

Original Article



# Determination of cervical vertebral maturation using machine learning in lateral cephalograms

Shahab Kavousinejad<sup>1,2</sup>, Asghar Ebadifar<sup>1</sup>, Azita Tehrani<sup>1</sup>, Farzan Zakermashhadi<sup>3</sup>, Kazem Dalaie<sup>1,2\*</sup>

<sup>1</sup>Dentofacial Deformities Research Center, Research Institute for Dental Sciences, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>2</sup>Department of Orthodontics, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>3</sup>School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

## ARTICLE INFO

### Article History:

Received: March 1, 2024

Accepted: October 25, 2024

ePublished: December 14, 2024

### Keywords:

Cervical vertebra dimensions,  
Growth modification treatment,  
Machine learning, Skeletal age

## Abstract

**Background.** The accurate timing of growth modification treatments is crucial for optimal results in orthodontics. However, traditional methods for assessing growth status, such as hand-wrist radiographs and subjective interpretation of lateral cephalograms, have limitations. This study aimed to develop a semi-automated approach using machine learning based on cervical vertebral dimensions (CVD) for determining skeletal maturation status.

**Methods.** A dataset comprising 980 lateral cephalograms was collected from the Department of Orthodontics, Shahid Beheshti Dental School in Tehran, Iran. Eight landmarks representing the corners of the third and fourth cervical vertebrae were selected. A ratio-based approach was employed to compute the values of C3 and C4, accompanied by the implementation of an auto\_error\_reduction (AER) function to enhance the accuracy of landmark selection. Linear distances and ratios were measured using the dedicated software. A novel data augmentation technique was applied to expand the dataset. Subsequently, a stacking model was developed, trained on the augmented dataset, and evaluated using a separate test set of 196 cephalograms.

**Results.** The proposed model achieved an accuracy of 99.49% and demonstrated a loss of 0.003 on the test set.

**Conclusion.** By employing feature engineering, simplified landmark selection, AER function, data augmentation, and eliminating gender and age features, a model was developed for accurate assessment of skeletal maturation for clinical applications.

## Introduction

The timing of growth modification treatments is crucial for achieving optimal results. The peak of mandibular growth represents the ideal time for intervention.<sup>1</sup> Skeletal age determination is an important method for assessing growth status in orthodontics.<sup>2-4</sup> However, chronological age does not always correlate well with skeletal age,<sup>5-7</sup> leading to the introduction of alternative methods for skeletal age assessment.<sup>8-11</sup> While hand-wrist radiographs are considered the gold standard for skeletal age determination,<sup>12</sup> their limited use in dentistry is due to concerns about excessive radiation exposure.<sup>13-15</sup> In dentistry, evaluating cervical vertebra maturation (CVM) on lateral cephalograms is the most common approach to assessing skeletal age, as it is easy to perform and provides valuable information for initial diagnosis in orthodontics.<sup>10,16-18</sup>

However, interpreting lateral cephalograms for CVM analysis can be challenging due to variations in image clarity and the absence of a definitive cutoff point between CVM stages.<sup>18,19</sup> By incorporating a quantitative approach,

we can enhance our understanding of the patient's skeletal maturation, ultimately leading to more effective treatment outcomes. Moreover, several studies have reported low inter- and intra-observer agreement, indicating that the CVM method lacks reliability and reproducibility.<sup>1,20-23</sup> These limitations arise from the qualitative nature of parameters assessed in the CVM method, such as the amount of concavity and the shape of cervical vertebrae, highlighting the need for quantitative approaches. Quantitative methods have been developed to address these limitations, focusing on measuring cervical vertebra dimensions (CVD) to determine skeletal age.<sup>9,24,25</sup> A strong and statistically significant correlation between CVM and CVD has been demonstrated.<sup>25</sup> In this method, the six groups (CVM method) were divided into three groups. Groups 3 and 4 in the CVM method (group 2 in the CVD method) are associated with the mandibular peak growth period.<sup>25,26</sup> Therefore, we used the three-class method (pre-peak, peak, and post-peak) in the present study.

In recent years, the rapid advancement of imaging technologies, coupled with the increasing complexity

\*Corresponding author: Kazem Dalaie, Email: [Kazemdalaie@gmail.com](mailto:Kazemdalaie@gmail.com)

© 2024 The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of interpretation, has sparked a surge of interest among researchers in exploring the potential application of artificial intelligence (AI) in orthodontics. AI can potentially assist orthodontists in diagnosing and predicting outcomes with high accuracy and reduce time compared to trained dentists.<sup>27</sup> Several studies have evaluated the accuracy of deep learning models in determining CVM stages. Atici et al<sup>28</sup> and Khazaei et al<sup>29</sup> achieved accuracy rates ranging from 75% to 82%. Kök et al<sup>30</sup> compared deep learning models with machine learning models and concluded that deep learning models outperformed machine learning models. However, considering the novel methodology employed to measure CVD in the present study, integrating feature engineering and feature selection into machine learning models is expected to yield significantly higher accuracy than deep learning models. This study aimed to determine skeletal maturation status using machine learning algorithms based on quantitative measurements of CVD obtained from lateral cephalograms. By leveraging the potential of AI, this research aimed to enhance the accuracy of skeletal age assessment in orthodontics.

## Methods

### Data collection and dataset preparation

In this study, 980 digital cephalograms were collected from 6-17-year-olds. The cephalograms were collected from existing files in the Department of Orthodontics, Shahid Beheshti Dental School, Tehran, Iran. Inclusion criteria consisted of high-quality cephalograms and cervical vertebrae and the absence of specific syndromes

