Systematic Review Article

Neural networks for classification of cervical vertebrae maturation: a systematic review

Reji Mathew[a]; Stephen Palatinus[b]; Soumya Padala[c]; Abdulrahman Alshehri[d]; Wael Awadh[d]; Shilpa Bhandi[e]; Jacob Thomas[f]; Shankargouda Patil[g]

ABSTRACT

Objective: To assess the accuracy of identification and/or classification of the stage of cervical vertebrae maturity on lateral cephalograms by neural networks as compared with the ground truth determined by human observers.

Materials and Methods: Search results from four electronic databases (PubMed [MEDLINE], Embase, Scopus, and Web of Science) were screened by two independent reviewers, and potentially relevant articles were chosen for full-text evaluation. Articles that fulfilled the inclusion criteria were selected for data extraction and methodologic assessment by the QUADAS-2 tool.

Results: The search identified 425 articles across the databases, from which 8 were selected for inclusion. Most publications concerned the development of the models with different input features. Performance of the systems was evaluated against the classifications performed by human observers. The accuracy of the models on the test data ranged from 50% to more than 90%. There were concerns in all studies regarding the risk of bias in the index test and the reference standards. Studies that compared models with other algorithms in machine learning showed better results using neural networks.

Conclusions: Neural networks can detect and classify cervical vertebrae maturation stages on lateral cephalograms. However, further studies need to develop robust models using appropriate reference standards that can be generalized to external data. (*Angle Orthod.* 2022;92:796–804.)

KEY WORDS: Artificial intelligence; Cervical vertebrae maturation; Machine learning; Neural networks; Skeletal maturity

INTRODUCTION

Artificial intelligence (AI) has powered voice-activated personal assistants and self-driving cars from the pages of science fiction to reality. A similar incursion into medicine has seen AI used to diagnose cancer, predict survival, simulate the spread of disease, visualize musculoskeletal tissue, and predict hospital

[a] Associate Professor, Department of Oral and Maxillofacial Radiology, College of Dental Medicine, Midwestern University, Downers Grove, Illinois, USA.

[b] Professor and Director of Clinical Faculty, Dental Institute, College of Dental Medicine, Midwestern University, Downers Grove, Illinois, USA.

[c] Assistant Professor and Director of Orthodontics, Rush Orthodontics and Craniofacial Center, Department of Plastic Surgery and Reconstruction, Rush University Medical Center, Chicago, Illinois, USA.

[d] Assistant Professor, Division of Orthodontics, Department of Preventive Dental Sciences, College of Dentistry, Jazan University, Jazan, Saudi Arabia.

[e] Assistant Professor, Department of Restorative Dental Sciences, Division of Operative Dentistry, College of Dentistry, Jazan University, Jazan, Saudi Arabia.

[f] Private Practice, Cochin, Kerala, India.

[g] Adjunct Faculty, College of Dental Medicine, Roseman University of Health Sciences, South Jordan, Utah and Professor, Centre of Molecular Medicine and Diagnostics (COMManD), Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India.

Corresponding author: Dr Shankargouda Patil, College of Dental Medicine, Roseman University of Health Sciences, South Jordan, Utah 84095
(email:spatil@roseman.edu)

Accepted: June 2022. Submitted: March 2022.
Published Online: September 7, 2022

attendance.[1–5] Since one of its first uses in medicine as a rule-based expert system that advised antimicrobial therapy,[6] medical applications of AI have grown tremendously. In the oral health sciences, reviews explain the use of AI in dentistry, orthodontics, dental imaging, and cleft lip and/or palate.[7–10]

A popular form of AI is machine learning, which provides computers knowledge through data and observations without explicit programming. Such systems improve with experience and more data. Extensive digital data help create machine-learning models with very high accuracy. A subset of machine learning, deep learning, has greater flexibility and can extract abstract features from raw data.[11] It relies on multiple processing layers to detect features in a hierarchical structure.[12] In medical imaging, machine-learning models can analyze chest radiographs to detect numerous lesions.[13] They can make comprehensive radiologic examinations, sometimes outperforming resident doctors.[14–16]

Neural networks (NNs) are deep-learning algorithms that structurally emulate the neural connections in the human body.[17] Each perceptron (the basic unit) receives one or more input from a previous layer. Each input value has a weight (or strength) assigned to it. The total input is a sum of all input values subject to the weights assigned to each. This is processed via a mathematical function to provide an output from the perceptron.[18] This can be a final output value or an input value for the next layer of the NN.

