The Saudi Dental Journal (2023) 35, 487–497

King Saud University

The Saudi Dental Journal

www.ksu.edu.sa
www.sciencedirect.com

REVIEW ARTICLE

Evaluation of deep learning and convolutional neural network algorithms accuracy for detecting and predicting anatomical landmarks on 2D lateral cephalometric images: A systematic review and *meta*-analysis

Jimmy Londono [a], Shohreh Ghasemi [b], Altaf Hussain Shah [c], Amir Fahimipour [d], Niloofar Ghadimi [e], Sara Hashemi [f], Zohaib Khurshid [g,h], Mahmood Dashti [i,*]

[a] *FACP, Professor and Director of the Prosthodontics Residency Program and the Ronald Goldstein Center for Esthetics and Implant Dentistry, Dental College of Georgia at Augusta University, Augusta, GA, United States*
[b] *Department of Oral and Maxillofacial Surgery, The Dental College of Georgia at Augusta University, Augusta, GA, United States*
[c] *Special Care Dentistry Clinics, University Dental Hospital, King Saud University Medical City, Riyadh, Saudi Arabia*
[d] *School of Dentistry, Faculty of Medicine and Health, Westmead Centre for Oral Health, The University of Sydney, NSW 2145, Australia*
[e] *Department of Oral and Maxillofacial Radiology, Dental School, Islamic Azad University of Medical Sciences, Tehran, Iran*
[f] *School of Dentistry, Isfahan University of Medical Sciences, Isfahan, Iran*
[g] *Department of Prosthodontics and Dental Implantology, King Faisal University, Al-Ahsa 31982, Saudi Arabia*
[h] *Center of Excellence for Regenerative Dentistry, Department of Anatomy, Faculty of Dentistry, Chulalongkorn University, Bangkok 10330, Thailand*
[i] *School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

Received 16 March 2023; revised 15 May 2023; accepted 17 May 2023
Available online 23 May 2023

KEYWORDS

Machine learning;
Convolutional neural network;

Abstract *Introduction:* Cephalometry is the study of skull measurements for clinical evaluation, diagnosis, and surgical planning. Machine learning (ML) algorithms have been used to accurately identify cephalometric landmarks and detect irregularities related to orthodontics and dentistry. ML-based cephalometric imaging reduces errors, improves accuracy, and saves time.

* Corresponding author at: School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, District 1, Daneshjou Blvd, Tehran, Tehran Province, Iran.
E-mail address: dashti.mahmood72@gmail.com (M. Dashti).

Peer review under responsibility of King Saud University. Production and hosting by Elsevier

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.sdentj.2023.05.014

Artificial intelligence;
Lateral cephalometry;
Orthodontics;
Accuracy

Method: In this study, we conducted a *meta*-analysis and systematic review to evaluate the accuracy of ML software for detecting and predicting anatomical landmarks on two-dimensional (2D) lateral cephalometric images. The *meta*-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for selecting and screening research articles. The eligibility criteria were established based on the diagnostic accuracy and prediction of ML combined with 2D lateral cephalometric imagery. The search was conducted among English articles in five databases, and data were managed using Review Manager software (v. 5.0). Quality assessment was performed using the diagnostic accuracy studies (QUADAS-2) tool.

Result: Summary measurements included the mean departure from the 1–4-mm threshold or the percentage of landmarks identified within this threshold with a 95% confidence interval (CI). This *meta*-analysis included 21 of 577 articles initially collected on the accuracy of ML algorithms for detecting and predicting anatomical landmarks. The studies were conducted in various regions of the world, and 20 of the studies employed convolutional neural networks (CNNs) for detecting cephalometric landmarks. The pooled successful detection rates for the 1-mm, 2-mm, 2.5-mm, 3-mm, and 4-mm ranges were 65%, 81%, 86%, 91%, and 96%, respectively. Heterogeneity was determined using the random effect model.

Conclusion: In conclusion, ML has shown promise for landmark detection in 2D cephalometric imagery, although the accuracy has varied among studies and clinicians. Consequently, more research is required to determine its effectiveness and reliability in clinical settings.

Contents

1. Introduction

Utilizing oral radiology can be lucrative in various fields of dentistry, such as endodontics, periodontology, and orthodontics (Abdinian and Baninajarian 2017, Mehdizadeh et al., 2022). Cephalometry is the study of skull dimensions using linear and angular measurements of anatomical and constructed landmarks on standardized two-dimensional (2D) lateral head films. The linear and angular measurements from cephalometry can be used in facial recognition and forensic identification (Hlongwa 2019). However, cephalometry is used most frequently in orthodontics and oral surgery for the diagnosis of malocclusion and treatment planning. It is used in combination with facial form evaluation and model analysis to identify the location of skeletal and dental anomalies that can be improved with braces and/or surgery (Durão et al., 2013).

Currently, detecting irregularities related to orthodontics and dentistry has become possible owing to advancements in artificial intelligence (AI) (Pattanaik 2019). AI technology has been incorporated into cephalometry to resolve accurate diagnosis and surgical planning issues (Shin and Kim 2022). Cephalometry combined with AI may be able to assist practitioners with determination of bone age, extraction decisions, orthognathic surgical prediction, and temporomandibular bone segmentation (Mohammad-Rahimi et al., 2021, Mehdizadeh et al., 2022, Ebadian et al., 2023). Cephalometry

