Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) 2023 24(11):974-984
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.B2300244

Development and accuracy of artificial intelligence-generated prediction of facial changes in orthodontic treatment: a scoping review

Jiajun ZHU[1], Yuxin YANG[2], Hai Ming WONG[2✉]

[1]*Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Zhejiang Provincial Clinical Research Center for Oral Diseases, Key Laboratory of Oral Biomedical Research of Zhejiang Province, Cancer Center of Zhejiang University, Engineering Research Center of Oral Biomaterials and Devices of Zhejiang Province, Hangzhou 310000, China*
[2]*Faculty of Dentistry, The University of Hong Kong, Hong Kong, China*

Abstract: Artificial intelligence (AI) has been utilized in soft-tissue analysis and prediction in orthodontic treatment planning, although its reliability has not been systematically assessed. This scoping review was conducted to outline the development of AI in terms of predicting soft-tissue changes after orthodontic treatment, as well as to comprehensively evaluate its prediction accuracy. Six electronic databases (PubMed, EBSCO*host*, Web of Science, Embase, Cochrane Library, and Scopus) were searched up to March 14, 2023. Clinical studies investigating the performance of AI-based systems in predicting post-orthodontic soft-tissue alterations were included. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) and Joanna Briggs Institute (JBI) appraisal checklist for diagnostic test accuracy studies were applied to assess risk of bias, while the Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) assessment was conducted to evaluate the certainty of outcomes. After screening 2500 studies, four non-randomized clinical trials were finally included for full-text evaluation. We found a low level of evidence indicating an estimated high overall accuracy of AI-generated prediction, whereas the lower lip and chin seemed to be the least predictable regions. Furthermore, the facial morphology simulated by AI via the fusion of multimodality images was considered to be reasonably true. Since all of the included studies that were not randomized clinical trials (non-RCTs) showed a moderate to high risk of bias, more well-designed clinical trials with sufficient sample size are needed in future work.

Key words: Facial morphology; Soft-tissue changes; Artificial intelligence (AI); Orthodontic treatment

1 Introduction

Improving the facial esthetics has become a common reason for patients to seek orthodontic treatment. The visualized treatment objective (VTO) is an indispensable metric for treatment planning and patient management. Over the years, great attention was given to soft-tissue alterations after orthognathic surgery (ter Horst et al., 2021) as well as growth development (Moon et al., 2022). However, few studies have been concerned with the impact of orthodontic treatment on facial soft tissue. Due to the minor craniofacial involvement and complex biomechanisms, the accurate prediction of post-orthodontic facial appearance is still clinically challenging.

In the past decades, the facial profile was believed to adapt to the underlying dentoalveolar structures at an empirical ratio (Ricketts, 1960; Holdaway, 1983), which in turn gave rise to the emergence of VTO software and its widespread use (Sample et al., 1998; Toepel-Sievers and Fischer-Brandies, 1999). However, skepticism about the reliability and validity of this proportional analysis method merged when the prediction error of measure metrics exceeded the clinical acceptable range of 2 mm in several cases (Sample et al., 1998; Zhang et al., 2019). On the other hand, the facial morphology was found to be influenced by multiple factors including age, gender, inherent lip thickness and stress, type of malocclusion, as well as direction of teeth and jaw movement (Wen et al., 2019; Bral et al., 2020; Lim et al., 2022). This

✉ Hai Ming WONG, wonghmg@hku.hk
Hai Ming WONG, https://orcid.org/0000-0003-3411-6442

Received Apr. 10, 2023; Revision accepted June 7, 2023;
Crosschecked July 28, 2023; Published online Aug. 7, 2023

indicated a high probability that the relationship of soft- and hard-tissue changes was non-linear, and patient-specific factors should be taken into account to enhance the prediction accuracy.

Furthermore, data were mainly collected from lateral cephalometric radiographs and clinical assessment, and the deficiency of 2-dimensional (2D) radiographs in detecting abnormalities in coronal and transversal directions, image distortion, as well as ghosting of bilateral structures was known to reduce the reliability of such data (Moyers and Bookstein, 1979). Furthermore, 2D images used for soft-tissue assessment in conventional approaches frequently compressed and omitted details of facial features (Rongo et al., 2020; Graf et al., 2022). With the advent of three-dimensional (3D) imaging technology, the reconstruction of maxillofacial anatomical structures became possible (Stratemann et al., 2008). However, the analysis of 3D image data is challenging. Finite element analysis (FEA) was firstly applied for building up individualized 3D facial model to simulate orthodontic treatment outcomes, and the preliminary results indicated clinically acceptable accuracy (Chen et al., 2012). Nevertheless, FEA was used to process cone-beam computed tomography (CBCT) data through mathematical transformation, which is time-consuming and may induce morphological distortions. Thus, an optimized tool was needed to fuse the multidimensional images with effectiveness and efficiency.

Proposed in 1950s, artificial intelligence (AI) has advanced to the commercial scale with recent breakthroughs in machine learning. This could be mainly attributed to the evolution of deep learning (DL) algorithms based on artificial neural networks (ANNs). In 1998, ANNs were utilized as the first AI technique in orthodontic research to analyze human craniofacial growth (Lux et al., 1998). The main shortcoming of ANNs is that they cannot receive image data. Meanwhile, their effectiveness and efficiency have yet to be improved (Shen et al., 2017; Abiodun et al., 2019). Thereafter, to enhance the capability of AI in biomedical image analysis, modifications were made on the basis of the classic model of multilayer perceptron. Deep neural networks (DNNs), convolutional neural networks (CNNs), as well as generative adversarial networks (GANs) were constructed (Schwendicke et al., 2019; Pan, 2021, 2022; Vaz and Balaji, 2021; Zhang et al., 2021). To date, DL algorithms have already shown superiority in image detection, characterization, segmentation, and simulation (Howard, 2019). Moreover, automated 3D image management can be implemented with increased precision and efficiency (Howard, 2019; Liu et al., 2021; Tong, 2022).

