GMR

The Original

ARTIFICIAL INTELLIGENCE IN ORTHODONTIC ASSESSMENT: ACCURACY OF SEGMENTATION AND LANDMARK LOCALIZATION: A SYSTEMATIC REVIEW

[1*]Dr. Adheena Nambiar, [2]Dr. Hemamalini D, [3]Dr.Shahul Hameed Faizee, [4]Dr.Xavier Dhayananth

[1]Post graduate, Dept of Orthodontics and Dentofacial Orthopedics, Sathyabama dental college and general hospital, Chennai, Sathyabama University Email ID- adheenanambiar100@gmail.com, Orcid ID: 0009-0001-8711-8014
[2]Assistant professor, Dept of Orthodontics and Dentofacial Orthopedics, Sathyabama dental college and general hospital, Chennai, Sathyabama University, Email ID- drhemaortho@gmail.com
[3]Professor and Head of the Department, Dept of Orthodontics and Dentofacial Orthopedics, Sathyabama dental college and general hospital, Chennai , Sathyabama University, Email ID- sfaizee@hotmail.com
[4]Associate Professor, Dept of Orthodontics and Dentofacial Orthopedics, Sathyabama dental college and general hospital, Chennai, Sathyabama University Email ID- drxavy@gmail.com
*Corresponding author: Dr.Adheena Nambiar, Email id- adheenanambiar100@gmail.com, Orcid id: 0009-0001-8711-8014

ABSTRACT
Introduction
Artificial intelligence is a fast-developing technical field in which machines simulate the human intellect. it works on the principle of recognising the pattern of the data and then making the decision with small human assistance. The advanced techniques of AI, which are being used in orthodontics more often. Due to the gap in the literature, the generalisability of the findings is limited. Thus, these factors emphasise the need for a comprehensive review of the use of AI in orthodontic assessment.
AIM: The aim of the current review is to evaluate the analytical performance and reliability of AI-based tools in orthodontic patients compared with traditional diagnostic techniques.
Material and methods: In the current review, a thorough search was conducted in the following databases: PubMed (MEDLINE), Cochrane CENTRAL, Embase, ProQuest, and OVID from inception till 30th August 2024. The MeSH keywords such as AI, machine learning, deep learning, neural networks, convolutional neural networks, automated diagnosis, computer-aided diagnosis, orthodontics, cephalometric analysis, malocclusion, automated landmark detection, dental imaging, diagnostic imaging, diagnosis, analysis, CBCT, and radiographic analysis were used in combination. QUADAS II tool was used to assess the accuracy of the methodology of the included studies.
Results: The records identified were 91. Only 7 studies were included after thorough screening and removal of duplicates.
Conclusion: AI-based tools have been proven to be promising adjuncts for improving efficiency and consistency, as these tools have high accuracy in orthodontic diagnostic tasks such as landmark detection and CBCT analysis.

KEYWORDS: Artificial intelligence; Machine learning; Deep learning; Orthodontic assessment; Segmentation accuracy; LANDMARK localisation.

INTRODUCTION
In recent years, the incorporation of AI into healthcare has transformed several domains, including radiology, oncology, pathology, and diagnostic imaging, by improving accuracy, efficiency, and consistency in clinical decision-making [1]. Artificial intelligence (AI) is an expeditiously developing field of technology that allows machines to imitate human intellect by identifying intricate patterns, learning from data, and making well-informed judgments with minimal input from humans. [2] Two major domains of AI are machine learning and expert systems. The goal of machine learning is to "learn" from training data to enhance its capabilities. The biological neural system of the human brain serves as an inspiration for artificial neural networks (ANNs), a subdomain of machine learning. For large-scale data analysis of complex relationships, ANNs have been commonly used [3]. In computer vision tasks like classification and segmentation, which are utilized in orthodontic evaluation, ANNs with several hidden layers—often referred to as deep learning—have shown remarkable performance. Deep learning is gaining popularity because of its advanced model training techniques, increased computational performance, and great feasibility.
One of the most widely used methods, such as deep learning methods (DL), convolutional neural networks (CNNs), performs very well while analyzing high-resolution images. The intricacy of the images can be reduced by the convolution process, which makes CNNs efficient for multiple functions such as identifying objects, shapes, and patterns. These sub-domains of AI have been used vigorously in orthodontics. These systems can automate key diagnostic tasks such as cephalometric landmark detection, growth assessment, dental arch evaluation, and three-dimensional (3D) image segmentation from CBCT (Cone-Beam Computed Tomography) scans [4].