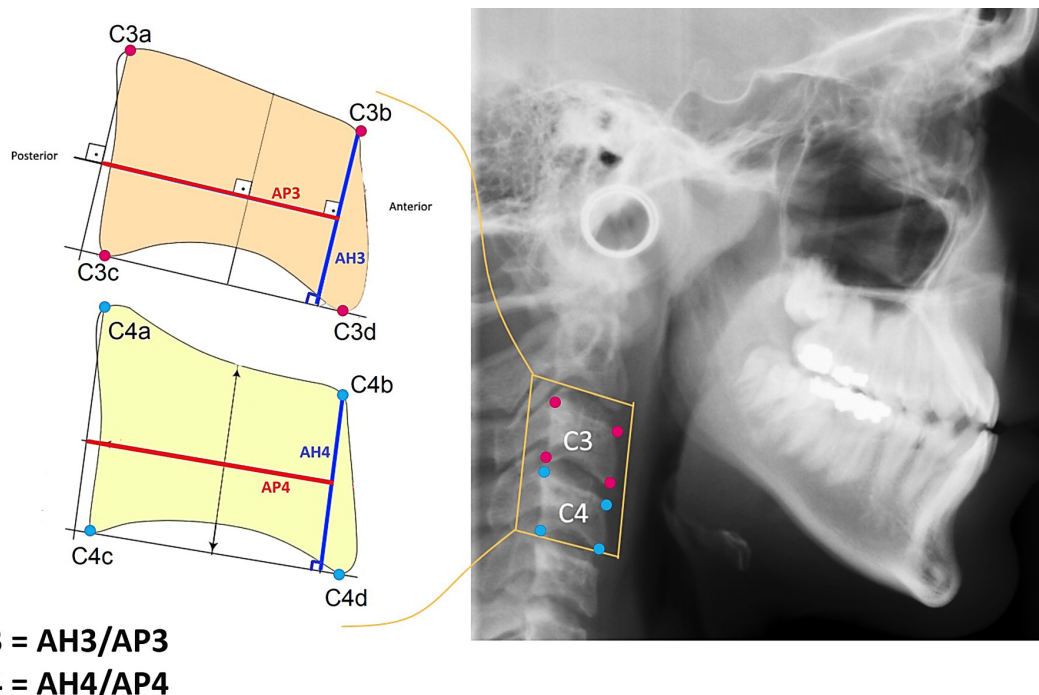
and systemic problems in patients. Each cephalogram was randomly assigned a unique identifier in the format of a letter and a value (e.g., A0). To perform feature engineering, the ratio of CVD was calculated according to the following formula.<sup>25</sup> This method is described in Figure 1.

$$C_3 = \frac{AH_3}{AP_3}, C_4 = \frac{AH_4}{AP_4}$$

In this method, eight landmarks representing the corners of the third and fourth cervical vertebrae were meticulously selected. In cases where the corners exhibited curvature, the midpoint of the curve was selected. A software application was developed using the C# programming language to facilitate the measurement of linear distances and ratios. The cephalograms were subsequently imported into the software, where they underwent resizing to achieve a uniform width of 2000 pixels while preserving the original aspect ratio. This resizing operation was necessary to standardize the pixels for subsequent steps. To calculate the lengths, the pixel count between the selected landmarks (X and Y coordinates) was measured using the following formula:

$$\text{Length} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The values of C3 and C4 were calculated for each sample. By employing the ratio-based approach, inherent variations in magnification associated with diverse radiographic views and x-ray devices were



**Figure 1.** The selected landmarks, indicated by red and blue dots, represent the corners of the third and fourth cervical vertebrae, respectively. Eight landmarks were selected. For C4, a red line connects the midpoint of the perpendicular line from C4a to the line C4c-C4d with the midpoint of the perpendicular line from C4b to the line C4c-C4d. The length of this line represents the value of AP4. The blue line, AH4, shows the perpendicular line from C4b to the line C4c-C4d. The value of C4 is presented as the ratio AH4/AP4. A similar method was applied to the third cervical vertebra, determining the value corresponding to C3.

effectively eliminated. An innovative AER function was implemented within the software framework to enhance the accuracy of landmark selection by the researcher (Figure 2). Within this function, the coordinates of each selected landmark within the software were subjected to random displacements spanning 1 to 4 pixels in both the X and Y directions relative to the original landmark. Subsequently, the values of C3 and C4 were computed for each iteration of the AER function, which was repeated a thousand times (as a loop). Ultimately, the average values of C3 and C4 were derived as the output for each sample. This sophisticated approach substantially reduced the error stemming from discrepancies in landmark selection across researcher iterations, as the selection area encompassed a set of randomly distributed landmarks within a maximum radius of 4 pixels. AER function is a probabilistic average of surrounding landmarks.

The data, including age, gender, C3 and C4 values, were placed in a CSV file. A three-class label column called "Maturation" was considered in this CSV file. To prevent bias, blind labeling was performed, meaning the expert determining the class of each sample was unaware of the features of each sample. Initially, labeling was done based on the CVM method for each sample's cephalometric measurements by an orthodontist. Using this method, six classes were identified (CVS1 to CVS6). In the next step, the final classification (maturation) was determined as follows:

Class 1: Pre-peak of mandibular growth (CVS1 and CVS2 classes)

Class 2: Peak of mandibular growth (CVS3 and CVS4 classes)

Class 3: Post-peak of mandibular growth (CVS5 and CVS6 classes)

Two other important indices, SumC3C4 and C3C4, were calculated for each sample based on the following formulas and included in the dataset:

$$\text{SumC3C4} = C_3 + C_4$$

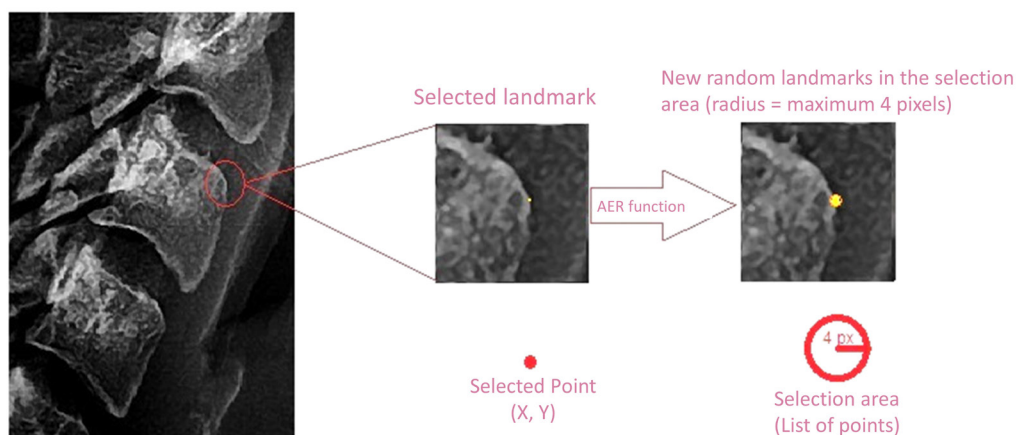
$$C_3C_4 = C_3 \times C_4$$

Therefore, the final dataset included age, gender, C3, C4, SumC3C4, C3C4, and maturation. The project was coded in Python programming language using the Jupyter Notebook environment (version 6.4.12). The following Python libraries were used: Scikit-learn (sklearn), CatBoost, LightGBM, and XGBoost.