In the scope of facial soft-tissue evaluation in orthodontic diagnosis, AI also provides opportunities such as integrating CBCT, photograph and facial scan using DL, solving the non-linear problems with neural networks, and making individualized assessment possible (Khanagar et al., 2021; Subramanian et al., 2022). Existing research has reported that AI-powered systems surpass conventional methods and subjective perceptions in both efficiency and accuracy of soft-tissue prediction following orthodontic treatment (Park et al., 2022). However, to the best of our knowledge, its reliability has not been systematically assessed.

This scoping review was performed in a systematic manner to outline the development of AI in predicting facial soft-tissue changes after orthodontic treatment, as well as to comprehensively evaluate its accuracy of prediction. Finally, the clinical usability of AI-based system for soft-tissue prediction and simulation was appraised.

2 Methods

The present scoping review has been registered in the International Prospective Register of Systematic Reviews (PROSPERO, #CRD42022347928) and was conducted and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021).

2.1 Eligibility criteria

The inclusion and exclusion criteria concerning the participants, interventions and comparators, outcomes, and type of study were set as follows.

2.1.1 Participants

Studies on patients who received orthodontic treatment were included. There was no restriction on the age and gender of participants or the orthodontic treatment method applied. Studies in which participants underwent combined orthognathic-orthodontic treatment or were followed up without any active treatment were excluded.

2.1.2 Interventions and comparators

We considered that the AI technique should be applied for soft-tissue assessment and prediction. Any other approach utilized to predict soft-tissue changes was regarded as the comparator. Studies that did not use the AI technique were excluded.

2.1.3 Outcomes

The accuracy of prediction was reported properly. They could be investigated qualitatively and/or quantitatively. Studies not reporting the accuracy of prediction were excluded.

2.1.4 Type of study

We only included original clinical studies with full text available. Publications that failed to report integrated original research, for instance, review, comments, research protocol, abstract, book chapter, and interview, were excluded.

2.2 Data sources

We identified eligible studies by searching six electronic databases, including PubMed, EBSCO*host*, Web of Science, Embase, Cochrane Library, and Scopus. Afterwards, we performed an additional manual search of relevant publications from the reference lists of included studies. The final search was run up to March 14, 2023.

2.3 Search strategy

The search terms consisted of four parts: (1) automatic or "artificial intelligence" or "deep learning" or "machine learning" or "neural networks" or computer; (2) orthodontic, (3) soft tissue, and (4) prediction or forecasting. Two stages of search were carried out (Table 1). Electronic searches were limited to English or Chinese language without restrictions on publication period.

Table 1 Search strategy applied in the selected electronic databases

Stage	Search strategy
1	#1: automatic or "artificial intelligence" or "deep learning" or "machine learning" or "neural networks" or computer
	#2: orthodontic
	#3: soft tissue
	#4: prediction or forecasting
2	#1 and #2 and #3 and #4 and "Chinese [Filter] or English [Filter]"

2.4 Selection process

Articles retrieved from the electronic search were imported into EndNote software (EndNote X9, Ver. 9.3.1). After discarding the duplicates, the titles and abstracts of remaining publications were screened by two reviewers (Zhu JJ and Yang YX) independently according to the eligibility criteria. Subsequently, full texts of included articles were obtained and evaluated, and additional publications were sorted manually. Any disagreements between the reviewers were solved by discussion first to reach consensus. If the dissension persisted, a senior researcher (Wong HM) was consulted to reach the final decision.

2.5 Data collection process

The data collection process was conducted by the two reviewers. All information extracted from the main texts, tables, figures, and appendices was tabulated in Microsoft Excel file. In case of any missing data, the authors of the study would be contacted.

2.6 Data items

The following data were collected from the included studies: first author, year of publication, country, study design, sample size, age of patients, diagnosis of malocclusion, orthodontic treatment method, image modality employed for soft-tissue evaluation, approach of landmark localization, prediction method, assessment, outcomes on accuracy of prediction, and clinical implication. The main outcomes included prediction error and level of accuracy, and the additional outcomes included morphological evaluation and subjective perceptions.

2.7 Study risk of bias assessment and reporting bias assessment

The methodological quality of included studies was independently evaluated by the two reviewers. The assessment tools applied were the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) (Whiting et al., 2011) and Joanna Briggs Institute (JBI) critical appraisal checklist for diagnostic test accuracy studies (Campbell et al., 2020). Any disagreements were solved by discussion to reach consensus. If this was not possible, the article would be submitted to the senior researcher for further assessment.

2.8 Effect measures and synthesis methods

Meta-analysis was not planned in this review. The main and additional outcomes acquired from the included studies were summarized. On the basis of synthetic evaluation, the development of AI in predicting soft-tissue changes after orthodontic treatment was outlined, and the accuracy of AI-generated prediction was comprehensively evaluated.

2.9 Certainty assessment

The Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) was applied for the certainty assessment of evidence (Ryan and Hill, 2016). The rating was decided by the two reviewers independently. If necessary, the senior researcher was involved to make the final decision.

3 Results

3.1 Study selection

A total of 2500 items were initially retrieved from the electronic database search. After discarding duplicates, a total of 2002 records were screened, and 1805 of them were excluded according to the titles and abstracts. The full texts of the remaining 197 articles were obtained and assessed by the two reviewers independently. Then, 193 studies were further excluded based on the exclusion criteria. Finally, four studies (Nanda et al., 2015; Park et al., 2021, 2022; Tanikawa and Yamashiro, 2021) were included in this review. The study selection process is shown in Fig. 1.