In dentistry, and more specifically orthodontics, AI has shown potential to revolutionise diagnostic and treatment-planning protocols. Conventional methods of orthodontic diagnosis rely on cephalometric tracing, plaster model analysis, clinical examination, and interpretation of two-dimensional radiographs [5]. Although they are widely practised, they are more time-consuming and often subject to intra- and inter-examiner variability, leading to inconsistent diagnoses for even well-trained orthodontists and potentially suboptimal treatment outcomes, specifically in the cephalometric analysis or landmark identification. Cephalometric analysis is prone to two types of mistakes: identification errors resulting from inaccurate landmark, tracing, and measurement identification, and projection errors resulting from projected X-ray images derived from 3D objects [6]. In such cases, Convolutional neural networks (CNN) and deep learning have demonstrated unexpected results in computer vision applications, which might be used for medical imaging classification, detection, and semantic segmentation. In some studies, CNNs have been used to detect anatomical landmarks with accuracy, while deep learning algorithms have demonstrated the ability to interpret complex imaging data and predict treatment outcomes based on historical datasets [6-9].

Several studies have evaluated these applications. A study conducted by Alqahtani et al. (2023) on the use of CNN-based models in tooth segmentation and classification, the result showed excellent segment performance with a high intersection of Union i.e 0.99, with a high recall rate and precision, i.e 99 %, respectively. [10]. Similarly, Wang et al. (2024) reported high accuracy in mandibular landmark detection using a PointRend deep learning mechanism and automated segmentation in automatic cephalometric analysis [11]. The reviews by Kazimierczak et al. (2024) and Hendrickx et al. (2024) also corroborated the analytical performance of AI tools, especially in cephalometric analysis [12,13]. Therefore, most of the existing literature has various pitfalls, such as study design, smaller sample size, and varying levels of validation, which cause the generalizability of the findings to be very restricted [14,15-17]. This gap in the literature emphasises the need for a comprehensive review that includes the recent evidence and the studies highlighting the clinical applicability incorporates the most recent evidence and includes studies emphasising clinical applicability, which can be deduced by interpreting the technical performance, such as segmentation accuracy, landmark localisation error in relation to the diagnosis.

Though there are various promising features of AI-based tools, there are also various challenges of concern. The quality depends on the large data for training purposes, the requirement of a diverse population/data for validation, and privacy concerns, over dependency on an automated system without clinician oversight [18,19]. AI performance remains very restricted in the cases of clinical judgment, treatment planning, or extraction decisions. [20,21].

In orthodontic diagnosis, artificial intelligence is used widely, but the existing literature shows the major focus of the AI-based tools on tasks like landmark detection. The systematic analysis of AI analytical performance and its clinical use in comparison to conventional methods has been limited. Hence, the aim of the current review is to evaluate the diagnostic precision and dependability of AI-based tools in orthodontic patients in comparison to traditional diagnostic techniques.

MATERIALS AND METHODS

PRISMA 2020 (Preferred Reporting Items for Systematic Review and Meta-Analyses) guidelines were used to conduct the systematic review [22]. The PROSPERO registration number of the present review was **CRD42024547157**.

Review Question

How accurate are AI-based tools for the analysis of the orthodontic image, and how useful it is clinically?

Information Sources and Search Strategy

A thorough search was conducted in the following databases: PubMed (MEDLINE), Cochrane CENTRAL, Embase, ProQuest, and OVID from inception till 30th August 2024. The MeSH keywords such as AI, machine learning, deep learning, neural networks, convolutional neural networks, automated diagnosis, computer-aided diagnosis, orthodontics, cephalometric analysis, malocclusion, automated landmark detection, dental imaging, diagnostic imaging, diagnosis, analysis, CBCT, and radiographic analysis were used in combination. [Table 1/Fig 1]:

