Current neural networks demonstrate potential in automated cervical vertebral maturation stage classification based on lateral cephalograms

DOI:
10.1016/j.jebdp.2023.101928

Review Analysis and Evaluation

DECLARATIVE TITLE: Current neural networks demonstrate potential in automated cervical vertebral maturation stage classification based on lateral cephalograms

REVIEWERS: Lingyun Cao, Hong He, Fang Hua

PURPOSE/QUESTION: How accurate are the current neural networks for cervical vertebral maturation stage classification based on lateral cephalograms?

ARTICLE TITLE AND BIBLIOGRAPHIC INFORMATION
Neural networks for classification of cervical vertebrae maturation: a systematic review. Mathew R, Palatinus S, Padala S, Alshehri A, Awadh W, Bhandi S, Thomas J, Patil S. Angle Orthod. 2022 Nov 1;92(6):796-804.

Strength of Recommendation Taxonomy (SORT) Grading

STRENGTH OF RECOMMENDATION GRADE:

Grade C	Consensus, disease-oriented evidence, usual practice, expert opinion, or case series for studies of diagnosis, treatment, prevention, or screening

LEVEL OF EVIDENCE:

Level 2	Limited-quality evidence

SOURCE OF FUNDING: No financial support was reported.

TYPE OF STUDY/DESIGN: Systematic review

KEY WORDS: Neural networks, Artificial intelligence, Cervical vertebral maturation, Bone age, Orthodontics

SUMMARY

Selection Criteria

The authors searched four databases (MEDLINE, Scopus, Embase, and Web of Science) for diagnostic accuracy studies published in English. The search strategy was a combination of terms related to cervical vertebrae maturation (CVM) and neural networks (NNs), including “skeletal maturity”, “bone age”, “artificial intelligence”, “machine learning” and others. Experimental studies, prospective/retrospective observational studies, randomized controlled trials, case-control studies, and cohort studies were considered eligible. Abstracts, opinions, conference papers, reviews, and studies that did not use NNs were excluded. The study screening process was performed in two stages, each by two independent reviewers.

Key Study Factors

The review was conducted to compare the diagnostic accuracy of NNs and the ground truth determined by human observers. Among the 8 included studies, 6 utilized equally distributed cervical vertebrae maturation stage (CVS) data as the training datasets. Three studies used the radiograph image as the input data, while 5 utilized manually labeled datasets with measurements. The input measurements included linear measurements in both the vertical and horizontal directions, as well as ratios derived from them. Besides, 2 studies applied cross-validation method for the training and testing datasets, and 6 studies used separate datasets. Three studies utilized pre-developed convolutional NNs, and modified them to suit the input. Six studies created new NNs that were specifically designed for CVS classification, 1 of which used both pre-developed NN and newly developed NN. In all 8 included studies, human observers classified CVS according to Hassel method[1] or the method modified by Baccetti et al.[2]

Main Outcome Measures

The main outcome of this review was the accuracy of CVS classification based on lateral cephalograms, reported as the level of agreement between NNs and the reference standard.

Main Results

The findings of included studies were presented in a descriptive manner and no meta-analysis was carried out. The reported accuracy varied significantly across these studies, ranging from less than 50% to over 95%. Notably, 5 studies reported accuracy levels of above 90%, while 1 study reported an accuracy of 58.3% and another reported 62.5%. Furthermore, 1 study exclusively reported accuracy values for different CVS (range, 47.4% to 93%) but did not provide an overall accuracy figure. Comparison between NNs and other forms of artificial intelligence (AI), such as Bayes models, was conducted in 3 studies and they all concluded that NNs were more accurate. Additionally, 1 study found that NN had the most stable results compared to other AI algorithms.

According to the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) assessment tool, 7 studies that lacked evaluation of inter-observer or intra-observer agreement were considered having some concerns for bias in reference standard. Two studies without separate test dataset were considered having some concerns for bias in index test. All the 8 included studies presented low concerns regarding their applicability.

Conclusions

The authors concluded that NNs can successfully classify various stages of CVM based on lateral cephalograms. However, the accuracy of NNs showed significant variations across different studies.