3.2 Characteristics of included studies

Table 2 lists the information collected from the included studies in chronological order. The period of publication was from 2015 to 2022. All of the investigations were retrospective and carried out in Asian

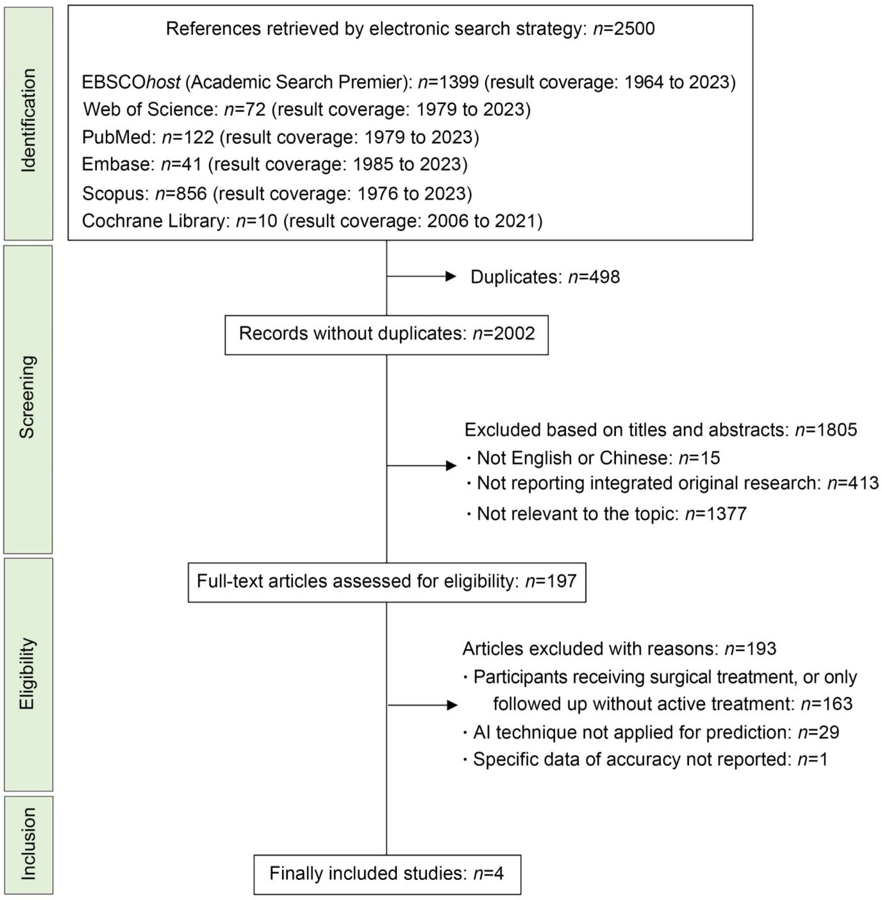

Fig. 1 Study selection process. AI: artificial intelligence.

Table 2 Characteristics of the included studies

Authors, year	Country	Study design	Sample size	Age (years)	Malocclusion	Orthodontic treatment method	Image modality	Landmark localization
Nanda et al., 2015	India	C	40	19.75 (Ex.); 18.75 (non-Ex.)	N/A	Ex. (4/4, 4/5, 5/5) and non-Ex. with FOT	LC	M
Park et al., 2021	Korea	C-S	284	>18	Class II malocclusion	Non-Ex., MMD	LC	M
Tanikawa and Yamashiro, 2021	Japan	C-S	65	15.6 (12.0–37.0)	Positive overjet; mandibular deviation of ≤5 mm	Ex. with FOT	LC and 3D facial images	N/A
Park et al., 2022	Korea	C-S	312	27.7±10.2 (18.0–62.0)	Skeletal discrepancy: Class I (52.2%), Class II (39.4%), Class III (8.4%); Lip incompetency: 13.5% present, 86.5% not present	FOT	CBCT	M

Authors, year	Prediction method	Assessment and outcomes	Clinical implication
Nanda et al., 2015	SRA	ULCC: R^2=60.40% (4/4 Ex.), 42.70% (4/5 Ex.), 30.00% (5/5 Ex.), 4.10% (non-Ex.); LLCC: R^2=7.30% (4/4 Ex.), 95.30% (4/5 Ex.), 32.00% (5/5 Ex.), 49.00% (non-Ex.)	The ANNs-based system achieved higher accuracy of prediction than SRA.
	ANNs	ULCC: PE%=(29.6±25.1)%, (5.9%–72.6%); LLCC: PE%=(7.0±11.0)%, (3.0%–37.2%)	
Park et al., 2021	CNNs	PE1≤1.00 mm: 0.67±1.49 (A'), 0.83±1.12 (Sn), 0.25±0.60 (Pn), 0.13±0.21 (Cm); 1.00 mm<PE1≤2.00 mm: 1.70±1.99 (ULP), 1.44±1.26 (Ls), 1.12±1.11 (Stms); 2.00 mm<PE1≤3.0.0 mm: 2.01±2.13 (LLP), 2.26±2.74 (B'), 2.18±1.81 (Stmi); PE1>3.00 mm: 3.35±3.35 (Pog'), 4.44±2.94 (Me')	The accuracy of CNNs-generated prediction was higher for the nasal and upper lip areas, while it was lower for the lower lip and menton area.
Tanikawa and Yamashiro, 2021	GMMs+ DNNs	PE2 (mm): 0.69±0.22 (0.24–1.77) among points; 0.69±0.18 (0.30–1.02) among patients LoA: 92% (PE2<1.00 mm), 100% (PE2<2.00 mm) Md: maximum errors were observed in lower lip area	The prediction accuracy of DNNs-based system was clinically acceptable.
Park et al., 2022	cGANs	Md: similar features were showed on predicted and real face; relatively large deviation was observed in menton area PE3 (mm): 0.80±0.56 (Sn), 1.00±0.88 (ULP), 1.00±0.67 (LLP), 1.50±1.31 (Pog'), 1.30±0.96 (Ch), 1.50±1.08 (Ck), 1.20±1.01 (overall) LoA (PE≤2.00 mm): 95.5% (Sn), 86.4% (LLP), 84.1% (ULP), 81.8% (Ch), 72.2% (Ck), 68.2% (Pog'), 80.8% (overall) S: (1) selection ratio for predicted face: 26.0% (45°), 44.6% (90°), 34.4% (45°+90°); (2) did not recognize 17.9% (10/56), recognized but unable to distinguish 46.4% (26/56), distinguished well 35.7% (20/56)	The prediction accuracy of cGANs-based system was clinically acceptable.

