Accuracy of artificial intelligence in the prediction of cervical vertebrae maturation stages in orthodontics: a systematic review

Abstract

Objective
To assess the ability of artificial intelligence in evaluating cervical vertebrae maturation stages to enhance orthodontic diagnosis considering as main outcome the accuracy of the AI software.

Materials and methods
A search was conducted of 3 databases (Cochrane Library, PubMed/MEDLINE, EMBASE) to identify studies focusing on the ability of atificial intelligence in correctly evaluating the cervical vertebrae maturation stages. Databases were searched including articles until March 2024 only published in English. The Preferred Reporting Items for Reporting Systematic Reviews and Meta Analyses (PRISMA) protocol was adopted, two independent reviewers screened the articles and the agreement was defined by Kappa statistic. The quality of the studies was assessed through the New Castle-Ottawa scale. Due to heterogeneity of data a meta-analysis could not be performed.

Results
The search initially returned 2.953 results and after removing duplicated the number dropped to 1.104. At the end, a total of 7 studies were included in this review. It was evident that AI systems are very good in performing the screening among big amount of data, capable of differentiating what the operator often can not evaluate.

Conclusion
AI can be considered a powerful tool in helping the orthodontic diagnosis since these softwares can manage a big amount of data and perform always the same but on the other hand training of both clinicians and devices is of detrimental importance to overcome the phenomenon of overfitting and instrumental mistakes by the clinicians.

DOI

10.23805/JO.2025.710

Authors

C. Navone
T. Doldo*

Department of Orthodontics, University of Siena, Siena, Italy.

* Corresponding Author

Keywords

Cervical vertebrae maturation stages, pubertal peak, artificial intelligence, tailored treatment plan.

INTRODUCTION

Dentists and orthodontists are involved in the development of dentition and more broadly of the dentofacial complex(1), therefore diagnosis and treatment planning of an orthodontic patient must include knowledge in craniofacial growth and dental development(2). Moreover changes in skeletal and soft tissue integument during adolescent growth must be included in the diagnostic assessment.

Growth prediction is needed for various reasons such as to intercept and correct the malocclusion, as a tool for orthodontic planning, to predict the response to a particular treatment, as patients educational aid and for planning the retention period. The main difference between treating an adult and a child is the chance of growth since in younger patients the maxilla and mandible are still growing; for this reason, an orthopedic treatment allows the orthodontist to monitor and manipulate growth, taking advantage of the growth spurt and recommend a future treatment if necessary.

At present, no method is available to accurately predict accurately the amount, direction and timing of facial growth so usually orthodontists make assumptions referring to average patterns and considering that the patient will follow the same direction and amount of growth during the orthodontic treatment.

Up to now the most reliable method of assessing skeletal maturity with respect to growth for orthodontic purposes has been the radiological assessment: the hand-wrist (HW) radiographic assessment and the cervical vertebrae (CV) assessment. The use of the maturation of cervical vertebrae as an assessment of growth may be very useful since these bones are already visible on the lateral cephalogram, excluding the issue of additional radiations and since the interpretation of the HW x-rays can be an additional misleading factor in the diagnosis.

Certain types of treatment should ideally be performed in certain growth stages, for example: the facial mask is ideal for use at a young age, that is, the cervical vertebral stage 2 (CVS2), while orthognathic surgery and implant positioning are not undertaken until growth ceases (CVS6)(3). For this reason, the CVM stage can be a useful indicator in all ages and for a wide range of orthodontic treatments rather than just for functional treatment.

What is evident from the methods that are usually employed to establish growth patterns is the subjectivity since the consideration relays completely on the clinician and this may lead to error in the evaluation and therefore in the treatment planning especially in the case of a junior orthodontist. What is needed today that the technology reached also our field is a system that can analyse and evaluate a big amount of data in an objective way. Orthodontics has had an incredible development in terms of available technologies with the advent of digital systems such as cone-beam computed tomography, intraoral scanners and new software(4). With the advent of artificial intelligence (AI) and its integration in multiple aspects of the profession great improvements happened in diagnosis, treatment planning, assessment of growth and development, assessment of treatment progress and results, monitoring phase also at distance and long-term follow-up.

