, since CVM staging could be assessed in lateral cephalograms that are required for orthodontic diagnosis, CVM staging has been gaining more and more popularity among orthodontists. CVM is determined through the morphological changes of the bodies of the second, third and fourth cervical vertebrae (C2–C4) on lateral cephalograms [8]. However, it is difficult and time-consuming for practitioners to determine skeletal maturation and growth spurt through CVM staging correctly.

Nowadays, artificial intelligence (AI) has gained a giant leap in dentistry, assisting clinicians in a variety of fields, e.g., detection of periapical lesions and root fractures, optimizing implant designs, diagnosis of oral cancer [9–11]. Deep learning is a key field of AI, which uses a learning model to extract features of the labelled dataset and eventually can predict labels on a new dataset [12]. Mimicking the way that human brain neurons signal to another, neural networks are widely used in deep learning.

Over the past decade, several researchers have explored the promising application of AI in analysis of cephalometric images. Hwang et al. used a customized You-Only-Look-Once version 3 algorithm (YOLOv3) to detect 80 landmarks on 1028 cephalograms. The mean detection error between AI and human was up to 1.46 ± 2.97 mm [13]. Larson et al. used a deep residual network to estimate skeletal maturity on pediatric hand radiographs and reported a similar accuracy to an expert radiologist [14]. Moreover, Kok et al. and Amasya et al. pioneered the applications of neural networks in CVM assessment [15,16]. However, they included 300 and 647 images in the aforementioned two studies, respectively. Nevertheless, they didn't list the distribution of the dataset. We inferred that there were 50 and 108 images on average for AI learning for each CVM stage. Considering the short duration of growth and development, the number of CS 3 which means the growth spurt was insufficient for AI to learn. Therefore, the generalization of their results is limited by small sample sizes ($n=300$ and 647 for Kok et al. and Amasya et al., respectively) and only studying the agreement between AI and actual CVM staging.

Therefore, in this study, we developed an AI system to automatically assess CVM based on a larger sample size ($n=1080$) and to assess other indices (e.g., sensitivity), in order to comprehensively evaluate the generalization potential of an AI system for CVM staging.

2. Materials and Methods

This study was approved by Ethical Committee of West China Hospital of Stomatology, Sichuan University. For better understanding, we reported this study according to a checklist prompted by Schwendicke et al. [17]. The experimental design of the study is summarized in Figure 1. Briefly, 1080 images were selected and labelled by human examiners. After pre-processing, they were divided into training and testing dataset. The training dataset of 980 labelled images was input to an AI-based system for machine learning. Finally, 100 images were used to test the AI system and statistical analysis was conducted to evaluate AI performance.

2.1. Patients and Dataset

Cephalometric radiographs of patients with a chronological age between 6 and 22 years old were obtained from Department of oral radiology, West China Hospital of Stomatology, Sichuan University. Only images with clearly identified contours of the second (C2), third (C3) and fourth (C4) cervical vertebrae were included. Patients with congenital diseases were excluded from the dataset. In case of several images from the same patient, we deleted the repeated ones based on their medical record number. Therefore, the images in the training set and the testing set were from different patients. Finally, 1080 images (jpg format) were included and randomly assigned into training and testing dataset (Table 1).