### Data preprocessing and feature selection

The dataset had no missing values, and the sample sizes for each class were balanced. The labels were converted from qualitative (pre-peak, peak, post-peak) to quantitative, representing three classes: 1, 2, and 3. The dataset was randomly split into two sets: train (80%) and test (20%). Figure 3 visually presents the three-dimensional distribution of the training data categorized by class before and after data augmentation. The x, y, and z axes correspond to the values C3, C4, and SumC3C4, respectively. Additionally, the size of each sample (represented as a sphere) is determined by the weighted value of C3C4.

To enhance the generalization and accuracy of our machine learning model, introduce data diversity, and mitigate overfitting, we developed a data augmentation technique called CVD\_Generator specifically for the training data. This method involves generating random values within the range of minimum and maximum values for C3 and C4 for each class, considering their distributions within the training data. As a result, 1000 new samples were created for each class, with random values of C3 and C4 based on their respective classifications. Additionally, the values of SumC3C4 and C3C4 were calculated for each newly generated sample. The CVD\_Generator method is defined as follows:



**Figure 2.** The AER function's operational procedure is depicted. We anticipated an operator error of up to 4 pixels during point selection, which the function systematically addresses. Within this function, a thousand randomly generated points are automatically distributed across the designated selection area. The corresponding values of C3 and C4 were calculated for each generated point. This iterative process is repeated a thousand times, resulting in a dataset of a thousand C3 and C4 values for each sample. Finally, the average values of C3 and C4 were computed and considered as the definitive C3 and C4 values for each sample, respectively.

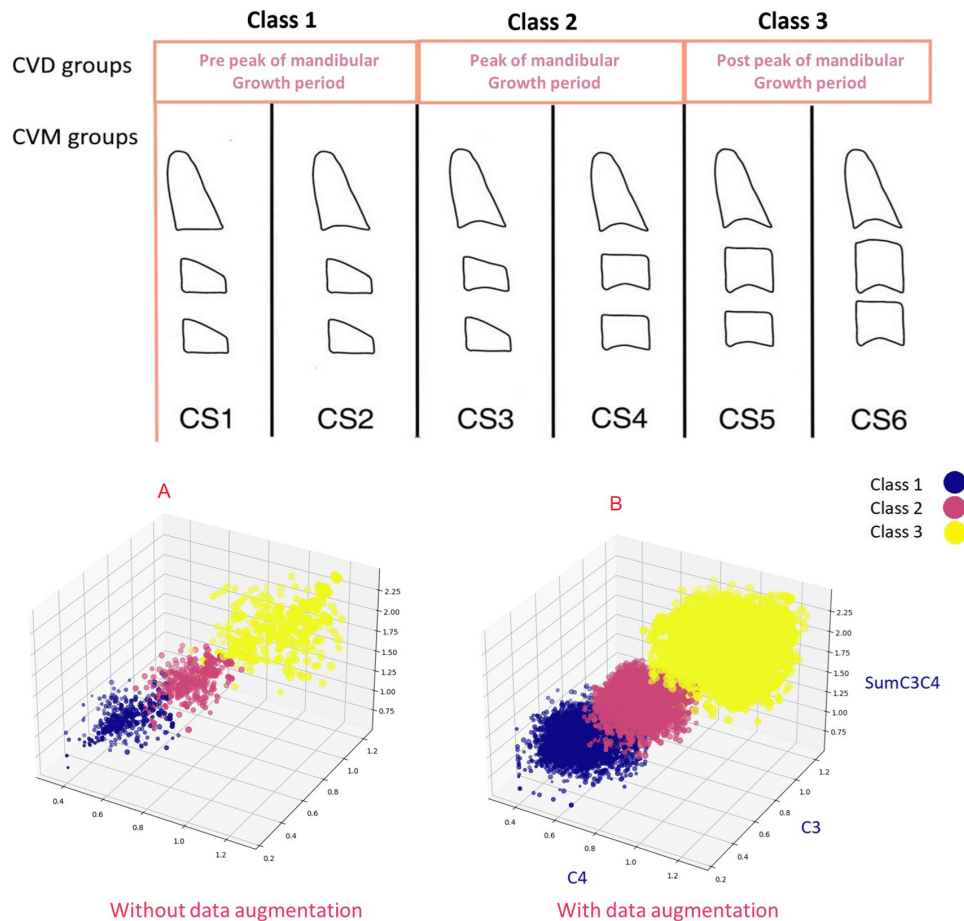


Figure 3. Visualization of the training dataset in 3D format before (A) and after (B) data augmentation

$n = 1000, G = \{1, 2, 3\}, C = \{C3, C4\}$

Where  $n$  represents the number of new samples, and  $G$  is the set of groups defined based on the maturation column. The values in  $C$  include the set of columns  $C3$  and  $C4$ . For each group  $g_i \in G$  and each value  $j \in \{1, 2, \dots, n\}$ , a new sample was generated as follows:

$$S_{ij} = \left\{ \begin{array}{l} C_3 : v_{ij3}, C_4 : v_{ij4}, SumC3C4 : v_{ij3} + v_{ij4}, \\ C3C4 : v_{ij3} \times v_{ij4}, Maturation : g_i \end{array} \right\}$$

Where  $S_{ij}$  is a new row added to the data frame, and  $v_{ijk}$  is a new random value of the  $C_k$  column within the range of values for that group. Then, the age and gender values were also randomly generated based on their respective distributions in each class for the new samples.

To reduce the model's reliance on age and gender and mitigate bias, we exclusively focused on the independent variables ( $X$ ):  $C3, C4, SumC3C4,$  and  $C3C4$ , while considering the three-class label as the dependent variable ( $Y$ ). As the selected features possess the same nature (dimension ratio), we refrained from applying any data normalization techniques (e.g., employing the StandardScaler method) or feature scaling to them.