An NN typically has an input layer, one or more hidden layers, and an output layer. The input layer has one or more perceptrons, and each receives one input. In subsequent layers, one perceptron may get multiple inputs. The input variables are scaled to a value ranging between the assigned all and none values (such as between 0 and 1) in the input layer. This is in contrast to a binary input of all or none (i.e., 0 or 1). These graded values are processed by the network to provide a graded output.[18] Therefore, the output is ordinal data, making NNs a good choice for multicategory classification problems.

Lateral cephalograms can be used to determine the peak pubertal growth of the jaws depending on skeletal maturity.[19] Individuals treated with functional orthopedic appliances before attaining skeletal maturity achieve optimal growth and develop a harmonious relationship between the jaws.[20] These appliances are cost-effective, in contrast to surgical repositioning at later stages, which, additionally, has a high morbidity. Skeletal maturity is determined by a technique initially explained by Lamparski[21] and later modified by Hassel and Farman.[22] They used changes in the shape of the second to fifth cervical vertebrae for assessment. This method was modified further by Baccetti et al.[23] to include the shape of the inferior border of individual vertebrae and limited to the fourth cervical vertebrae as they are seen with a protective radiation collar, and the stage can be defined based on a single radiograph.

A drawback of the (CVM) method is the need for specialized training that is not provided to general dentists, and orthodontists may also find it difficult to distinguish the shapes of vertebrae.[24] Thus, patients who can benefit from treatment with myofunctional appliances may not get timely referral to a specialist and ultimately need surgery later. AI can detect features that may not be obvious to human observers. Automating or augmenting the process of detecting cervical vertebrae maturation stages with AI can help referral of suitable patients for orthopedic or myofunctional therapy and train or aid residents in diagnosis. The aim of this review was to assess the accuracy of identification and/or classification of the stage of cervical vertebrae maturity on lateral cephalograms by NNs compared with the ground truth determined by human observers.

MATERIALS AND METHODS

This systematic review was conducted according to the PRISMA guidelines. The question posed was, "What is the accuracy of neural networks in detecting cervical maturation stages on lateral cephalometric radiographs compared to human observers?"

Search Strategy and Study Selection

A search of PubMed (MEDLINE), Scopus, Embase, and Web of Science was conducted in November 2021. The queries used to search each database are listed in Table 1. The results obtained from each database were exported to EndNote online (Clarivate Analytics). After duplicates were removed, two reviewers (Dr Mathew and Dr Awadh) independently screened the remaining articles based on their titles and abstracts. Potentially relevant studies and those with insufficient information were selected for a full-text reading by two independent reviewers (Dr Alshehri and Dr Padala). Studies that satisfied the inclusion criteria were selected for data extraction.

Inclusion Criteria

- Population: Lateral cephalograms with varying patient age that captured the cervical vertebrae
- Intervention: Using an NN to identify or classify the stage of skeletal maturity with inputs from a lateral cephalogram
- Comparator/reference: A human observer's identification and classification of the cervical vertebrae to assess skeletal maturity based on radiographic characteristics

Table 1. Search Queries for Each Database

S.No	Database	Query	Results
1	Embase	('cervical vertebra' OR 'bone age' OR 'bone maturation') AND ('artificial intelligence'/exp OR 'artificial intelligence' OR 'machine learning':jt OR 'neural network':au)	96
2	Web Of Science	(artificial intelligence OR machine learning OR neural network) AND (cervical vertebrae OR cervical bone OR skeletal maturity)	109
3	Scopus	ALL (artificial AND intelligence OR machine AND learning OR algorithm) AND (cervical AND vertebrae OR skeletal AND maturation OR maturation OR maturity)	154
4	PubMed	(((((((artificial intelligence) OR (machine learning)) OR (neural network))) AND ((cervical vertebrae) OR (skeletal maturity))) AND (((((artificial intelligence) OR (machine learning)) OR (neural network))) AND ((cervical vertebrae) OR (skeletal maturity))))) AND (radiography OR cephalogram)	66

- Outcome: Accuracy of the automated classification of cervical vertebral stages using an NN compared with a human observer
- Study type: Studies that compare the use of automated diagnostic systems with the predetermined human observer results were considered; experimental studies, randomized controlled trials, prospective and retrospective observational studies, case-control studies, and cohort studies were included

Exclusion Criteria

Conference papers, abstracts, opinion pieces, literature or systematic reviews, and studies that used any form of AI other than NNs were not included. Studies with full text unavailable in English were excluded.

