RESEARCH PAPER

Development of an Artificial Intelligence Algorithm Using Machine Learning for the Automatic Evaluation of Cervical Vertebral Maturation Status on Digital Lateral Cephalograms

Dr. Yadav D[1*], Dr. Paul R[2], Dr. Sharma IK[3], Dr. Saini P[4], Dr. Gautom R[5]

[1*]*Professor, Department of Orthodontics & Dentofacial Orthopedics, Inderprastha College & Hospital, Email: deepti.yadav@yahoo.com.*

[2]*Principal & Head of the Department, Department of Orthodontics & Dentofacial Orthopedics, Inderprastha College & Hospita.*

[3]*Reader, Department of Orthodontics & Dentofacial Orthopedics, Inderprastha College & Hospital*

[4]*Senior Lecturer, Department of Orthodontics & Dentofacial Orthopedics, Inderprastha College & Hospital.*

[5]*Post Graduate Student, Department of Orthodontics & Dentofacial Orthopedics, Inderprastha College & Hospital.*

ABSTRACT

Introduction: Accurate assessment of skeletal maturity, particularly through Cervical Vertebral Maturation (CVM) staging, is essential for optimal orthodontic diagnosis and growth-modification treatment timing. While manual CVM evaluation using lateral cephalograms is standard, it is inherently subjective and prone to inter-observer variability. This study aimed to develop and validate an artificial intelligence (AI) algorithm to automate CVM staging, enhancing objectivity, consistency, and clinical workflow efficiency.

Aim: To develop and evaluate a machine learning-based AI algorithm for the automatic evaluation of cervical vertebral maturation on digital lateral cephalograms and **t**o compare the AI model's performance with manual CVM staging performed by expert orthodontists.

Method: A retrospective study was conducted using 1000 lateral cephalograms (patients aged 9–20 years). The dataset was annotated according to the McNamara and Franchi CVM staging system (CS1–CS6) and partitioned into training (70%), validation (20%), and test (10%) sets. A ResNet-50 convolutional neural network was trained and optimized using PyTorch Model performance was assessed through stage-wise accuracy, sensitivity, specificity, positive and negative predictive values, and agreement with manual staging by orthodontists.

Results: The AI model demonstrated high stage-wise classification accuracy: 100% for CS1 and CS6, and 87.5%–94.4% for intermediate stages (CS2–CS5). Overall agreement with manual staging was strong (Chi-square = 136.392, $p < 0.001$), with most misclassifications occurring between adjacent stages. Diagnostic metrics were robust, with specificity ≥92.8% across all stages and a negative predictive value of 100% for CS1. The model maintained consistent performance across the maturation continuum without statistically significant stage-dependent error ($p = 0.47$).

Conclusion: The developed AI algorithm reliably automates CVM staging with accuracy comparable to expert orthodontists. It shows potential to reduce observer variability, improve diagnostic consistency, and streamline clinical workflows. Future multi-center validation, integration of longitudinal data, and real-time clinical implementation are recommended to enhance generalizability and clinical utility.

Keywords: Artificial Intelligence, Machine Learning, Neural Networks, Cervical Vertebral Maturation (CVM) / CVM staging, Skeletal maturity.

How To Cite This Article: Yadav D, Paul R, Sharma Ik, Saini P, Gautom R. Development Of An Artificial Intelligence Algorithm Using Machine Learning For The Automatic Evaluation Of Cervical Vertebral Maturation Status On Digital Lateral Cephalograms. Int J Drug Deliv Technol. 2026;16(26s):92-98. Doi: 10.25258/ijddt.16.26s.9

INTRODUCTION

The accurate assessment of skeletal maturity is a fundamental principle in orthodontic treatment, particularly in dentofacial orthopedic correction. The stage of facial growth[1], especially mandibular development, plays a crucial role in diagnosis, prognosis, and treatment planning[2]. Growth modification therapies such as functional appliances and maxillary expansion are most effective when initiated during the peak circumpubertal growth period. Improper timing of these interventions can lead to prolonged treatment duration, compromised outcomes, or the need for surgical correction of jaw discrepancies[3]. Therefore, accurate identification of the pubertal growth spurt is essential for effective and efficient orthodontic treatment.

Traditionally, skeletal maturity has been assessed using indicators such as dental development and hand–wrist radiographs[4]. Hand–wrist radiographs have long been considered the reference standard due to their strong correlation with pubertal growth velocity. However, their clinical use requires an additional radiographic procedure, increasing patient radiation exposure as well as treatment time and cost.