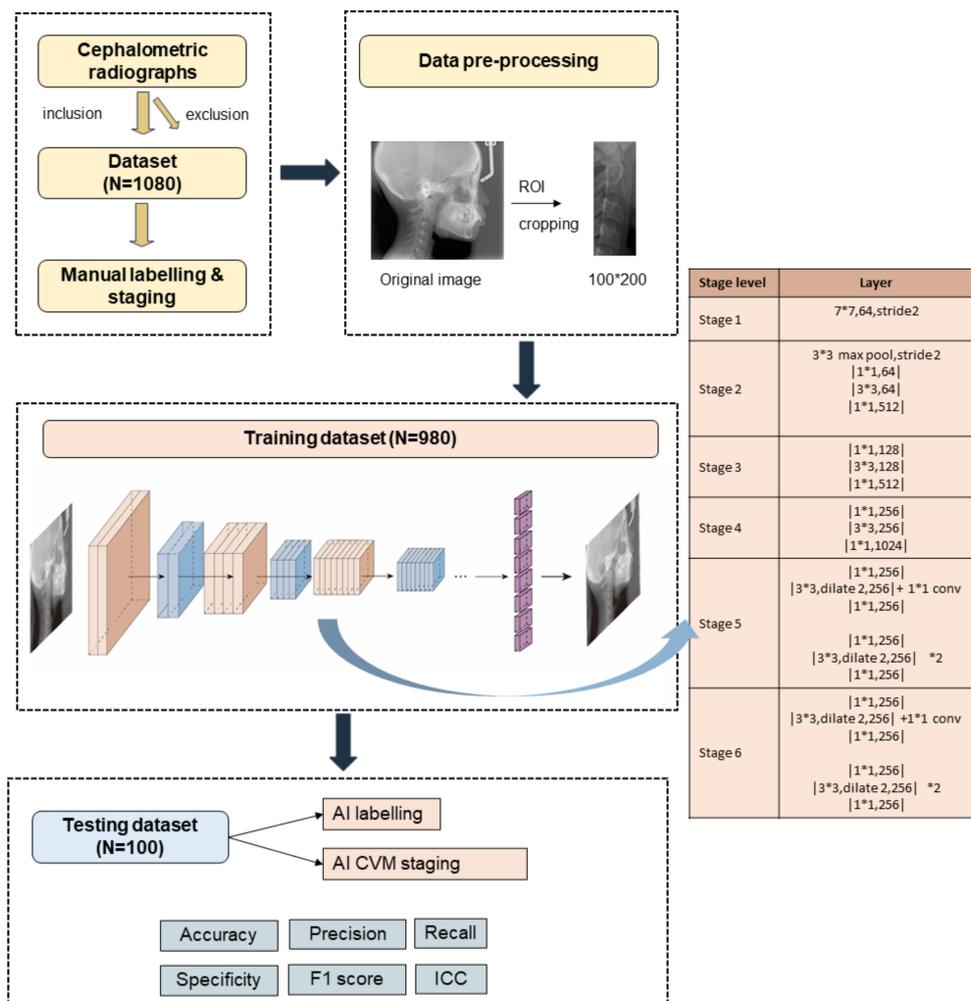


Figure 1. The experimental design of the study. Step 1: inclusion and exclusion. Step 2: data pre-processing. Step 3: model training and testing. Step 4: performance evaluation.

Table 1. The age, gender and CVM staging distribution of the dataset.

	Training	Testing
Age	12.02 ± 4.71	14.27 ± 4.78
Gender (M/F)	432/548	43/57
CS1	242	13
CS2	214	11
CS3	164	5
CS4	108	24
CS5	52	12
CS6	200	35
Total	980	100

2.2. Manual CVM Staging

The CVM staging (CS) of 1080 images were identified by two examiners independently and in duplicate (J.Z. & H.Z., who had three years' experience in CVM assessment) according to Baccetti methods [8]. Briefly, the CVM assessment was based on the morphology changes of C2, C3 and C4. At CS1, the lower border of C2, C3 and C4 were flat and presented no concavity. Moreover, the shape of C3 and C4 were trapezoid in shape. With growth and development, the concavity of lower border of C2, C3, C4 became more obvious and the height of C3 and C4 increased, therefore the shape of C3, C4 changed to horizontal rectangles, squares or vertical rectangles. Disagreements were resolved by a

third examiner (H.L., who had 10 years' relevant experience). The final CVM stages were served as a gold standard. The CS of both the training and testing datasets are displayed in Table 1. For inter-rater reliability, the kappa value was 0.86 and the intraclass correlation efficient (ICC) value was 0.98.

2.3. Manual Labelling

The training and testing dataset was uploaded to an open-source annotation tool named LabelMe (<https://github.com/wkentaro/labelme>; accessed on 18 February 2021). Then, as presented in Figure 2, two reference landmarks and thirteen anatomic landmarks were manually labelled on each cephalometric image by an examiner (J.Z.) in duplicate after a three-month interval. The distance between two reference points is 10 mm. These landmarks were used for linear and ratio measurements. The definitions of landmarks and measurements are shown in Table 2. Then the labelled films were saved as .json format and input to the model for training and testing.

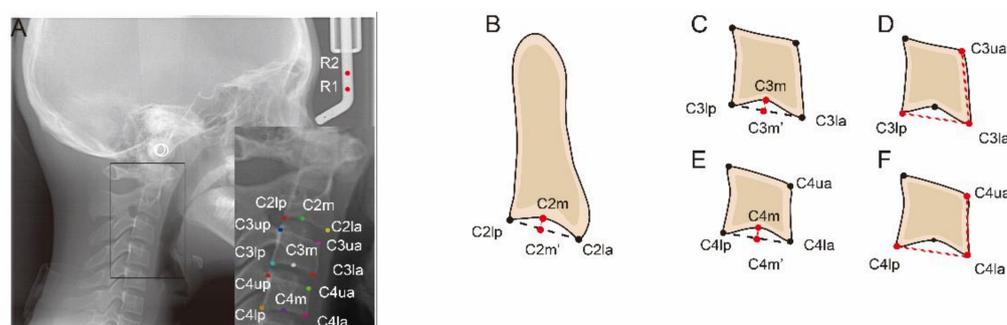


Figure 2. Schematic representation of anatomic landmarks and liner measurements for the determination of cervical vertebral morphology. (A) Two reference points (R1 and R2) and thirteen anatomic landmarks of C2, C3 and C4. The distance between the two reference points (R1 and R2) is 10 mm. (B) $C2Conc = C2m - C2m'$. (C) $C3Conc = C3m - C3m'$. (D) $C3BAR = (C3lp - C3la)/(C3ua - C3la)$. (E) $C4Conc = C4m - C4m'$. (F) $C4BAR = (C4lp - C4la)/(C4ua - C4la)$.

Table 2. Landmarks and measurements used to determine cervical vertebral morphology.