C: cohort study; C-S: cross-sectional study; Ex.: extraction orthodontic treatment, or tooth extracted; Non-Ex.: non-extraction orthodontic treatment, or tooth not extracted; N/A: not available or not applicable; 4/4 Ex.: all first premolars; 4/5 Ex.: upper first and lower second premolars; 5/5 Ex.: all second premolars; FOT: fixed orthodontic treatment; MMD: maxillary molar distalization; LC: X-ray lateral cephalogram; M: manually; 3D: 3-dimensional; CBCT: cone-beam computed tomography; SRA: stepwise regression analysis; ANNs: artificial neural networks; CNNs: convolutional neural networks; GMMs: geometric morphometric methods; DNNs: deep neural networks; cGANs: conditional generative adversarial networks; R^2: explained variance; ULCC: upper lip curvature change; LLCC: lower lip curvature change; PE%: percentage of mean prediction error; PE1: prediction error, defined as Euclidean distance between the predicted distribution and the actual post-treatment points; A': soft tissue A point; Sn: subnasale; Pn: pronasale; Cm: columella; ULP: upper lip point; Ls: labrale superius; Stms: stomion superius; LLP: lower lip point; B': soft tissue B point; Stmi: stomion inferius; Pog': soft-tissue pogonion; Me': soft-tissue menton; PE2: prediction error, defined as the difference between the predicted and actual post-treatment coordinates of the semi-landmarks along the Z-axis; LoA: level of accuracy, defined as the percentage of PE within a certain range; Md: morphological differences, shown as a color probability map and/or color distance map; PE3: prediction error, defined as the mean absolute closest distance between the predicted and actual post-treatment surfaces in areas surrounding the landmarks including ULP, LLP, Sn, Pog', and the average of the left and right cheilion as well as cheek; Ch: cheilion; Ck: cheek; S: survey in which experienced orthodontists were asked to (1) select the expected post-treatment face, and (2) distinguish simulated versus real face.

countries, including India, Korea, and Japan. We found variation in the study design, characteristics of participants, and orthodontic treatment method. The sample size ranged from 40 to 312.

3.3 Image modality and method of landmark localization

A lateral cephalogram was adopted for soft-tissue evaluation in three studies (Nanda et al., 2015; Park et al., 2021; Tanikawa and Yamashiro, 2021). Among them, one study utilized the combination of 2D radiograph and 3D facial image (Tanikawa and Yamashiro, 2021). In the fourth study (Park et al., 2022), CBCT was employed as the only image material. On the other hand, landmarks were mainly localized in manual mode in three papers (Nanda et al., 2015; Park et al., 2021, 2022), while one study (Tanikawa and Yamashiro, 2021) did not clarify the localization method.

The facial profile parameters and indicators applied in the included studies were shown in Fig. 2. The labrale superius (or upper lip point (ULP)) and labrale inferius (or lower lip point (LLP)) were identified in all four studies, whilst subnasale and soft-tissue pogonion were used in three of them (Park et al., 2021, 2022; Tanikawa and Yamashiro, 2021). In addition, Nanda et al. (2015) and Park et al. (2021) gave attention to the alteration of soft tissue A point (A') and B point (B'). Besides traditional midsaggital landmarks, parasaggital landmarks can be identified on 3D images. Bilateral cheilion was analyzed by both studies utilizing 3D imaging technology (Tanikawa and Yamashiro, 2021; Park et al., 2022). Other points, including porion, exocanthion, endocanthion, alar curvature, and cheek, were also investigated.

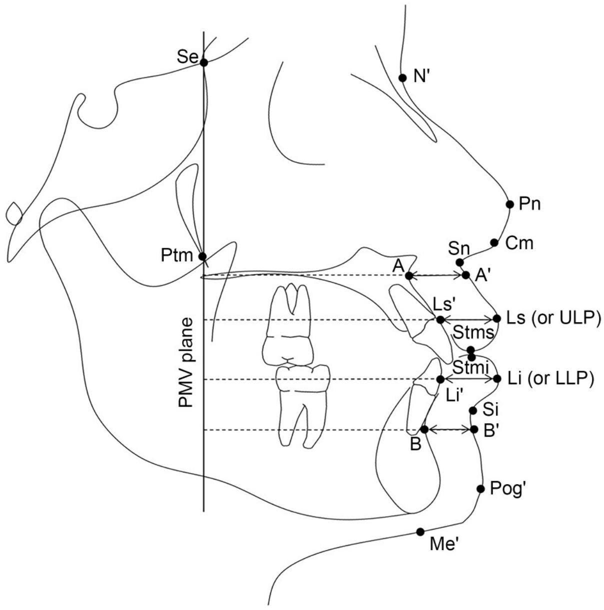

Fig. 2 Facial profile parameters and indicators applied in the included studies. The solid line represents reference plane; the double-headed arrow represents linear parameters. The upper lip curvature was calculated as the difference between upper lip thickness at point Ls (Ls–Ls') and at point A (A–A'); the lower lip curvature was calculated as the difference between lower lip thickness at point Li (Li–Li') and at point B (B–B'). Se: sphenoethmoidal point, the intersection of the greater wings of the sphenoid with the floor of the anterior cranial fossa; Ptm: the inferior and most posterior point on the anterior outline of the pterygomaxillary fissure; PMV plane: pterygomaxillary vertical plane; N': soft tissue N point; Pn: pronasale; Cm: columella; Sn: subnasale; A: subspinale point; A': soft tissue A point; Ls: labrale superius; Ls': projected Ls, point constructed where a line perpendicular to the PMV plane and passing through Ls intersects the hard-tissue outline; ULP: upper lip point; Stms: stomion superius; Li: labrale inferius; Li': projected Li, point constructed where a line perpendicular to the PMV plane and passing through Li intersects the hard-tissue outline; LLP: lower lip point; Stmi: stomion inferius; Si: mentolabial sulcus; B: supramentale point; B': soft tissue B point; Pog': soft-tissue pogonion; Me': soft-tissue menton.