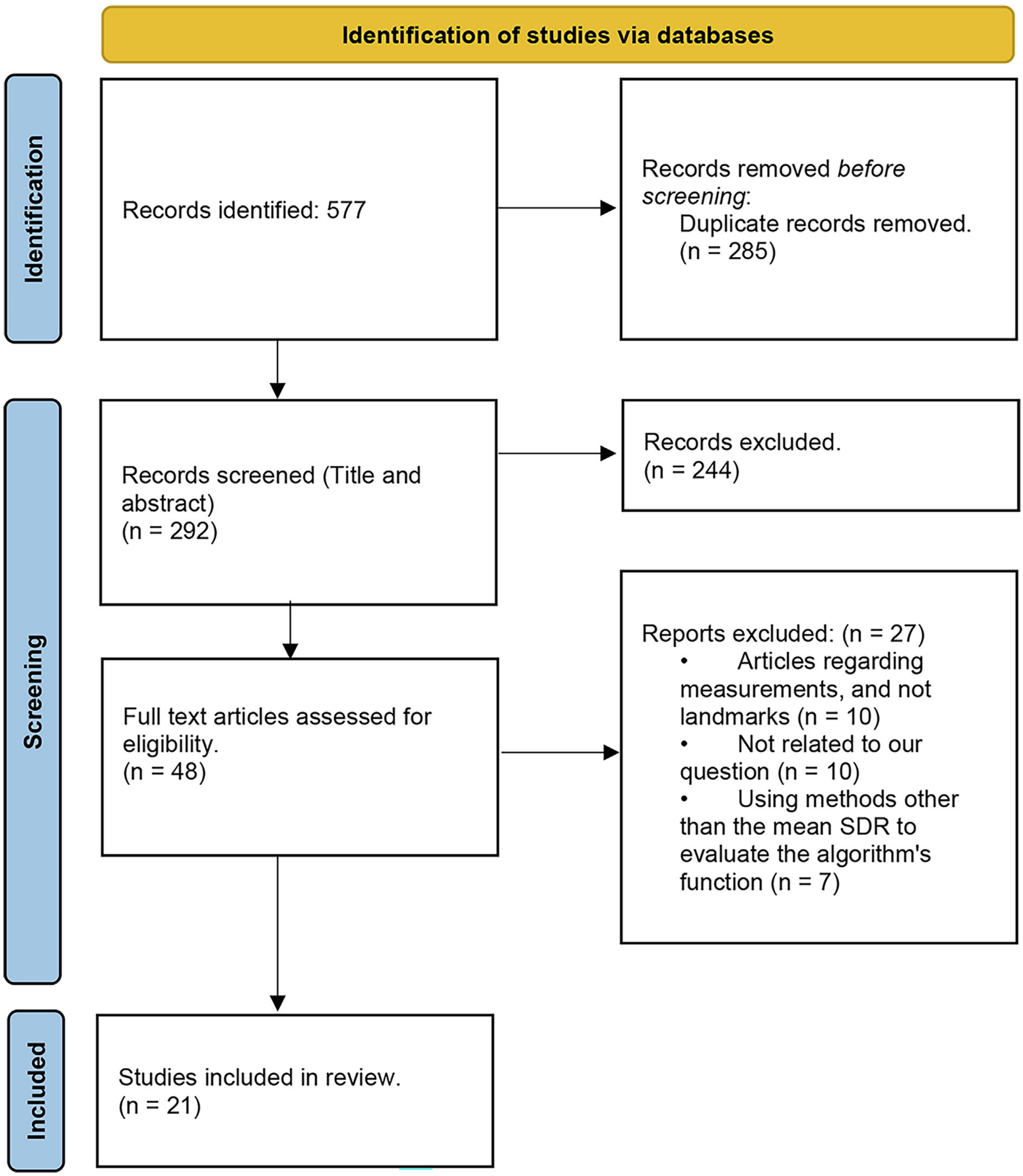

Fig. 1 PRISMA flowchart for screening and selection of standardized research articles.

and AI are often combined with other diagnostic tools, such as facial form analysis and model analysis; thus, the time-consuming task of orthodontic diagnosis can be made more efficient, accurate, and objective (Ruizhongtai Qi 2020).

Decision-making models can hopefully be used in computerized analysis to acquire accurate and consistent data in a timely fashion and then utilize this data to formulate treatment strategies. This type of computerized diagnosis and treatment planning is still in its infancy despite several technical advancements in AI (Juneja et al., 2021). This technology would be a major advancement for diagnosis since the introduction of cephalometry by Broadbent and Hofrath in the 1930s (Helal et al., 2019, Park and Pruzansky 2019, Palomo et al., 2021, Tanna et al., 2021).

In the last few decades, ML approaches have been implicated in anatomical landmarks detection, computerized diagnosis, and data mining related to medical assessments. ML algorithms have commonly been used extensively for decision-making and in various fields to solve real-world data-related issues (Bollen 2019, Jodeh et al., 2019). Research has indicated that cephalometric analysis provides detailed images of anatomical structural points. This improves reliability by maximizing the identifying points' accuracy (Kök et al., 2019). However, there is still uncertainty regarding the accuracy of cephalometric imaging results in detecting anatomical landmarks; thus, the algorithm's accuracy is unclear and should be addressed by analyzing previous studies. In this study, we conducted a *meta*-analysis and systematic review to assess the accuracy of machine learning (ML) software for detecting and predicting anatomical landmarks on 2D lateral cephalometric images.

2. Materials and methods

The *meta*-analysis conducted in this study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009) for extracting, selecting, and screening the included research articles (Fig. 1). After the initial screening phase, the study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) with code CRD42023399216 (Alshamrani et al., 2022). The population, intervention, control, and outcomes (PICO) question was as follows:

Is 2D lateral cephalometric imagery suitable for detecting and predicting anatomical landmarks using ML software? What is the accuracy?

2.1. Eligibility criteria

The *meta*-analysis included the following inclusion criteria: (1) studies employing the diagnostic accuracy and prediction of ML, (2) evaluation and assessment of 2D cephalometric imagery analysis, such as 2D lateral radiographs with relevant landmarks that provide detection and prediction accuracy, (3) reporting the outcome as the mean successful detection rate (SDR), (4) those published after 2000 until February 2023, as we expected ML-related data to be included, and (5) articles published only in the English language. Only studies that met the above criteria were included.