Landmarks/ Measurements	Definition
C2lp	The most posterior point of C2 on the lower border
C2la	The most anterior point of C2 on the lower border
C2m	The deepest point of the concavity at the lower border of C2
C2Conc	The distance between C2m and the line connecting C2lp and C2la
C3up	The most posterior point of C3 on the upper border
C3ua	The most anterior point of C3 on the upper border
C3lp	The most posterior point of C3 on the lower border
C3la	The most anterior point of C3 on the lower border
C3m	The deepest point of the concavity at the lower border of C3
C3Conc	The distance between C3m and the line connecting C3lp and C3la
C3BAR	Ratio between the length of the base (distance C3lp – C3la) and the anterior height (distance C3ua – C3la) of the body of C3.
C4up	The most posterior point of C4 on the upper border
C4ua	The most posterior point of C4 on the upper border
C4lp	The most posterior point of C4 on the lower border
C4la	The most anterior point of C4 on the lower border
C4m	The deepest point of the concavity at the lower border of C4
C4Conc	The distance between C4m and the line connecting C4lp and C4la
C4BAR	ratio between the length of the base (distance C4lp – C4la) and the anterior height (distance C4ua – C4la) of the body of C4

2.4. Model Training and Testing

To reduce interference from other anatomic structures, a final ROI (region of interest) included all part of C2–C4 was cropped on images with a size of $100^* \times 200$ pixels (Figure 1). The preprocessing images were input to a convolutional neural network (CNN) for training and testing. The experiments were performed on Intel core i7 quadra core processor with Nvidia 1080 graphics card. The models were developed with the pytorch libraries using the Python programming language. We used a Detnet architecture with relu activation function, adam optimization and MSE loss function to conduct machine learning. Based on resnet50, Detnet introduced the extra stages in the backbone and was more powerful in locating large objects and finding small objects [18]. The model consisted of 58 convolution layers and 1 fully connected layer. In the training stage, we set the training epochs to 200 and batch size as 32. The learning rate was set as 0.0001 and epsilon was 1×10^{-8} . After training on 980 radiographic images, the CNN model achieved satisfactory results for new images. To test AI performance, 100 new images were uploaded. Automatically, the CNN model labelled 2 reference landmarks and 13 anatomic landmarks on each image. These landmarks were output in fully connected layer. Afterwards, the linear and ratio measurements (C2conc, C3conc, C4conc, C3BAR, C4BAR) were calculated and the CVM stage was output according to a workflow shown in Figure 3. In brief, these measurements were input and estimated with thresholds from Baccetti's original data [8].

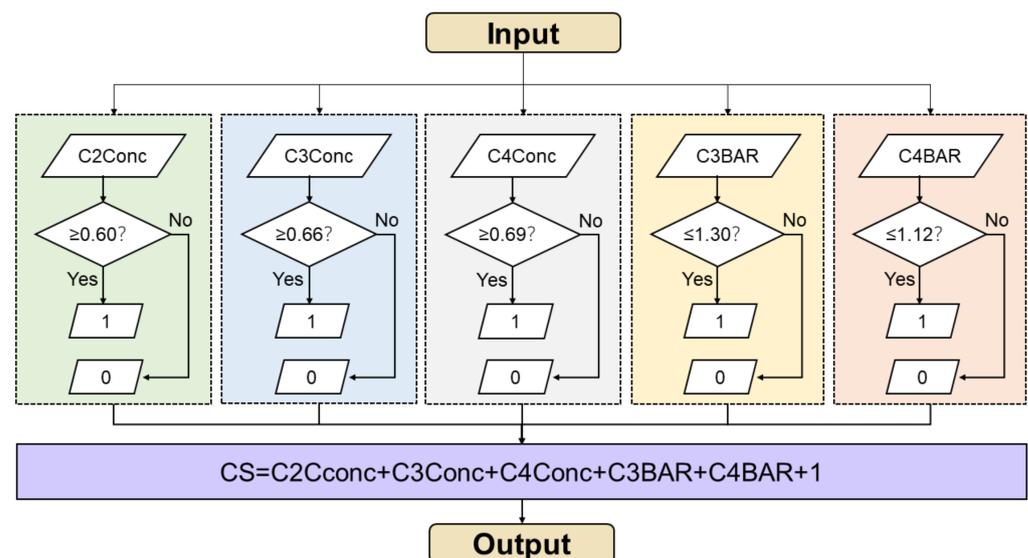


Figure 3. Workflow of the AI system to assess the CVM stage. Measurements are input and reassigned according to thresholds. The CVM stages were calculated and output.

2.5. Statistical Analysis

All the statistical analysis were performed using SPSS Statistics Version 22.0. The inter-rater reliability of the manual CVM staging was determined through kappa coefficient and intra-class correlation coefficient (ICC) test. The intra-rater reliability of the labelling work was calculated through mean difference and the mean value of labelling was served as gold standard. The disagreement between AI and manual labelling were calculated in terms of distances measured in millimeter scales. Moreover, both linear measurements (C2Conc, C3Conc, C4Conc) and ratio measurements (C3BAR, C4BAR) were performed. The differences of the aforementioned measurements between AI and manual performance were compared. As for CVM staging difference between AI and gold standard, the overall accuracy and ICC were calculated and the accuracy of each stage was calculated in precision, recall, specificity, F1 score [19]. The formulas of these metrics are listed as follows:

Accuracy = $\frac{TP+TN}{TP+TN+FN+FP}$, which is a general evaluation of AI performance.

Precision = $\frac{TP}{TP+FP}$ = positive predictive value (PPV), which presents the ability of AI to correctly predict positives.

