



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE

ARTIFICIAL INTELLIGENCE IN ORTHOGNATHIC SURGERY: A
REVIEW OF DIAGNOSTIC SUPPORT, VIRTUAL PLANNING, AND
SOFT TISSUE OUTCOME PREDICTION

DOI

[https://doi.org/10.31435/ijitss.3\(51\).2026.6581](https://doi.org/10.31435/ijitss.3(51).2026.6581)

RECEIVED

24 June 2026

ACCEPTED

10 September 2026

PUBLISHED

15 September 2026

LICENSE



The article is licensed under a **Creative Commons Attribution 4.0 International License**.

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

ARTIFICIAL INTELLIGENCE IN ORTHOGNATHIC SURGERY: A REVIEW OF DIAGNOSTIC SUPPORT, VIRTUAL PLANNING, AND SOFT TISSUE OUTCOME PREDICTION

Kinga Bukala (Corresponding Author, Email: kinga.bukala@gmail.com)

Medical University of Lodz, Lodz, Poland

ORCID ID: 0009-0003-9587-4826

Szymon Klimaszewski

Private dental practice, Poland

ORCID ID: 0009-0001-3505-2797

Natalia Dejewska

Bieleński Hospital named after Rev. Jerzy Popiełuszko, Warsaw, Poland

ORCID ID: 0009-0006-4806-6780

Urszula Gadomska

Bieleński Hospital named after Rev. Jerzy Popiełuszko, Warsaw, Poland

ORCID ID: 0009-0005-8381-9955

Jowita Wiktoria Maksymiuk

Bieleński Hospital named after Rev. Jerzy Popiełuszko, Warsaw, Poland

ORCID ID: 0009-0004-3130-9913

Mateusz Onopiuk

Bieleński Hospital named after Rev. Jerzy Popiełuszko, Warsaw, Poland

ORCID ID: 0009-0003-9916-1292

Zuzanna Wiktoria Szumska

Powiat Hospital in Otwock, Otwock, Poland

ORCID ID: 0009-0009-5287-9946

Mikołaj Daniluk

Central Clinical Hospital, Warsaw, Poland

ORCID ID: 0009-0005-6458-373X

Jonasz Żuk

Central Clinical Hospital, Warsaw, Poland

ORCID ID: 0009-0001-7500-7982

ABSTRACT

Objectives. Orthognathic treatment of dentofacial deformities is highly experience-dependent, and the prediction of post-operative facial soft tissue change has historically been the least reliable step of surgical planning. This review synthesises the current applications, reported performance, and limitations of artificial intelligence (AI) across the orthognathic care pathway, with particular emphasis on diagnostic support, virtual surgical planning, and soft tissue outcome prediction.

Methods. A narrative review of peer-reviewed studies applying machine learning (ML), deep learning (DL), and related methods to diagnosis, patient qualification, cephalometric analysis, treatment planning, soft tissue prediction, and outcome evaluation in orthognathic surgery was conducted. Findings were organised thematically and appraised against commonly reported metrics.

Results. AI is most mature for automated cephalometric landmarking and for screening the need for surgery, where convolutional neural networks and ensemble classifiers approach expert-level agreement. For soft tissue prediction, DL methods that learn nonlinear bone-to-soft-tissue mappings achieve clinically relevant accuracy that exceeds traditional proportional and software-based simulations, with region-specific surface errors typically in the sub-millimetre to low-millimetre range. Applications in planning, outcome verification, and patient communication are emerging but less validated.

Conclusions. AI provides meaningful decision support across the orthognathic pathway and improves soft tissue simulation, yet evidence remains limited by small single-centre datasets, weak external validation, and restricted interpretability. Multicentre data, prospective validation, and explainable, clinician-supervised models are required before routine clinical adoption.

KEYWORDS

Artificial Intelligence, Deep Learning, Orthognathic Surgery, Cephalometric Analysis, Soft Tissue Prediction, Virtual Surgical Planning

CITATION

Bukała, K., Klimaszewski, S., Dejewski, N., Gadomska, U., Maksymiuk, J. W., Onopiuk, M., Szumska, Z. W., Daniluk, M., & Żuk, J. (2026). Artificial intelligence in orthognathic surgery: A review of diagnostic support, virtual planning, and soft tissue outcome prediction. *International Journal of Innovative Technologies in Social Science*, 3(51). [https://doi.org/10.31435/ijitss.3\(51\).2026.6581](https://doi.org/10.31435/ijitss.3(51).2026.6581)

COPYRIGHT

© The author(s) 2026. This article is published as open access under the **Creative Commons Attribution 4.0 International License (CC BY 4.0)**, allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

Orthognathic surgery is the surgical correction of skeletal dentofacial deformities and is typically delivered as a combined orthodontic–surgical treatment involving orthodontists and oral and maxillofacial surgeons. Its goals extend beyond achieving a stable, functional occlusion to restoring facial balance, airway function, and the patient’s psychosocial well-being. Because the discipline sits at the interface of dental occlusion, skeletal morphology, and facial soft tissue, decision-making must integrate clinical examination, photographs, dental models or intraoral scans, and two- and three-dimensional radiographic imaging. This multidimensional and multidisciplinary character makes the workflow accurate but also time-consuming, technique-sensitive, and dependent on the experience of the clinician.

Dentofacial deformities arise from disproportionate growth of the maxilla, the mandible, or both, and may present as anteroposterior discrepancies, vertical excess or deficiency, transverse problems, or facial asymmetry. They affect mastication, speech, and airway patency, and frequently carry a psychosocial burden related to facial appearance. Definitive correction usually follows a staged pathway: comprehensive records and diagnosis, pre-surgical orthodontic decompensation, surgical planning, the osteotomies themselves—most commonly Le Fort I maxillary osteotomy, bilateral sagittal split mandibular osteotomy, and genioplasty, alone or in combination—and post-surgical orthodontic finishing. Each stage generates quantitative data, including cephalometric measurements, three-dimensional surface scans, dental models, and tomographic volumes, that are well suited to computational analysis. This data-rich character helps to explain why AI has been investigated at almost every point of the pathway.