To overcome these limitations, the Cervical Vertebral Maturation (CVM) method was developed[5]. This method evaluates morphological changes in the second, third, and fourth cervical vertebrae (C2, C3, and C4), which are visible on routine lateral cephalometric radiographs taken for orthodontic diagnosis. The CVM method proposed by Baccetti, Franchi, and McNamara classifies skeletal maturation into six stages (CS1–CS6), representing pre-pubertal (CS1–CS2), circumpubertal (CS3–CS4), and post-pubertal (CS5–CS6) growth phases[6]. Because it utilizes an already available radiograph, CVM staging has become widely adopted in orthodontic practice.

Despite its advantages, CVM assessment remains subjective and dependent on examiner experience. Meanwhile, advancements in Artificial Intelligence (AI)—the science of creating intelligent machines[7]—have enabled computers to analyze complex medical images. AI includes Machine Learning (ML), which allows systems to learn patterns from data, and Deep Learning (DL), a subset of ML that uses neural networks to automatically extract features from images[8]. The development of AI models typically involves data collection, annotation, preprocessing, model design, training, and validation[9].

In healthcare, AI has evolved into practical applications such as Clinical Decision Support Systems (CDSS), which assist clinicians in diagnosis and treatment planning[10,11,12]. In dentistry and orthodontics, AI has already been applied to tasks including automated cephalometric landmark identification and analysis of digital scans[13].

Given the subjectivity of CVM assessment and the growing capability of AI-based image analysis, this study aims to develop, validate, and evaluate a deep learning model for the automated classification of Cervical Vertebral Maturation stages from digital lateral cephalometric radiographs. By integrating clinical orthodontic knowledge with modern deep learning techniques, this research seeks to provide a more objective and reliable tool to support orthodontic diagnosis and treatment planning.

AIM OF THE STUDY:

The aim of this study is to develop and evaluate an artificial intelligence algorithm using machine learning techniques for the automatic evaluation of cervical vertebral maturation status on digital lateral cephalograms.

NULL HYPOTHESIS:

There is no statistically significant difference in the accuracy of cervical vertebral maturation status evaluation between the developed artificial intelligence algorithm using machine learning and traditional manual evaluation methods.

METHOD:

Sample Composition

This study was set up in Inderprastha Dental College, Ghaziabad. This study included existing lateral cephalogram radiographs data from 1000 untreated patients aged from 9 to 20 years of age visiting the Department of Orthodontics & Dentofacial Orthopedics. Inclusion criterion for the study group was Lateral cephalometric radiographs in which cervical vertebrae (C2 to C4) visible were included. Cephalometric radiographs of individuals wearing retainers, orthodontic appliances or necklaces, earrings and Non-standard images (such as those captured with incorrect head positions), as well as low-quality images (including blurred images) were excluded.

Study Method:

Cervical Vertebral Maturation (CVM) staging was performed using the McNamara and Franchi method (CS1–CS6), selected for its reliability and correlation with hand-wrist analysis. An examiner trained by an expert orthodontist classified all images.

Table No. 1: Six stages of CVM using McNamara & Franchi Method

Sl no.	CVM stages	Description
1	CS1	The lower border of all three vertebrae (C2 – C4) are flat. The bodies of C3 and C4 are trapezoidal (the superior edge of the vertebral body tapers from posterior to anterior).
2	CS2	Concavity at the lower edge of C2. The bodies of the C3 and C4 are trapezoidal.
3	CS3	Concavity at the lower edge of C2 and C3. The bodies of C3 and C4 can be trapezoidal or horizontally rectangular.
4	CS4	A concavity exists at the lower edge of C2, C3, and C4. The bodies of the C3 and C4 are horizontally rectangular.
5	CS5	The concavities at the lower edges of C2, C3, and C4 are still visible. There is at least one body from C3 and C4, which is square. If it is not square, one of the other bodies is still a horizontal rectangle.
6	CS6	The concavities at the lower edges of C2, C3, and C4 are still visible. There is at least one body from C3 and C4, which is square. If it is not square, one of the other bodies is still a vertical rectangle.

Annotation:
Images were annotated on Roboflow software, where cervical vertebrae (C2–C4) were outlined with a polygon tool. Each CVM stage was color-coded.