COMMENTARY AND ANALYSIS

In the field of orthodontics and dentofacial orthopedics, proper timing of treatments is crucial for achieving the most favorable response. However, relying solely on chronological age is inadequate since craniofacial growth is influenced by many genetic and environmental factors.[3] Although hand-wrist radiographs had traditionally been used as the gold standard to assess the bone age, one of its main drawbacks is the additional radiation exposure to the patient.[1-3] The CVM method, first introduced by Lampalski,[4] is based on the morphology of the second, third, and fourth cervical vertebrae (C2/C3/C4).[1,2] Since these structures are visible on the lateral cephalograms and lateral cephalograms are routinely prescribed for orthodontic treatment, the CVM method is considered practical and safe for the patients.[1,2]

Over the past few decades, artificial intelligence (AI) has made remarkable strides in the field of medical imaging.[5,6] In the realm of dental imaging, AI has been applied to automate various clinical processes, such as caries detection, cephalometric analysis, and tumor classification.[7-9] NNs, as the subfield of AI, are algorithms that mimic the biological neural structure and are capable of learning to discriminate important patterns from various resources including radiograph images.[10-12] By training with the input/output data set by human experts, NNs are designed to provide stable, accurate, and rapid interpretations, thereby enhancing the objectivity and efficiency of clinical processes.[13,14] In orthodontics, NNs have been applied for landmark detection on cephalograms, and have achieved clinically acceptable results.[7] For cervical maturation assessment, studies have demonstrated that NNs have good consistency with human examiners in landmark labelling and CVS determination.[3,13]

The systematic review of Mathew et al.[15] focused on the research progress of NNs in automatic CVS determination based on lateral cephalograms. This is an innovative project as it may be the first systematic review regarding this specific topic. The remarkable stability and accuracy exhibited by NNs set them apart from other classical AI algorithms.[11,13,16] By focusing on NNs specifically and excluding other classical types of algorithms, the scope of this systematic review is reasonable and relevant. The utilization of QUADAS-2 for evaluating the quality and applicability of diagnostic accuracy studies is also commendable.

However, some limitations exist in this review. Firstly, its search strategy and eligibility criteria may not be adequate for the research question. Given the interdisciplinary nature of this review, it is imperative to search IEEE (Institute of Electrical and Electronics Engineers) Xplore, an important database in the fields of computer science. IEEE Xplore contains articles, proceedings, and standards in fields of computer science and engineering. For example, one study indexed in IEEE Xplore and seems to meet the eligibility criteria,[17] was not included in the review. The NN developed in this study realized the automation of bone age estimation by focusing on the concavity of C3, and has achieved a correlation of around 0.9. Similarly, conference papers are valued greatly in the field of computer science. However, conference papers were excluded from the review. This may raise the risk of overlooking existing research findings and arriving at an incomplete body of evidence. For instance, a conference paper by Makaremi et al.[18] in 2019, which developed a fully-automated CVS determination NN based on 300 lateral cephalograms with a 90% accuracy, was not included in the systematic review.

Second, the authors did not carry out quantitative syntheses. Pooled accuracy of the NNs can provide a more comprehensive and precise assessment of their capability in automatic CVS determination. Moreover, subgroup analysis could be carried out, accounting for the different types of NNs applied or different training datasets used in the included studies. In a previous similar review by Schwendicke et al.[19], meta-analyses and subgroup analyses were performed, revealing that NNs generally exhibited higher accuracy for landmark detection on 2D lateral cephalograms in comparison to 3D radiographs.

Third, some of the conclusions may not be accurate. The authors stated that studies using newly developed NNs demonstrated higher accuracy compared to those using pre-developed NNs, and NNs developed with linear and ratio inputs (secondary handcrafted features) outperformed those developed using radiographic inputs directly. However, out of the 8 studies included, 3 employed pre-developed NNs, which also utilized radiographs as inputs.[3,20,21] The remaining 5 studies developed new NN models and employed linear and ratio inputs.[11-13,16,22] While both "inputs" and "NN model types" were factors that could impact accuracy, it is challenging to determine which factor actually mattered. Moreover, among the 8 studies analyzed, the results of 4 did not support the authors' statements. Two studies achieved over 90% accuracy using radiographs as direct inputs.[3,21] In contrast, another 2 studies achieved less than 60% accuracy with new NNs using secondary linear and ratio values as inputs.[16,22] Therefore, further research is needed to assess the impact of the "inputs" and "NN model types" on the accuracy of the NNs.