Several steps of this pathway are known to be vulnerable to inter- and intra-operator variability. The decision of whether a borderline patient should be treated with orthodontic camouflage or with combined

surgery is one example; cephalometric landmark identification, which underpins almost every quantitative analysis, is another, as manual landmarking is laborious and shows measurable variability between examiners. The most challenging task, however, is the prediction of how the facial soft tissue envelope will respond to the underlying skeletal movement. Conventional prediction tools embedded in commercial planning software have frequently relied on the assumption that hard- and soft-tissue displacements are linearly proportional, with operator-specified ratios for which the supporting evidence is limited. As a result, post-operative facial appearance has remained difficult to communicate to patients with confidence during planning.

Historically, clinicians attempted to anticipate the facial result using two-dimensional methods: manual or computerised profile prediction (profilograms) and video-imaging systems that morphed a lateral photograph according to the planned movements. These tools were useful for communication but were constrained by their two-dimensional nature and by the same proportional assumptions, and they could not faithfully represent asymmetry or the full three-dimensional envelope of the face. The shift towards three-dimensional imaging and data-driven prediction is, in large part, an attempt to overcome these long-standing limitations.

Over the last decade, artificial intelligence (AI) has advanced rapidly in medicine and dentistry, driven by greater computing power, the routine accumulation of digital clinical data, and progress in machine learning (ML) and deep learning (DL). In imaging-intensive specialties, convolutional neural networks (CNNs) and, more recently, transformer- and graph-based architectures have demonstrated strong performance in detection, segmentation, classification, and regression tasks. Dentistry has seen AI applied to caries detection, radiographic diagnosis, cephalometric landmarking, and the assessment of treatment need, and these capabilities map naturally onto the quantitative and image-rich nature of orthognathic surgery (Motamedian et al., 2025).

It is helpful to distinguish the principal families of AI methods encountered in this literature. Machine learning refers broadly to algorithms that learn patterns from data; classical examples used in orthognathic research include logistic regression, support vector machines, decision trees, and ensemble methods such as random forests and gradient boosting, typically applied to tabular inputs such as cephalometric measurements. Deep learning is a subfield in which multilayer neural networks learn hierarchical features directly from raw data, and convolutional neural networks are particularly effective for images, forming the workhorse of landmark detection and segmentation. More recent architectures extend this toolkit: transformer models capture long-range relationships and have been applied to planning framed as a regression task, graph neural networks operate naturally on the meshes and point clouds that represent facial and skeletal anatomy, and generative models, including diffusion models, can synthesise plausible post-operative images. Understanding which method has been used, and on what data, is essential to interpreting the performance that any study reports.

Against this background, a growing body of work has examined whether AI can support clinicians at each stage of orthognathic care: diagnosis and patient qualification, cephalometric analysis, treatment planning and virtual surgical planning (VSP), soft tissue outcome prediction, and post-operative outcome evaluation and communication. These studies are heterogeneous in their data, algorithms, reference standards, and outcome metrics, which makes it difficult for clinicians to judge where AI is genuinely useful and where claims outrun the evidence.

The aim of this review is therefore to provide a structured, clinically oriented synthesis of AI applications across the orthognathic surgery pathway, with particular emphasis on diagnostic support, virtual planning, and soft tissue outcome prediction. The review summarises the principal methods that have been applied, the performance that has been reported, and the limitations that currently constrain translation into routine practice, in order to help clinicians critically appraise the technology and to identify priorities for future research.

2. Methodology

This work is a narrative, thematically structured review rather than a systematic review or meta-analysis. The literature was identified through searches of the PubMed, Scopus, Embase, and Web of Science databases, supplemented by hand-searching the reference lists of relevant reviews and primary studies. Search terms combined concepts relating to the surgical domain (“orthognathic surgery,” “dentofacial deformity,” “cephalometric,” “virtual surgical planning,” “soft tissue prediction”) with terms describing the methods of interest (“artificial intelligence,” “machine learning,” “deep learning,” “neural network,” “convolutional neural network,” “transformer”).

Studies were considered relevant if they applied AI, ML, or DL methods to one or more of the following tasks in the context of orthognathic surgery or the closely related field of surgical orthodontics: diagnosis and the decision to operate, classification of skeletal pattern, automatic cephalometric landmarking and analysis, bone segmentation, treatment and surgical planning, prediction of skeletal or soft tissue outcomes, and post-operative outcome evaluation. Only English-language, peer-reviewed publications were included. Editorials and non-peer-reviewed sources were excluded. Because the purpose of the review is interpretive and clinically oriented, no formal risk-of-bias scoring or quantitative pooling was undertaken; instead, the methodological strength and limitations of the evidence are discussed narratively.

Findings were organised into five thematic domains corresponding to the stages of the orthognathic pathway (Figure 1). Within each domain, the principal AI approaches are described together with the performance metrics most frequently reported in the source studies. These include the mean radial error (MRE) and successful detection rate (SDR) for landmarking; accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC) for classification tasks; and the mean surface or landmark distance, expressed in millimetres, for soft tissue and skeletal prediction. Where studies reported numerical performance, representative values are cited to give the reader a concrete sense of current capability, while recognising that direct comparison across studies is limited by differences in datasets and reference standards.