Studies were excluded if they (1) already conducted a systematic review and *meta*-analysis or scoping review, (2) reported other methods for the function of the algorithms rather than SDR, (3) were studies of cephalometry-irrelevant landmarks or used other methods for non-radiographic data, or (4) were articles published in other languages.

2.2. Research strategy and screening

The search and screening of research articles were systematically performed using five databases, including PubMed, Scopus, Scopus Secondary, Embase, and Web of Science (WOS), for studies published from January 2000 to February 2023 in English. The *meta*-analysis utilized PRISMA systematic review and *meta*-analysis guidelines for screening and selecting the included studies. The overall search was designed to analyze the different publications across different disciplines; the keywords for each database are outlined in Table 1.

The titles and abstracts were screened independently by reviewers, and the third reviewer resolved disagreements. All included studies met the eligibility criteria in full and were those for which the full text was available.

Table 1 Keywords for each database.

Database	Keyword	Result
Pubmed	("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "Deep Learning"[Mesh]) AND ("lateral cephalometry" OR "lateral cephalometric")	129
Scopus	(TITLE-ABS-KEY ("Artificial Intelligence") OR TITLE-ABS-KEY ("Machine Learning") OR TITLE-ABS-KEY ("Neural Networks") OR TITLE-ABS-KEY ("Deep Learning")) AND (TITLE-ABS-KEY ("Cephalometry") OR TITLE-ABS-KEY ("lateral cephalometry") OR TITLE-ABS-KEY ("lateral cephalometric"))	193
Scopus secondary	(TITLE-ABS-KEY ("Artificial Intelligence") OR TITLE-ABS-KEY ("Machine Learning") OR TITLE-ABS-KEY ("Neural Networks") OR TITLE-ABS-KEY ("Deep Learning")) AND (TITLE-ABS-KEY ("Cephalometry") OR TITLE-ABS-KEY ("lateral cephalometry") OR TITLE-ABS-KEY ("lateral cephalometric"))	5
Embase	('artificial intelligence'/exp OR 'artificial intelligence' OR 'machine learning'/exp OR 'machine learning' OR 'artificial neural network'/exp OR 'artificial neural network' OR 'neural networks'/exp OR 'neural networks' OR 'deep learning'/exp OR 'deep learning') AND ('cephalometry'/exp OR cephalometry OR 'lateral cephalometry' OR 'lateral cephalometric')	191
WOS	(ALL = ("Artificial Intelligence" OR "Machine Learning" OR "Neural Networks" OR "Deep Learning")) AND ALL = ("Cephalometry" OR "lateral cephalometry" OR "lateral cephalometric")	59

2.3. Data collection and synthesis

The information extracted from research papers is displayed in Table 1. The extracted information was based on study characteristics, including author, year of publication, country of study, imagery (2D lateral radiographs), objective, number of landmarks detected, and findings, as shown in Table 2. Studies were fully extracted if the article mentioned several test datasets or models.

2.4. Quality assessment

The quality assessment of diagnostic accuracy studies (QUADAS-2) tool (Whiting et al., 2011) was utilized to evaluate risk bias, which accounted for risk bias (data selection, index test, and reference test) and applicability concerns (no flow or timing, data selection, index test, and reference test). Two reviewers assessed the bias risk in the included studies and interpreted the results.

Table 2 Data extraction.

Author/year	Country	Architecture	Objective	Sample size	SDR (successful detection rate)
Alshamrani et al. (2022) (Alshamrani et al., 2022)	Saudi Arabia	CNN (autoencoder-based Inception layers)	Generate a Bjork–Jarabak and Ricketts cephalometrics automatically.	100	Basic autoencoder model trained on Set 1 2.0 mm: 64% 2.5 mm: 69% 3.0 mm: 72% 4.0 mm: 77%
				150	Model autoencoder wider Paddup box set 2 2.0 mm: 71% 2.5 mm: 75% 3.0 mm: 78% 4.0 mm: 84%
El-Fegh et al. (2008) (El-Fegh et al., 2008)	Libya/ Canada	CNN	A new approach to cephalometric X-ray landmark localization	> 80	2.0 mm: 91%
El-Feghi et al. (2003) (El-Feghi et al., 2003)	Canada	MLP	A novel algorithm based on the use of the Multi-layer Perceptron (MLP) to locate landmarks on a digitized X-ray of the skull	134	2.0 mm: 91.6%
Hwang et al. (2021) (Hwang et al., 2021)	South Korea	CNN (YOLO version 3)	To compare an automated cephalometric analysis based on the latest deep learning method	200	2.0 mm: 75.45% 2.5 mm: 83.66% 3.0 mm: 88.92% 4.0 mm: 94.24%
Jiang et al. (2023) (Jiang et al., 2023)	China	CNN (A cascade framework "CephNet")	Utilizing artificial intelligence (AI) to achieve automated landmark localization in patients with various malocclusions	259	1.0 mm: 66.15% 2.0 mm: 91.73% 3.0 mm: 97.99%
Kafieh et al. (2009) (Kafieh et al., 2009)	Iran	ASM	As a new method for automatic landmark detection in cephalometry, they propose two different methods for bony structure discrimination in cephalograms.	63	1.0 mm: 24.00% 2.0 mm: 61.00% 5.0 mm: 93.00%
Kim et al. (2020) (Kim et al., 2020)	South Korea	CNN	Develop a fully automated cephalometric analysis method using deep learning and a corresponding web-based application that can be used without high-specification hardware.	100	2.0 mm: 84.53% 2.5 mm: 90.11% 3.0 mm: 93.21% 4.0 mm: 96.79%
Kim et al. (2021) (Kim et al., 2021)	South Korea	CNN	Propose a fully automatic landmark identification model based on a deep learning algorithm using real clinical data	50	2.0 mm: 64.30% 2.5 mm: 77.30% 3.0 mm: 85.50% 4.0 mm: 95.10%
Lee et al. (2020) (Lee et al., 2020)	South Korea	BCNN	Develop a novel framework for locating cephalometric landmarks with confidence regions	250	2.0 mm: 82.11% 2.5 mm: 88.63% 3.0 mm: 92.28% 4.0 mm: 95.96%
Oh et al. (2021) (Oh et al., 2020)	South Korea	CNN	They proposed a novel framework DACFL that enforces the FCN to understand a much deeper semantic representation of cephalograms	150	2.0 mm: 86.20% 2.5 mm: 91.20% 3.0 mm: 94.40% 4.0 mm: 97.70%
				100	2.0 mm: 75.90% 2.5 mm: 83.40% 3.0 mm: 89.30% 4.0 mm: 94.70%
Ramadan et al. (2022) (Ramadan et al., 2022)	Saudi Arabia	CNN	Detection of the cephalometric landmarks automatically	150	2.0 mm: 90.39% 3.0 mm: 92.37%
				100	2.0 mm: 82.66% 3.0 mm: 84.53%
Song et al. (2020) (Song et al., 2020)	Japan	CNN (with a backbone of ResNet50)	A two-step method for the automatic detection of cephalometric landmarks	150	2.0 mm: 86.40% 2.5 mm: 91.70% 3.0 mm: 94.80% 4.0 mm: 97.80%
				100	2.0 mm: 74.00% 2.5 mm: 81.30% 3.0 mm: 87.50%