Besides, it is important to note that the utilization of those secondary handcrafted features can only be considered "semi-automated" since it is based on cephalometric data (not cephalograms *per se*), which still necessitates manual tracing of landmarks. Whereas algorithms developed based on cephalograms directly, which do not require additional tracing, could be considered as true "fully-automated" approach.

Additionally, an important conclusion that might have been missed in this systematic review is the variation in accuracy of the NNs for different CVS. Out of the 8 studies included, 5 of them reported the lowest accuracy at CVS3 or CVS4, while the highest accuracy was obtained at CVS1 or CVS6.[3,11,12,20,21] The pubertal growth spurt, which takes place between CVS3 and CVS4, results in the increased variation in cervical vertebrae morphology.[21] As a consequence, it becomes more challenging to differentiate between stages and achieve accurate results. On the other hand, CVS1 and CVS6, representing minimal and maximal maturity, exhibited more stability and are easier to differentiate.[14] This suggests that future studies need to consider providing additional training for CVS3 and CVS4 specifically.

In summary, NNs have demonstrated some potential in aiding orthodontists with automated CVS determination based on lateral cephalograms. Routinely produced reports on CVS may play an important role in orthodontic diagnosis and treatment planning, reminding clinicians to evaluate and consider patients' bone age, and helping those with inadequate experience to achieve accurate CVS determination as well as correct clinical decision. However, it is important to notice that the accuracy of NNs has shown notable variations across different studies and for different CVS. For clinicians, results obtained from the current NNs should be interpreted cautiously, particularly for the determination of CVS3 and CVS4 in which case additional human validation may be necessary.

REVIEWER NAME and CONTACT INFORMATION:

Lingyun Cao, BDS, MSc

Department of Orthodontics, State Key Laboratory of Oral & Maxillofacial Reconstruction and Regeneration, Key Laboratory of Oral Biomedicine Ministry of Education, Hubei Key Laboratory of Stomatology, School & Hospital of Stomatology, Wuhan University, Wuhan, China

Hong He, BDS, MSc, PhD

Department of Orthodontics, State Key Laboratory of Oral & Maxillofacial Reconstruction and Regeneration, Key Laboratory of Oral Biomedicine Ministry of Education, Hubei Key Laboratory of Stomatology, School & Hospital of Stomatology, Wuhan University, Wuhan, China, drhehong@whu.edu.cn

Fang Hua, BDS, MSc, PhD

Center for Orthodontics and Pediatric Dentistry at Optics Valley Branch, Center for Evidence-Based Stomatology, State Key Laboratory of Oral & Maxillofacial Reconstruction and Regeneration, Key Laboratory of Oral Biomedicine Ministry of Education, Hubei Key Laboratory of Stomatology, School & Hospital of Stomatology, Wuhan University, Wuhan, China; Division of Dentistry, School of Medical Sciences, The University of Manchester, Manchester, UK, huafang@whu.edu.cn