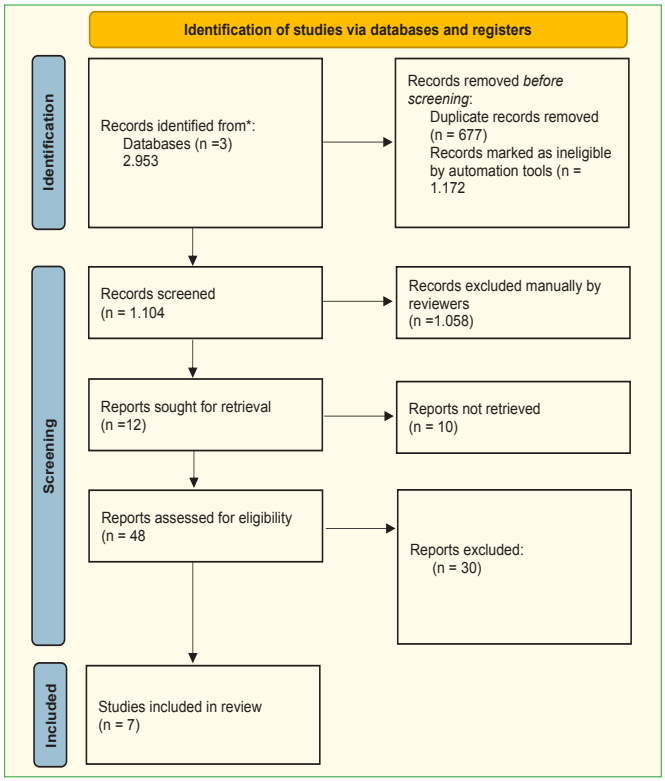

Fig. 1.
PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only

Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.
This work is licensed under CC BY 4.0.
To view a copy of this license, visit
https://creativecommons.org/licenses/by/4.0/

The aim of this systematic review is to systematically review the current knowledge about AI software in evaluating the cervical vertebrae maturation (CVM) stages and to appraise their performance in terms of accuracy.

MATERIALS AND METHODS

This systematic review was conducted following PRISMA (Preferred Reported Items for Systematic Review and Meta-analysis) guidelines(5).

The PROSPERO registration number of our review was:

- Focused question: "How can AI correctly establish the CVM stages to help clinicians in creating a tailored treatment plan and in acting at the right time?" This

work performs a qualitative analysis of the studies regarding the accuracy of AI in determining the CVM stages. A meta-analysis couldn't be performed due to heterogeneity of data.
- The electronic search strategy focused on the following keywords: "Orthodontics" MeSH OR "Orthodontic diagnosis" MeSH AND "Artificial Intelligence" MeSH AND "cervical vertebrae maturation stages" NOT "systematic review" (tiab) OR "meta-analysis (tiab).

The search was based on the PICO (population, intervention, comparison and outcome) elements:
Population: orthodontic and non-orthodontic patients' two-dimensional images and three- dimensional images (periapical, bitewing, orthopantomography, cephalometric exam and cone- beam computed tomography).
Intervention: AI techniques (deep learning, image processing, decision trees, convolutional neural networks, machine learning) applied in predicting CVM stages.
Comparison: automatic algorithms, image analysis, classic models, dentist opinions.
Outcomes: analysis of AI accuracy and performance.

Study design type

For this review the authors decided to include observational studies (cohort studies, case- control studies and cross-sectional studies) and experimental studies (randomized controlled trials, controlled clinical trials) enrolling orthodontic and non-orthodontic patients, published in English.

Inclusion criteria

- only studies with human subjects;
- only published in English;
- full-text;
- studies relevant to fulfill the research questions.

Exclusion criteria

- Case reports;
- review articles;
- animal studies;
- grey literature;
- letters to editors;
- commentaries.

Search methodology

A detailed search was conducted through the following electronic databases: Cochrane Library, PubMed/ MEDLINE, EMBASE, until March 2024 with language restrictions (only English).
A reference management tool (rayyan.qcri.org) was used for initial reference entry and elimination duplicates. Title, abstract and full-text screening were conducted using a specific web-based application for systematic reviews. Titles and abstracts were independently screened and assessed for eligibility by two reviewers (CN and CM) through Rayyan. Full-text papers meeting the inclusion criteria were evaluated in duplicate by the same two reviewers. Any disagreement regarding their eligibility was resolved by consensus (and the agreement between the reviewers was assessed by Kappa statistic). The Cohen's Kappa coefficient that defines the inter-rater agreement was k=0.897 indicating an almost perfect agreement. The following information were retrieved from all the eligible studies: author(s), year of publication, PMID, type of study design, sample size, age range of participants, AI technique used, type of results evaluated, published conclusions (Table 1). Relevant data were systematically extracted from eligible studies.