Throughout, a deliberate distinction is drawn between technical performance, such as a sub-millimetre landmark error, and clinical usefulness, such as a demonstrated improvement in decision quality, planning time, or patient outcome. The two are related but not equivalent, and most of the available evidence speaks to the former. This framing is carried into the Discussion to identify where the field most needs to mature before routine adoption can be justified.

3. Results

AI applications were grouped into five domains spanning the orthognathic pathway: (1) diagnosis and patient qualification; (2) cephalometric analysis and anatomical landmarking; (3) treatment planning and virtual surgical planning; (4) soft tissue outcome prediction; and (5) outcome evaluation and patient communication. Figure 1 summarises representative AI tasks within each domain together with the cross-cutting requirements that condition their clinical use, and Table 1 provides an overview of representative studies and their reported performance.

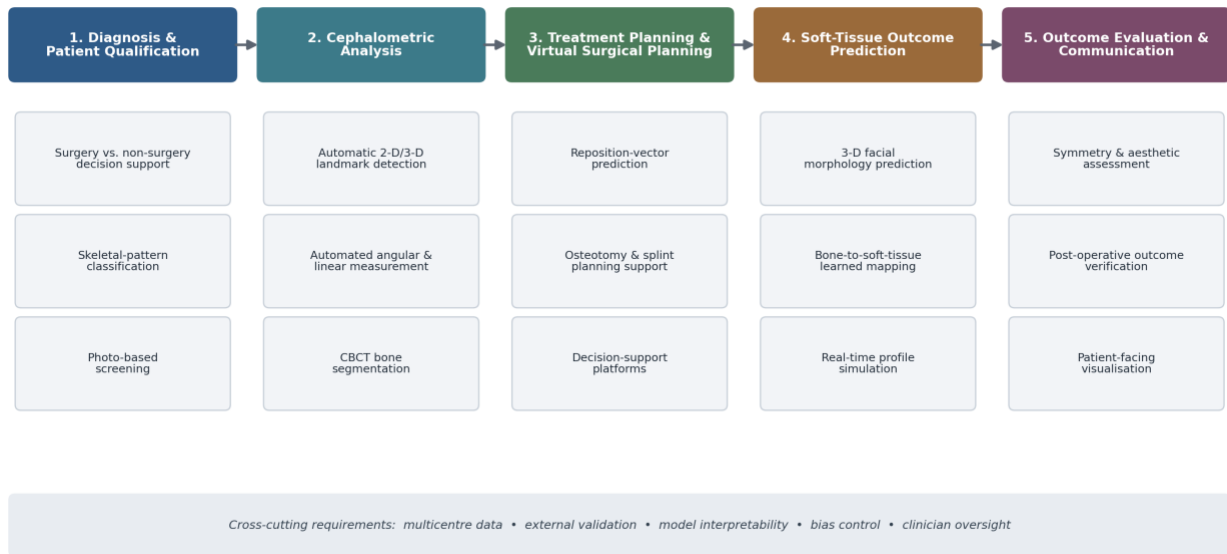


Fig. 1. Artificial intelligence applications across the orthognathic surgery workflow. Each stage lists representative AI-supported tasks; the lower band highlights cross-cutting requirements that condition safe clinical translation.

Table 1. Overview of representative artificial intelligence applications across the orthognathic surgery pathway and their reported performance.

Pathway stage	Representative AI task	Methods / models	Reported performance
Diagnosis & qualification	Surgery vs. non-surgery decision; need-for-surgery screening	Random forest, logistic regression, CNN (ResNet)	AUC up to ~0.94; sensitivity ~75–95.5%
Cephalometric analysis	Automatic 2-D/3-D landmark detection; bone segmentation	CNN, fully convolutional networks	DL landmark error ~0.5–1.9 mm; high SDR <2 mm
Treatment planning / VSP	Reposition-vector prediction; planning support	Transformer regression, ML frameworks	Landmark-level error ~5.4 mm; high diagnostic PPV/NPV
Soft tissue prediction	Bone-to-soft-tissue mapping; 3-D facial morphology	DL, geometric DL, morphable models, FEM-DL	Mean surface error ~1.0–1.6 mm (region-dependent)
Outcome evaluation / communication	Symmetry assessment; post-operative outcome prediction	Transfer learning, graph neural networks, diffusion models	Reported high agreement; early-stage validation

Note. PPV = positive predictive value; NPV = negative predictive value; SDR = successful detection rate; VSP = virtual surgical planning; FEM = finite element method; DL = deep learning. Values are representative figures drawn from the cited studies and are not directly comparable across datasets.

3.1. Diagnosis and Patient Qualification

The first point at which AI can support orthognathic care is the decision of whether a patient with a skeletal discrepancy should undergo surgery or be managed by orthodontic camouflage. This decision is multifactorial and, in borderline cases, strongly experience-dependent. Lee et al. (2022) developed ML classifiers for the surgery versus non-surgery decision in a sample of 196 skeletal Class III patients, training random forest and logistic regression models on cephalometric variables. Both models discriminated surgical from non-surgical cases with high separability, the random forest reaching an AUC of approximately 0.94 on the test set, and the features driving the prediction-overjet, the Wits appraisal, lower incisor inclination, and the Holdaway H angle-coincided with parameters that clinicians weigh heavily, lending the model face validity.

Beyond tabular classifiers, image-based DL has been used to determine the need for surgery directly from radiographs or photographs. In a synthesis of the field, Motamedian et al. (2025) reported that diagnostic models specifying the need for orthognathic surgery achieved sensitivities in the range of roughly 75% to 95.5%, with CNN architectures applied to lateral cephalograms performing well in screening tasks. Photograph-based CNNs have also been explored as low-cost screening tools, classifying soft tissue profiles that are likely to require surgery with accuracies approaching 90% in single-centre datasets. Collectively, these studies suggest that AI can act as a triage and second-opinion aid that standardises the qualification decision, reduces subjectivity in borderline cases, and supports less experienced clinicians, while the definitive treatment decision remains with the clinical team.