(*continued on next page*)

Table 2 (*continued*)

Author/year	Country	Architecture	Objective	Sample size	SDR (successful detection rate)
					4.0 mm: 94.30%
Song et al. (2021) (Song et al., 2021)	Japan/ China	CNN (Deep convolutional neural networks)	A coarse-to-fine method to detect cephalometric landmarks	150	2.0 mm: 85.20% 2.5 mm: 91.20% 3.0 mm: 94.40% 4.0 mm: 97.20%
				100	2.0 mm: 72.20% 2.5 mm: 79.50% 3.0 mm: 85.00% 4.0 mm: 93.50%
Song et al. (2020) (Song et al., 2019)	Japan/ China	CNN (Resnet50)	A semi-automatic method for detection of cephalometric landmarks using deep learning.	150	2.0 mm: 85.00% 2.5 mm: 90.70% 3.0 mm: 94.50% 4.0 mm: 98.40%
				100	2.0 mm: 81.80% 2.5 mm: 88.06% 3.0 mm: 93.80% 4.0 mm: 97.95%
Tanikawa et al. (2009) (Tanikawa et al., 2009)	Japan	N/A	Evaluate the reliability of a system that performs automatic recognition of anatomic landmarks and adjacent structures on lateral cephalograms using landmark-dependent criteria unique to each landmark	65	88.00%
Ugurlu, (2022) (Uğurlu 2022)	Turkey	CNN	Develop an artificial intelligence model to detect cephalometric landmark, automatically enabling the automatic analysis of cephalometric radiographs	180	2.0 mm: 76.20% 2.5 mm: 83.50% 3.0 mm: 88.20% 4.0 mm: 93.40%
Wang et al. (2018) (Wang et al., 2018)	China	Multiscale decision tree regression voting using SIFTbased patch	Develop a fully automatic system of cephalometric analysis, including cephalometric landmark detection and cephalometric measurement in lateral cephalograms.	150	2.0 mm: 73.37% 2.5 mm: 79.65% 3.0 mm: 84.46% 4.0 mm: 90.67%
				165	2.0 mm: 72.08% 2.5 mm: 80.63% 3.0 mm: 86.46% 4.0 mm: 93.07%
Yao et al. (2022) (Yao et al., 2022)	China	CNN	Develop an automatic landmark location system to make cephalometry more convenient	100	1.0 mm: 54.05% 1.5 mm: 91.89% 2.0 mm: 97.30% 2.5 mm: 100.00% 3.0 mm: 100.00% 4.0 mm: 100.00%
Yoon et al. (2022) (Yoon et al., 2022)	South Korea	CNN (EfficientNetB0 (Eff-UNet B0) model)	Evaluate the accuracy of a cascaded two-stage (CNN) model in detecting upper airway soft tissue landmarks in comparison with the skeletal landmarks on lateral cephalometric images	100	1.0 mm: 74.71% 2.0 mm: 93.43% 3.0 mm: 97.29% 4.0 mm: 98.71%
Yue et al. (2006) (Yue et al., 2006)	China	ASM	Craniofacial landmark localization and structure tracing are addressed in a uniform framework.	86	2.0 mm: 71.00% 4.0 mm: 88.00%
Zeng et al. (2021) (Zeng et al., 2021)	China	CNN	A novel approach with a cascaded three-stage convolutional neural networks to predict cephalometric landmarks automatically.	150	2.0 mm: 81.37% 2.5 mm: 89.09% 3.0 mm: 93.79% 4.0 mm: 97.86%
				100	2.0 mm: 70.58% 2.5 mm: 79.53% 3.0 mm: 86.05% 4.0 mm: 93.32%

CNN: convolutional neural network, ASM: Active shape model, BCNN: Bayesian Convolutional Neural Networks, MLP: Multi-layer Perceptron.

2.5. Summary measures and data synthesis

To be considered for *meta*-analysis, a study had to report either the deviation from a 1-, 2-, 3-, and 4-mm estimated error criterion (in mm) or the percentage of landmarks accurately predicted within this 1-, 2-, 3-, and 4-mm prediction error thresholds (Higgins and Thompson 2002). Our final measurements were the mean deviation from the 1-, 2-, 3-, and 4-mm thresholds (in mm) or the percentage of landmarks identified within the 1-, 2-, 3-, and 4-mm thresholds, both with their 95% confidence intervals (CI). The *meta*-analysis was conducted using Review Manager version 5.0, and heterogeneity was evaluated using Cochrane's Q and I^2 statistics using the random effect model (Viechtbauer 2010).