Data Extraction

Two reviewers (Dr Alshehri, Dr Thomas) independently chose the studies eligible for data extraction according to the inclusion criteria, regardless of the qualitative assessment. Details from the studies were extracted and summarized in tabular form.

Methodologic Assessment

The review focused on the accuracy of diagnosing CVM stages on radiographs by NNs with the diagnosis by human observers as the reference. Therefore, the Cochrane tool for diagnostic test accuracy, the latest version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2), was used for the methodologic assessment of studies.[25] The assessment was carried out by two independent reviewers (Dr Palatinus, Dr Bhandi). The results were discussed among both reviewers, and a third reviewer was consulted for arbitration.

RESULTS

An initial search that was adapted according to each database provided 425 articles across all searched databases. The search queries and results are displayed in Table 1. The initial screening of the title and abstract by both reviewers was in almost perfect agreement (Cohen's Kappa = 0.93). Disagreements were arbitrated by a third reviewer (Dr Patil). After screening, 11 articles were considered for full-text evaluation; finally, eight articles were found eligible for inclusion in the review.[26–33] The reviewers were in complete agreement. Three articles that were excluded either did not use NNs or did not report on accuracy.[34–36] The PRISMA flow diagram is shown in Figure 1. The eligible studies were selected for data extraction and qualitative assessment. Most studies published the development of an NN model for classifying cephalograms according to the stages of CVM. One study compared the assessment of lateral cephalometric radiographs by a previously described model to four human observers. The data obtained from the studies are presented in Table 2.

Data Sets

The number of samples varied from 72 to 1870 among the studies. The studies used locally sourced data sets of different sizes when developing the models. The data were obtained from university settings; however, Makaremi et al.[33] did not mention the source. The data sets were split into the training and testing data in five studies.[26,28,30,31,33] The training data set was used to develop and validate the model. The age range of the data in all studies was adequate to include all of the CVM stages. Six studies used equal distribution of training data across stages.[26,27,29–33]

Reference Standards

The human observers who classified the lateral cephalograms used two methods: the Hassel and Farman method and the method modified by Baccetti et al.[22,37] The experience of the human observers and their field of expertise varied. Interobserver agreement was calculated in two studies and was poor in one study.[26,29] Five studies used a single expert's judgment as a reference standard.[27,28,30–32]