### Model architecture

Figure S1 visually represents the data preprocessing process, model architecture, and model testing. In the initial stage, we employed the 5-fold cross-validation technique along with grid search and genetic algorithms (for MLP) to determine the optimal hyperparameters for each stage 1 model. The hyperparameters of each model were adjusted accordingly. Subsequently, each model underwent individual training using the training data and evaluation using the test data. Consequently, the hyperparameters for the base models were appropriately configured. The fine-tuned models were then integrated as base models within the stacking model (depicted as stage 1 in Figure S1). The stacking model was trained using the training data. The fundamental concept of model stacking involves training multiple diverse base models and combining their predictions through the training of a meta model. The meta model generates the final prediction by considering the predictions made by the base models. The base models underwent training based on 5-fold cross-validation (CV) on the data and forwarded their predictions to the final estimator. In our proposed model, we used logistic regression as the final estimator, employing default hyperparameters. The 5-fold CV technique divides the data into 5 subsets, using one-fifth as the test data and the remaining 4 subsets as the



training data in each iteration. The final prediction is then obtained by averaging the results of these 5 iterations. The architecture of the stacking model is depicted in Figure S1.

### Base models

- ExtraTreesClassifier
- Multi-layer perceptron (MLP)
- XGBClassifier
- CatBoostClassifier
- LGBMClassifier (LightGBM)
- VotingClassifier

### Final model

- Classification Meta Model: As the ultimate estimator in the stacking model, logistic regression was used, providing a reliable and interpretable prediction for the ensemble model.

### Evaluation

In evaluating our model on the test data, we employed various metrics to assess its effectiveness. These metrics encompass accuracy, precision, F1 score, recall, log loss, Jaccard, and the confusion matrix.

**Log loss:** This function gauges the performance of a classification model by computing the negative logarithm of the predicted probability for the correct label. In the provided formula,  $N$  denotes the number of samples,  $y_{ij}$  represents the true label for sample  $i$  and class  $j$ , and  $p_{ij}$  corresponds to the predicted probability for sample  $i$  and class  $j$ .

$$L(y, p) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^3 y_{ij} \log(p_{ij})$$

**Accuracy:** Accuracy measures the overall correctness of the model's predictions, calculated as the ratio of correct predictions to the total number of predictions.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- TP (True Positive): The number of positive instances correctly predicted as positive
- FP (False Positive): The number of negative instances incorrectly predicted as positive
- TN (True Negative): The number of negative instances correctly predicted as negative
- FN (False Negative): The number of positive instances incorrectly predicted as negative

**Precision:** Precision assesses the model's capability to accurately predict positive samples, computed as the ratio of true positive predictions to the total number of predicted positives.

$$precision = \frac{TP}{TP + FP}$$

**Recall:** Recall evaluates the model's ability to correctly identify all positive samples, expressed as the ratio of true positive predictions to the total number of true positives.

$$recall = \frac{TP}{TP + FN}$$

**F1 score:** The F1 score represents a measure of the balance between precision and recall, computed as the harmonic mean of precision and recall.

$$F1score = 2 \times \frac{precision \times recall}{precision + recall}$$

Additionally, as an extra objective in our study, we used the following formula to determine the cutoff points between the three classes:

$$\text{Class\_1\_range} = \{x \in X | y(x) = 1\}, \quad \text{Class\_2\_range} = \{x \in X | y(x) = 2\}, \quad \text{Class\_3\_range} = \{x \in X | y(x) = 3\}$$

$$\text{cut\_point\_1} = \frac{\max(\text{Class\_1\_range}) + \min(\text{Class\_2\_range})}{2}$$

The first and second cutoff points were determined using the following formula.

$$\text{cut\_point\_2} = \frac{\max(\text{Class\_2\_range}) + \min(\text{Class\_3\_range})}{2}$$

To ensure the presence of a cutoff point that includes both the SumC3C4 and C3C4 features, we devised the following formula:

$$\text{Final cutoff point\_1} = \frac{\text{cutoff\_1 (C3C4)}}{\text{cutoff\_1 (SumC3C4)}} \times 100$$

$$\text{Final cutoff point\_2} = \frac{\text{cutoff\_2 (C3C4)}}{\text{cutoff\_2 (SumC3C4)}} \times 100$$

### Results

Table 1 shows the means and standard deviations of dataset features. Figure 4a visualizes the correlation among the initial features, indicating a weak correlation with "gender" and strong positive correlations among the other features. Table 2 presents the results of fine-tuning various base models, with the stacking model outperforming base models (99.49% accuracy, 0.003 log loss). The confusion matrix for the proposed model on the test data is shown in Figure 4b. Sensitivity analysis in Figure 4c highlights the highest importance of "C3C4" and the least importance of "C4." Table 3 displays cutoff points between the three groups based on "SumC3C4" and "C3C4" features. The dataset ranges from a minimum SumC3C4 value of 0.62 to a maximum of 2.38 and a minimum C3C4 value of 0.09 to a maximum of 1.42.