3. Results

3.1. Identified studies

The *meta*-analysis yielded approximately 577 research articles on the accuracy of ML algorithms for detecting and predicting anatomical landmarks from the abovementioned databases. According to the inclusion criteria, 48 papers were determined to be relevant, reliable, and in line with the study's objectives. Through the exclusion criteria, 27 papers were eliminated of the 48 studies. Approximately 21 of the remaining articles met the aforementioned criteria and were included.

The reasons for exclusion were as follows:

- Studies on measurements and not landmarks = 10
- Those not related to our question = 10
- Those using methods other than the mean SDR to evaluate the algorithm's function = 7

3.2. Descriptive analysis of identified studies

Among the 577 studies selected, 21 articles were included in the data extraction phase. These studies were model-based studies conducted in Korea, Saudi Arabia, Iran, Israel, Canada, Bosnia, China, Turkey, the USA, and Italy, representing different world regions. Furthermore, they included studies on ML cephalometric landmark detection through CNN, and the outcomes were successful detection rates.

3.3. Risk of bias

The risk of bias in the included studies was assessed using the QUADAS-2 tool in two main domains: risk of bias and applicability issues. The risk of bias assessment demonstrated that some of the included articles exhibit a high risk of bias in data selection (n = 11, 52.38%), reference tests (n = 6, 28.57%), index tests (n = 1.4, 76%), and timing (n = 2.9, 52%). The majority of the presented studies had applicability issues for data selection (n = 5.23, 8%), reference tests (n = 0), and index tests (n = 2.95, 2%). A detailed assessment of the risk of bias and applicability concerns is provided in Table 3.

3.4. Architecture of AI

The majority of the included studies use various modalities of CNNs as the architecture for detecting landmarks on radiographs (n = 15, 71.4%) followed by the active shape model (ASM) at 9% (n = 2). Further information is provided in Table 2.

3.5. Successful detection rates

Twenty-one of the included studies reported the SDR of anatomical landmarks in different ranges. Most studies reported the SDR for the range of 2 mm (n = 20, 95.2%). In addition, 13 of the included studies reported the SDR for the 2.5-mm range (61.9%), 16 studies reported the SDR for the 3-mm range (76.2%), 15 studies reported the SDR for the 4-mm range (71.4%), and 3 studies reported the SDR for the 1-mm range (14.2%). The pooled SDR for the 1-mm, 2-mm, 2.5-mm, 3-mm, and 4-mm ranges were 65%, 81%, 86%, 91%, and 96%, respectively, the supplementary files for Figures 2-6. Table 4 presents further findings of each *meta*-analysis.

4. Discussion

This study's systematic review revealed that ML algorithms for anatomical landmarking of 2D cephalometric images have been implicated as an active radiography resource, as 20 of 21 are studies that reported accuracy, which were typically published between 2006 and 2023. Fifteen studies used varied modalities of CNN, and six studies utilized other AI architectures, such as ASM and Bayesian convolutional neural networks (BCNN). Most of the studies reported SDR for the 2-mm (95.2%), 2.5-mm (61.9%), 3-mm (76.2%), and 4-mm (71.4%) ranges. The overall reported SDR for the 1-mm range was 65% followed by 81% for 2 mm, 86% for 2.5 mm, 91% for 3 mm, and 96% for 4 mm.

Even though these assessments are based on landmarks, it is impossible to systematically determine a total systematic error from landmark machine translation errors. The overall standard deviation might be decreased or increased based on landmark coordinate values, which alters the therapeutic relevance of the findings. Consequently, there is a shortage of data on the diagnosis accuracy of computerized three-dimensional (3D) cephalometry.

Another study found that, compared to other radiographic techniques, cephalograms provide quantitative and qualitative results for anatomical landmark detection (Bichu et al., 2021, Joda and Pandis 2021, Liu et al., 2021, Auconi et al., 2022). Skeletal landmark detection improves the accuracy of quantitative analyses as it identifies reference points. Thus, the landmarks' precise source must be determined to produce relevant results. The current study assessed research that utilized 2D cephalometric images and ML for landmark detection.

The efficacy of ML, as demonstrated by experimental trials, has transformed the implications of ML for cephalometric analysis. However, it requires considerable attention due to the association of certain challenges in orthodontics and other medical assessments. One such difficulty is the presence of

Table 3 Bias risk assessment.

Authors	Year	Risk of bias				Applicability concerns		
		Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Kim et al. (Kim et al., 2021)	2021	Low	Low	Low	Low	Low	Low	Low
Kafieh et al. (Kafieh et al., 2009)	2009	High	Low	High	Unclear	Low	Low	High
Oh et al. (Oh et al., 2020)	2021	Low	Low	Low	Low	Low	Low	Low
Ramadan et al. (Ramadan et al., 2022)	2022	High	Low	Low	Low	High	Low	Low
El-Fegh (El-Fegh et al., 2008)	2008	High	Low	Low	High	High	Low	Low
El-Feghi et al. (El-Feghi et al., 2003)	2003	High	Low	Low	High	High	Low	Low
Lee et al. (Lee et al., 2020)	2020	Low	Low	Low	Low	Low	Low	Low
Kim et al. (Kim et al., 2020)	2020	Low	Low	Low	Low	Low	Low	Low
Alshamrani et al. (Alshamrani et al., 2022)	2022	High	Low	High	Low	High	Low	Low
Hwang et al. (Hwang et al., 2021)	2021	Low	Unclear	Low	Low	Low	Unclear	Low
Jiang et al. (Jiang et al., 2023)	2022	Low	Low	Low	Low	Low	Low	Low
Song et al. (Song et al., 2020)	2020	High	Low	Low	Low	Low	Low	Low
Song et al. (Song et al., 2019)	2019	High	Low	High	Unclear	High	Low	High
Tanikawa et al. (Tanikawa et al., 2009)	2009	Low	Low	High	Low	Low	Low	Low
Yao et al. (Yao et al., 2022)	2022	Low	Low	Low	Low	Low	Low	Low
Wang et al. (Wang et al., 2018)	2018	High	Low	Low	Unclear	Low	Low	Low
Yue et al. (Yue et al., 2006)	2006	High	Low	Low	Low	Low	Low	Low
Yoon et al. (Yoon et al., 2022)	2022	Low	Low	High	Low	Low	Low	Low
Song et al. (Song et al., 2021)	2021	High	High	Low	Low	Low	Unclear	Low
Zeng et al. (Zeng et al., 2021)	2021	High	Low	Low	Low	Low	Low	Low
Ugurlu (Uğurlu 2022)	2022	Low	Low	High	Low	Low	Low	Low

Table 4 Meta-analysis results.