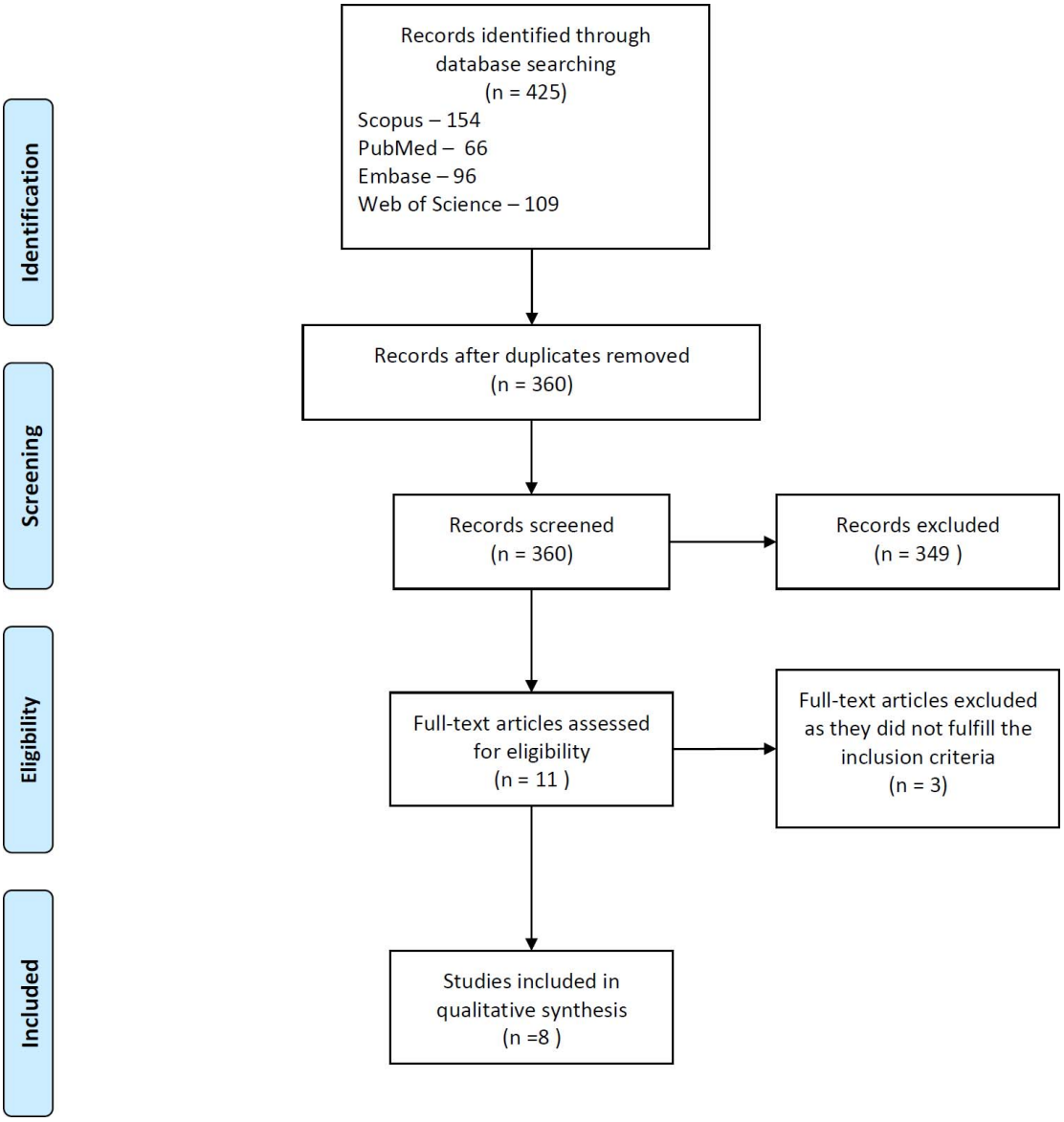

Figure 1. PRISMA flow chart of study selection.

Input Features

Five studies used manually labeled data sets with measurements.[28–32] Three used the image of the radiograph as input.[26,27,33] Measurements used as inputs were linear measurements in the vertical and horizontal directions and ratios derived from them. One study that used photographs cropped the region of interest on the lateral cephalograms and used filters on the image.[33] One study devised a region-of-interest detector and a segmenter module for the photographs to be used with the NN instead of manual measurements.[26]

Neural networks

Three studies used previously trained convolutional NNs (a type of NN) for identification of radiographs.[26,27,33] Two modified them to suit the input provided.[26,27] Six studies developed an NN specifically to classify radiographs.[28–33] Makaremi et al.[33] used both pretrained networks and developed a NN to classify radiographs. The number of hidden layers and perceptrons in the hidden layers of the networks varied. The pretrained networks adapted for this task had more than 50 layers depending on the network. The NNs devised by the researchers solely for the classification task had one hidden layer or simple six-layered architectures.