**Discussion**

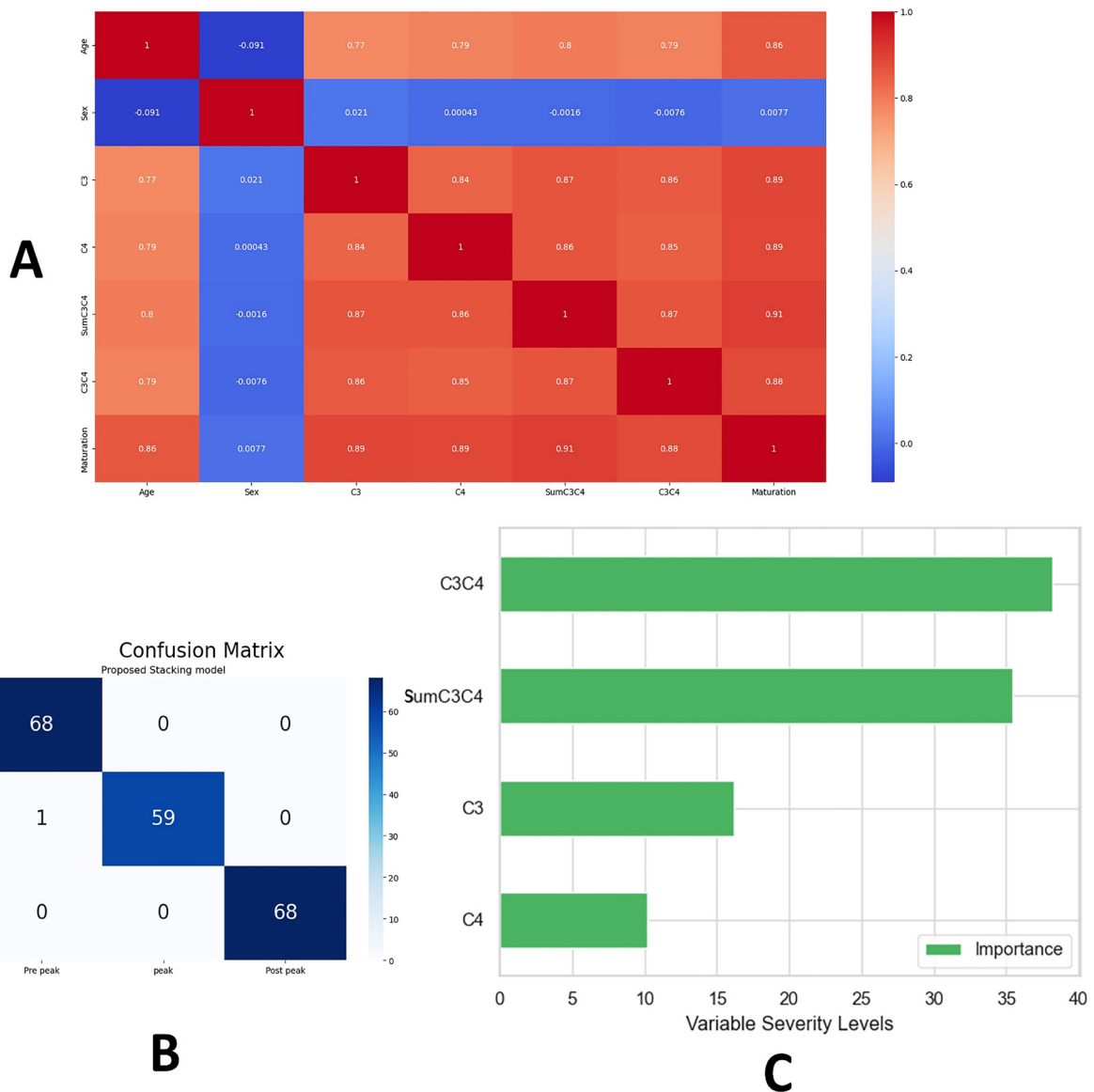
The present study aimed to assess skeletal maturation using cervical vertebrae. We gained valuable insights into feature engineering, correlation visualization, model fine-tuning, sensitivity analysis, and classification cutoff points through data analysis and model evaluation. The proposed model achieved 99.49% accuracy and a test set loss of 0.003, outperforming base models. This highlights

the effectiveness of combining multiple models to improve skeletal maturation prediction accuracy.

The relationship between hand-wrist radiographs and skeletal age is well established.<sup>10,31,32</sup> Kim et al<sup>33</sup> found that an ensemble model of eight machine learning models achieved the highest accuracy of 43% in predicting hand-wrist maturation stages based on cervical vertebrae from lateral cephalograms. However, the CVM method on

**Table 1.** The mean and standard deviation for C3C4, SumC3C4, C4, C3, and age in the initial dataset

Maturation	Age (mean ± SD)	Gender (Counts)	C3 (mean ± SD)	C4 (mean ± SD)	SumC3C4 (mean ± SD)	C3C4 (mean ± SD)
Pre peak (n=326)	8.61 ± 1.65	M:192 F:134	0.52 ± 0.08	0.49 ± 0.08	1.01 ± 0.14	0.26 ± 0.07
Peak (n=326)	11.53 ± 1.37	M:178 F:148	0.74 ± 0.08	0.68 ± 0.08	1.42 ± 0.14	0.51 ± 0.1
Post peak (n=328)	16.1 ± 1.23	M:194 F:134	1.01 ± 0.14	0.94 ± 0.12	1.96 ± 0.23	0.97 ± 0.23



**Figure 4.** A) Heatmap of the correlation among the initial dataset features. B) Confusion matrix of the proposed model on the test data. C) The importance of each feature for the proposed model

lateral cephalograms is widely recognized as a reliable approach for determining skeletal age.<sup>16</sup>

Two studies compared the performance of deep learning models with human visual analysis and reported a low agreement percentage of 58%, possibly due to small sample sizes or the specific AI models used.<sup>34,35</sup> Khazaei et al<sup>29</sup> increased the sample size to 1846 patients in a study using CNN models, resulting in a higher agreement percentage. However, the highest accuracy was still relatively low in the three-group classification at 82%. Similarly, Atici et al<sup>28</sup> used data augmentation and found their CNN model superior to other deep learning models, but the accuracy remained below 83% for females and 75% for males. In contrast, Seo et al<sup>36</sup> achieved a higher average accuracy of 95.6% using a deep learning approach and image segmentation on a relatively large sample size of 900 participants for bone age estimation based on the CVM method. Therefore, one strength of our study is the use of a large sample size and a novel data augmentation approach.

What sets our study apart is its higher accuracy, using only eight vertebral reference points and four linear measurements. In contrast, Amasya et al<sup>37</sup> compared five machine learning models in CVM analysis using 26 marked landmarks and evaluating 54 features on each lateral cephalogram, and their results indicated that ANN had the highest agreement of 86.93% with visual analysis. Additionally, Xie et al<sup>38,39</sup> achieved accuracies of 87% and 88% in two separate studies by considering various parameters such as chronological age, C3 height (H3), and the ratio of posterior height to lower width of C4 (PH4/LW4). K ok et al<sup>40</sup> evaluated 24 ANN models with 27 vertebral reference points and 32 linear measurements, with the best model achieving an accuracy of 94.27% using 32 linear measurements and age. The highest accuracy with the fewest linear measurements (13) was 86.87%. Therefore, the advantages of our study include higher accuracy, fewer landmarks, AER function, data augmentation, and feature engineering.