Data Partition:
The dataset was randomly split into:
- Training set: Used to teach the model. (70%)
- Validation set: It was a portion of original dataset that was used during model training to tune and evaluate model's performance before testing it on completely unseen data. (20%)
- Test set: Used only at the end to evaluate final performance. (10%)

Pre-Processing:
Images were standardized by:
- Cropping to a Region of Interest (ROI) of C2–C4
- Resizing to 100×200 pixels
- Applying noise reduction and contrast enhancement

Model Initialization & Development:
A ResNet-50 convolutional neural network (CNN) was implemented using PyTorch. The model was designed to detect two key morphological features:
1. Presence/absence of inferior border concavities in C2, C3, C4
2. Shape changes of C3 and C4 (trapezoidal → rectangular → square)

Training Process:
- Loss Function: Cross-entropy loss measured prediction error.
- Optimization: Stochastic Gradient Descent (SGD) adjusted model weights via backpropagation.
- Training was conducted over multiple epochs until loss minimized.

Model Evaluation:
After training, the model was saved and evaluated on the test set using metrics such as accuracy and precision.

Manual Tracing:
50 unseen radiographs were manually traced by experienced orthodontists. The AI's CVM stage predictions were compared against these manual annotations.

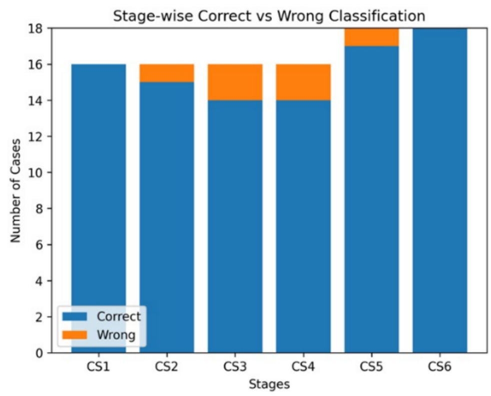
Data Analysis:
Data were analyzed using **SPSS v27.0**. Descriptive statistics, Chi-square tests, and a **5% significance level** were applied. Performance metrics were calculated to evaluate AI vs. manual staging.

RESULTS:
The results of the study demonstrate that the ResNet-50 deep learning model developed for the automated staging of Cervical Vertebral Maturation (CVM) achieved a high level of performance across key evaluation metrics. On the independent test set of 100 images, the model exhibited robust stage-wise classification, with perfect accuracy (100%) for both the initial CS1 and final CS6 stages.

Table 1: Stage-Wise Classification Accuracy of the AI Model in Cervical Vertebral Maturation Staging (n=100)

Stage	Correct	Wrong	Chi-Square value	p-value, S/NS
CS1	16 (100%)	0 (0%)	4.57	0.47, NS
CS2	15 (93.75%)	1 (6.25%)		
CS3	14 (87.5%)	2 (12.5%)		
CS4	14 (87.50%)	2 (12.5%)		
CS5	17 (94.44%)	1 (5.56%)		
CS6	18 (100%)	0 (0%)		

p ≤ 0.05 – Significant, CI = 95 %

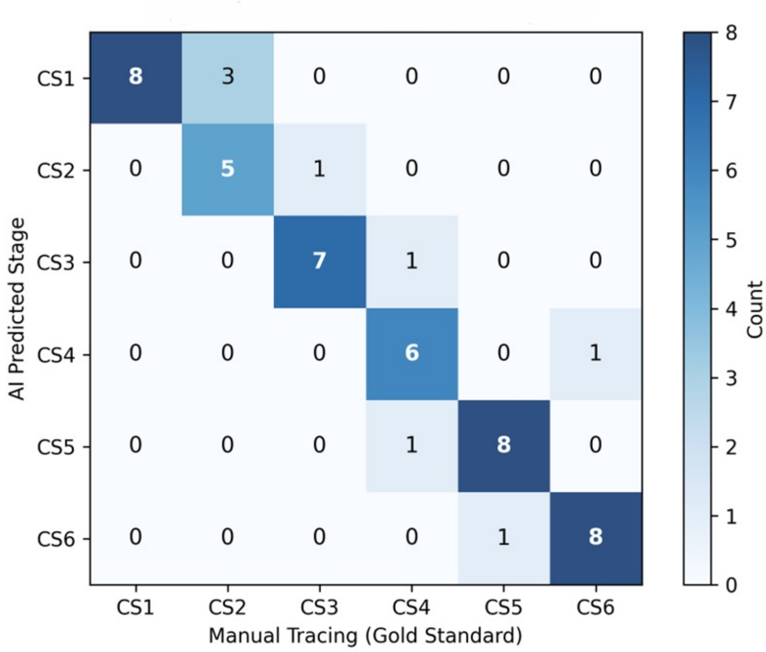


Graph 1: Stage-Wise Classification Accuracy of the AI Model in Cervical Vertebral Maturation Staging (n=100)