Related diagnostic tasks have also been automated. Classifiers have been trained to assign skeletal sagittal and vertical patterns from cephalometric data, to estimate treatment need, and to support the extraction-versus-non-extraction decision in orthodontics, tasks that frequently precede or accompany the surgical decision. A consistent finding across these studies is that the variables the models weight most heavily-sagittal discrepancy indices such as the Wits appraisal and ANB angle, the maxillomandibular ratio, overjet, and incisor inclination-mirror the parameters experienced clinicians use, which both increases confidence in the models and limits the risk that they encode spurious correlations (Lee et al., 2022; Motamedian et al., 2025). In a multidisciplinary clinic, such tools are most realistically deployed not to replace the decision but to flag borderline cases for joint orthodontic-surgical review and to document the reasoning behind a recommendation, contributing to consistency and auditability.

3.2. Cephalometric Analysis and Anatomical Landmarking

Cephalometric landmark identification is the quantitative foundation of orthognathic diagnosis and planning, yet manual landmarking is time-consuming and variable. Deep learning has produced the most mature AI capability in this domain. In a systematic review and meta-analysis of DL for cephalometric landmark detection, Schwendicke et al. (2021) found that CNN-based methods achieved consistently high accuracy on two-dimensional lateral radiographs, with the majority of studies reporting detection within the clinically accepted two-millimetre threshold for a large proportion of landmarks. The same review cautioned that most evidence concerned 2-D imagery, that three-dimensional data were comparatively sparse, and that the overall body of work carried a high risk of bias, underlining the need to demonstrate robustness and generalisability.

Fully automated, web-based cephalometric analysis pipelines have since been described (Kim et al., 2020), and independent evaluations of contemporary DL methods have reported performance comparable to human examiners (Hwang et al., 2021). Importantly, however, studies of commercial AI cephalometric services indicate that, at present, a human examiner's review is still required to correct occasional gross errors, so AI is best positioned as an accelerator of, rather than a replacement for, expert analysis. Three-dimensional landmarking on cone-beam computed tomography (CBCT) is more demanding than its 2-D counterpart because of the larger search space and memory requirements. Dot et al. (2022) demonstrated automatic 3-D cephalometric landmarking via deep learning, and joint approaches that segment the craniomaxillofacial skeleton while localising landmarks, such as the context-guided fully convolutional networks described by Zhang et al. (2020), report errors that for many landmarks fall below two millimetres. Accurate, reproducible, and rapid landmarking and segmentation are prerequisites for the planning and prediction tasks discussed below.

From a methodological standpoint, the field has progressed from heatmap-regression CNNs on two-dimensional cephalograms to encoder-decoder and detection architectures, and, for volumetric data, to three-

dimensional convolutional and reinforcement-learning approaches that localise landmarks within CBCT volumes. Joint segmentation-and-landmarking pipelines are attractive because the segmented skeleton is itself required for planning, allowing a single model to serve several downstream tasks (Zhang et al., 2020). Adjacent structures relevant to orthognathic surgery, such as the pharyngeal airway, have also been segmented automatically, which is clinically important because maxillomandibular movements alter airway dimensions and are relevant to obstructive sleep apnoea. Despite this progress, two practical points temper enthusiasm: accuracy degrades for anatomically ambiguous or low-contrast landmarks, and reported errors are sensitive to the choice of reference standard, so that apparently small differences between studies may reflect annotation protocols as much as algorithmic quality. The pragmatic consequence is that automated analysis is best used to produce a first-pass result that the clinician verifies, rather than an unsupervised measurement.

The clinical attraction of automated analysis is twofold: consistency and speed. Because an algorithm applies the same criteria to every image, it removes much of the intra- and inter-examiner variability inherent in manual landmarking, and it returns a complete analysis in seconds rather than the many minutes that careful manual tracing requires. Independent evaluations nonetheless caution that current tools are not infallible and that expert review remains necessary to detect occasional gross misplacements (Kim et al., 2020; Hwang et al., 2021). The most defensible workflow therefore couples automated speed with human judgement: the algorithm proposes, the clinician confirms or corrects, and the time saved is reinvested in interpretation and planning rather than in measurement.

3.3. Treatment Planning and Virtual Surgical Planning

Virtual surgical planning has become a standard component of contemporary orthognathic surgery, allowing osteotomies, segment repositioning, and surgical splints to be designed in a three-dimensional digital environment. AI is increasingly being applied to automate or accelerate the more repetitive and experience-dependent elements of this process. Knoop et al. (2019) described an ML framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery that, applied to Le Fort I planning, distinguished patients requiring surgery with high predictive values—reported positive and negative predictive values of approximately 87.5% and 98.3%, respectively—and generated patient-specific rather than generic predicted outcomes.

More recent work has framed surgical planning as a regression problem, in which the model predicts the displacement vectors required to reposition the bony segments. Transformer-based architectures that take three-dimensional cephalometric features as input have been used to predict reposition vectors, with reported landmark-level accuracy of the order of several millimetres, and analyses of these models have highlighted that facial soft tissue features contribute meaningfully to the prediction of skeletal movements. Explainable, decision-support platforms that combine reinforcement learning with interpretability components have also been proposed to support personalised orthodontic–orthognathic planning. While such systems remain experimental, they point towards a future in which AI proposes a draft plan that the surgeon reviews and refines, shortening planning time and standardising routine cases. Accurate planning is, however, inseparable from the question that matters most to patients: how the face will look afterwards.