Recall = $\frac{TP}{TP+FN}$ = sensitivity, which reflects ability of AI to find all the positive samples.

Specificity = $\frac{TN}{TN+FP}$, which reflects the ability of AI to find all the negative samples.

F1 score = $\frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$, which weight precision and recall harmoniously to completely evaluate AI performance.

TP (true positive) is the number of AI correctly predicted images and FP (false positive) is the number of AI incorrectly predicted image while human classified as positive in a binary task. In similar, TN (true negative) and FN (false negative) are the number of AI correctly predicted images and incorrectly predicted image while human classified as negative, respectively.

3. Results

3.1. Evaluation of Labelling

The mean differences between the first and second manual labelling were 0.48 ± 0.12 mm. Moreover, the mean differences between the manual labelling (gold standard) and the AI labelling were 0.36 ± 0.09 mm. Two examples of the landmarks that AI (points in red) and human (points in green) has labelled were displayed in Figure 4. To visualize and evaluate the error pattern in two-dimensional space, scattergrams were depicted. As displayed in Figure 5, the AI and human labelling were well matched. The differences between the first and second manual labelling and between the manual and AI labelling for each point are detailed in Table 3.

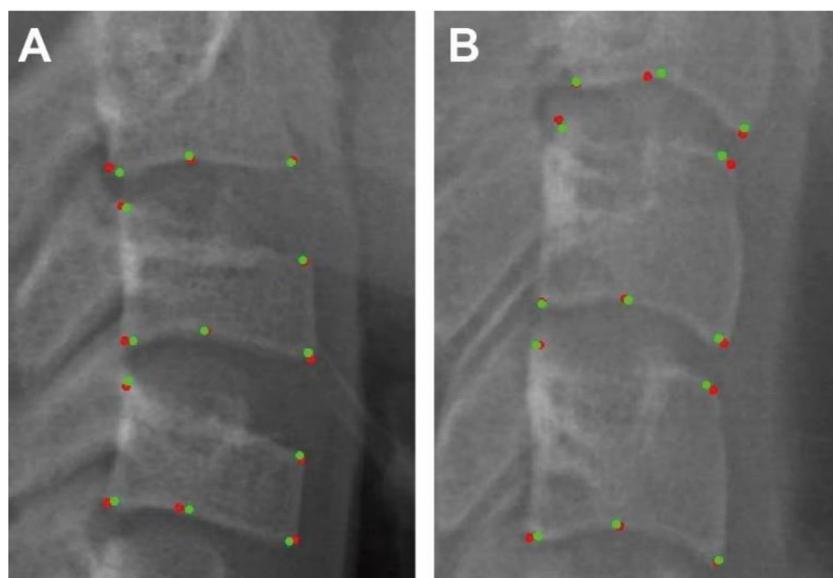


Figure 4. CNN output results: Anatomic landmarks that AI (points in red) and human (points in green) labelled in testing dataset. (A). AI and human labelled landmarks for CS 3 (B). AI and human labelled landmarks for CS6.