Accuracy for the intermediate stages ranged from 87.5% (CS3 and CS4) to 94.44% (CS5), and a Chi-square analysis confirmed no statistically significant variation in performance across the six stages ($\chi^2 = 4.57$, *p* = 0.47), indicating consistent predictive capability throughout the maturation spectrum.

Table 2: Comparison of stage prediction between Artificial intelligence algorithm and manual evaluation of cervical vertebral maturation on digital lateral cephalograms. (n=50)

Stage	CS1	CS2	CS3	CS4	CS5	CS6	Chi-Square value	p-value, S/NS
CS1	8 (16%)	3 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	136.392	**<0.001,HS**
CS2	0 (0%)	5 (10%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)		
CS3	0 (0%)	0 (0%)	7 (14%)	1 (2%)	0 (0%)	0 (0%)		
CS4	0 (0%)	0 (0%)	0 (0%)	6 (12%)	0 (0%)	1 (2%)		
CS5	0 (0%)	0 (0%)	0 (0%)	1 (2%)	8 (16%)	0 (0%)		
CS6	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2%)	8 (16%)		



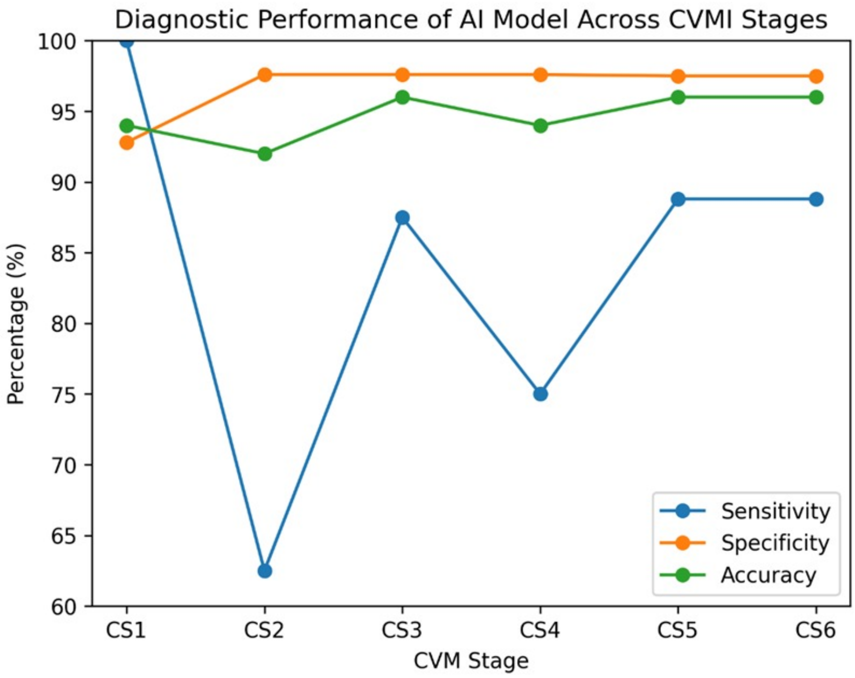
Graph 2: Comparison of stage prediction between Artificial intelligence algorithm and manual evaluation of cervical vertebral maturation on digital lateral cephalograms. (n=50)

A critical comparison against manual evaluations by expert orthodontists on a separate set of 50 radiographs revealed substantial agreement, with an overall concordance rate of 84%. The majority of disagreements were clinically minor, consisting of adjacent-stage misclassifications, which accounted for 87.5% of all errors. Statistical analysis of the confusion matrix, however, revealed a significant, non-random pattern in the discrepancies ($\chi^2(15) = 136.392$, *p* $< .001$), suggesting a systematic bias in the model, often toward predicting a slightly more advanced stage than the human evaluator.