REFERENCES

1. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. *Am J Orthod Dentofacial Orthop.* 1995;107:58-66. doi:10.1016/s0889-5406(95)70157-5
2. Baccetti T, Franchi L, McNamara JA. The Cervical Vertebral Maturation (CVM) Method for the Assessment of Optimal Treatment Timing in Dentofacial Orthopedics. *Semin Orthod.* 2005;11:119-129. doi:10.1053/j.sodo.2005.04.005
3. Makaremi M, Lacaule C, Mohammad-Djafari A. Deep Learning and Artificial Intelligence for the Determination of the Cervical Vertebra Maturation Degree from Lateral Radiography. *Entropy (Basel).* 2019;21:1222. doi:10.3390/e21121222
4. Lamparski, DG. Skeletal Age Assessment Utilizing Cervical Vertebrae. Master Science Thesis. 1972
5. Lin L, Tang B, Cao L, Yan J, Zhao T, Hua F, He H. The knowledge, experience, and attitude on artificial intelligence-assisted cephalometric analysis: Survey of orthodontists and orthodontic students. *Am J Orthod Dentofacial Orthop.* 2023;S0889-5406(23)00398-0. doi:10.1016/j.ajodo.2023.07.006
6. Law M, Seah J, Shih G. Artificial intelligence and medical imaging: applications, challenges and solutions. *Med J Aust.* 2021;214(10):450-452.e1. doi:10.5694/mja2.51077.
7. Cao L, He H, Hua F. Deep Learning Algorithms Have High Accuracy for Automated Landmark Detection on 2d Lateral Cephalograms. *J Evid Based Dent Pract.* 2022;22(4):101798. doi:10.1016/j.jebdp.2022.101798.
8. Kuhnisch J, Meyer O, Hesenius M, Hickel R, Gruhn V. Caries Detection on Intraoral Images Using Artificial Intelligence. *J Dent Res.* 2022;101(2):158-165. doi:10.1177/00220345211032524.
9. Putra RH, Doi C, Yoda N, Astuti ER, Sasaki K. Current applications and development of artificial intelligence for digital dental radiography. *Dento maxillo facial radiology.* 2022;51:20210197. doi:10.1259/dmfr.20210197.

10. Zhao T, Zhou J, Yan J, Cao L, Cao Y, Hua F, He H. Automated Adenoid Hypertrophy Assessment with Lateral Cephalometry in Children Based on Artificial Intelligence. *Diagnostics (Basel).* 2021;11(8):1386. doi:10.3390/diagnostics11081386.
11. 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. *Dento maxillo facial radiology.* 2020;49:20190441. doi:10.1259/dmfr.20190441.
12. Kok 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.
13. Kok H, Izgi MS, Acilar AM. Evaluation of the Artificial Neural Network and Naive Bayes Models Trained with Vertebra Ratios for Growth and Development Determination. *Turk J Orthod.* 2020;34(1):2-9. doi: 10.5152/TurkJOrthod.2020.20059.
14. Zhou J, Zhou H, Pu L, Gao Y, Tang Z, Yang Y, You M, Yang Z, Lai W, Long H. Development of an Artificial Intelligence System for the Automatic Evaluation of Cervical Vertebral Maturation Status. *Diagnostics (Basel)*. 2021;11(12):2200. doi:10.3390/diagnostics11122200.
15. Mathew R, Palatinus S, Padala S, et al. Neural networks for classification of cervical vertebrae maturation: a systematic review. *Angle Orthod.* 2022;92(6):796-804. doi:10.2319/031022-210.1.
16. Kok H, Acilar AM, Izgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog Orthod.* 2019;20(1):41. doi:10.1186/s40510-019-0295-8.
17. Moraes DR, Casati JPB, Rodrigues ELL. Analysis of polynomial behavior of the C3 cervical concavity to bone age estimation using artificial neural networks. *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*2013:1-6.
18. Makaremi M, Lacaule C, Mohammad-Djafari A. Determination of the Cervical Vertebra Maturation Degree from Lateral Radiography. *Proceedings (Basel).* 2019;33:30. doi:10.3390/proceedings2019033030.
19. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, Demarco F, Krois J. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Investig.* 2021;25(7):4299-4309. doi:10.1007/s00784-021-03990-w.
20. Kim EG, Oh IS, So JE, Kang J, Le VNT, Tak MK, Lee DW. Estimating Cervical Vertebral Maturation with a Lateral Cephalogram Using the Convolutional Neural Network. *J Clin Med.* 2021;10(22):5400. doi:10.3390/jcm10225400.
21. Seo H, Hwang J, Jeong T, Shin J. Comparison of Deep Learning Models for Cervical Vertebral Maturation Stage Classification on Lateral Cephalometric Radiographs. *J Clin Med.* 2021;10(16):3591. doi:10.3390/jcm10163591.
22. Amasya H, Cesur E, Yildirim 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-e179. doi:10.1016/j.ajodo.2020.08.014.