Performance of the NNs

The accuracy with which the systems could classify the radiographs differed. Accuracy was the agreement

Table 2. Summary of Data Extracted From Selected Studies[a]

S.No	Author, Year	Country	Age Range, y	Sample Size	CVM Method Used	Inputs	Reference Standards/Comparisons	Outcome
1	Kim et al., 2021[26]	Korea	6–18	Training: 600 images Testing: 120 images	Baccetti et al.	Images of lateral cephalograms equally distributed across stages	Two specialists	The combination of the CNN with a region-of-interest detector and segmentor module was significantly more accurate (62.5%) than without them.
2	Seo et al., 2021[27]	Korea	6–19	600 lateral cephalograms	Baccetti and Franchi	Cropped images of lateral cephalograms equally distributed across stages displaying the inferior border of C2 to C4	One radiologist	A pretrained network, Inception-ResNet-v2, had relatively high accuracy of 0.941 $\pm$ 0.018 when adapted. It also had the highest recall and precision scores among all pretrained models tested.
3	Amasya et al., 2020[28]	Turkey	10–30	72 images	Baccetti and Franchi	Manually labeled image data set with 54 features and ratios with equal distribution across stages	Three dentomaxillofacial radiologists and an orthodontist	Interobserver agreement between researchers and the ANN model was substantial to almost perfect ($w\kappa = 0.76$–0.92). Percentage agreements between the ANN model and each researcher were 59.7%, 50%, 62.5%, and 61.1%.
4	Amasya et al., 2020[29]	Turkey	10–30	647 images (498 for training and 149 for testing)	Baccetti and Franchi	Manually labeled image data set with 54 features and results of the evaluation by a clinical decision support system	Expert visual evaluation	Percentage agreement between the model and the visual analysis of the researcher was 86.93%, which was the highest among all models tested.
5	Kök et al., 2020[30]	Turkey	8–17	419 individuals	Hassel and Farman	Measurements used in different combinations for seven neural networks	Human observer's classification	Highest classification accuracy was obtained from the model that used all 32 measurements and age as inputs. The overall accuracy was 94.2% for this model on the test data set.
6	Kök et al., 2020[31]	Turkey	8–17	360 individuals	—	Measurements on the second, third, fourth, and fifth vertebrae used in different combinations as inputs for four models	Human observer's classification	Highest accuracy obtained with one of the neural networks was 0.95 when the training and test data were split into a ratio of 70%:30%.
7	Kök et al., 2019[32]	Turkey	8–17	300 individuals	Hassel and Farman	Linear measurements performed on second, third, and fourth cervical vertebrae	Orthodontist	The neural network model had the second highest accuracy values for determining individual stages, except the fifth stage second-highest accuracy values (93%, 89.7%, 68.8%, 55.6%, 47.4%, and 78%) but was the most stable among all algorithms tested.

Table 2. Continued

S.No	Author, Year	Country	Age Range, y	Sample Size	CVM Method Used	Inputs	Reference Standards/Comparisons	Outcome
8	Makaremi et al., 2019[33]	France	Not mentioned	1870 cephalograms	Baccetti and Franchi	Cropped images without filters and cropped images processed with mean, median, and entropy filter	Human observer	The pretrained models were not as effective as the neural network made by the researchers. The accuracy of the neural network did not exceed 90% on test images. The accuracy improved with more images and preprocessing with the entropy filter.

[a] ANN indicates artificial neural network; CNN, convolutional neural network.

between the NN and the reference standard. The studies that developed new models showed an overall accuracy of greater than 90% for most models, but when a previously developed model was compared with human observers, the agreement ranged from 50% to 62%.

Studies that compared NNs to other forms of AI such as k-nearest neighbor, Naïve Bayes, support vector machine, logistic regression, random forest, and decision trees, used for the same task, found NNs were more accurate.[28,32] The NN used by Kök et al.[32] displayed the most stable results compared with other systems.

Qualitative Assessment

The studies were assessed according to the QUADAS-2 tool (Table 3). The risk of bias and the applicability of both tests and the patients were examined. The flow and timing of the tests was also examined. All studies presented low concerns regarding the applicability of the patient data sets and the tests.

The risk of bias presented some concerns in all studies. Studies with no separate test data set to evaluate the accuracy of the NN classification were considered as having some concerns for bias in the index test. In case of the reference standards, studies that had poor interobserver or intraobserver agreement for the data or did not evaluate this parameter were considered at risk of bias in reference standards. The risk of bias according to each domain is presented in Figure 2.