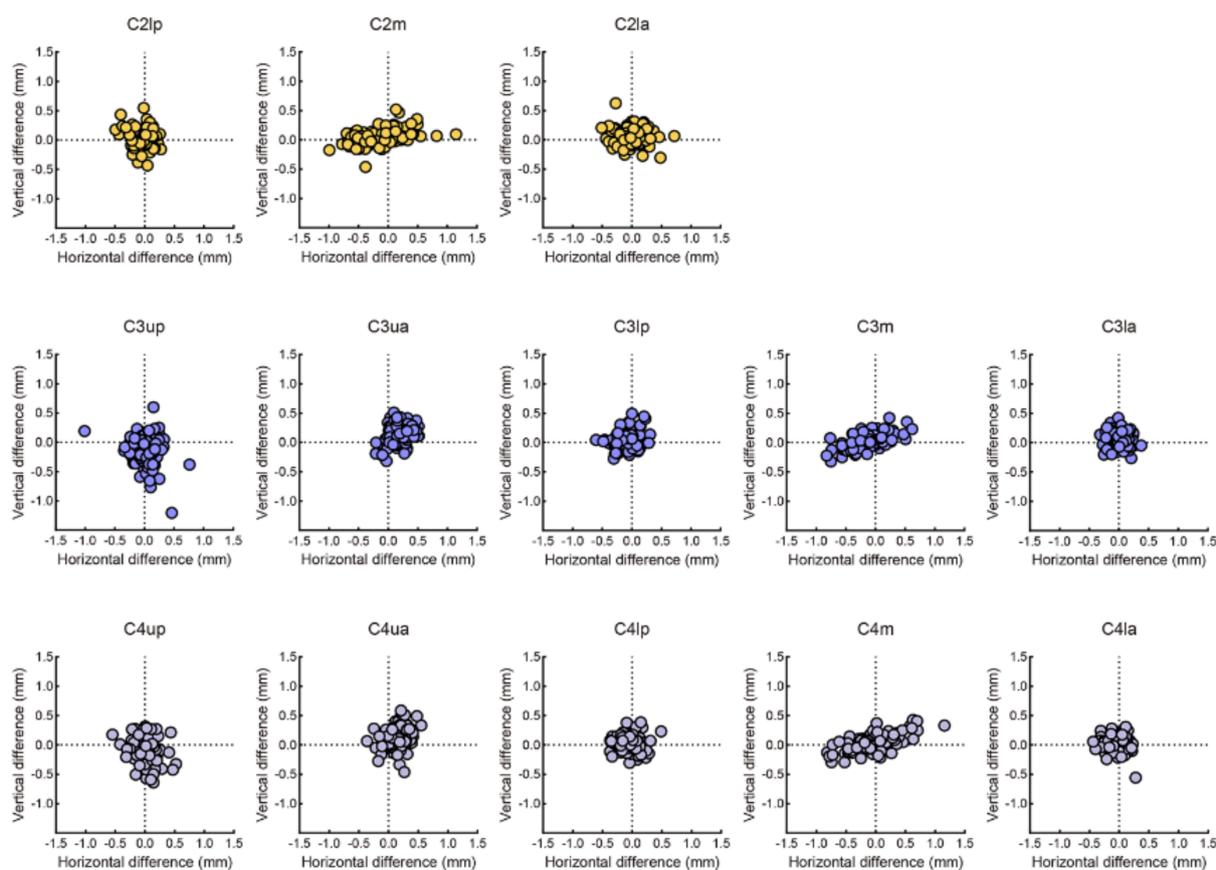


Figure 5. Scattergram representation of the labelling difference between AI and human. The mean differences between AI and human of C2 landmarks (line 1) were 0.29 ± 0.25 mm, 0.47 ± 0.33 mm, 0.35 ± 0.27 mm, respectively. The mean differences between AI and human of C3 landmarks (line 2) were 0.48 ± 0.38 mm, 0.42 ± 0.24 mm, 0.24 ± 0.15 mm, 0.37 ± 0.28 mm, 0.24 ± 0.17 mm, respectively. The mean differences between AI and human of C4 landmarks (line 3) were 0.42 ± 0.36 mm, 0.43 ± 0.27 mm, 0.25 ± 0.21 mm, 0.46 ± 0.27 mm, 0.27 ± 0.18 mm, respectively.

Table 3. Landmarks and measurements used to determine cervical vertebral morphology.

Landmarks	Between Human Examiners		Between AI and Human Examiner	
	Mean \pm SD (mm)		Mean \pm SD (mm)	
C2lp	0.38	0.23	0.29	0.25
C2lm	0.70	0.44	0.47	0.33
C2la	0.41	0.28	0.35	0.27
C3up	0.56	0.40	0.48	0.38
C3ua	0.54	0.28	0.42	0.24
C3lp	0.41	0.24	0.24	0.15
C3m	0.57	0.37	0.37	0.28
C3la	0.32	0.19	0.24	0.17
C4up	0.50	0.29	0.42	0.36
C4ua	0.53	0.30	0.43	0.27
C4lp	0.37	0.21	0.25	0.21
C4m	0.62	0.45	0.46	0.27
C4la	0.35	0.20	0.27	0.18
Total	0.48	0.12	0.36	0.09

3.2. Evaluation of Measurements

The linear and ratio measurements were calculated according to manual- and AI-labelled anatomic points respectively. As displayed in Figure 6, the results revealed that they were well-matched and fitted into a linear function ($Y = X$) ($R^2 = 0.93, 0.94, 0.93, 0.97$

and 0.95, respectively; all $p < 0.001$). Moreover, the ICC for all the five measurements were greater than 0.90, indicating that AI performed comparably with human examiners.

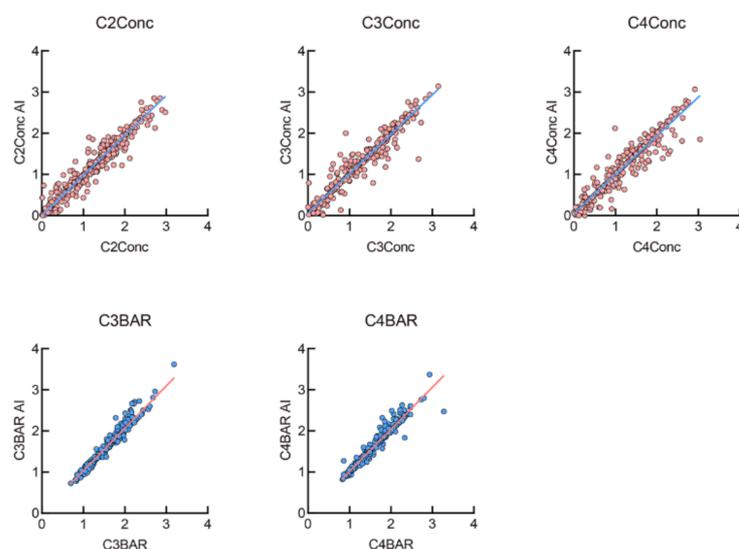


Figure 6. The linear and ratio measurements difference between AI and human. They were well-matched and fitted into $Y = X$, $R^2 = 0.93, 0.94, 0.93, 0.97$ and 0.95 , respectively; all $p < 0.001$).