In practice, virtual surgical planning comprises several sequential steps—segmentation of the maxilla, mandible, and dentition; establishment of the final occlusion; specification of segment movements; and design of the surgical splints or cutting guides—each of which is a candidate for automation. AI has been applied to the most repetitive of these, particularly segmentation and the proposal of movement vectors, with the aim of reducing the considerable time that fully manual planning demands. Because the predicted skeletal movement and the predicted soft tissue response are interdependent, the more advanced systems treat them jointly rather than in isolation, optimising bony repositioning with explicit reference to the resulting facial profile (Ma, Lian, et al., 2023). The realistic near-term role of these systems is to generate a defensible draft plan, with an associated outcome simulation, that the surgeon evaluates, edits, and approves—standardising routine cases and freeing expert time for complex ones, rather than planning autonomously.

The downstream products of planning—the final occlusal set-up and the surgical splints or patient-specific guides—are themselves increasingly digital and manufactured by computer-aided design and manufacturing. Automating the occlusal set-up is difficult, because it must respect cusp–fossa relationships and arch coordination, but partial automation of segmentation and of the geometric steps that follow can compress a planning session that might otherwise occupy a substantial part of a clinician’s day. When AI-proposed movements are exported into established planning environments, the resulting splints and guides can be

produced by the same additive-manufacturing pipelines already in clinical use, so that algorithmic assistance is layered onto, rather than substituted for, a validated workflow. This incremental model of adoption-automating discrete, checkable steps within an existing pipeline-is more likely to gain clinical acceptance than an end-to-end system that plans without human checkpoints.

3.4. Soft Tissue Outcome Prediction

Predicting the facial soft tissue response to skeletal movement is both the most clinically important and the most technically difficult task in orthognathic planning. Traditional prediction relied on proportional relationships between hard- and soft-tissue displacement, often implemented in commercial three-dimensional planning software. Validation studies of such software have reported clinically usable but imperfect accuracy: for example, evaluations of three-dimensional soft tissue prediction for Le Fort I osteotomy and for orthognathic cases using widely used commercial tools have shown regionally variable errors, with the lips and chin being particularly difficult to predict (Resnick et al., 2017; Elshebiny et al., 2019). A central limitation of these approaches is the assumption of a fixed, largely linear bone-to-soft-tissue ratio, which does not capture the nonlinear, region-specific, and patient-specific behaviour of facial tissues.

Deep learning offers a way to learn this nonlinear mapping directly from data. ter Horst et al. (2021) applied deep learning to predict soft tissue change after mandibular advancement surgery within a three-dimensional virtual planning context, while Lampen et al. (2022) combined deep learning with biomechanical modelling of facial tissue deformation to improve the physical plausibility of predictions. Geometric and bidirectional approaches have further advanced the field: Ma, Xiao, et al. (2023) used geometric deep learning to simulate post-operative facial appearance efficiently from planned bony movements, and Ma, Lian, et al. (2023) proposed a bidirectional model that jointly predicts facial and bony shapes, reflecting the coupled nature of skeletal and soft tissue change. Tanikawa and Yamashiro (2021) developed AI systems combining geometric morphometric methods with learning-based prediction to estimate three-dimensional facial morphology after orthognathic surgery and orthodontic treatment, explicitly moving beyond the proportional-ratio assumption embedded in conventional software.

A practical advantage of several recent systems is speed: methods that combine statistical shape models, principal component analysis, and neural networks can predict soft tissue outcomes from three-dimensional photographs in near real time, supporting interactive planning. Reported accuracies for DL-based soft tissue prediction are clinically relevant and generally superior to linear methods, with mean surface errors commonly around one millimetre and region-specific errors ranging from sub-millimetre values over rigid, well-constrained areas such as the nose to larger errors of approximately one and a half millimetres over mobile, anatomically complex regions such as the chin and lips. Table 2 contrasts the principal families of soft tissue prediction methods and their characteristic accuracy. Two caveats recur across this literature: the ground truth itself is imperfect, because the actual post-operative result reflects the executed rather than the planned movement and is influenced by swelling, healing, and time; and most models are trained and tested on single-centre data, leaving their generalisability to other populations and imaging set-ups largely unproven.

The methods used to learn the bone-to-soft-tissue relationship are diverse. Some combine statistical or morphable shape models with principal component analysis to constrain the space of plausible faces before a neural network predicts the deformation, which improves stability when training data are limited. Others operate directly on dense surface meshes or point clouds using geometric deep learning, preserving the spatial structure of the face, while finite-element and other biomechanical formulations introduce physical priors so that predicted deformations remain anatomically credible (Lampen et al., 2022). A recurring theme is that accuracy is markedly region-dependent: predictions over rigid, well-supported regions such as the forehead and nasal dorsum are typically more accurate than those over mobile, soft regions such as the lips and chin, where tissue thickness, muscle tone, and surgical detail exert greater influence. This pattern is observed with both conventional software and AI methods, but the better deep learning systems narrow the error in the difficult regions relative to linear approaches. For the clinician, the key advance is conceptual as much as numerical: by abandoning the fixed-ratio assumption in favour of a learned, patient-specific, nonlinear mapping, AI aligns the prediction more closely with the biology of facial change.

Quantitative results reported for these systems, while not directly comparable, give a consistent picture. Machine learning frameworks applied to computer-assisted planning have reported mean soft tissue prediction errors of approximately one millimetre, alongside high positive and negative predictive values for the underlying surgical decision (Knoops et al., 2019). Deep learning systems built on three-dimensional

photographs or surface scans report comparable mean errors, with the best accuracy over stable regions falling well below one millimetre and the largest errors concentrated, as expected, around the lips and chin (ter Horst et al., 2021; Tanikawa & Yamashiro, 2021). Crucially, several studies show that AI predictions lie closer to the actual post-operative scan than either a generic mean face or a conventionally simulated result, which is the comparison that matters clinically. What remains underexplored is prospective validation: almost all of this evidence is retrospective, and the predictive value of these tools in routine, real-time planning has yet to be demonstrated at scale.