Database	**Search Strategy**
PubMed	(“Artificial Intelligence” OR “Machine Learning” OR “Deep Learning” OR “Neural Networks” OR “Convolutional Neural Network” OR “Automated Diagnosis” OR “Computer-Aided Diagnosis” OR “Deep Neural Network”) AND (“Orthodontics” OR “Cephalometric Analysis” OR “Malocclusion” OR “Automated Landmark Detection” OR “Dental Imaging”) AND (“Diagnosis” OR “Image Analysis” OR “CBCT” OR “Radiographic Analysis”)
Cochrane CENTRAL	(“Artificial Intelligence” OR “Machine Learning” OR “Deep Learning” OR “Neural Networks” OR “Convolutional Neural Network” OR “Automated Diagnosis” OR “Deep Neural Network” OR “Computer-Aided Diagnosis”) AND (“Orthodontics” OR “Cephalometric Analysis” OR “Malocclusion” OR “Dental Imaging”) AND (“Diagnosis” OR “Image Analysis” OR “CBCT” OR “Radiographic Analysis”)
Embase (via Elsevier)	(‘artificial intelligence’/exp OR ‘machine learning’/exp OR ‘deep learning’/exp OR ‘neural network’/exp OR ‘automated diagnosis’ OR ‘computer-aided diagnosis’) AND (‘orthodontics’/exp OR ‘cephalometric analysis’ OR ‘malocclusion’/exp OR ‘dental imaging’/exp) AND (‘diagnosis’/exp OR ‘image analysis’ OR ‘CBCT’ OR ‘radiographic analysis’)
ProQuest	(“Artificial Intelligence” OR “Machine Learning” OR “Deep Learning” OR “Neural Networks” OR “Deep Neural Network” OR “Automated Diagnosis” OR “Computer-Aided Diagnosis”) AND

	(“Orthodontics” OR “Cephalometric Analysis” OR “Malocclusion” OR “Dental Imaging” OR “Computer Vision”) AND (“Diagnosis” OR “Image Analysis” OR “CBCT” OR “Radiographic Analysis”)
Ovid (MEDLINE)	(“Artificial Intelligence” OR exp Artificial Intelligence/ OR “Machine Learning” OR exp Machine Learning/ OR “Deep Learning” OR “Deep Neural Network” OR “Neural Networks” OR exp Neural Networks/ OR “Automated Diagnosis” OR “Computer-Aided Diagnosis”) AND (“Orthodontics” OR exp Orthodontics/ OR “Cephalometric Analysis” OR “Malocclusion” OR exp Malocclusion/ OR “Dental Imaging” OR exp Radiography, Dental/) AND (“Diagnosis” OR exp Diagnosis/ OR “Image Analysis” OR “CBCT” OR “Radiographic Analysis”)

Table 1/Fig 1: A sample search strategy used in this review

Eligibility Criteria (PICOS Framework)
The PICOS framework is:

- **Population**: orthodontic patients diagnosed by using radiographs, cephalograms, or clinical records.
- **Intervention**: Machine learning, Convolution neural networks, and deep learning (AI-based tools).
- **Comparison**: Radiographs, CBCT, expert orthodontist assessment, and clinician consensus.
- **Outcomes**: landmark localisation error, mean error, reliability, and its clinical applicability.
- **Study Design**: Clinical studies, validation studies, diagnostic accuracy studies, and observational cross-sectional studies. Although randomised controlled trials (RCTs) were considered, none met the inclusion criteria.

As per the PICOS framework, this review included orthodontic patients (Population), examined AI-based diagnostic tools (Intervention), compared them with conventional methods (Comparison), and evaluated analytical performance and reliability (Outcomes) based on clinical trials and observational studies (Study Design). Therefore, the scope of articles is limited to this criterion and will not include non-comparing articles. This improves the rigour of the review, despite the possibility of omitting certain evidence.

Inclusion Criteria:

- Original studies evaluating AI-based tools for orthodontic diagnosis
- Studies comparing AI methods with traditional/manual diagnostic approaches
- Studies reporting quantitative diagnostic accuracy metrics
- Only English-language studies were included because translation resources were unavailable, and the majority of orthodontic AI literature is published in English.

Exclusion Criteria:

- Review articles, conference abstracts, editorials, or commentaries
- In vitro, animal, or simulation-only studies
- Studies not involving orthodontic diagnostic tasks

Study Selection and Analysis
For all the retrieved records, two calibrated reviewers screened the study titles and abstracts independently. These articles were then reviewed for inclusion based on the eligibility criteria. Disagreements have been resolved by conferring with a third reviewer. 91 records were identified. 82 records remained after the removal of 9 duplicates. Following title screening, 43 records were excluded. The abstracts of the remaining 39 studies were assessed, of which 26 were excluded. Full texts of 13 studies were reviewed, and 6 studies were excluded. Finally, 7 studies were included in the present review. [Table2/Fig 2]