3.3. Evaluation of AI Staging

For manual CVM staging, our results revealed that the inter-rater reliability was 0.86, indicative of perfect agreement between the two examiners. With the manual CVM staging as the gold standard, the 71% general accuracy was observed for AI CVM staging. In terms of F1 score, CS6 ranked the highest accuracy (85%) and CS1 the second. CS3, which indicated the growth spurt, was the lowest (31%). The overall ICC value was 0.98, indicating that the manual and AI CVM staging was in perfect agreement. In particular, the precision, recall and specificity ranging from 25–100%, 36–100% and 84–100% respectively. More details were listed in Table 4.

Table 4. The evaluation of CVM staging through AI.

	Precision/PPV	Recall/Sensitivity	Specificity	F1 Score
CS1	0.67	0.92	0.93	0.77
CS2	1.00	0.36	1.00	0.53
CS3	0.25	0.40	0.94	0.31
CS4	0.83	0.63	0.96	0.71
CS5	0.46	1.00	0.84	0.63
CS6	1.00	0.74	1.00	0.85
Accuracy			0.71	
ICC			0.98	

4. Discussion

The assessment of growth and development plays a critical role in orthodontic treatment planning for growing patients. Appropriate growth prediction and prudent treatment timing benefit adolescents with skeletal discrepancy through growth modification by utilizing growth potential. Several indicators have been used to determine the growth potential, i.e., chronological age, body height, sexual maturity, dental age and skeletal age. Great individual variations have been discovered in chronological age, increase in body height and sexual maturity [20,21]. As for dental age, dentition phase and dental maturity are often used to assess skeletal maturity. However, a large body of evidence indicated that dental age is not recommended to assess skeletal maturity or to determine growth spurt [22–24]. Due to high reliability, skeletal age is more popular in the dental community. Larson et al.

applied CNN network in hand-wrist radiographs method and reported a good agreement. Compared with the hand-and-wrist method [25,26], CVM staging method based on cervical vertebrae is more popular among orthodontists for no additional radiographic requirements. Since the introduction of a modified CVM staging method by Baccetti et al. [8], this CVM staging has been consistently proved to be reliable [25,27,28].

The CVM stages are determined through the morphological changes of the contours of cervical vertebrae (C2–C4). With the aid of deep learning algorithm, CVM assessment can be more accurate and efficient. As pixel values could be digitally coded, radiology images are easily translated into computer language [29]. Several seminal studies have successfully translated AI applications into identifying cephalometric landmarks [30,31]. The first step of the AI CVM staging was to identify the contours of the cervical vertebrae. Our results revealed that the AI labelling was in almost perfect with the gold standard (manual labelling), with the mean error being 0.36 mm that was even smaller than that (0.48 mm) between the two manual labellings. This suggested that AI labelling was more consistent with manual labelling. It has been revealed that automatic identification of landmarks is considered to be successful if the difference between AI and human is less than 2 mm [32], indicating that the AI labelling in our present study was accurate. Moreover, the mean error (0.36) between the AI and manual labelling was smaller than that reported by Hwang et al. [13] (1.46 mm), which could be attributed to the fact that only cervical vertebral points were labelled in this present study while all cephalometric points were identified in the previous study (Hwang et al.). The precise point labelling ensured the accuracy and precision of linear measurements. Likewise, we found that the five linear and ratio measurements evaluating the morphology of cervical vertebrae through the AI algorithm were highly consistent with those through manual methods.

In this study, we developed a CNN model to label the contours of cervical vertebrae and to assess the CVM staging. As an optimal feature extractor applied at image positions, CNN is highly efficient for image processing [33,34]. The architecture of a typical CNN contains three layers: convolutional layers, pooling layers and fully connected layers. In convolutional layers, features of each image were extracted and organized in feature maps. These similar features were merged into new feature maps in pooling layers. Finally, all features were connected and classified in fully connected layers to output a predicted image [35].

Although six maturation stages are defined by the CVM staging according to different morphological shapes of cervical vertebrae, CVM stages are sometimes still difficult to differentiate since the morphological changes of cervical vertebrae are continuous rather than incremental. Thus, CS1 (no development) and CS6 (maturity) stages are easier to identify while the CS2–CS5 (in the process of development) stages are more difficult to differentiate. Consistently, our results revealed that AI performed best at CS1 (F1 score: 77%) and CS6 (F1 score: 85%). CS3 (growth spurt) showed the lowest F1 score (31%). Another explanation is the insufficient training set of CS3, for growth spurt is short and difficult to encounter in clinical practice. This finding was in accordance with that in Kok et al. [13]. Moreover, we found that the overall ICC between the AI staging and the gold standard was 0.98, indicating that the CNN model in this present study was accurate and precise in the assessment of CVM staging.

One of the limitations of our study was the size of testing dataset. In our study, we used 10% hold-out validation to test AI performance. This may result in over-fitting since the dataset was not properly distributed [36]. This may explain high precision and recall were detected for CS5 (recall: 100%) and CS6 (precision: 100%, specificity: 100%) stages. Therefore, we used F1 score to comprehensively evaluated AI performance.

Another limitation was the number of examiners in labelling. All the images were labelled by examiner 1. To keep labelling consistency, it was done twice after three months. The mean differences between the first and second manual labelling indicated a well consistency. Considering the labelling error, the midpoint of two manual labellings were saved as gold standard.