Table 2. Comparison of principal approaches to facial soft tissue outcome prediction in orthognathic surgery.

Approach	Principle	Characteristics and reported accuracy
Proportional / commercial software	Fixed, largely linear bone-to-soft-tissue ratios in 3-D planning packages	Clinically usable but region-variable; lips and chin least accurate (Resnick et al., 2017; Elshebiny et al., 2019)
Biomechanical / finite-element with DL	Physics-based tissue deformation accelerated or refined by learning	Improved physical plausibility of deformation (Lampen et al., 2022)
Deep learning bone-to-soft-tissue mapping	Neural networks learn nonlinear displacement mapping from data	Mean surface error ~1 mm; outperforms linear methods (ter Horst et al., 2021; Ma, Lian, et al., 2023)
Geometric / morphometric deep learning	Geometric DL or morphometrics predict 3-D facial morphology	Patient-specific 3-D morphology prediction (Tanikawa & Yamashiro, 2021; Ma, Xiao, et al., 2023)

3.5. Outcome Evaluation and Patient Communication

AI also contributes after planning, in the objective evaluation of results and in communication with patients. Facial symmetry, a key aesthetic outcome, has been assessed using transfer-learning models that compare pre- and post-operative facial morphology (Lin et al., 2021), offering a reproducible alternative to subjective rating. Generative approaches have begun to address outcome verification directly: Kim et al. (2025) combined graph neural networks with diffusion models to predict post-operative lateral cephalograms, illustrating how generative AI may be used both to anticipate and to audit surgical results. Perhaps the most immediate clinical value, however, lies in communication. A credible, patient-specific visualisation of the likely facial outcome supports informed consent, helps to align patient expectations with what surgery can realistically achieve, and can strengthen shared decision-making-provided that the limitations and uncertainty of the prediction are made explicit to the patient.

AI has additionally been used to quantify facial aesthetics and attractiveness, providing a reproducible, if necessarily reductive, complement to clinician and patient ratings, and to support the objective auditing of results against the surgical plan. Such capabilities have value beyond research: a documented, quantitative comparison of planned and achieved outcomes supports clinical governance and may have medico-legal utility, while an honest, uncertainty-aware visualisation of the likely facial result is a powerful aid to shared decision-making. The same tools carry a risk, however, if patients interpret a simulated face as a guarantee; clear communication that the prediction is probabilistic, and that biological variation, healing, and surgical execution all influence the final result, is therefore an ethical as well as a clinical necessity.

The practical value of these capabilities is clearest in the consultation itself. Near-real-time simulation allows the clinician to show a patient how alternative surgical movements would alter the face, turning an abstract treatment plan into something the patient can see and respond to, and supporting a more genuinely shared decision. Used well, this can improve the calibration of patient expectations, which is among the strongest determinants of satisfaction after orthognathic surgery. Used poorly-without conveying uncertainty,

or by presenting an idealised single image-it risks overpromising. The responsible deployment of outcome simulation therefore depends as much on how the prediction is communicated as on how accurate it is, a point that links the technical literature directly to the ethics of consent.

3.6. Predicting Skeletal Outcomes, Stability, and Complications

A further domain, distinct from soft tissue prediction, concerns the anticipation of skeletal outcomes and their stability. Generative models have been used to predict the post-operative lateral cephalogram directly from pre-operative records, providing a hard-tissue counterpart to facial soft tissue simulation; Kim et al. (2025) combined graph neural networks with diffusion models for this purpose, illustrating how the predicted skeletal position can be visualised before surgery. Such predictions are relevant not only for planning but also for setting expectations about the radiographic result and for teaching, since they make the mechanical consequences of a planned movement explicit.

Two clinically important questions remain comparatively underserved by AI: long-term skeletal stability and the risk of relapse, and the prediction of surgical complications. Scoping reviews note that, although diagnosis, planning, and soft tissue prediction have attracted most attention, complication assessment has also been explored, albeit in fewer studies and with smaller samples (Motamedian et al., 2025). Identifying which patients are at higher risk of relapse after large advancements, or of specific peri-operative complications, would have clear clinical value, but it demands longitudinal data that are harder to assemble than the cross-sectional images used for landmarking. As such datasets accumulate, outcome- and risk-prediction models represent a natural and high-value extension of the work reviewed here.

4. Discussion

4.1. Clinical Implications and Current Maturity

Taken together, the evidence indicates that AI has reached different levels of maturity across the orthognathic pathway. Automated cephalometric landmarking and the screening of surgical need are comparatively well developed, with multiple studies reporting accuracy approaching that of human experts. Treatment planning, soft tissue prediction, and outcome verification are advancing quickly but rest on a thinner and more heterogeneous evidence base. The clearest clinical gains are in efficiency and standardisation: AI can shorten laborious tasks such as landmarking and segmentation, reduce inter-operator variability, provide objective decision support in borderline qualification cases, and generate patient-specific outcome simulations that improve communication and consent.

Beyond the individual clinic, AI may influence access to expertise. Tasks that currently require a highly experienced orthodontist or surgeon-reliable cephalometric analysis, sound qualification of borderline cases, and credible outcome simulation-could, if validated tools become available, be supported in settings where such expertise is scarce. This prospect is attractive for equity of care but raises the stakes for generalisability and bias, because a model trained predominantly on one population may perform less well on the very groups that stand to benefit most. It also reframes professional education: clinicians will increasingly need to understand the assumptions, inputs, and failure modes of the tools they supervise, making critical appraisal of AI a core clinical competency rather than a specialist interest.