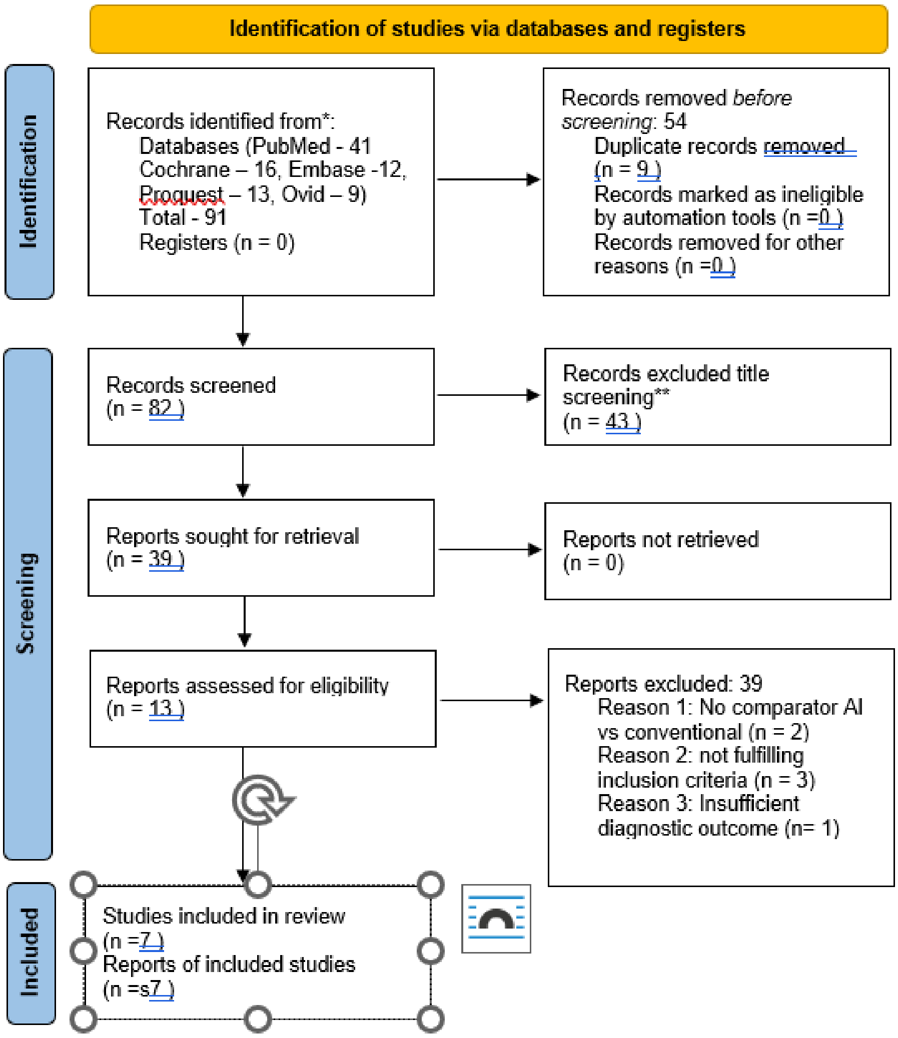

Table 2/Fig 2 :PRISMA flow diagram

The complete search strategy was adapted for each database. Manual screening of references and citation tracking were also performed to identify additional eligible studies. .

Data Extraction

The studies obtained were deduplicated using Zotero reference management software. Data from each included study were independently extracted using a standardised form. The data were extracted independently by two reviewers. The following details were recorded: A standardised method was used for data extraction to ensure standardisation among all included studies. Study attributes like author, year of publication, country, and study design had been among those extracted. Additionally, information about population characteristics and sample size was collected. The type of artificial intelligence model and its architecture, such as U-Net models or convolutional neural networks, were mentioned. Each study's specific diagnostic task, such as CBCT segmentation or cephalometric landmark identification, was recorded. The comparators used are usually the conventional standards. The data extracted for the outcome were error rates, reliability metrics, and mean deviation. The main findings were also outlined.
The reviewers would resolve the disagreement through discussion. The extracted data was tabulated and explained.