3.4 Prediction method

All of the included studies developed DL-powered prediction systems based on different neural networks. In 2015, ANNs were first used to estimate the lip curvature changes (Nanda et al., 2015) and compared with stepwise regression analysis (SRA). Afterwards, CNNs with U-net structure were adopted to predict the profile alteration at the 2D level (Park et al., 2021). Approximately at the same time, 3D facial morphology was simulated using DNNs combined with landmark-based geometric morphometric methods (GMMs) (Tanikawa and Yamashiro, 2021). Recently, researchers constructed conditional GANs (cGANs) to optimize the 3D simulation of soft-tissue changes (Park et al., 2022).

3.5 Assessment and outcomes

3.5.1 Prediction error

The landmark-based prediction error was clearly stated in two of the included studies (Park et al., 2021, 2022), which measures the distance between the predicted and actual positions of points in the coordinate system established by AI. The outcomes showed that the absolute error tended to increase from midface portion to submental region (Park et al., 2021, 2022). More specifically, the prediction error for the nasal and upper lip areas was generally within 2 mm, while for the soft-tissue pogonion and menton areas, it could exceed 3 mm (Park et al., 2021, 2022).

Meanwhile, the other two studies only reported some overall results. In the ANNs-based system, the percentage of mean error in predicting the upper lip curvature change and lower lip curvature change was 29.6% and 7.0%, respectively. The outcomes were apparently closer to the actual values in comparison with the explained variance showed in SRA (Nanda et al., 2015). The overall error of DNNs-generated prediction was (0.69±0.22) mm among points and (0.69±0.18) mm among patients (Tanikawa and Yamashiro, 2021).

3.5.2 Level of accuracy

The level of accuracy was determined in two of the included studies (Tanikawa and Yamashiro, 2021; Park et al., 2022) according to the magnitude of prediction error, where less than 2 mm was considered clinically acceptable. Extraordinary high level of accuracy was achieved using DNNs, that is, 100% of cases showed a prediction error of <2 mm, while 92% cases showed <1 mm (Tanikawa and Yamashiro, 2021). The overall accuracy of cGANs-based system was 80.8% (Park et al., 2022). The highest level of accuracy was found in the subnasale (95.5%), followed by LLP (86.4%), ULP (84.1%), and cheilion (81.8%). An apparently low level of accuracy was found in the soft-tissue pogonion and cheek (Park et al., 2022).

3.5.3 Morphological differences

With the assistance of AI, surface deviation can be identified and visualized as a color distance map (Tanikawa and Yamashiro, 2021; Park et al., 2022). Through the superimposition of the predicted and real 3D facial forms, a significant difference was indicated in the lower lip (Tanikawa and Yamashiro, 2021) and mentolabial sulcus (Park et al., 2022), which was consistent with the quantitative outcomes.

3.5.4 Subjective perceptions

Park et al. (2022) conducted an online survey to investigate the opinions of experienced orthodontists toward the simulated faces. The panelists were asked to select the expected post-treatment face as well as to distinguish the predicted versus the real face. The results indicated that cGANs-generated simulation might be less "ideal" but provide "real enough" outcome for most of the orthodontists.

3.6 Risk of bias assessment

Methodological quality assessment was performed using QUADAS-2 (Table 3) and the JBI critical appraisal tool (Table 4). All of the included studies showed a moderate to high risk of bias.

Table 3 Risk of bias assessment according to QUADAS-2

Study	Risk of bias assessment				Applicability concern		
	Patient selection	Index test(s)	Reference standard	Flow and timing	Patient selection	Index test(s)	Reference standard
Nanda et al., 2015	H	H	H	H	L	H	H
Park et al., 2021	H	H	U	L	L	L	U
Tanikawa and Yamashiro, 2021	H	L	U	U	L	H	U
Park et al., 2022	H	L	U	L	L	L	U

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies-2; H: high; L: low; U: unclear.

Table 4 Risk of bias assessment according to the JBI critical appraisal checklist for diagnostic test accuracy studies

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Nanda et al., 2015	N	N	U	N	N	N	N	Y	N	Y
Park et al., 2021	N	N	Y	N/A	N	N/A	N/A	N/A	Y	Y
Tanikawa and Yamashiro, 2021	N	N	U	N/A	Y	N/A	N/A	N/A	U	Y
Park et al., 2022	N	N	Y	N/A	Y	N/A	N/A	N/A	Y	Y

JBI: Joanna Briggs Institute; Q1–Q10: questions 1–10 in JBI critical appraisal checklist; N: no; Y: yes; U: unclear; N/A: not available or not applicable.

3.7 Certainty assessment

The certainty of evidence was assessed according to the criteria in the GRADE system. Since all of the included studies were not randomized clinical trials (non-RCTs) and identified moderate to high risk of bias, the rating started as low level of efficacy. No serious inconsistency, indirectness, imprecision, or publication bias was found. Meanwhile, the criteria for upgrading were not fulfilled. Thus, the overall strength of evidence was kept as the baseline rating of low certainty.

4 Discussion

After screening 2002 nonduplicated articles, only four non-RCTs with moderate to high risk of bias were included in this scoping review. In spite of the low level of efficacy, existing evidence indicated that the AI-based system should be clinically acceptable to predict soft-tissue changes after orthodontic treatment. The findings are discussed in detail below.