Table 3: Comprehensive Diagnostic Performance of the AI Algorithm

Variable		Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
CS1		100	92.8	72.7	100	94
CS2		62.5	97.6	83.3	93.1	92
CS3		87.5	97.6	87.5	97.6	96
CS4		75	97.6	85.7	95.3	94
CS5		88.8	97.5	88.8	97.5	96
CS6		88.8	97.5	88.8	97.5	96

PPV = Positive Predictive Value; NPV = Negative Predictive Value



Graph 3: Comprehensive Diagnostic Performance of the AI Algorithm

Diagnostically, the model demonstrated exceptionally high and consistent specificity, ranging from 92.8% to 97.6% across all stages, indicating a strong ability to correctly identify cases that were not a particular stage. Sensitivity was perfect for CS1 (100%) and remained high for most other stages, though it was lowest for CS2 (62.5%). The overall stage-wise classification accuracy was high, ranging from 92% to 96%. Furthermore, the model's Negative Predictive Value (NPV) was perfect for CS1 and exceeded 93% for all other stages, underscoring its reliability in ruling out a maturation stage. The Positive Predictive Value (PPV) was highest for the later CS5 and CS6 stages (88.8%). In summary, the AI algorithm achieved a high degree of accuracy and diagnostic consistency, showing particular strength in specificity and NPV, which supports its potential utility as an objective and efficient decision-support tool in clinical orthodontic practice.

DISCUSSION:

The developed deep learning model demonstrated robust and consistent performance in the automated classification of all six CVM stages (CS1-CS6). Achieving perfect accuracy (100%) for the initial (CS1) and final (CS6) stages, and high accuracy (87.5% to 94.44%) for the intermediate transitional stages, the model's reliability was statistically uniform across the maturation continuum, as confirmed by a non-significant Chi-square test ($\chi^2 = 4.57$, *p* = 0.47). This stage-wise consistency is a significant strength, indicating that the model did not degrade in performance at any specific developmental phase.

The superior accuracy for CS1 and CS6 can be attributed to their anatomically distinct and unambiguous morphological features—such as uniformly flat vertebral borders in CS1 and definitive concavities with square/vertical shapes in CS6—which present minimal inter-stage ambiguity and are highly amenable to feature extraction by convolutional neural networks (CNNs). This finding aligns with established literature, where early and late stages with high morphological contrast are consistently classified with greater accuracy than the subtle, transitional intermediate stages (CS3-CS4)[14]. The model's performance situates it within the higher tier of current AI implementations for CVM staging, as documented in recent systematic reviews which report a broad accuracy range of approximately 60% to over 90% across different studies and architectures[15].

A critical validation of the model's clinical utility was its high concordance (84%) with manual staging by expert orthodontists. This strong agreement validates the model's capacity to emulate expert visual assessment, a cornerstone of current clinical practice. Importantly, the pattern of disagreement was clinically interpretable and non-random. The vast majority of misclassifications (87.5%) were confined to adjacent stages (e.g., CS2–CS3, CS3–CS4), a pattern frequently reported in the literature and reflective of the inherent subjectivity and morphological continuum in CVM assessment[14]. This "ordinal integrity"—where errors are typically off by only one stage—is clinically reassuring, as misclassifications between adjacent stages have a lesser impact on growth timing decisions compared to gross errors spanning multiple maturation phases[16]. The model's ability to minimize such gross misclassifications underscores its potential to reduce the inter-observer variability that plagues manual CVM assessment[2], thereby introducing greater objectivity and consistency into the diagnostic process.

The model exhibited a strong diagnostic profile, characterized by exceptionally high and consistent specificity (92.8%–97.6%) and Negative Predictive Value (NPV) across all stages, including a perfect 100% NPV for CS1. This indicates a superior ability to correctly rule out a specific CVM stage when it is absent, minimizing false-positive diagnoses. This characteristic is particularly valuable in clinical screening and triage, where confidently excluding a pre-pubertal status (CS1) can streamline patient management[17]. The high specificity aligns with findings from other studies employing CNN and ensemble models, which similarly report specificity metrics above 94-95%[18]. In contrast, the model's sensitivity and Positive Predictive Value (PPV) showed more variability, with a notably lower PPV for CS1 (72.7%). This is a recognized challenge in multi-class CVM systems, often attributable to class imbalance or the subtle morphological overlap between the earliest stages (CS1 and CS2)[19]. This pattern reflects the intrinsic difficulty automated systems face in distinguishing stages with less definitive anatomical boundaries, a limitation also highlighted in other investigations[20].