Risk of Bias Assessment

The **Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2)** tool was used to assess the methodological quality. [21]. Though QUADAS-2 cannot assess the quality of the methodology based on AI tools accurately. Key domains assessed were patient selection, Index test, reference standard, and flow and timing.
The QUADAS-2 framework was used to classify the risk of bias as low, high, or unclear.
The bias assessment was done by two reviewers, independently. (Table 4/Fig 4)
GRADE certainty of evidence was assessed and tabulated (Table 5/Fig 5)

Data Synthesis

A meta-analysis was not feasible because of the disparity in AI models, imaging modalities, evaluation, reporting, and diagnostic tasks. Landmark identification and anatomical segmentation findings and their reliability were summarised from all the studies. The findings and their interpretation were presented in a tabulated form. AI-based tools need to be validated for their performance in comparison to the conventional orthodontic assessment. The validation studies present critical evidence of the AI model accuracy, which is a must requirement for assessing its clinical use.

RESULTS

Study Selection

Total number of 7 studies fulfilled the inclusion criteria.

Study Characteristics

The studies included were seven, which were published between 2021 and 2024. The type of studies included were observational diagnostic accuracy or validation studies. The sample sizes of the patients or images were 30 to 1785. Almost six studies used CBCT with AI tools for diagnosing anatomical segmentation or landmark identification. [10,11,23-26]. In one study, a CT scan was used as input [27].

The AI models employed included convolutional neural networks (CNNs), U-Net architectures, Point Rend mechanisms, and deep neural patchworks. Reference standards for comparison included manual tracings by orthodontic experts or peer-reviewed diagnosis panels [19]. Characteristics of the study are summarised in Table 3/Fig 3.

Author	Population	Intervention	Control	Outcome (Accuracy in)	Study Design
Ayidh Alqahtani et al. (2023)[10]	CBCT scans of 1780 teeth	Multiclass CNN-based tool	CBCT (manual labelling)	99% (segmentation accuracy))	Analytical performance study
Preda et al., (2022)[23]	144 CBCT scans	3D U-Net (CNN) model	CBCT (manual segmentation)	92.60% (Anatomical Segmentation)	Analytical performanceStudy
Shimizu et al., (2022)[24]	967 CBCT scans	AI systems 1 & 2	Peer team diagnosis and planning	Moderate (Diagnosis); Low (Treatment Planning)	Validation Study
Tanikawa et al., (2021)[25]	1785 CBCT scans	AI system for cephalometric recognition	Actual landmarks	85–91% (Landmark Identification)	Cross-sectional Validation Study
Tao et al., (2023)[26]	70 CBCT scans	AI system	Actual landmarks	74–83% (Landmark Identification)	Landmark localization accuracy Study
Wang et al., (2024)[11]	410 CBCT scans	PointRend deep learning mechanism	Human identification	98.00% (Landmark Identification)	Landmark localization Accuracy Study
Weingart et al. (2023)[27]	CT scans of 30 adults	Deep neural patchworks	Human identification	1.94±1.45 mm; Human: 1.32±1.08 mm (Landmark Identification)	Comparative Accuracy Study

Table 3/Fig 3 : Characteristics of included studies and analytical performance outcomes

CBCT: Cone Beam Computed Tomography;
CNN: Convolutional Neural Network

Narrative Synthesis

A narrative synthesis of the included studies demonstrated that artificial intelligence-based CNN deep learning models exhibit high accuracy in image segmentation, with values ranging from 92.6% to 99.0 %, whereas landmark identification showed extensive variability [10,23]. Wang et al (2024) showed an accuracy of 98%, whereas other studies, such as Tao et al., showed lower accuracy (74-83%)[11,26]. Weingart conducted a comparative accuracy study that highlighted the variability present across different anatomical landmarks [27]. A study by Shimizu et al showed that diagnostic utility is moderate with lower consistency in treatment planning in comparison to clinical experts [24]. Overall, higher accuracy was noticed in CNN-based anatomical segmentation than in landmark identification.

Risk of Bias

Risk of bias was assessed using the Cochrane **QUADAS-2** tool. Low to moderate risk was present in all the studies, such as patient selection and index test conduct [10,11,23,27]. However, concerns were noted regarding the lack of blinding and the inconsistent reporting of reference standards in two studies [26,27]. The overall risk of bias across studies is visually summarised in Table 4/Fig 4.

Overview of Methodological Quality

Considering all studies included, the studies have demonstrated generally good methodological quality. Low risk was present in four studies. Strengths of these studies were appropriate patient selection protocol, suitable reference standards, and acceptable flow and timing. Methodological limitations were seen in two studies, which were rated as having a moderate overall risk of bias. These concerns were about unclear reporting of patient selection, index test interpretation, and the flow and timing of procedures.