The results revealed a weak correlation between the “gender” variable and other features, while strong positive correlations were observed among the remaining features. These findings suggest that the “gender” variable may have limited influence on skeletal maturation assessment, while the other features exhibit interdependencies that can be leveraged for accurate evaluation. In this study, we aimed to develop a skeletal maturation assessment model free from gender and age bias. To achieve this, we excluded the gender feature from the model input, as it showed no significant correlation with other factors. Additionally, we removed the age feature to ensure that our model solely relies on the geometric dimensions of the third and fourth vertebrae. Consequently, when the model is deployed in the application software, we may confidently avoid the influence of chronological age on skeletal maturation status, even if an individual presents with a higher chronologic age but has delayed skeletal maturation due to factors such as illness, syndrome, or vitamin D deficiency.<sup>41</sup>

We used feature engineering and machine learning techniques to evaluate skeletal maturation based on cervical vertebrae. We focused on the changing dimensions of the third and fourth vertebrae as important features through feature engineering, which may explain the lower accuracies observed in CNN studies. By considering the variability of the anterior border of the third and fourth vertebrae during the 6-stage cervical vertebral maturation process, we emphasized the length features of AH3 and AH4. To standardize radiographs, we used ratios by dividing these values by AP3 and AP4. Consequently, the values of C3 and C4 contain valuable information about skeletal growth features. We also introduced the features

**Table 3.** The cutoff points for class separation based on C3C4 and SumC3C4 features

Class separation	Cutoff point SumC3C4	Cutoff point C3C4	Final cutoff point
Between Class 1 and Class 2	1.22	0.365	29.91
Between Class 2 and Class 3	1.56	0.62	39.74

**Table 2.** Model performance metrics on the test data and fine-tuned hyperparameters for base models

Model	Accuracy (%)	Precision*	Recall*	F1 score*	Log loss	Hyperparameters (fine-tuned)	
Support vector machine	96.94	0.97	0.97	0.97	0.08	Kernel = 'rbf', C = 800	
K-nearest neighbors	97.96	0.98	0.98	0.98	0.19	n_neighbors = 4	
Random forest	98.98	0.99	0.99	0.99	0.06	n_estimators = 211, max_depth = 13	
Extra trees classifier	97.44	0.98	0.97	0.97	0.10	n_estimators = 63, max_depth = 12	
Base models	Multi-layer perceptron	97.45	0.98	0.97	0.97	0.09	hidden_layer_sizes = (20,), learning_rate_init = 0.01 (adaptive), activation = logistic, max_iter = 300, solver = 'adam', alpha = 0.0001
	XGB classifier	98.98	0.99	0.99	0.99	0.02	n_estimators = 100, max_depth = 3, learning_rate = 0.2, subsample = 0.9, colsample_bytree = 0.85
	CatBoost classifier	97.42	0.98	0.98	0.98	0.09	Iterations = 100, learning_rate = 0.1, depth = 5
	LGBM classifier	98.98	0.99	0.99	0.99	0.04	n_estimators = 1000, learning_rate = 0.1, max_depth = 10
	Voting classifier	98.47	0.99	0.98	0.98	0.09	Default
Final model	Proposed Stacking model	<b>99.49</b>	1.0	0.99	0.99	0.003	final_estimator (logisticRegression)

\* Weighted average

SumC3C4 and C3C4 to represent an individual's current peak growth status within a specified range. C3C4 had the most significant impact on the classification model among these features. Multiplying features together may increase their importance, suggesting the advantage of generating new features by combining existing ones in other machine learning studies with numerous features in this field.

This study introduces a new method for selecting cervical vertebra landmarks, using a simplified process and fewer landmarks. Instead of using many landmarks, we only selected eight landmarks from the cervical vertebrae on lateral cephalometric radiographs, making the process faster and more user-friendly. This approach offers a more efficient alternative to previous machine learning studies on cervical vertebrae. The core of our proposed method is the AER function, which reduces researcher error in landmark selection. We predicted a four-pixel operator error within the AER function. We standardized the width of all cephalometric images while maintaining the aspect ratio and performed landmark selection three times on 20 randomly selected samples. On average, the coordinates of three points for each landmark fell within a four-pixel radius. By automatically executing the AER function, we calculated the values of C3 and C4 a thousand times for each sample. This procedure improved calculation accuracy, reduced bias, and minimized landmark selection errors.

By maintaining the data distribution, our data augmentation approach effectively generated additional samples, enhancing the diversity present in the data. This approach helped avoid overfitting, enhancing the model's ability to generalize. Additionally, by increasing the quantity of training data within each class, our augmentation method provided the model with more samples to learn from patterns.

Unlike previous studies, our method eliminated the need to determine the curvature of the inferior border and focused on the correlation between vertebral dimensions and stages of CVM. The classification model used in the study did not rely on chronological age, enhancing confidence in the results' validity and accuracy. The study suggested that the optimal timing for growth modification in the CVM method is between CS3 and CS4. However, according to the three-class cervical vertebral maturation method, the middle of group 2 is considered the best treatment timing. This implies that the patient's current skeletal position can be visually represented as resembling [Figure S2](#).

## Conclusion

The proposed model achieved an accuracy of 99.49% in evaluating skeletal maturation based on cervical vertebrae. Overall, by employing feature engineering, simplified landmark selection, AER function, data augmentation, and the elimination of gender and age features, a model has been developed for accurate assessment of skeletal maturation for clinical applications.

## Acknowledgments

We would like to thank the Dentofacial Deformities Research Center at Shahid Beheshti University of Medical Sciences for their support and contributions to this research. We thank the faculty members, researchers, and staff for their guidance and expertise.

## Authors' Contribution

**Conceptualization:** Shahab Kavousinejad.

**Data curation:** Azita Tehranchi.

**Formal analysis:** Shahab Kavousinejad, Asghar Ebadifar.

**Funding acquisition:** Kazem Dalaie.

**Investigation:** Asghar Ebadifar, Farzan Zakermashhadi.

**Methodology:** Shahab Kavousinejad, Asghar Ebadifar.

**Project administration:** Kazem Dalaie.

**Resources:** Kazem Dalaie.

**Supervision:** Kazem Dalaie.

**Validation:** Azita Tehranchi.

**Writing-original draft:** Shahab Kavousinejad.

**Writing-review & editing:** Shahab Kavousinejad, Asghar Ebadifar, Azita Tehranchi, Farzan Zakermashhadi, Kazem Dalaie.

## Competing Interests

None.