CONCLUSION:

1. The study successfully developed and validated a ResNet-50 deep learning model capable of automatically classifying Cervical Vertebral Maturation (CVM) stages (CS1-CS6) from digital lateral cephalograms.
2. The model demonstrated high diagnostic efficacy, achieving stage-wise accuracy ranging from 87.5% to 100% and overall accuracy up to 96% on the test set.
3. The AI model's predictions showed a highly significant agreement ($p < 0.001$) with manual staging performed by expert orthodontists, validating its capacity to emulate expert clinical assessment.
4. The model exhibited a strong and consistent diagnostic profile, characterized by high specificity (≥92.8%) and Negative Predictive Value (NPV) across all stages, indicating superior reliability in ruling out specific maturation stages.
5. Key limitations were identified, including the single-center, retrospective design which may affect generalizability, and a tendency for reduced sensitivity/PPV in morphologically similar intermediate stages (CS2-CS4), reflecting a known challenge in CVM assessment.

The model holds significant promise for enhancing diagnostic consistency, improving workflow efficiency, and serving as an accessible aid and educational tool in diverse clinical and training settings.

REFERENCES

1. Paul, R., Bhattacharjee, A., Yadav, D., & Gulia, V. (2023). Growth indicators in orthodontics. Healtalk Publication.
2. Baccetti T, Franchi L, McNamara JA Jr. The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics. Semin Orthod. 2005;11(3):119–129.
3. Perinetti G, Contardo L, Primozic J. The diagnostic performance of dental maturity for identification of the circumpubertal growth phases: a meta-analysis. Prog Orthod. 2016;17(1):1–10.
4. Hägg U, Taranger J. Skeletal stages of the hand and wrist as indicators of the pubertal growth spurt. Acta Odontol Scand. 1980;38(3):187–200.
5. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae [doctoral dissertation]. Pittsburgh (PA): University of Pittsburgh; 1972.
6. 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):316–323.
7. McCarthy J. What is artificial intelligence? Basic questions. Stanford University; 2007.
8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.
9. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge (MA): MIT Press; 2016
10. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3(1):1–10.
11. Paul R, Yadav D, Sharma IK, Gupta M, Gulia V, Darang A. Comparative evaluation of CephNinja for iOS/Android and Nemoceph for computer for cephalometric analysis: a study to evaluate the diagnostic performance for CephNinja for cephalometric analysis. South East Eur J Public Health. 2025;:3480–3494.
12. Golchha Vineet., et al. "Accuracy and Efficiency of Model Analysis Using an App Based Model Analysis Software ". Acta Scientific Dental Sciences 6.7 (2022): 140-146.
13. Park JH, Hwang HW, Moon JH, et al. Automated identification of cephalometric landmarks: Part 1—Comparisons between the latest deep-learning methods YOLOv3 and SSD. Angle Orthod. 2021;91(3):317–324.
14. 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.
15. San Román P, Palma JC, Oteo MD, Nevado E. Skeletal maturation determined by cervical vertebrae development. Eur J Orthod. 2002;24(3):303–311
16. Li H, Li H, Yuan L, Liu C, Xiao S, Liu Z, et al. The psc-CVM assessment system: a three-stage type system for CVM assessment based on deep learning. BMC Oral Health. 2023;23(1):557.
17. Nogueira-Reis F, Cascante-Sequeira D, Farias-Gomes A, de Macedo MMG, Watanabe RN, Santiago AG, et al. Determination of the pubertal growth spurt by artificial intelligence analysis of cervical vertebrae maturation in lateral cephalometric radiographs. Oral Surg Oral Med Oral Pathol Oral Radiol. 2024 Aug;138(2):306-315.

18. Radwan MT, Sin Ç, Akkaya N, Vahdettin L. Artificial intelligence-based algorithm for cervical vertebrae maturation stage assessment. Orthod Craniofac Res. 2023 Aug;26(3):349 355.
19. Mohammad-Rahimi H, Motamadian SR, Nadimi M, Hassanzadeh-Samani S, Minabi MAS, Mahmoudinia E, Lee VY, Rohban MH. Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study. Korean J Orthod. 2022 Mar 25;52(2):112-122.
20. Atici S, Ansari R, Allareddy V, Suhaym O, Cetin A, Elnagar M. Fully automated determination of the cervical vertebrae maturation stages using deep learning with directional filters. PLoS One. 2022;17(8):e0269198.