Diameter range	Detection percentage	95% confidence interval	I2	P-value heterogeneity
1 mm	65%	54–76	83.27	0.01
2 mm	81%	78–85	87.83	0.00
2.5 mm	86%	83–89	91.38	0.00
3 mm	91%	88–93	93.44	0.00
4 mm	96%	94–97	90.47%	0.00

"black-box" characteristics in ML, which necessitates improving the visuals and gaining the confidence of physicians and patients before the clinical implementation of ML (Su et al., 2020, Du et al., 2022). Moreover, trial techniques are needed to manage bias risk. For instance, performing consistency evaluations is crucial to assess consistency. Allocation plans also need to be free of personal prejudices. Furthermore, a few other issues, such as a reliability crisis, underfitting, and inadequacy of data, have limited the use of ML in cephalometry (Asiri et al., Tandon et al., 2020, Palanivel et al., 2021, Tanikawa et al., 2021).

Montufar et al. (Montúfar et al., 2018) conducted automatic cephalometric analysis for landmark detection using cone beam computed tomography (CBCT) images and an active surface AI model. They determined the accuracy of this process to be 3.64 mm on average at 18 anatomical points.

Several studies have reported the risk of more errors while detecting irregular structures through cephalometric analysis. Patcas et al. (Patcas et al., 2019) conducted a 2D hybrid cephalometric analysis to acquire CBCT images. Approximately 18 anatomical landmark points were identified with a mean error of 2.51 mm via holistic three-dimensional cephalometric detection. Yu et al. (Yu et al., 2014) evaluated the accuracy of cephalometric analysis using the ML method and reported the interaction between landmark detection and facial attractiveness identification algorithms.

Similarly, Patcas et al. (Patcas et al., 2019) performed a study using AI to assess the accuracy of landmark detection through cephalometric analysis before treatment or decision-making before surgery. For approximately 146 patients that underwent orthognathic surgery, their starting and final images were evaluated using algorithms for facial beauty and appearance. Their study suggested that patients undergoing orthognathic surgery might be assessed for facial symmetry and chronological age using ML.

This *meta*-analysis had several limitations. First, we focused on ML to detect anatomical landmarks, and a comparison with automated landmarking procedures was not conducted.

Second, we excluded several studies, following the inclusion criteria, e.g., those utilizing DL to detect cephalometric analysis, whose full texts were unavailable and did not comprehensively address the study objectives. Third, a variety of risk biases existed in the included studies. Data selection produced limited and potentially unrepresentative groups; most studies utilized the same dataset. Conclusive evidence for predictive data value was relatively poor, particularly for 3D images; images in the test dataset typically were from only a few individuals. Fourth, as previously stated, the limited generalizability was because only a few researchers tested the established DL models on fully independent datasets, such as those from different centers, people, or image processors. Finally, most studies relied on precision estimations rather than other, obviously comparable outcome measures, such as variations in millimeters, pixels, or percentages (primarily as a result of our inclusion criteria) (Gupta et al., 2016).

The use of an ML tool in primary care and its impact on diagnostic and treatment practices, the efficacy, and safety were not documented as additional outcomes that would have been relevant to physicians, patients, or other users. Future research should consider expanding the outcome set and thoroughly testing the applicability of DL in various contexts and situations (e.g., observational studies in clinical care and randomized controlled trials). Of note, the criteria for AI-based cephalometric evaluations could change based on the resulting treatment decisions.

One of the limitations of this study was not including books, other types of literature, and articles that were not in English. To obtain a more accurate outcome, further studies should include more databases, such as Google Scholar, and gray literature.

5. Conclusion

This study demonstrated that ML is significant for detecting landmarks through cephalometric 2D imagery. Most included studies focused on 2D imagery generated by automated cephalometric analysis of ML, which shows promise for the future. The accuracy of landmark detection using ML was heterogeneous across the included studies; however, the accuracy rates of clinicians varied significantly. While generally consistent, the overall evidence shows low generalizability and consistent accuracy, and the clinical utility of ML has yet to be demonstrated. The use of AI for accurately detecting cephalometric landmarks with the extremely low certainty of the findings is intriguing. However, future research should focus on establishing its efficacy and reliability in various samples.

CRediT authorship contribution statement

Jimmy Londono: Conceptualization, Writing – review & editing. **Shohreh Ghasemi:** Writing – original draft, Writing – review & editing. **Altaf Hussain Shah:** Investigation, Writing – original draft. **Amir Fahimipour:** Methodology, Formal analysis. **Niloofar Ghadimi:** Methodology, Writing – original draft. **Sara Hashemi:** Methodology, Formal analysis. **Zohaib Khurshid Sultan:** Investigation, Writing – original draft. **Mahmood Dashti:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sdentj.2023.05.014.

References

Abdinian, M., Baninajarian, H., 2017. The accuracy of linear and angular measurements in the different regions of the jaw in cone-beam computed tomography views. Dental Hypotheses. 8, 100. https://doi.org/10.4103/denthyp.denthyp_29_17.

Alshamrani, K., Alshamrani, H., Alqahtani, F.F., et al, 2022. Automation of Cephalometrics Using Machine Learning Methods. Comput. Intell. Neurosci. 2022, 3061154. https://doi.org/10.1155/2022/3061154.