Study	**Patient Selection**	**Index Test**	**Reference Standard**	**Flow & Timing**	**Overall Risk**
Alqahtani et al. (2023)	Low	Low	Low	Low	Low
Preda et al. (2022)	Low	Low	Low	Low	Low
Shimizu et al. (2022)	Low	Unclear	Low	Unclear	Moderate
Tanikawa et al. (2021)	Low	Low	Low	Low	Low
Tao et al. (2023)	Unclear	Low	Low	Unclear	Moderate
Wang et al. (2024)	Low	Low	Low	Low	Low
Weingart et al. (2023)	Unclear	Unclear	Low	Unclear	Moderate

Table 4/Fig 4: Risk of bias assessment of included studies using QUADAS-2 tool

Overall, studies showed a low to moderate risk of bias. But some studies had unclear sampling techniques, and the patient selection scope was linked to low risk due to a lack of blinding and clarity regarding inclusion and timing between the index and reference tests. Publication bias was not assessed as the number of included studies was small.

Domain	**Assessment**	**Judgment**	**Explanation**
Risk of Bias	QUADAS-2	Not serious	Four studies had low risk of bias across all domains, while two had moderate risk due to unclear patient selection, index test, and flow/timing. No study showed high risk of bias.
Inconsistency	Accuracy outcomes	Serious	Considerable variation was observed in performance metrics (74–99% accuracy), reflecting heterogeneity in AI architectures, sample sizes, and diagnostic tasks.
Indirectness	Population and outcomes	Not serious	Studies directly evaluated AI-based analysis of CBCT/CT images for orthodontic and craniofacial diagnostic applications relevant to the review question.
Imprecision	Sample size and estimates	Not serious	Most studies included moderate-to-large datasets (70–1785 scans/teeth), providing reasonably precise estimates of performance.
Publication Bias	Suspected	Undetected	The small number of studies limits formal assessment; however, no clear evidence of publication bias was identified.
Large Effect	Upgrade	Yes (+1)	Several studies demonstrated very high analytical performance accuracy (>95%), suggesting a substantial effect compared with conventional assessment methods.

Table 5/Fig 5: GRADE Evidence assessment

Overall GRADE Certainty of Evidence was judged as moderate, as seen in Table 5.

DISCUSSION

This systematic review assessed the analytical performance and clinical applicability, which is inferred through the technical performance of artificial intelligence (AI)-based tools in orthodontic diagnosis. The findings suggest that AI, particularly deep learning models or convolutional neural networks (CNNs), demonstrates high accuracy in objective diagnostic tasks like tooth segmentation, cephalometric landmark detection, and CBCT interpretation. However, its performance declines in more complex and subjective tasks, such as treatment planning and clinical decision-making [10,23,24,27]. According to the studies included in the current review, CNN-based tools exhibit 90% accuracy in the anatomical segmentation, whilst in comparison with clinicians, landmark detection showed more variability and precision gaps. [10,11,23,25-27].

The reason behind these findings may be a smaller sample size and AL algorithms with less optimization. This may be due to factors such as a smaller dataset size and less optimised AI algorithms. This significant variation in accuracy highlights that a model design, dataset size, and validation methods influence the AI-based tools.

However, depending on the data and task, the effectiveness of AI tools differed widely. In diagnostic listing, AI can help to a significant extent, but AI fails miserably in treatment planning. AI can help in interpreting the clinical applicability in the cases of anatomic segmentation and landmark detection. The important limitation underneath: AI can do object-based image analysis but is unable to make a clinical judgement, and experience [24].

In comparison with previous reviews in the literature, the findings of the present review are consistent with the findings of Kazimierczak et al. and Hendrickx et al., who suggested that AI performs remarkably in landmark detection but highlighted the necessity of better external validation and generalizability.[12,13]. The growing use of AI in orthodontics