5. Conclusions

Taken together, we suggest that the AI algorithm described in this study is accurate and reliable in identifying the contours of cervical vertebrae and in CVM staging, with high accuracy (F1 score up to 85%) and perfect agreement with gold standard (ICC = 0.98).

Author Contributions: Conceptualization, J.Z. and H.L.; methodology, J.Z. and H.L.; validation, J.Z. and H.Z.; formal analysis, L.P.; investigation, Y.G., Y.Y. and Z.T.; resources, M.Y.; data curation, M.Y.; writing—original draft preparation, J.Z.; writing—review and editing, H.L.; supervision, W.L.; project administration, Z.Y.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (NSFC, No. 82071147), Research Grant of Health Commission of Sichuan Province (Nos. 19PJ233 and 20PJ090), Sichuan Science and Technology Program (Nos. 2018JY0558 & 2021YJ0428), CSA Clinical Research Fund (CSA-02020-02), and Research and Develop Program, West China Hospital of Stomatology, Sichuan University (No. LCYJ2020-TD-2).

Institutional Review Board Statement: This study was approved by Ethical Committee of West China Hospital of Stomatology, Sichuan University (Protocol number: WCHSIRB-D-2020-151; date of approval: 5 May 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data sharing not applicable.

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

References

1. Zou, J.; Meng, M.; Law, C.S.; Rao, Y.; Zhou, X. Common dental diseases in children and malocclusion. *Int. J. Oral Sci.* **2018**, *10*, 7. [[CrossRef](#)]
2. Shen, L.; He, F.; Zhang, C.; Jiang, H.; Wang, J. Prevalence of malocclusion in primary dentition in mainland China, 1988–2017: A systematic review and meta-analysis. *Sci. Rep.* **2018**, *8*, 4716. [[CrossRef](#)]
3. Akbari, M.; Lankarani, K.B.; Honarvar, B.; Tabrizi, R.; Mirhadi, H.; Moosazadeh, M. Prevalence of malocclusion among Iranian children: A systematic review and meta-analysis. *Dent. Res. J.* **2016**, *13*, 387–395. [[CrossRef](#)]
4. Long, H.; Wang, Y.; Jian, F.; Liao, L.N.; Yang, X.; Lai, W.L. Current advances in orthodontic pain. *Int. J. Oral Sci.* **2016**, *8*, 67–75. [[CrossRef](#)] [[PubMed](#)]
5. Fleming, P.S. Timing orthodontic treatment: Early or late? *Aust. Dent. J.* **2017**, *62*, 11–19. [[CrossRef](#)] [[PubMed](#)]
6. Liu, L.; Zhan, Q.; Zhou, J.; Kuang, Q.; Yan, X.; Zhang, X.; Shan, Y.; Lai, W.; Long, H. A comparison of the effects of Forsus appliances with and without temporary anchorage devices for skeletal Class II malocclusion. *Angle Orthod.* **2020**, *91*, 255–266. [[CrossRef](#)] [[PubMed](#)]
7. Fishman, L.S. Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand-wrist films. *Angle Orthod.* **1982**, *52*, 88–112. [[CrossRef](#)]
8. Baccetti, T.; Franchi, L.; McNamara, J.A. The Cervical Vertebral Maturation (CVM) Method for the Assessment of Optimal Treatment Timing in Dentofacial Orthopedics. *Semin. Orthod.* **2005**, *11*, 119–129. [[CrossRef](#)]
9. Nagendrababu, V.; Aminoshariae, A.; Kulild, J. Artificial Intelligence in Endodontics: Current Applications and Future Directions. *J. Endod.* **2021**, *47*, 1352–1357. [[CrossRef](#)]
10. Revilla-León, M.; Gómez-Polo, M.; Vyas, S.; Barmak, B.A.; Galluci, G.O.; Att, W.; Krishnamurthy, V.R. Artificial intelligence applications in implant dentistry: A systematic review. *J. Prosthet. Dent.* **2021**. [[CrossRef](#)]
11. Khanagar, S.B.; Naik, S.; Al Kheraif, A.A.; Vishwanathaiah, S.; Maganur, P.C.; Alhazmi, Y.; Mushtaq, S.; Sarode, S.C.; Sarode, G.S.; Zanza, A.; et al. Application and Performance of Artificial Intelligence Technology in Oral Cancer Diagnosis and Prediction of Prognosis: A Systematic Review. *Diagnostics* **2021**, *11*, 1004. [[CrossRef](#)]
12. Schwendicke, F.; Chaurasia, A.; Arsiwala, L.; Lee, J.H.; Elhennawy, K.; Jost-Brinkmann, P.G.; Demarco, F.; Krois, J. Deep learning for cephalometric landmark detection: Systematic review and meta-analysis. *Clin. Oral Investig.* **2021**, *25*, 4299–4309. [[CrossRef](#)]
13. Hwang, H.W.; Park, J.H.; Moon, J.H.; Yu, Y.; Kim, H.; Her, S.B.; Srinivasan, G.; Aljanabi, M.N.A.; Donatelli, R.E.; Lee, S.J. Automated identification of cephalometric landmarks: Part 2- Might it be better than human? *Angle Orthod.* **2020**, *90*, 69–76. [[CrossRef](#)]
14. Larson, D.B.; Chen, M.C.; Lungren, M.P.; Halabi, S.S.; Stence, N.V.; Langlotz, C.P. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. *Radiology* **2018**, *287*, 313–322. [[CrossRef](#)]
15. Kok, H.; Acilar, A.M.; Izgi, M.S. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog. Orthod.* **2019**, *20*, 41. [[CrossRef](#)] [[PubMed](#)]