## Ethical Approval

This study was approved by Shahid Beheshti University of Medical Sciences (IR.SBMU.DRC.REC.1402.126).

## Funding

Dentofacial Deformities Research Center, Research Institute of Dental Sciences, School of Dentistry, Shahid Beheshti University of Medical Sciences.

## Supplementary File

Supplementary file contain Figures S1 and S2.

## References

- Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method: poor reproducibility. *Am J Orthod Dentofacial Orthop.* 2009;136(4):478.e1-478.e7. doi: [10.1016/j.ajodo.2007.08.028](https://doi.org/10.1016/j.ajodo.2007.08.028).
- Cameriere R, Giuliadori A, Zampi M, Galić I, Cingolani M, Pagliara F, et al. Age estimation in children and young adolescents for forensic purposes using fourth cervical vertebra (C4). *Int J Legal Med.* 2015;129(2):347-55. doi: [10.1007/s00414-014-1112-z](https://doi.org/10.1007/s00414-014-1112-z).
- Hauspie RC, Cameron N, Molinari L. *Methods in Human Growth Research.* Cambridge University Press; 2004.
- Thevissen PW, Kaur J, Willems G. Human age estimation combining third molar and skeletal development. *Int J Legal Med.* 2012;126(2):285-92. doi: [10.1007/s00414-011-0639-5](https://doi.org/10.1007/s00414-011-0639-5).
- Demirturk Kocasarac H, Altan AB, Yerlikaya C, Sinanoglu A, Noujeim M. Correlation between spheno-occipital synchondrosis, dental age, chronological age and cervical vertebrae maturation in Turkish population: is there a link? *Acta Odontol Scand.* 2017;75(2):79-86. doi: [10.1080/00016357.2016.1255352](https://doi.org/10.1080/00016357.2016.1255352).
- 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(6):364-8. doi: [10.5455/aim.2015.23.364-368](https://doi.org/10.5455/aim.2015.23.364-368).
- Sokic E, Tiro A, Sokic-Begovic E, Nakas E. Semi-automatic assessment of cervical vertebral maturation stages using cephalograph images and centroid-based clustering. *Acta*