4.2. Limitations of the Evidence and Technical Challenges

These promising results must, however, be read against substantial methodological limitations. Many studies are retrospective and based on small, single-centre datasets, with reference standards derived from a limited number of expert annotators. Systematic appraisals of the field have repeatedly highlighted heterogeneity in data, algorithms, and outcome measures, together with a high risk of bias and a scarcity of external validation (Schwendicke et al., 2021; Motamedian et al., 2025). For soft tissue prediction in particular, defining a reliable ground truth is intrinsically difficult, because the recorded post-operative result reflects the movement actually achieved rather than the movement planned, and is further modified by oedema, healing, and the passage of time. Reported accuracy figures should therefore be interpreted as indicative rather than definitive, and direct comparison across studies is rarely justified.

Several technical challenges follow from these limitations. The annotation burden for three-dimensional data is considerable, and large, well-curated, publicly available datasets are scarce, which constrains both training and benchmarking. Models trained on a single population or imaging set-up are vulnerable to domain shift when applied elsewhere, and class imbalance is common in surgical-decision tasks where surgical cases may be under-represented. Three-dimensional tasks remain harder than their two-dimensional equivalents because of higher dimensionality and computational cost. Addressing these issues will require multicentre data sharing, standardised reporting in line with frameworks for prediction-model and AI studies, and rigorous external and ideally prospective validation that measures clinical impact rather than only technical accuracy.

4.3. Interpretability, Ethics, and Clinical Integration

Interpretability and trust represent a further barrier to adoption. Many high-performing models are effectively black boxes, which is problematic in a surgical context where clinicians are accountable for decisions and patients are entitled to an explanation. Evidence that commercial AI cephalometric tools still require human correction reinforces the case for a human-in-the-loop model in which AI proposes and the clinician disposes. Explainable AI techniques, calibrated uncertainty estimates, and transparent reporting of failure modes are likely to be prerequisites for regulatory clearance and clinical acceptance. Alongside these technical questions sit ethical and governance considerations: the risk of biased performance across ethnic groups and facial morphologies, the protection of sensitive imaging data, informed consent for AI-assisted planning, and clear lines of clinical accountability.

Regulatory status is an increasingly important consideration. Software that informs surgical decisions is generally treated as a medical device, and clinical deployment therefore entails formal evaluation, clearance, and ongoing post-market surveillance, together with documentation of the training data, intended use, and performance boundaries. Many of the systems described in the research literature are prototypes that have not undergone such evaluation, and the gap between a promising research result and a regulated clinical tool is substantial. Transparent reporting aligned with recognised guidelines for prediction-model and AI studies would ease this transition by making validation, limitations, and intended use explicit from the outset.

4.4. Future Directions

Looking ahead, the most valuable developments are likely to be multimodal and integrative. Fusing complementary data-lateral cephalograms, CBCT volumes, intraoral scans, and three-dimensional facial photographs-should yield more robust predictions than any single modality, and embedding validated models within existing VSP software would place decision support directly in the surgeon's workflow. Generative models offer a route to richer, patient-specific outcome simulation, while extending prediction beyond immediate morphology to longer-term stability, relapse, and complication risk would address questions that matter greatly to patients but remain poorly served by current tools. Realising this potential depends less on novel architectures than on disciplined validation, data sharing, and integration into clinical practice.

Two technical trajectories deserve particular attention. The first is multimodal fusion: because skeletal position, dental occlusion, and the soft tissue envelope each carry only partial information, models that jointly ingest cephalograms, CBCT volumes, intraoral scans, and three-dimensional facial photographs are likely to predict outcomes more robustly than any single source, and early multimodal frameworks already point in this direction. The second is generative modelling. Diffusion and related generative methods can synthesise full post-operative images rather than only point estimates and, combined with calibrated uncertainty, could present clinicians and patients with a range of plausible outcomes rather than a single deterministic face (Kim et al., 2025). If paired with rigorous validation, such approaches may also help to model rarer presentations for which training data are scarce, although they introduce their own risk of producing convincing but inaccurate images.

Successful translation will ultimately depend on factors beyond accuracy. Models must be integrated into the software environments clinicians already use, validated on the imaging hardware of the deploying centre, and supported by clear protocols specifying when an AI output should be overridden. The economic case-whether the time saved and the consistency gained justify the cost of acquisition, validation, and maintenance-remains largely unquantified, as do the training needs of clinicians who must supervise these tools competently. Addressing these implementation questions is as important as further algorithmic

refinement, and their neglect is a plausible reason why strong reported accuracy has so far translated into limited routine use.

Finally, the limitations of this review should be acknowledged. It is a narrative rather than a systematic synthesis; studies were selected for their relevance and illustrative value rather than through an exhaustive, pre-registered protocol, and no formal quality scoring or quantitative pooling was performed. The performance figures cited are drawn from heterogeneous studies and are intended to convey the order of magnitude of current capability rather than to support direct comparison. The field is also moving quickly, so specific results will date; the broad conclusions regarding relative maturity, evidentiary gaps, and translational requirements are, however, likely to remain valid in the near term.

5. Conclusions

Artificial intelligence offers clinically meaningful support across the orthognathic surgery pathway. It is most mature in automated cephalometric landmarking and in screening the need for surgery, where performance approaches expert agreement, and it is advancing rapidly in treatment planning, soft tissue prediction, and outcome evaluation. For the prediction of post-operative facial soft tissue-historically the least reliable step of planning-deep learning methods that learn nonlinear bone-to-soft-tissue relationships achieve sub-millimetre to low-millimetre accuracy and consistently outperform traditional proportional and software-based simulations, while also enabling near-real-time, patient-specific visualisation that strengthens patient communication.