Quality Assessment

The methodological quality of studies included in this review was assessed using the Newcastle-Ottawa Scale (NOS) for observational studies. Studies were scored

Author and year	PMID	Study design	Sample size	Age range of participants	Test	Control	AI model	Classification of stages	Outcome
Fernanda Nogueira-Reis et al. 2024	38553310	retrospective, cross-sectional	600 lateral cephalometries	from 6 to 17 years	4 CNN models	human examiners	convolutional neural network	Classification into 6 CVM stages and difficulty level	75% of accuracy in determining the CVM stages, the accuracy values indicate a high proportion of correct predictions of the total predisctions made by the model
Jing Zhou et al. 2021	34943436	retrospective	1080 lateral cephalometries	from 6 to 22 years	1 CNN model	human examiners	convolutional neural network	Classification into 6 CVM stages	71% of accuracy
Haizhen Li et al. 2022	35612567	retrospective	6079 lateral cephalometries	from 5 to 18 years	4 CNN models	human examiners	VGG16, GoogLeNet, DenseNet161, ResNet152	Classification into 6 CVM stages	ResNet152 proved to be the best model with a total accuracy of 67.06%
Salih Furkan Atici et al. 2022	35776715	retrospective	1018 lateral cephalometries	from. 4 to 29 years	1 CNN model	human examiners	convolutional neural network with directional filters	Classification into 5 and 6 CVM stages	75.11% accuracy for six classes classification and 84.63% for five classes classification
Hatice Kok et al. 2019	31728776	retrospective	300 individuals	from 8 to 17 yeras	7 algorithms	human examiners	k-nearest neighbors, Naive Bayes, decision tree, artificial neural	Classification into 6 CVM stages	kNN and Log.Regr. algorithms had the lowest accuracy values, while SVM, RF, Tree, and NB algorithms had varying accuracy values so ANN could be the
Salih Furkan Atici et al. 2023	36855827	retrospective	1018 lateral cephalometries	NR	DL network and a CNN	human examiners	deep leaning network and a streuctured deep convolutional network	Classification into 6 CVM stages	82.35% accuracy in female patients and 75.0% in male patients
Hossein Mohammad-Rahimi et al. 2021	35321950	retrospective	890 lateral cephalometries	NR	AI model	human examiners	AI model	Classification into 6 CVM stages and 3 degrees of pubertal sprout: pre-pubertal,	61.62% of accuracy in CVM classification and 82.83% in detecting pubertal stages

Table 1. Study Characteristics

as low risk of bias (RoB) (7–9 stars), moderate RoB (4–6 stars) and high RoB (1–3 stars). Criteria for qualitative assessment comprised the following items: sample selection, comparability and exposure.
Each of the items was assessed and graded (1 or 2 points) according to the suggested criteria. In this analysis, studies with NOS scores of 1–3, 4–6 and 7–9 were defined as of low, intermediate and high quality, respectively (Table 2).

Study	Selection				Comparability		Exposure			Total 9/9	Risk of bias (RoB)
					Comparability of cases and controls						
	Is case definition adequate?	Representativeness of the cases	Selection of controls	Definition of controls	Main factor	Additional factor	Assessment of exposure	Same method of ascertainment for test and control	Non-response rate		
Rahimi et al.	1	1	1	0	1	1	1	1	0	7/9	Low
Atici et al.	1	1	1	1	1	1	1	1	0	8/9	Low
Kok et al.	1	1	1	1	1	1	1	1	0	8/9	Low
Atici et al.	1	1	1	0	1	1	1	1	0	7/9	Low
Li et al.	1	1	1	0	1	1	1	1	0	7/9	Low
Zhou et al.	1	1	1	1	1	1	1	1	0	8/9	Low
Nogueira-Reis et al.	1	1	1	1	1	1	1	1	0	8/9	Low

Table 2. NOS Scale

RESULTS

The electronic literature search was performed among 3 different databases: PubMed/MEDLINE, Cochrane and Embase. The search initially returned 2.953 results and after removing duplicated the number dropped to 1.104. Two independent reviewers (C.N. and C.M.) screened those 1.104 articles by title and abstract and a total of 48 studies were examined more in detail and one of them was excluded because no full text was available. All the studies about vertical maturation stages (seven studies) were retrospective and achieved overall good results regarding AI performance Each study used the six-stages maturation division and three of them used also another classification. All the studies had as control group human examiners and assessed agreement between them with the Kappa statistic. Overall it can be stated that AI devices have a classification accuracy that ranges from 63% to 85%. These percentages are promising but the limitations rely in the fact that in order to perform at their best these devices need to be trained with very big amount of data.