Asiri, S.N., Tadlock, L.P., Schneiderman, E., et al., Applications of artificial intelligence and machine learning in orthodontics. APOS Trends in Orthodontics. 10, https://doi.org/10.25259/APOS_117_2019.

Auconi, P., Gili, T., Capuani, S., et al, 2022. The validity of machine learning procedures in orthodontics: what is still missing? J Pers Med. 12. https://doi.org/10.3390/jpm12060957.

Bichu, Y.M., Hansa, I., Bichu, A.Y., et al, 2021. Applications of artificial intelligence and machine learning in orthodontics: a scoping review. Prog. Orthod. 22, 18. https://doi.org/10.1186/s40510-021-00361-9.

Bollen, A.-M., 2019. Cephalometry in orthodontics: 2D and 3D. Am. J. Orthod. Dentofac. Orthop. 156, 161.

Du, W., Bi, W., Liu, Y., et al, 2022. Decision Support System for Orthgnathic diagnosis and treatment planning based on machine learning. Res Sq.

Durão, A.R., Pittayapat, P., Rockenbach, M.I., et al, 2013. Validity of 2D lateral cephalometry in orthodontics: a systematic review. Prog. Orthod. 14, 31. https://doi.org/10.1186/2196-1042-14-31.

Ebadian, B., Fathi, A., Tabatabaei, S., 2023. Stress distribution in 5-Unit fixed partial dentures with a pier abutment and rigid and nonrigid connectors with two different occlusal schemes: a three-dimensional finite element analysis. Int. J. Dent. 2023, 3347197. https://doi.org/10.1155/2023/3347197.

El-Fegh, I., Galhood, M., Sid-Ahmed, M., et al, 2008. Automated 2-D cephalometric analysis of X-ray by image registration approach based on least square approximator. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2008, 3949–3952. https://doi.org/10.1109/iembs.2008.4650074.

El-Feghi, I., Sid-Ahmed, M.A., Ahmadi, M., 2003. Automatic identification and localization of craniofacial landmarks using multi layer neural network. Int. Conf. Medical Image Comput. Computer-Assisted Intervention.

Gupta, A., Kharbanda, O.P., Sardana, V., et al, 2016. Accuracy of 3D cephalometric measurements based on an automatic knowledge-based landmark detection algorithm. Int. J. Comput. Assist. Radiol. Surg. 11, 1297–1309. https://doi.org/10.1007/s11548-015-1334-7.

Helal, N.M., Basri, O.A., Baeshen, H.A., 2019. Significance of cephalometric radiograph in orthodontic treatment plan decision. J. Contemp. Dent. Pract. 20, 789–793.

Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558. https://doi.org/10.1002/sim.1186.

Hlongwa, P., 2019. Cephalometric analysis: manual tracing of a lateral cephalogram. S Afr. Dent. J. 74. https://doi.org/10.17159/2519-0105/2019/v74no6a6.

Hwang, H.W., Moon, J.H., Kim, M.G., et al, 2021. Evaluation of automated cephalometric analysis based on the latest deep learning method. Angle Orthod. 91, 329–335. https://doi.org/10.2319/021220-100.1.

Jiang, F., Guo, Y., Yang, C., et al, 2023. Artificial intelligence system for automated landmark localization and analysis of cephalometry. Dentomaxillofac. Radiol. 52, 20220081. https://doi.org/10.1259/dmfr.20220081.

Joda, T., Pandis, N., 2021. The challenge of eHealth data in orthodontics. Am. J. Orthod. Dentofac. Orthop. 159, 393–395. https://doi.org/10.1016/j.ajodo.2020.12.002.

Jodeh, D.S., Kuykendall, L.V., Ford, J.M., et al, 2019. Adding depth to cephalometric analysis: comparing two- and three-dimensional angular cephalometric measurements. J. Craniofac. Surg. 30, 1568–1571. https://doi.org/10.1097/scs.0000000000005555.

Juneja, M., Garg, P., Kaur, R., et al, 2021. A review on cephalometric landmark detection techniques. Biomed. Signal Process. Control 66, 102486.

Kafieh, R., Sadri, S., Mehri, A., et al., 2009. Discrimination of bony structures in cephalograms for automatic landmark detection. In: Advances in Computer Science and Engineering: 13th International CSI Computer Conference, CSICC 2008 Kish Island, Iran, March 9-11, 2008 Revised Selected Papers, Springer.

Kim, M.-J., Liu, Y., Oh, S.H., et al, 2021. Automatic cephalometric landmark identification system based on the multi-stage convolutional neural networks with CBCT combination images. Sensors 21, 505.

Kim, H., Shim, E., Park, J., et al, 2020. Web-based fully automated cephalometric analysis by deep learning. Comput. Methods Programs Biomed. 194,. https://doi.org/10.1016/j.cmpb.2020.105513 105513.

Kök, H., Acilar, A.M., İzgi, M.S., 2019. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. Prog. Orthod. 20, 41. https://doi.org/10.1186/s40510-019-0295-8.

Lee, J.H., Yu, H.J., Kim, M.J., et al, 2020. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. BMC Oral Health 20, 270. https://doi.org/10.1186/s12903-020-01256-7.

Liu, J., Chen, Y., Li, S., et al, 2021. Machine learning in orthodontics: challenges and perspectives. Adv. Clin. Exp. Med. 30, 1065–1074. https://doi.org/10.17219/acem/138702.

Mehdizadeh, M., Tavakoli Tafti, K., Soltani, P., 2022. Evaluation of histogram equalization and contrast limited adaptive histogram equalization effect on image quality and fractal dimensions of digital periapical radiographs. Oral. Radiol. https://doi.org/10.1007/s11282-022-00654-7.

Mohammad-Rahimi, H., Nadimi, M., Rohban, M.H., et al, 2021. Machine learning and orthodontics, current trends and the future opportunities: a scoping review. Am. J. Orthod. Dentofac. Orthop. 160 (170–192), e174.