DISCUSSION

This review examined the accuracy of NNs in classifying lateral cephalograms according to the stages of cervical vertebrae maturation. In the eight included articles, NN models were either adapted from previously trained models or devised specifically for the task of classifying CVM. They showed high accuracy during development, being more than 90% accurate at classifying the stages. The agreement between individual human observers and the NN's classification ranged from 50% to 62% in one study.[29]

The data used for development of the models was obtained ethically. All studies used cross-sectional data to identify the CVM stage. It was classified into training and testing sets by some studies,[26,28,30,31,33] whereas others studies[27,32] used cross-validation to determine the system's performance. Validation helps prevent overfitting of the model to a data set and helps make a generalized model.[38] In fivefold cross-validation, used by most included studies, the data are split into five sets. One is used as a validation test set,

Table 3. Methodologic Assessment of Studies According to the QUADAS Tool[a]

	Risk of Bias				Applicability Concern		
Author, Year	Patient Selection	Index Test (Neural Network Classification)	Reference Test (Human Expert's Classification)	Flow and Timing	Patient Selection	Index Test (Neural Network Classification)	Reference Test (Human Expert's Classification)
Kim et al., 2021[26]	L	L	?	L	L	L	L
Seo et al., 2021[27]	L	?	?	L	L	L	L
Amasya et al., 2020[28]	L	L	L	L	L	L	L
Amasya et al., 2020[29]	L	L	?	L	L	L	L
Kök et al., 2020[30]	L	L	?	L	L	L	L
Kök et al., 2020[31]	L	L	?	L	L	L	L
Kök et al., 2019[32]	L	?	?	L	L	L	L
Makaremi et al., 2019[33]	L	L	?	L	L	L	L

[a] ? indicates unclear; L, low.

whereas the remaining four are used for training. The process is repeated by changing the training and validation sets during each iteration of data. It helps adjust the weights assigned to the inputs for the perceptrons to map an accurate output. Validation minimizes the errors in a system by repeatedly adjusting the weights for inputs until there is a minimum error in the output with no significant change in the weights.

The validation data set is seen by the NN before and is an unreliable assessment of the NN's generalization. Generalization of an NN is its ability to correctly provide an output for data that were not a part of its creation or training. Thus, unseen test data provide a better evaluation of its accuracy. It is believed that to provide generalizability, the number of training examples provided to the model should be more than the parameters.[11] This is also influenced by the size and type of training data.[39] The studies used test and training data from the same study center. A better test of the model's practical performance would be on data provided from a different setting. The models can be shared with researchers or data obtained from remote institutions to better establish their performance. It is also possible to compare the performance of different models on the same test data sets.

The inputs used by the studies were linear measurements, ratios, or images. The models developed with linear inputs and ratios performed better than those that used only photographic or linear inputs. This could be due to the increased noise in photographs, since training on preprocessed images helped improve the output. The inputs must accurately represent the environment in which the NN operates. Makaremi et al.[33] showed that NNs performed better when the data were distributed evenly across all CVM stages. Most included studies specified having an equal or near equal distribution in the training data. Cervical vertebrae maturation stage was classified using the description by Hassel and Farman and Baccetti et al.[22,23] Both methods were considered reliable in predicting the pubertal growth spurt.[40] Systems developed using both methods had similar accuracy in classifying lateral cephalograms. The method of Baccetti et al.[23] may be more applicable across different data sets, as it can be used regardless of a radiation collar.

The cervical vertebrae maturation methods have inherent drawbacks. Studies show that the method is reproducible, but there may be a poor correlation between the cervical stages and peak mandibular growth on longitudinal assessment.[41,42] Previous stud-