and treatment planning is also acknowledged by the other reviews [15,16]. A wide variance (1-4mm) in landmark localisation due to variations in imaging modalities, algorithm tweaking, and training datasets
was suggested by Londono et al. [20]. A similar disparity was found in the present review, which suggests that orthodontic AI research continues to require standardized datasets and procedures [17]. The difference in the results of these studies in comparison to the present review is because of the narrower eligibility criteria focusing on peer-reviewed, updated evidence, deep learning–based diagnostic accuracy, and comparative studies.
The present review consists of various limitations that need to be acknowledged. While the review strictly followed the review protocol registered in PROSPERO, the small size and heterogeneous nature of sample sizes found in the included studies may affect the statistical robustness of the outcomes. However, this limitation is the outcome of a well-adhered protocol. Secondly, the variation of comparator methods does introduce clinical and methodological heterogeneity, which can have an impact on the pooled interpretation of diagnostic performance. Including studies published in the English language might create language bias and limit the comprehensiveness of evidence retrieval. Other limitations are the dependence on expert orthodontist assessment and clinician consensus as reference standards, which creates potential variability that will affect the reliability of the accuracy of AI results. Most of the included studies have an absence of external validation in research and restricted generalizability of findings beyond the specific datasets used. Despite the review question focused on the clinical utility, few studies directly assessed patient-centered outcomes or real-world integration; this was mostly deduced from technical results such as segmentation accuracy and landmark localisation inaccuracy.
Nonetheless, this has resulted in suggestions that standard protocols be used in future research.
Due to the variation in the studies, quantitative synthesis could not be performed, and diagnostic performance metrics (ROC curves, area under the curve (AUC), and predictive values) were not published for these articles. Variations were also seen in the type of AI models used. This aspect made it difficult to make a direct comparison in the studies, as there is a lack of validation procedures and standardised measures.

Future directions

AI has several advantages: rapid analysis, reproducibility, and less human error in standardised diagnostic tasks. However, the use of AI tools in clinical workflows must be done carefully. Unresolved issues include data privacy, algorithm disclosure, and excessive dependence on AI predictions [18,19]. Additionally, a large number of the included studies had low methodology, a considerable risk of bias, and inadequate data on reference standards, calibration, and blinding.
The significant potential of AI-based tools has high analytical performance and segmentation accuracy in several studies, and is to be included in the dental and craniofacial diagnostic arena. However, the studies included were retrospective; therefore, the ability to evaluate the accuracy of AI systems in routine cases, when the variation in variability in patient demographics, imaging techniques, and clinical decision-making may affect the outcomes. Also, the limited generalisability of results, and a larger number of prospective studies should be conducted to confirm the clinical utility. However, these AI-based tools have a lot of potential to reduce the burden of clinicians while ensuring consistency and reducing the probability of diagnostic errors. Due to these limitations, further research is required, specifically in treatment planning and landmark identification in complex cases. Larger multicentric studies with standardized validation procedures and expert clinician comparisons are thus required in the future.

CONCLUSION

AI-based tools show high accuracy in image segmentation and landmark detection tasks and CBCT analysis, making them a useful adjunct for increasing efficiency and consistency. Their role in intricate tasks requiring clinical judgment, such as treatment planning, remains limited or restricted. Further research is required for the validation of AI-based tools. Before standard clinical integration, further validation through standardized, real-world studies is needed.
From the viewpoint of clinical research, the present review has highlighted the limitations of the present AI models and the importance of standardising methods in further investigations.

Limitations of Included Studies

- Direct comparisons were impeded in AI model designs, datasets, and performance indicators.
- Included studies might show bias as QUADAS 2 does not fully address AI-specific bias.
- Many studies lacked long-term validation or real-world clinical implementation.

Acknowledgments:

The authors would like to acknowledge all researchers whose studies were included in this systematic review.

Conflict of Interest:

No conflict of interest

Funding:

No external funding was received.

REFERENCES

1. Vodanović M, Subašić M, Milošević D, Savić Pavičin I. Artificial intelligence in medicine and dentistry. *Acta Stomatol Croat.* 2023;57(1):70–84

2. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: Transforming the practice of medicine. *Future Healthc J.* 2021;8(2): e188–94.
3. Soori M, Arezoo B, Dastres R. Artificial intelligence, machine learning and deep learning in advanced robotics: A review. *Cogn Robot.* 2023; 3:54–70.
4. Baxi S, Shadani K, Kesri R, Ukey A, Joshi C, Hardiya H. Recent advanced diagnostic aids in orthodontics. *Cureus.* 2022;14(11): e31921.
5. Durão AP, Morosolli A, Pittayapat P, Bolstad N, Ferreira AP, Jacobs R. Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: A comparative study. *Imaging Sci Dent.* 2015;45(4):213–20.
6. Choi YJ, Lee KJ. Possibilities of artificial intelligence use in orthodontic diagnosis and treatment planning: Image recognition and three-dimensional VTO. *Semin Orthod.* 2021;27(2):121–9.
7. Uğurlu M. Performance of a convolutional neural network-based artificial intelligence algorithm for automatic cephalometric landmark detection. *Turk J Orthod.* 2022;35(2):94–100.
8. Lahoud P, Jacobs R, Boisse P, EzEldeen M, Ducret M, Richert R. Precision medicine using patient-specific modelling: State of the art and perspectives in dental practice. *Clin Oral Investig.* 2022;26(8):5117–28.
9. Alam MK, Alanazi DSA, Alruwaili SRF, Alderaan RAI. Assessment of AI models in predicting treatment outcomes in orthodontics. *J Pharm Bioallied Sci.* 2024;16(Suppl 1): S540–2.
10. Alqahtani KA, Jacobs R, Smoulders A, Van Gerven A, Willems H, Shujaat S, et al. Deep convolutional neural network-based automated segmentation and classification of teeth with orthodontic brackets on cone-beam computed tomographic images: A validation study. *Eur J Orthod.* 2023;45(2):169–74.
11. Wang Y, Wu W, Christelle M, Sun M, Wen Z, Lin Y, et al. Automated localisation of mandibular landmarks in the construction of the mandibular median sagittal plane. *Eur J Med Res.* 2024;29(1):84.
12. Kazimierczak N, Kazimierczak W, Serafin Z, Nowicki P, Nożewski J, Janiszewska-Olszowska J. AI in orthodontics: Revolutionising diagnostics and treatment planning—A comprehensive review. *J Clin Med.* 2024;13(2):344.
13. Hendrickx J, Gracea RS, Vanheers M, Winderickx N, Preda F, Shujaat S, et al. Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. *Eur J Orthod.* 2024;46(4): cjae029.
14. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, et al. Deep learning for cephalometric landmark detection: Systematic review and meta-analysis. *Clin Oral Investig.* 2021;25(7):4299–309.
15. Dipalma G, Inchingolo AD, Inchingolo AM, Piras F, Carpentiere V, Garofoli G, et al. Artificial intelligence and its clinical applications in orthodontics: A systematic review. *Diagnostics (Basel).* 2023;13(24):3677.
16. Junaid N, Khan N, Ahmed N, Abbasi MS, Das G, Maqsood A, et al. Development, application, and performance of artificial intelligence in cephalometric landmark identification and diagnosis: A systematic review. *Healthcare (Basel).* 2022;10(12):2454.
17. Monill-González A, Rovira-Calatayud L, Oliviera NG, Ustrell-Torrent JM. Artificial intelligence in orthodontics: Where are we now? A scoping review. *Orthod Craniofac Res.* 2021;24(Suppl 2):6–15.
18. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical considerations in artificial intelligence in healthcare: Who takes responsibility? *Front Surg.* 2022; 9:862322.
19. Thirumoorthy S, Gopal S. Diagnostic accuracy of AI in orthodontic extraction decisions: Are we ready to let Mr Data run our enterprise? *Evid Based Dent.* 2023;24(1):17–8.
20. Londono J, Ghasemi S, Hussain Shah A, Fahimipour A, Ghadimi N, Hashemi S, et al. Evaluation of deep learning and convolutional neural network algorithms' accuracy for detecting anatomical landmarks on 2D lateral cephalometric images: A systematic review and meta-analysis. *Saudi Dent J.* 2023;35(5):487–97.
21. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529–36.
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ.* 2021;372: n71.
23. Preda F, Morgan N, Van Gerven A, Nogueira-Reis F, Smoulders A, Wang X, et al. Deep convolutional neural network-based automated segmentation of the maxillofacial complex from cone-beam computed tomography: A validation study. *J Dent.* 2022; 124:104238.
24. Shimizu Y, Tanikawa C, Kajiwara T, Nagahara H, Yamashiro T. The validation of orthodontic artificial intelligence systems that perform orthodontic diagnoses and treatment planning. *Eur J Orthod.* 2022;44(4):436–44.
25. Tanikawa C, Lee C, Lim J, Oka A, Yamashiro T. Clinical applicability of automated cephalometric landmark identification: Part I—patient-related identification errors. *Orthod Craniofac Res.* 2021; 24:43–52.
26. Tao T, Zou K, Jiang R, He K, He X, Zhang M, et al. Artificial intelligence-assisted determination of available sites for palatal orthodontic mini implants based on palatal thickness through CBCT. *Orthod Craniofac Res.* 2023;26(3):491–9.
27. Weingart JV, Schlager S, Metzger MC, Brandenburg LS, Hein A, Schmelzeizen R, et al. Automated detection of cephalometric landmarks using deep neural patchworks. *Dentomaxillofac Radiol.* 2023;52(6):20230059.