At present, however, AI should be regarded as a decision-support adjunct rather than a replacement for clinical judgement. The supporting evidence is constrained by small single-centre datasets, weak external validation, imperfect ground truth, and limited interpretability. Before routine clinical adoption, the field requires multicentre and public datasets, standardised reporting, prospective validation of clinical impact, and explainable, clinician-supervised models. If these conditions are met, AI is well placed to make orthognathic care more efficient, more consistent, and more clearly communicated to the patients it serves.

Conflicts of Interest: No conflicts of interest to declare.

REFERENCES

1. Dot, G., Schouman, T., Chang, S., Rafflenbeul, F., Kerbrat, A., Rouch, P., et al. (2022). Automatic 3-dimensional cephalometric landmarking via deep learning. *Journal of Dental Research*, 101(11), 1380–1387.
2. Elshebiny, T., Morcos, S., Mohammad, A., Queresby, F., & Valiathan, M. (2019). Accuracy of three-dimensional soft tissue prediction in orthognathic cases using Dolphin three-dimensional software. *Journal of Craniofacial Surgery*, 30(2), 525–528.
3. Hwang, H.-W., Moon, J.-H., Kim, M.-G., Donatelli, R. E., & Lee, S.-J. (2021). Evaluation of automated cephalometric analysis based on the latest deep learning method. *The Angle Orthodontist*, 91(3), 329–335.
4. Kim, H., Shim, E., Park, J., Kim, Y.-J., Lee, U., & Kim, Y. (2020). Web-based fully automated cephalometric analysis by deep learning. *Computer Methods and Programs in Biomedicine*, 194, 105513.
5. Kim, I. H., Jeong, J., Kim, J. S., Lim, J., Cho, J. H., Hong, M., et al. (2025). Predicting orthognathic surgery results as postoperative lateral cephalograms using graph neural networks and diffusion models. *Nature Communications*, 16(1), 2586.
6. Knoop, P. G. M., Papaioannou, A., Borghi, A., Breakey, R. W. F., Wilson, A. T., Jeelani, O., et al. (2019). A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. *Scientific Reports*, 9, 13597. <https://doi.org/10.1038/s41598-019-49506-1>
7. Lampen, N., Kim, D., Fang, X., Xu, X., Kuang, T., Deng, H. H., et al. (2022). Deep learning for biomechanical modeling of facial tissue deformation in orthognathic surgical planning. *International Journal of Computer Assisted Radiology and Surgery*, 17(5), 945–952.
8. Lee, H., Ahmad, S., Frazier, M., Dundar, M. M., & Turkkahraman, H. (2022). A novel machine learning model for class III surgery decision. *Journal of Orofacial Orthopedics / Fortschritte der Kieferorthopädie*. Advance online publication. <https://doi.org/10.1007/s00056-022-00421-7>

9. Lin, H.-H., Chiang, W.-C., Yang, C.-T., Cheng, C.-T., Zhang, T., & Lo, L.-J. (2021). On construction of transfer learning for facial symmetry assessment before and after orthognathic surgery. *Computer Methods and Programs in Biomedicine*, 200, 105928. <https://doi.org/10.1016/j.cmpb.2021.105928>
10. Ma, L., Lian, C., Kim, D., Xiao, D., Wei, D., Liu, Q., et al. (2023). Bidirectional prediction of facial and bony shapes for orthognathic surgical planning. *Medical Image Analysis*, 83, 102644.
11. Ma, L., Xiao, D., Kim, D., Lian, C., Kuang, T., & Liu, Q. (2023). Simulation of postoperative facial appearances via geometric deep learning for efficient orthognathic surgical planning. *IEEE Transactions on Medical Imaging*, 42(2), 336–345. <https://doi.org/10.1109/TMI.2022.3180078>
12. Motamedian, S. R., Mohaghegh, S., Niazmand, M., Mohammad-Rahimi, H., Ahmadi, N., Yaseri, M., & Haeri Boroojeni, H. S. (2025). Application of artificial intelligence in orthognathic surgery: A scoping review. *BioMed Research International*, 2025, 8284581. <https://doi.org/10.1155/bmri/8284581>
13. Resnick, C. M., Dang, R. R., Glick, S. J., & Padwa, B. L. (2017). Accuracy of three-dimensional soft tissue prediction for Le Fort I osteotomy using Dolphin 3D software: A pilot study. *International Journal of Oral and Maxillofacial Surgery*, 46(3), 289–295. <https://doi.org/10.1016/j.ijom.2016.10.016>
14. Schwendicke, F., Chaurasia, A., Arsiwala, L., Lee, J.-H., Elhennawy, K., Jost-Brinkmann, P.-G., Demarco, F., & Krois, J. (2021). Deep learning for cephalometric landmark detection: Systematic review and meta-analysis. *Clinical Oral Investigations*, 25(7), 4299–4309. <https://doi.org/10.1007/s00784-021-03990-w>
15. Tanikawa, C., & Yamashiro, T. (2021). Development of novel artificial intelligence systems to predict facial morphology after orthognathic surgery and orthodontic treatment in Japanese patients. *Scientific Reports*, 11(1), 15853. <https://doi.org/10.1038/s41598-021-95002-w>
16. ter Horst, R., van Weert, H., Loonen, T., Bergé, S., Vinayahalingam, S., & Baan, F. (2021). Three-dimensional virtual planning in mandibular advancement surgery: Soft tissue prediction based on deep learning. *Journal of Cranio-Maxillofacial Surgery*, 49(9), 775–782. <https://doi.org/10.1016/j.jcms.2021.04.001>
17. Zhang, J., Liu, M., Wang, L., Chen, S., Yuan, P., Li, J., et al. (2020). Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization. *Medical Image Analysis*, 60, 101621.