16. Amasya, H.; Cesur, E.; Yıldırım, D.; Orhan, K. Validation of cervical vertebral maturation stages: Artificial intelligence vs human observer visual analysis. *Am. J. Orthod. Dentofac. Orthop.* **2020**, *158*, e173–e179. [[CrossRef](#)] [[PubMed](#)]
17. Schwendicke, F.; Singh, T.; Lee, J.H.; Gaudin, R.; Chaurasia, A.; Wiegand, T.; Uribe, S.; Krois, J. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *J. Dent.* **2021**, *107*, 103610. [[CrossRef](#)]
18. Li, Z.; Peng, C.; Yu, G.; Zhang, X.; Deng, Y.; Sun, J. DetNet: A Backbone network for Object Detection. *arXiv* **2018**, arXiv:1804.06215.
19. Lee, K.S.; Jung, S.K.; Ryu, J.J.; Shin, S.W.; Choi, J. Evaluation of Transfer Learning with Deep Convolutional Neural Networks for Screening Osteoporosis in Dental Panoramic Radiographs. *J. Clin. Med.* **2020**, *9*, 392. [[CrossRef](#)]
20. Hägg, U.; Taranger, J. Maturation indicators and the pubertal growth spurt. *Am. J. Orthod.* **1982**, *82*, 299–309. [[CrossRef](#)]
21. Baccetti, T.; Franchi, L.; De Toffol, L.; Ghiozzi, B.; Cozza, P. The diagnostic performance of chronologic age in the assessment of skeletal maturity. *Prog. Orthod.* **2006**, *7*, 176–188. [[PubMed](#)]
22. Franchi, L.; Baccetti, T.; De Toffol, L.; Polimeni, A.; Cozza, P. Phases of the dentition for the assessment of skeletal maturity: A diagnostic performance study. *Am. J. Orthod. Dentofac. Orthop.* **2008**, *133*, 395–400. [[CrossRef](#)]
23. Perinetti, G.; Contardo, L.; Gabrieli, P.; Baccetti, T.; Di Lenarda, R. Diagnostic performance of dental maturity for identification of skeletal maturation phase. *Eur. J. Orthod.* **2012**, *34*, 487–492. [[CrossRef](#)] [[PubMed](#)]
24. Bittencourt, M.V.; Cericato, G.; Franco, A.; Girão, R.; Lima, A.P.B.; Paranhos, L. Accuracy of dental development for estimating the pubertal growth spurt in comparison to skeletal development: A systematic review and meta-analysis. *Dentomaxillofac. Radiol.* **2018**, *47*, 20170362. [[CrossRef](#)]
25. Lai, E.H.; Liu, J.P.; Chang, J.Z.; Tsai, S.J.; Yao, C.C.; Chen, M.H.; Chen, Y.J.; Lin, C.P. Radiographic assessment of skeletal maturation stages for orthodontic patients: Hand-wrist bones or cervical vertebrae? *J. Formos. Med. Assoc.* **2008**, *107*, 316–325. [[CrossRef](#)]
26. Flores-Mir, C.; Burgess, C.A.; Champney, M.; Jensen, R.J.; Pitcher, M.R.; Major, P.W. Correlation of skeletal maturation stages determined by cervical vertebrae and hand-wrist evaluations. *Angle Orthod.* **2006**, *76*, 1–5. [[CrossRef](#)]
27. Cericato, G.O.; Bittencourt, M.A.; Paranhos, L.R. Validity of the assessment method of skeletal maturation by cervical vertebrae: A systematic review and meta-analysis. *Dentomaxillofac. Radiol.* **2015**, *44*, 20140270. [[CrossRef](#)] [[PubMed](#)]
28. 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*, 125–128. [[CrossRef](#)]
29. Fazal, M.I.; Patel, M.E.; Tye, J.; Gupta, Y. The past, present and future role of artificial intelligence in imaging. *Eur. J. Radiol.* **2018**, *105*, 246–250. [[CrossRef](#)]
30. Hung, K.; Montalvo, C.; Tanaka, R.; Kawai, T.; Bornstein, M.M. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofac. Radiol.* **2020**, *49*, 20190107. [[CrossRef](#)]
31. 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. [[CrossRef](#)] [[PubMed](#)]
32. Leonardi, R.; Giordano, D.; Maiorana, F.; Spampinato, C. Automatic cephalometric analysis. *Angle Orthod.* **2008**, *78*, 145–151. [[CrossRef](#)] [[PubMed](#)]
33. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [[CrossRef](#)] [[PubMed](#)]
34. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
35. Kriegeskorte, N.; Golan, T. Neural network models and deep learning. *Curr. Biol.* **2019**, *29*, R231–R236. [[CrossRef](#)]
36. Yadav, S.; Shukla, S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. In Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC), Piscataway, NJ, USA, 27–28 February 2016; pp. 78–83. [[CrossRef](#)]