4.1 Characteristics of included studies

From the compiled information, the development of AI in terms of predicting post-orthodontic soft-tissue changes could be outlined. The first attempt was made in 2015 (Park et al., 2022). It took approximately six years for AI to automatically identify the cephalometric landmarks (Leonardi et al., 2009), which was mandatory to hard- and soft-tissue analysis in orthodontics. Subsequently, the investigation came to a halt until 2021 (Park et al., 2021; Tanikawa and Yamashiro, 2021). Overall, the research in this field has been sparse. It is noteworthy that all of the included studies were conducted in Asian countries. Other reviews also found that the majority of publications about the utilization of AI in orthodontics (Khanagar et al., 2021) and dentistry (Mörch et al., 2021) were from Asia. Moreover, inequal development and distinct concerns of AI application were revealed among different countries and regions.

4.2 Image modality

The lateral cephalogram has remained in use as the main image modality for soft-tissue analysis in AI-based systems (Nanda et al., 2015; Park et al., 2021; Tanikawa and Yamashiro, 2021). Due to technical reasons, image data could not be received directly by some DL frameworks (Pan, 2022), such as ANNs (Nanda et al., 2015) and DNNs (Tanikawa and Yamashiro, 2021). Preprocessing had to be performed to convert the image to tabular data before input (Shen et al., 2017; Abiodun et al., 2019). The procedure was apparently overly complex for 3D images. On the other hand, considering the radiation risk and high expense, 3D radiographies and images had not yet been adopted routinely in the clinic (Karatas and Toy, 2014; Scarfe et al., 2017). This may also explain why only a few clinical trials used 3D imaging for orthodontic diagnosis.

Nevertheless, with the rapid development of AI technology, optimized algorithms have been introduced to deal with image data. CNNs have shown excellent performance in landmark detection and facial feature classification since 2017 (Schwendicke et al., 2019; Zhang et al., 2021). One of the included studies constructed cGANs models by adding a convolutional discriminator onto the CNN structure (Park et al., 2022). Then, it was employed to analyze CBCT data and produced a simulation of 3D facial form with high resolution and definition (Park et al., 2022). Moreover, multimodality image fusion could be implemented by the aid of AI (Tanikawa and Yamashiro, 2021). Altogether, with the enhancement of the safety and accessibility of 3D medical imaging technology, more and more applications of AI on 3D soft-tissue analysis are expected in the future.

4.3 Accuracy of AI-generated prediction

The approach of hard- and soft-tissue evaluations transformed after the involvement of AI technique. Unlike the traditional cephalometric analysis that required plenty of work in angular and linear measurements, the AI-based system could localize each landmark through the establishment of a coordinate system (Park et al., 2021, 2022; Tanikawa and Yamashiro, 2021). The prediction error for each landmark can be determined by identifying the deviation between the coordinates of predicted and actual positions. Without setting reference plane and analysis indexes, this process minimizes measurement error and guarantees the homogeneity of results.

Both the quantitative and the qualitative results showed a general trend of prediction accuracy in relation to the proportion of face. Only one study indicated that the lower lip was more predictable than the upper lip (Nanda et al., 2015). Since only 40 participants were enrolled in this cohort study and a small sample size tended to cause overestimation, the outcomes should be interpreted with caution. Based on the other three studies, the synthetic evaluation suggested that AI-generated prediction for the nasal and upper lip areas has reached the clinically acceptable level of accuracy. On the other hand, the simulation of lower lip and chin was less accurate and might not fulfill the requirement of practical application (Park et al., 2021, 2022; Tanikawa and Yamashiro, 2021). Similar outcomes have been reported in previous studies using traditional mathematical approaches (Kassem and Marzouk, 2018; Zhang et al., 2019; Soheilifar et al., 2022). The unpredictability of soft tissue in the mandibular region may be due to several reasons. Soft-tissue adaptability to hard tissue varied in different facial parts. In the dynamic state, chin form and mentalis strain were affected by the horizontal relationship between the upper and lower jaws as well as lower facial height (Kasai, 1998). In addition, mandible rotation, movement of anterior teeth, and limitations in reproducibility of image along with other unknown factors all increased the difficulty of prediction for the lower lip and chin areas (Moyers and Bookstein, 1979; Zhang et al., 2019). Thus, to enhance the predictive power, it is necessary to deepen the understanding of orthodontic biomechanics and its effect on soft-tissue changes.

In general, all DL algorithms applied in the included studies displayed clinical usability at varying levels. As the most classic and concise multilayer perceptron, the ANNs-based system has already outperformed traditional mathematical models in soft-tissue prediction (Nanda et al., 2015). However, its accuracy varied greatly among different cases and was insufficient compared to lately modified neural networks. Tanikawa and Yamashiro (2021) constructed a DNN model with two rectified linear unit dense (ReDense) layers and one dropout layer to attain rapid computation without gradient vanishing, lower training loss, and avoidance of overfitting (Javid et al., 2021). Therein, the highest level of accuracy was reported among all included studies. However, prediction error of each landmark was not specified in the article, which may be due to the aforementioned limitation of the neural networks, that is, image data cannot be directly received. Thus, further research was needed to fully evaluate the forecasting capability of DNNs. Meanwhile, landmark-based measurement was accomplished by the AI system using U-net-based CNNs (Park et al., 2021) and cGANs (Park et al., 2022). According to the outcomes, cGANs showed obvious priority in 3D facial feature analysis and prediction. Like CNNs, cGANs were able to capture the spatial features of images and output the results based on very few training samples (Pan, 2021, 2022; Zhang et al., 2021). Moreover, 3D images with higher resolution and definition were produced with the optimization from discriminative filter (Park et al., 2022), which may also contribute to the higher prediction accuracy.

Furthermore, the simulation of facial form was also revolutionized with the AI technique. After training, the AI-based system could process and integrate multimodal images with the same coordinate system for further diagnostic analysis and prognosis forecasting (Tanikawa and Yamashiro, 2021; Park et al., 2022). Moreover, the simulation of predicted 3D facial morphology with high fidelity could be achieved (Tanikawa and Yamashiro, 2021; Park et al., 2022) and its clinical validation has been primarily confirmed via a survey that collected the subjective perceptions from experienced orthodontists (Park et al., 2022).