Moher, D., Liberati, A., Tetzlaff, J., et al, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151, (264–269). https://doi.org/10.7326/0003-4819-151-4-200908180-00135 w264.

Montúfar, J., Romero, M., Scougall-Vilchis, R.J., 2018. Automatic 3-dimensional cephalometric landmarking based on active shape models in related projections. Am. J. Orthod. Dentofac. Orthop. 153, 449–458. https://doi.org/10.1016/j.ajodo.2017.06.028.

Oh, K., Oh, I.-S., Le, V.N.T., et al, 2020. Deep anatomical context feature learning for cephalometric landmark detection. IEEE J. Biomed. Health Inform. 25, 806–817.

Palanivel, J., Davis, D., Srinivasan, D., et al, 2021. Artificial intelligence-creating the future in orthodontics-a review. J. Evol. Med. Dent. Sci. 10, 2108–2114.

Palomo, J.M., El, H., Stefanovic, N., et al., 2021. 3D Cephalometry. 3D Diagnosis and Treatment Planning in Orthodontics: An Atlas for the Clinician, pp. 93-127.

Park, J.H., Pruzansky, D.P., 2019. Imaging and Analysis for the Orthodontic Patient. Craniofacial 3D Imaging: Current Concepts in Orthodontics and Oral and Maxillofacial Surgery. 71-83.

Patcas, R., Bernini, D.A.J., Volokitin, A., et al, 2019. Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age. Int. J. Oral Maxillofac. Surg. 48, 77–83. https://doi.org/10.1016/j.ijom.2018.07.010.

Pattanaik, S., 2019. Evolution of cephalometric analysis of orthodontic diagnosis. Indian J. Forensic. Med. Toxicol. 13.

Ramadan, R., Khedr, A., Yadav, K., et al, 2022. Convolution neural network based automatic localization of landmarks on lateral x-ray images. Multimed. Tools Appl. 81. https://doi.org/10.1007/s11042-021-11596-3.

Ruizhongtai Qi, C., 2020. Deep learning on 3D data. 3D Imaging, Analysis and Applications. 513-566.

Shin, S., Kim, D., 2022. Comparative validation of the mixed and permanent dentition at web-based artificial intelligence cephalometric analysis. J. Korean Acad. Pedtatric. Dent. 49, 85–94.

Song, Y., Qiao, X., Iwmoto, Y., et al., 2019. Semi-automatic Cephalometric Landmark Detection on X-ray Images Using Deep Learning Method. In: International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery.

Song, Y., Qiao, X., Iwamoto, Y., et al, 2020. Automatic cephalometric landmark detection on X-ray images using a deep-learning method. Appl. Sci. 10, 2547.

Song, Y., Qiao, X., Iwamoto, Y., et al, 2021. An efficient deep learning based coarse-to-fine cephalometric landmark detection method. IEICE Trans. Inf. Syst. E104.D, 1359–1366. https://doi.org/10.1587/transinf.2021EDP7001.

Su, M., Feng, G., Liu, Z., et al, 2020. Tapping on the black box: how is the scoring power of a machine-learning scoring function dependent on the training set? J. Chem. Inf. Model. 60, 1122–1136.

Tandon, D., Rajawat, J., Banerjee, M., 2020. Present and future of artificial intelligence in dentistry. J. Oral Biol. Craniofac. Res. 10, 391–396.

Tanikawa, C., Yagi, M., Takada, K., 2009. Automated cephalometry: system performance reliability using landmark-dependent criteria. Angle Orthod. 79, 1037–1046. https://doi.org/10.2319/092908-508r.1.

Tanikawa, C., Kajiwara, T., Shimizu, Y., et al, 2021. Machine/deep learning for performing orthodontic diagnoses and treatment planning. Mach. Learn. Dentistry, 69–78.

Tanna, N.K., AlMuzaini, A., Mupparapu, M., 2021. Imaging in Orthodontics. Dent. Clin. N. Am. 65, 623–641. https://doi.org/10.1016/j.cden.2021.02.008.

Uğurlu, M., 2022. Performance of a convolutional neural network-based artificial intelligence algorithm for automatic cephalometric landmark detection. Turk. J. Orthod. 35, 94–100. https://doi.org/10.5152/TurkJOrthod.2022.22026.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36, 1–48. https://doi.org/10.18637/jss.v036.i03.

Wang, S., Li, H., Li, J., et al, 2018. Automatic analysis of lateral cephalograms based on multiresolution decision tree regression voting. J. Healthc. Eng. 2018, 1797502. https://doi.org/10.1155/2018/1797502.

Whiting, P.F., Rutjes, A.W., Westwood, M.E., et al, 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155, 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

Yao, J., Zeng, W., He, T., et al, 2022. Automatic localization of cephalometric landmarks based on convolutional neural network. Am. J. Orthod. Dentofac. Orthop. 161, e250–e259. https://doi.org/10.1016/j.ajodo.2021.09.012.

Yoon, H.J., Kim, D.R., Gwon, E., et al, 2022. Fully automated identification of cephalometric landmarks for upper airway assessment using cascaded convolutional neural networks. Eur. J. Orthod. 44, 66–77. https://doi.org/10.1093/ejo/cjab054.

Yu, X., Liu, B., Pei, Y., et al, 2014. Evaluation of facial attractiveness for patients with malocclusion: a machine-learning technique employing Procrustes. Angle Orthod. 84, 410–416. https://doi.org/10.2319/071513-516.1.

Yue, W., Yin, D., Li, C., et al, 2006. Automated 2-D cephalometric analysis on X-ray images by a model-based approach. I.E.E.E. Trans. Biomed. Eng. 53, 1615–1623. https://doi.org/10.1109/TBME.2006.876638.

Zeng, M., Yan, Z., Liu, S., et al, 2021. Cascaded convolutional networks for automatic cephalometric landmark detection. Med. Image Anal. 68,. https://doi.org/10.1016/j.media.2020.101904 101904.