- Stomatol Croat. 2012;46(4):280-90.
8. Cameron N. Can maturity indicators be used to estimate chronological age in children? *Ann Hum Biol.* 2015;42(4):302-7. doi: [10.3109/03014460.2015.1032349](https://doi.org/10.3109/03014460.2015.1032349).
  9. Chandrasekar R, Chandrasekhar S, Sundari KKS, Ravi P. Development and validation of a formula for objective assessment of cervical vertebral bone age. *Prog Orthod.* 2020;21(1):38. doi: [10.1186/s40510-020-00338-0](https://doi.org/10.1186/s40510-020-00338-0).
  10. 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(1):118-24. doi: [10.1043/0003-3219\(2004\)074<0118:Uosmbo>2.0.Co;2](https://doi.org/10.1043/0003-3219(2004)074<0118:Uosmbo>2.0.Co;2).
  11. Simmons K, Greulich WW. Menarcheal age and the height, weight, and skeletal age of girls age 7 to 17 years. *J Pediatr.* 1943;22(5):518-48. doi: [10.1016/s0022-3476\(43\)80022-6](https://doi.org/10.1016/s0022-3476(43)80022-6).
  12. 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(3):119-29. doi: [10.1053/j.sodo.2005.04.005](https://doi.org/10.1053/j.sodo.2005.04.005).
  13. Ferrillo M, Curci C, Rocuzzo A, Migliario M, Invernizzi M, de Sire A. Reliability of cervical vertebral maturation compared to hand-wrist for skeletal maturation assessment in growing subjects: a systematic review. *J Back Musculoskelet Rehabil.* 2021;34(6):925-36. doi: [10.3233/bmr-210003](https://doi.org/10.3233/bmr-210003).
  14. Liu N. Chronological age estimation of lateral cephalometric radiographs with deep learning. *ArXiv [Preprint]*. January 28, 2021. Available from: <https://arxiv.org/abs/2101.11805>.
  15. Szmraj 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;102:125-8. doi: [10.1016/j.ejrad.2018.03.012](https://doi.org/10.1016/j.ejrad.2018.03.012).
  16. Baccetti T, Franchi L, McNamara JA Jr. An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth. *Angle Orthod.* 2002;72(4):316-23. doi: [10.1043/0003-3219\(2002\)072<0316:Aivotc>2.0.Co;2](https://doi.org/10.1043/0003-3219(2002)072<0316:Aivotc>2.0.Co;2).
  17. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. *Am J Orthod Dentofacial Orthop.* 1995;107(1):58-66. doi: [10.1016/s0889-5406\(95\)70157-5](https://doi.org/10.1016/s0889-5406(95)70157-5).
  18. O'Reilly MT, Yanniello GJ. Mandibular growth changes and maturation of cervical vertebrae--a longitudinal cephalometric study. *Angle Orthod.* 1988;58(2):179-84. doi: [10.1043/0003-3219\(1988\)058<0179:Mgcamo>2.0.Co;2](https://doi.org/10.1043/0003-3219(1988)058<0179:Mgcamo>2.0.Co;2).
  19. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: general overview. *Korean J Radiol.* 2017;18(4):570-84. doi: [10.3348/kjr.2017.18.4.570](https://doi.org/10.3348/kjr.2017.18.4.570).
  20. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mouggiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging.* 2016;35(5):1207-16. doi: [10.1109/tmi.2016.2535865](https://doi.org/10.1109/tmi.2016.2535865).
  21. Hägg U, Taranger J. Maturation indicators and the pubertal growth spurt. *Am J Orthod.* 1982;82(4):299-309. doi: [10.1016/0002-9416\(82\)90464-x](https://doi.org/10.1016/0002-9416(82)90464-x).
  22. Hunter CJ. The correlation of facial growth with body height and skeletal maturation at adolescence. *Angle Orthod.* 1966;36(1):44-54. doi: [10.1043/0003-3219\(1966\)036<0044:Tcofgw>2.0.Co;2](https://doi.org/10.1043/0003-3219(1966)036<0044:Tcofgw>2.0.Co;2).
  23. Zhao XG, Lin J, Jiang JH, Wang Q, Ng SH. Validity and reliability of a method for assessment of cervical vertebral maturation. *Angle Orthod.* 2012;82(2):229-34. doi: [10.2319/051511-333.1](https://doi.org/10.2319/051511-333.1).
  24. Mito T, Sato K, Mitani H. Cervical vertebral bone age in girls. *Am J Orthod Dentofacial Orthop.* 2002;122(4):380-5. doi: [10.1067/mod.2002.126896](https://doi.org/10.1067/mod.2002.126896).
  25. Tehranchi A, Mahmoud M, Kavousinejad S. Quantitative determination of skeletal age using cervical vertebral dimensions. *Orthod Waves.* 2021;80(3):135-42. doi: [10.1080/13440241.2021.1952369](https://doi.org/10.1080/13440241.2021.1952369).
  26. Radwan MT, Sin Ç, Akkaya N, Vahdettin L. Artificial intelligence-based algorithm for cervical vertebrae maturation stage assessment. *Orthod Craniofac Res.* 2023;26(3):349-55. doi: [10.1111/ocr.12615](https://doi.org/10.1111/ocr.12615).
  27. Mohammad-Rahimi H, Nadimi M, Rohban MH, Shamsoddin E, Lee VY, Motamedian SR. Machine learning and orthodontics, current trends and the future opportunities: a scoping review. *Am J Orthod Dentofacial Orthop.* 2021;160(2):170-92.e4. doi: [10.1016/j.ajodo.2021.02.013](https://doi.org/10.1016/j.ajodo.2021.02.013).
  28. Atici SF, Ansari R, Allareddy V, Suhaym O, Cetin AE, Elnagar MH. AggregateNet: a deep learning model for automated classification of cervical vertebrae maturation stages. *Orthod Craniofac Res.* 2023;26 Suppl 1:111-7. doi: [10.1111/ocr.12644](https://doi.org/10.1111/ocr.12644).
  29. Khazaei M, Mollabashi V, Khotanlou H, Farhadian M. Automatic determination of pubertal growth spurts based on the cervical vertebral maturation staging using deep convolutional neural networks. *J World Fed Orthod.* 2023;12(2):56-63. doi: [10.1016/j.ejwf.2023.02.003](https://doi.org/10.1016/j.ejwf.2023.02.003).
  30. Kök H, İzgi MS, Acılar AM. Evaluation of the artificial neural network and Naive Bayes models trained with vertebra ratios for growth and development determination. *Turk J Orthod.* 2021;34(1):2-9. doi: [10.5152/TurkJOrthod.2020.20059](https://doi.org/10.5152/TurkJOrthod.2020.20059).
  31. Houston WJ. Relationships between skeletal maturity estimated from hand-wrist radiographs and the timing of the adolescent growth spurt. *Eur J Orthod.* 1980;2(2):81-93. doi: [10.1093/ejo/2.2.81](https://doi.org/10.1093/ejo/2.2.81).
  32. Şatir S, Büyükçavuş MH, Sari ÖF, Çimen T. A novel approach to radiographic detection of growth development period with hand-wrist radiographs: a preliminary study with ImageJ imaging software. *Orthod Craniofac Res.* 2023;26(1):100-6. doi: [10.1111/ocr.12584](https://doi.org/10.1111/ocr.12584).
  33. Kim DW, Kim J, Kim T, Kim T, Kim YJ, Song IS, et al. Prediction of hand-wrist maturation stages based on cervical vertebrae images using artificial intelligence. *Orthod Craniofac Res.* 2021;24 Suppl 2:68-75. doi: [10.1111/ocr.12514](https://doi.org/10.1111/ocr.12514).
  34. Akay G, Akcayol MA, Özdem K, Güngör K. Deep convolutional neural network-the evaluation of cervical vertebrae maturation. *Oral Radiol.* 2023;39(4):629-38. doi: [10.1007/s11282-023-00678-7](https://doi.org/10.1007/s11282-023-00678-7).
  35. 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(6):e173-9. doi: [10.1016/j.ajodo.2020.08.014](https://doi.org/10.1016/j.ajodo.2020.08.014).
  36. Seo H, Hwang J, Jung YH, Lee E, Nam OH, Shin J. Deep focus approach for accurate bone age estimation from lateral cephalogram. *J Dent Sci.* 2023;18(1):34-43. doi: [10.1016/j.jds.2022.07.018](https://doi.org/10.1016/j.jds.2022.07.018).
  37. Amasya H, Yıldırım D, Aydoğan 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. doi: [10.1259/dmfr.20190441](https://doi.org/10.1259/dmfr.20190441).
  38. Xie L, Tang W, Izadikhah I, Chen X, Zhao Z, Zhao Y, et al. Intelligent quantitative assessment of skeletal maturation based on multi-stage model: a retrospective cone-beam CT study of cervical vertebrae. *Oral Radiol.* 2022;38(3):378-88. doi: [10.1007/s11282-021-00566-y](https://doi.org/10.1007/s11282-021-00566-y).

- 
39. Xie L, Tang W, Izadikhah I, Zhao Z, Zhao Y, Li H, et al. Development of a multi-stage model for intelligent and quantitative appraising of skeletal maturity using cervical vertebrae cone-beam CT images of Chinese girls. *Int J Comput Assist Radiol Surg.* 2022;17(4):761-73. doi: [10.1007/s11548-021-02550-7](https://doi.org/10.1007/s11548-021-02550-7).
40. Kök H, Izgi MS, Acilar AM. Determination of growth and development periods in orthodontics with artificial neural network. *Orthod Craniofac Res.* 2021;24 Suppl 2:76-83. doi: [10.1111/ocr.12443](https://doi.org/10.1111/ocr.12443).
41. Azarbakhsh G, Iranparvar P, Tehranchi A, Moshfeghi M. Relationship of vitamin D deficiency with cervical vertebral maturation and dental age in adolescents: a cross-sectional study. *Int J Dent.* 2022;2022:7762873. doi: [10.1155/2022/7762873](https://doi.org/10.1155/2022/7762873).