To summarize, based on the existing literature, we demonstrated the great potential of AI in predicting and simulating post-orthodontic facial form. Admittedly, however, the current review has some limitations. Only four studies were included and this may not be sufficient to fully represent the accuracy of AI-generated prediction for soft-tissue changes. In addition, all of these studies were with moderate to high risk of bias, which may limit the confidence in the results. Last but not least, the sample size of each study was probably insufficient and may affect the generalizability of outcomes. The rapid evolution of AI technology involves transforming the mode of clinical diagnosis and helps advancing 3D VTO technologies, which are paramount to improve treatment planning and patient/parent management, and to enhance the facial esthetics more efficiently for patients receiving orthodontic treatment. Meanwhile, rigorous research has been rare; future work should be undertaken to carry out observational to RCT studies of sufficient sample size. Multicenter trials are also worth considering.

5 Conclusions

The AI technique has been applied to predict the post-orthodontic soft-tissue changes since 2015. We found a low level of evidence indicating that the overall accuracy of AI-generated prediction was estimated to be clinically acceptable, whereas the lower lip and chin areas seemed to be less predictable. Furthermore, the facial morphology simulated by AI via the fusion of multimodality images was considered to be reasonably true. However, since all of the included non-RCT studies demonstrated a moderate to high risk of bias, more well-designed clinical trials with sufficient sample size are needed in the future.

Data availability statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

The work was supported by the Research Grants Council of the Hong Kong, China (No. 17109619).

Author contributions

Jiajun ZHU was involved in conceptualization, methodology, investigation, formal analysis, validation, and writing. Yuxin YANG contributed to investigation and writing. Hai Ming WONG was involved in investigation, writing, supervision, and funding acquisition. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Jiajun ZHU, Yuxin YANG, and Hai Ming WONG declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Abiodun OI, Jantan A, Omolara AE, et al., 2019. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, 7:158820-158846. https://doi.org/10.1109/ACCESS.2019.2945545

Bral A, Olate S, Zaror C, et al., 2020. A prospective study of soft- and hard-tissue changes after mandibular advancement surgery: midline changes in the chin area. *Am J Orthod Dentofacial Orthop*, 157(5):662-667. https://doi.org/10.1016/j.ajodo.2019.05.022

Campbell JM, Klugar M, Ding S, et al., 2020. Chapter 9: Diagnostic test accuracy systematic reviews. *In*: Aromataris E, Munn Z (Eds.), JBI Manual for Evidence Synthesis. JBI, p.309-359. https://doi.org/10.46658/JBIMES-20-10

Chen S, Lou HD, Guo L, et al., 2012. 3-D finite element modelling of facial soft tissue and preliminary application in orthodontics. *Comput Methods Biomech Biomed Engin*, 15(3):255-261. https://doi.org/10.1080/10255842.2010.522188

Graf CC, Dritsas K, Ghamri M, et al., 2022. Reliability of cephalometric superimposition for the assessment of craniofacial changes: a systematic review. *Eur J Orthod*, 44(5): 477-490. https://doi.org/10.1093/ejo/cjab082

Holdaway RA, 1983. A soft-tissue cephalometric analysis and its use in orthodontic treatment planning. Part I. *Am J Orthod*, 84(1):1-28. https://doi.org/10.1016/0002-9416(83)90144-6

Howard J, 2019. Artificial intelligence: implications for the future of work. *Am J Ind Med*, 62(11):917-926. https://doi.org/10.1002/ajim.23037

Javid AM, Das S, Skoglund M, et al., 2021. A ReLU dense layer to improve the performance of neural networks. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, p.2810-2814. https://doi.org/10.1109/ICASSP39728.2021.9414269

Karatas OH, Toy E, 2014. Three-dimensional imaging techniques: a literature review. *Eur J Dent*, 8(1):132-140. https://doi.org/10.4103/1305-7456.126269

Kasai K, 1998. Soft tissue adaptability to hard tissues in facial profiles. *Am J Orthod Dentofacial Orthop*, 113(6):674-684. https://doi.org/10.1016/s0889-5406(98)70228-8

Kassem HE, Marzouk ES, 2018. Prediction of changes due to mandibular autorotation following miniplate-anchored intrusion of maxillary posterior teeth in open bite cases. *Prog Orthod*, 19:13. https://doi.org/10.1186/s40510-018-0213-5

Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, et al., 2021. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making – a systematic review. *J Dent Sci*, 16(1):482-492. https://doi.org/10.1016/j.jds.2020.05.022

Leonardi R, Giordano D, Maiorana F, 2009. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images. *J Biomed Biotechnol*, 2009:717102. https://doi.org/10.1155/2009/717102

Lim YN, Yang BE, Byun SH, et al., 2022. Three-dimensional digital image analysis of skeletal and soft tissue points A and B after orthodontic treatment with premolar extraction in bimaxillary protrusive patients. *Biology (Basel)*, 11(3):381. https://doi.org/10.3390/biology11030381

Liu CX, Kong DH, Wang SF, et al., 2021. Deep3D reconstruction: methods, data, and challenges. *Front Inform Technol Electron Eng*, 22(5):652-672. https://doi.org/10.1631/FITEE.2000068

Lux CJ, Stellzig A, Volz D, et al., 1998. A neural network approach to the analysis and classification of human craniofacial growth. *Growth Dev Aging*, 62(3):95-106.