DISCUSSION

Each study used the six- stages classification following the recent consensus about the classification(6, 7) in order to standardize the method. Three out of the seven studies selected used also another approach for the classification as seen in the study by Rahimi et al. (8) in which the images were categorized into three degrees on the basis of the growth sprout: pre-pubertal, growth sprut and post-pubertal and the justification for this further division is that one of the main applications of CVM classification is determining the best treatment timing for mandibular deficiencies. The model evaluated in the study reached an average classification accuracy of 61.62% in CVM stage classification with the best performance in classifying CS6 and an average classification accuracy of 82.83% in detecting pubertal stages. An evident limitation of the study was data imbalance as there were 43 samples in CS1 and 228 samples in CS5, but this is common also to the other studies. In this case to overcome this problem oversampling and data augmentation were used.

A simplified classification was used in the study by Nogueira-Reis (9) in which CS1 and CS2 were fused into Simplified stage 1 (SS1), CS3 and CS4 were fused into SS2 and CS5 and CS6 were fused into SS3. Moreover, the success rate was evaluated also considering full and cropped lateral cephalometric radiographs (LCRs). The accuracy for the original classification into six stages was of 80% for both full and cropped LCRs while in the case of the simplified classification it was of 74 % and 75% for full and cropped, respectively. It was evident that by simplifying the classification, cases with borderline characteristics were grouped and this favored the network evaluation metrics in the classification of the CVM stage concerning the pubertal growth peak. In the study by Atici et al. (10) the images were classified and divided into six-class CVM stages and five-class CVM stages in which CS1 and CS2 are merges into a single stage referred to as CVMS128 (the lower borders of all the three vertebrae are flat, with the possible exception of a concavity at the lower border of C2). In this study the model achieved 75.11% classification accuracy for six class and 84.63% in the five-stage classification, so the model performs better on 5- stage classification probably because the difference between images of CS1 and CS2 is the curvature in vertebrae in CS2 which is not a strong differentiator and increases the error. The other studies, by Zhou (11), by Kok (12), by Atici (13) and Li (14) classified the CVM stages only in a six- stages classification: In the study by Zhou et al. (11) an AI system was developed and its accuracy in CVM staging was 71% on images in which a ROI (region of interest) was identified, the limitation of this study was the size of the testing dataset (1080 cephalometric radiographs) because a lot of data are needed in order

to instruct the algorithm to perform better and also the number of examiners in labelling since the images were labelled only by examiner 1 (twice after three months). In this study the AI labelling was very consistent with the gold standard with a mean error of 0.36mm while the error between the two manual labeling was 0.48. In the study by Li et al. (14) a high number of cephalometric radiographs was used (6079 images) and classified by two experienced orthodontists. Also, a heat map was generated using class activation mapping (CAM) to highlight the regions which are mostly informative in distinguishing the CVM classification; in this study the area between C3 and C4 was activated when the CVM was assessing the images. The study however, achieved an accuracy of only 67.06% due to some limitations: the quality and quantity of the dataset and the impossibility of the CNN algorithm in identifying some special features related to the cervical stages.
In the study by Kok (12) seven algorithms of AI that are frequently used for classification were evaluated; confusion matrices were calculated for each algorithm and the most successful one was ANN while kNN and Log. Reg. had the lowest accuracy. Atici et al. (13) developed a parallel structured deep convolutional neural network (CNN) with a pre- processing layer that performs feature extraction. The name of the model is AggregateNet and with data augmentation it produced 82.35% accuracy in female patients and 75.0% in male patients, the data set was split by gender since male patients may experience a different rate of growth than female patients. Overall, it can be observed that these models perform well but with data augmentation, directional filters and chronological age input can perform better but very large samples are needed in order to properly train the models and prevent overfitting.

CONCLUSION

AI, through machine learning algorithms, is able to analyze large amounts of heterogeneous data, including X-ray images, anthropometric data, and clinical information, identifying complex patterns and non-obvious relationships between the different variables. This sophisticated processing capacity allows the development of predictive models capable of estimating skeletal and dental growth in specific individuals with high accuracy. Despite the challenges, AI offers several advantages over traditional methods for predicting skeletal and dental growth, these are:

- High accuracy: from this review we can say that in some cases AI models are able to estimate growth more accurately than orthodontists, allowing for more precise diagnosis and treatment planning. This was evident in the study by Zhang et al.(13) suggesting that orthodontists with less clinical experience tended to be overcautious in the prediction of mandibular growth making evident that subjectivity plays a major role in this setting while AI is not influenced by the human thinking.
- Personalized approach: AI models can be customized for each individual, considering their unique characteristics and providing more accurate predictions.
- Continuous learning ability: AI models can continuously learn and adapt to new data, constantly improving their predictive performance.