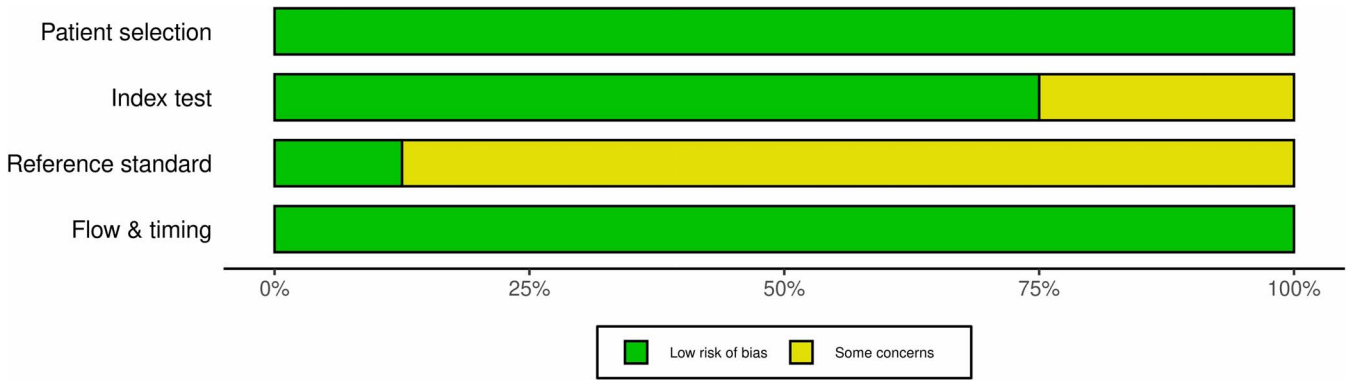

Figure 2. Risk of bias across different domains.

ies also raised concerns about the reproducibility of CVM stages.[43,44] Because the visual analysis of the researchers is considered as the ground truth for the model, it is imperative that this judgment is not biased by the analysis of a single expert. Using data with high inter- and intraobserver agreement could provide better training for NNs.

The ultimate aim of assessment of CVM stages is to determine skeletal maturity and multiple characteristics that can help identify it. Developing comprehensive models can help improve the estimation of the skeletal age. They can use additional inputs such as the chronological age, facial photographs, and secondary sexual characteristics, coupled with CVM indicators. Such comprehensive AI models may perform just as well or better than human practitioners.

CONCLUSIONS

- Neural networks can successfully classify the different stages of cervical vertebrae maturation from lateral cephalometric radiographs.
- The accuracy of the diagnosis probably varies due to different inputs used for developing the models and a lack of standardization of data with inter- and intraobserver agreement.
- Further studies can develop models considering these drawbacks.
- Generalization of the developed models can be tested using publicly available or cross-center data sets.

REFERENCES

1. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. *npj Digit Med*. 2019;2:1–7. doi:10.1038/s41746-019-0103-3
2. Hung M, Park J, Hon ES, et al. Artificial intelligence in dentistry: harnessing big data to predict oral cancer survival. *World J Clin Oncol*. 2020;11:918. doi:10.5306/wjco.v11.i11.918
3. García-Pola M, Pons-Fuster E, Suárez-Fernández C, Seoane-Romero J, Romero-Méndez A, López-Jornet P. Role of artificial intelligence in the early diagnosis of oral cancer: a scoping review. *Cancers*. 2021;13:4600. doi:10.3390/cancers13184600
4. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol*. 2019;213:506–513. doi:10.2214/AJR.19.21117
5. Tsui KL, Wong ZSY, Goldsman D, Edesess M. Tracking infectious disease spread for global pandemic containment. *IEEE Intelligent Systems*. 2013;28:60–64. doi:10.1109/MIS.2013.149
6. Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. *Comput Biomed Res*. 1975;8:303–320. doi:10.1016/0010-4809(75)90009-9
7. Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent*. 2019;91:103226. doi:10.1016/j.jdent.2019.103226
8. Dhillon H, Chaudhari PK, Dhingra K, et al. Current applications of artificial intelligence in cleft care: a scoping review. *Front Med*. 2021;8.:676490
9. Shan T, Tay FR, Gu L. Application of artificial intelligence in dentistry. *J Dent Res*. 2020;100:232–244. doi:10.1177/0022034520969115
10. Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry: a systematic review. *J Dent Sci*. 2021;16:508–522. doi:10.1016/j.jds.2020.06.019
11. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, Mass: MIT Press; 2016.
12. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J*. 2018;69:120–135. doi:10.1016/j.carj.2018.02.002
13. Jones CM, Buchlak QD, Oakden-Rayner L, et al. Chest radiographs and machine learning: past, present and future. *J Med Imaging Radiat Oncol*. 2021;65:538–544. doi:10.1111/1754-9485.13274
14. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018;15:e1002686. doi:10.1371/journal.pmed.1002686
15. Wu JT, Wong KCL, Gur Y, et al. Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents. *JAMA Netw Open*. 2020;3:e2022779. doi:10.1001/jamanetworkopen.2020.22779
16. Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health*. 2021;3:e496–e506. doi:10.1016/S2589-7500(21)00106-0
17. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol*. 1990;52:99–115.
18. Brickley MR, Shepherd JP, Armstrong RA. Neural networks: a new technique for development of decision support systems in dentistry. *J Dent*. 1998;26:305–309. doi:10.1016/s0300-5712(97)00027-4
19. Subramaniam P, Naidu P. Mandibular dimensional changes and skeletal maturity. *Contemp Clin Dent*. 2010;1:218–222. doi:10.4103/0976-237X.76387
20. Singh S, Singh M, Saini A, Misra V, Sharma VP, Singh GK. Timing of myofunctional appliance therapy. *J Clin Pediatr Dent*. 2010;35:233–240. doi:10.17796/jcpd.35.2.9572h13218806871
21. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae. *Am J Orthod*. 1975;67:458–459. doi:10.1016/0002-9416(75)90038-X
22. 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
23. 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:

316–323. doi:10.1043/0003-3219(2002)072<0316:AIVOTC>2.0.CO;2
24. Gray S, Bennani H, Farella M. Authors' response. *Am J Orthod Dentofacial Orthop*. 2016;150:7–8. doi:10.1016/j.ajodo.2016.04.013
25. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–536. doi:10.7326/0003-4819-155-8-201110180-00009
26. Kim EG, Oh IS, So JE, et al. Estimating cervical vertebral maturation with a lateral cephalogram using the convolutional neural network. *J Clin Med*. 2021;10:5400. doi:10.3390/jcm10225400
27. 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;10:3591. doi:10.3390/jcm10163591
28. 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. doi:10.1259/dmfr.20190441
29. 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:e173–e179. doi:10.1016/j.ajodo.2020.08.014
30. Kök H, Izgi MS, Acilar AM. Determination of growth and development periods in orthodontics with artificial neural network. *Orthod Craniofac Res*. 2021;4(suppl 2):76–83. doi:10.1111/ocr.12443
31. 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*. 2020;34:2–9. doi:10.5152/TurkJOrthod.2020.20059
32. 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;20:41. doi:10.1186/s40510-019-0295-8
33. 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*. 2019;21:1222. doi:10.3390/e21121222
34. Garza-Morales R, López-Irarragori F, Sanchez R. On the application of rough sets to skeletal maturation classification. *Artif Intell Rev*. 2016;45:489–508. doi:10.1007/s10462-015-9450-x
35. Xie L, Tang W, Izadikhah I, 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;38:378–388. doi:10.1007/s11282-021-00566-y
36. Kim DW, Kim J, Kim T, 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
37. 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
38. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop*. 2016;149:127–133. doi:10.1016/j.ajodo.2015.07.030
39. Arpit D, Jastrzębski S, Ballas N, et al. A closer look at memorization in deep networks. *arXiv:170605394 [cs, stat]*. Published July 1, 2017. Available at: http://arxiv.org/abs/1706.05394./ Accessed January 4, 2022.
40. Cericato GO, Bittencourt MAV, Paranhos LR. Validity of the assessment method of skeletal maturation by cervical vertebrae: a systematic review and meta-analysis. *Dentomaxillofac Radiol*. 2015;44:20140270. doi:10.1259/dmfr.20140270
41. Perinetti G, Primozic J, Sharma B, Cioffi I, Contardo L. Cervical vertebral maturation method and mandibular growth peak: a longitudinal study of diagnostic reliability. *Eur J Orthod*. 2018;40:666–672. doi:10.1093/ejo/cjy018
42. Cunha AC, Cevidanes LH, Sant'Anna EF, et al. Staging hand-wrist and cervical vertebrae images: a comparison of reproducibility. *Dentomaxillofac Radiol*. 2018;47:20170301. doi:10.1259/dmfr.20170301
43. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method: poor reproducibility. *Am J Orthod Dentofacial Orthop*. 2009;136:478.e1–e7. doi:10.1016/j.ajodo.2007.08.028
44. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Cervical vertebrae maturation method morphologic criteria: poor reproducibility. *Am J Orthod Dentofacial Orthop*. 2011;140:182–188. doi:10.1016/j.ajodo.2011.04.013