Moon JH, Kim MG, Hwang HW, et al., 2022. Evaluation of an individualized facial growth prediction model based on the multivariate partial least squares method. *Angle Orthod*, 92(6):705-713.
https://doi.org/10.2319/110121-807.1

Mörch CM, Atsu S, Cai W, et al., 2021. Artificial intelligence and ethics in dentistry: a scoping review. *J Dent Res*, 100(13):1452-1460.
https://doi.org/10.1177/00220345211013808

Moyers RE, Bookstein FL, 1979. The inappropriateness of conventional cephalometrics. *Am J Orthod*, 75(6):599-617.
https://doi.org/10.1016/0002-9416(79)90093-9

Nanda SB, Kalha AS, Jena AK, et al., 2015. Artificial neural network (ANN) modeling and analysis for the prediction of change in the lip curvature following extraction and non-extraction orthodontic treatment. *J Dent Specialities*, 3(2):130-139.
https://doi.org/10.5958/2393-9834.2015.00002.9

Page MJ, McKenzie JE, Bossuyt PM, et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71.
https://doi.org/10.1136/bmj.n71

Pan YH, 2021. Miniaturized five fundamental issues about visual knowledge. *Front Inform Technol Electron Eng*, 22(5): 615-618.
https://doi.org/10.1631/FITEE.2040000

Pan YH, 2022. On visual understanding. *Front Inform Technol Electron Eng*, 23(9):1287-1289.
https://doi.org/10.1631/FITEE.2130000

Park JH, Kim YJ, Kim J, et al., 2021. Use of artificial intelligence to predict outcomes of nonextraction treatment of Class II malocclusions. *Semin Orthod*, 27(2):87-95.
https://doi.org/10.1053/j.sodo.2021.05.005

Park YS, Choi JH, Kim Y, et al., 2022. Deep learning-based prediction of the 3D postorthodontic facial changes. *J Dent Res*, 101(11):1372-1379.
https://doi.org/10.1177/00220345221106676

Ricketts RM, 1960. Cephalometric synthesis: an exercise in stating objectives and planning treatment with tracings of the head roentgenogram. *Am J Orthod*, 46(9):647-673.
https://doi.org/10.1016/0002-9416(60)90172-X

Rongo R, Bucci R, Adaimo R, et al., 2020. Two-dimensional versus three-dimensional Fränkel Manoeuvre: a reproducibility study. *Eur J Orthod*, 42(2):157-162.
https://doi.org/10.1093/ejo/cjz081

Ryan R, Hill S, 2016. How to GRADE the quality of the evidence. Cochrane Consumers and Communication Group.
http://cccrg.cochrane.org/author-resources

Sample LB, Sadowsky PL, Bradley E, 1998. An evaluation of two VTO methods. *Angle Orthod*, 68(5):401-408.
https://doi.org/10.1043/0003-3219(1998)068<0401:AEOTVM>2.3.CO;2

Scarfe WC, Azevedo B, Toghyani S, et al., 2017. Cone Beam Computed Tomographic imaging in orthodontics. *Aust Dent J*, 62(Suppl 1):33-50.
https://doi.org/10.1111/adj.12479

Schwendicke F, Golla T, Dreher M, et al., 2019. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent*, 91:103226.
https://doi.org/10.1016/j.jdent.2019.103226

Shen DG, Wu GR, Suk HI, 2017. Deep learning in medical image analysis. *Annu Rev Biomed Eng*, 19:221-248.
https://doi.org/10.1146/annurev-bioeng-071516-044442

Soheilifar S, Soheilifar S, Afrasiabi Z, et al., 2022. Prediction accuracy of Dolphin software for soft-tissue profile in Class I patients undergoing fixed orthodontic treatment. *J World Fed Orthod*, 11(1):29-35.
https://doi.org/10.1016/j.ejwf.2021.10.001

Stratemann SA, Huang JC, Maki K, et al., 2008. Comparison of cone beam computed tomography imaging with physical measures. *Dentomaxillofac Radiol*, 37(2):80-93.
https://doi.org/10.1259/dmfr/31349994

Subramanian AK, Chen Y, Almalki A, et al., 2022. Cephalometric analysis in orthodontics using artificial intelligence – a comprehensive review. *Biomed Res Int*, 2022:1880113.
https://doi.org/10.1155/2022/1880113

Tanikawa C, Yamashiro T, 2021. Development of novel artificial intelligence systems to predict facial morphology after orthognathic surgery and orthodontic treatment in Japanese patients. *Sci Rep*, 11:15853.
https://doi.org/10.1038/s41598-021-95002-w

ter Horst R, van Weert H, Loonen T, et al., 2021. Three-dimensional virtual planning in mandibular advancement surgery: soft tissue prediction based on deep learning. *J Cranio-Maxillofac Surg*, 49(9):775-782.
https://doi.org/10.1016/j.jcms.2021.04.001

Toepel-Sievers C, Fischer-Brandies H, 1999. Validity of the computer-assisted cephalometric growth prognosis VTO (Visual treatment objective) according to ricketts. *J Orofac Orthop*, 60(3):185-194.
https://doi.org/10.1007/BF01365265

Tong X, 2022. Three-dimensional shape space learning for visual concept construction: challenges and research progress. *Front Inform Technol Electron Eng*, 23(9):1290-1297.
https://doi.org/10.1631/FITEE.2200318

Vaz JM, Balaji S, 2021. Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. *Mol Divers*, 25(3):1569-1584.
https://doi.org/10.1007/s11030-021-10225-3

Wen YF, Wong HM, McGrath CP, 2019. Developmental shape changes in facial morphology: geometric morphometric analyses based on a prospective, population-based, Chinese cohort in Hong Kong. *PLoS ONE*, 14(6):e0218542.
https://doi.org/10.1371/journal.pone.0218542

Whiting PF, Rutjes AWS, Westwood ME, et al., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, 155(8):529-536.
https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Zhang X, Mei L, Yan XY, et al., 2019. Accuracy of computer-aided prediction in soft tissue changes after orthodontic treatment. *Am J Orthod Dentofacial Orthop*, 156(6):823-831.
https://doi.org/10.1016/j.ajodo.2018.11.021

Zhang XB, Hu Y, Chen W, et al., 2021. 3D brain glioma segmentation in MRI through integrating multiple densely connected 2D convolutional neural networks. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 22(6):462-475.
https://doi.org/10.1631/jzus.B2000381