Nevertheless, these algorithms need to be accurately trained with a big number of data and in order to enhance their predictions they need as many records as possible. One of the main issues is the risk of overfitting, which occurs when a model overfits to training data and fails to generalize well to new data resulting in a model that can not make accurate predictions or conclusions from any data other than the training data. To prevent overfitting, regularization and cross-validation techniques must be adopted. As seen in the studies evaluated, algorithms that used regularization performed better than the others, for example the least absolute shrinkage and selection operator (LASSO). Healthcare professionals need to be accurately trained in order to use these new technologies in the proper way without falling into mistakes. AI, through the routine cephalometric radiographic exam, can help the orthodontist in identifying the growth sprout by evaluating the cervical vertebrae maturation stages and therefore leading to the best timing for the orthodontic treatment. By using the same x-ray, AI models can generate a skeletal growth prediction helping the orthodontist in the decision-making process in order to plan orthodontic or surgical treatments, optimizing esthetics and functional results. Moreover, AI could allow the early identification of children at risk of developing malocclusions or other dento-facial problems, allowing timely intervention with preventive or corrective treatments.
In conclusion, AI appears to be a precious ally for the dentistry of the future, offering new opportunities for the prevention, diagnosis and treatment planning of skeletal and dental growth problems. Integrating AI into clinical dental practice will require careful evaluation of its benefits and limitations, as well as adequate training of healthcare professionals. However, the potential of this technology is immense and paves the way for a future where skeletal and dental growth can be managed more effectively and individually, ensuring better oral health for all.

Funding sources/sponsors
This project does not receive any external funding.

Conflict of interest
None.

REFERENCES

1. William R. Proffit Henry W. Fields BELDMSarver. Contemporary Orthodontics . sixth.; 2019.
2. Nanda RS. The contributions of craniofacial growth to clinical orthodontics. Am J Orthod Dentofacial Orthop. 2000;117(5). doi:10.1016/S0889-5406(00)70197-1
3. Lucchese A, Bondemark L, Farronato M, et al. Efficacy of the Cervical Vertebral Maturation Method: A Systematic Review. Turk J Orthod. 2022;35(1):55-66. doi:10.5152/TurkJOrthod.2022.21003
4. Dipalma G, Inchingolo AD, Inchingolo AM, et al. Artificial Intelligence and Its Clinical Applications in Orthodontics: A Systematic Review. Diagnostics. 2023;13(24). doi:10.3390/diagnostics13243677
5. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ. 2021;372. doi:10.1136/bmj.n71
6. Baccetti T, Franchi L, McNamara JA. An Improved Version of the Cervical Vertebral Maturation (CVM) Method for the Assessment of Mandibular Growth. Angle Orthodontist. 2002;72(4). doi:10.1043/0003-3219(2002)072<0316:AIVOTC>2.0.CO;2
7. McNamara JA, Franchi L. The cervical vertebral maturation method: A user's guide. Angle Orthodontist. 2018;88(2):133-143. doi:10.2319/111517-787.1
8. Mohammad-Rahimi H, Motamadian SR, Nadimi M, et al. Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study. Korean J Orthod. 2022;52(2). doi:10.4041/kjod.2022.52.2.112
9. Nogueira-Reis F, Cascante-Sequeira D, Farias-Gomes A, 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;138(2):306-315. doi:10.1016/j.oooo.2024.02.017
10. Atici SF, Ansari R, Allareddy V, Suhaym O, Cetin AE, Elnagar MH. Fully automated determination of the cervical vertebrae maturation stages using deep learning with directional filters. PLoS One. 2022;17(7 July). doi:10.1371/journal.pone.0269198
11. Zhou J, Zhou H, Pu L, et al. Development of an artificial intelligence system for the automatic evaluation of cervical vertebral maturation status. Diagnostics. 2021;11(12).doi:10.3390/diagnostics11122200
12. Kök H, Acilar AM, İzgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. Prog Orthod. 2019;20(1). doi:10.1186/s40510-019-0295-8
13. 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(S1). doi:10.1111/ocr.12644
14. Li H, Chen Y, Wang Q, et al. Convolutional neural network-based automatic cervical vertebral maturation classification method. Dentomaxillofacial Radiology. 2022;51(6). doi:10.1259/dmfr.20220070