BIO Web of Conferences **214**, 01007 (2026) https://doi.org/10.1051/bioconf/202621401007
ICBBS 2025

Research Progress and Application of Artificial Intelligence in Cephalometric Analysis

Qiushi Tian[1*]

[1]School of Medicine, Nankai University, Tianjin, China

Abstract. Cephalometric analysis is a vital diagnostic tool in orthodontics and craniofacial surgery. It provides precise information on the relationship between the skull and teeth to guide treatment planning. Unlike traditional manual annotation methods, artificial intelligence (AI) approaches not only enable more efficient annotation but also mitigate result variability caused by differences in operator skill and experience. AI, particularly deep learning, excels at rapidly processing image while achieving accuracy comparable to that of experts. Recent studies confirm the effectiveness of convolutional neural networks and related architectures in automated landmark detection, structural segmentation, and intelligent measurement-based diagnostic support. When trained on large-scale annotated datasets, these models extract stable anatomical and pathological features. This enhances result reproducibility, reduces analysis time, and improves clinical efficiency. However, there are also some challenges persist, including inconsistent annotation protocols, limited model generalization across populations and imaging devices, and a lack of large-scale external validation. This review aims to summarize AI applications in cephalometric analysis, evaluate existing limitations, and explore future directions for establishing standardized, clinically reliable applications.

1 Introduction

Cephalometric analysis plays a key role in orthodontics and craniofacial surgery. As a two-dimensional image analysis tool, it can be used to evaluate craniofacial structural relationships, assist in diagnosing malocclusions, and develop individualized treatment plans. Its fundamental principle involves marking key anatomical landmarks on cephalometric X-ray images precisely to calculate parameters such as angles, distances, and proportions. This quantifies the relationships between structures like the maxilla, mandible, teeth, and soft tissues [1]. However, traditional cephalometric analysis relies on clinicians localizing anatomical landmarks manually, which is a time-consuming and inefficient process. Furthermore, this conventional approach still depends heavily on the clinician's expertise and proficiency, leading to variations in landmark annotation among different practitioners. Even experienced researchers face challenges in maintaining accuracy and consistency during big data analysis. This inconsistency undermines the reproducibility and clinical reliability of analyses. Consequently, automating cephalometric measurement to enhance efficiency and accuracy has become a critical focus of research and development in orthodontics.

In recent years, the rapid advancement of artificial intelligence (AI) and the application of deep learning (DL) technology have opened new avenues for automated cephalometric analysis. Deep learning automatically extracts image features through large-scale data learning and has demonstrated exceptional performance in anatomical landmark detection. It is gradually replacing traditional rule- or template-based methods, and among these models, convolutional neural networks (CNNs) are currently the most widely used. They progressively extract image information through multi-layer structures, facilitating data processing via neural networks and automatic data learning. This process simulates the human brain's image recognition process [2]. As research advances, more AI programs are being developed to automatically identify anatomical landmarks on cephalometric films. These systems operate rapidly, are highly repeatable, and can reduce human error. Studies indicate that AI models can locate certain landmarks with up to 98% accuracy, approaching the level of accuracy of manual annotation while significantly reducing analysis time and enhancing overall efficiency. The transition from manual to automated annotation boosts productivity, standardizes cephalometric applications, and enhances their controllability in clinical settings [3].

Despite significant progress, the clinical adoption of AI in cephalometric analysis faces challenges such as inconsistent data annotation standards, limited model generalization capabilities, and the interpretability of results. There are also difficulties in integrating AI with healthcare systems. Therefore, a systematic review of the current research landscape, key technological pathways, and clinical application outcomes is essential to analyze major existing issues and explore future development trends in this field. This review summarizes the latest advances in AI-based cephalometric analysis, covering core tasks such as

* Corresponding author: 2212704@mail.nankai.edu.cn

automatic landmark detection, structural segmentation, intelligent measurement, and auxiliary diagnosis. The review discusses the technical characteristics of deep learning models, their training and evaluation frameworks, and current challenges and future development directions.

2 Fundamentals of Cephalometric Analysis

Cephalometry is one of the most critical radiographic analysis methods in orthodontics. It quantitatively evaluates craniofacial skeletal and dental structures through radiographic images, providing a basis for diagnosis and treatment planning. In clinical practice, two-dimensional cephalometric X-rays (lateral and frontal views) are commonly used to evaluate sagittal and transverse craniofacial relationships; however, these images have limitations, such as structural overlap and distortion [4]. Advances in imaging technology have led to the application of three-dimensional imaging techniques (CT, CBCT, and MRI) to the analysis of complex malocclusions. Among these techniques, cone-beam computed tomography (CBCT) is considered a valuable supplement to two-dimensional imaging due to its lower radiation dose and ability to provide high-resolution three-dimensional structures [5].

Regarding specific methodologies, clinically prevalent analysis systems include Steiner, McNamara, and Ricketts. These approaches rely on identifying specific craniofacial anatomical landmarks, such as S, N, A, B, Pog, U1, and L1. Based on these landmarks, a series of key indicators can be derived. For example, SNA, SNB, and ANB reflect the anterior-posterior relationship of the jaws, FMA indicates the inclination of the mandibular plane, and IMPA assesses the axial inclination of the lower incisors. These indicators form the basis for evaluating craniofacial skeletal relationships and dental alignment characteristics [6].

The complete cephalometric analysis workflow involves four steps: image acquisition, landmark annotation, geometric measurement, and diagnostic interpretation. Manual annotation has long been regarded as the gold standard, but it is time-consuming and prone to interobserver variability [7]. However, with the advancement of digital technology and artificial intelligence, automated measurement has emerged as an alternative, enhancing efficiency and consistency while reducing human bias. This shift has laid the technical foundation for subsequent AI-based cephalometric studies and opened new pathways for clinical diagnosis and personalized treatment planning.

3 Fundamental Principles and Model Training Process of Medical Imaging AI

The core of medical imaging AI is data-driven feature learning. Through training on large-scale annotated images, models can extract stable anatomical and pathological features. These features can then be used to translate clinical tasks into various predictive outputs. In orthodontics and craniofacial image analysis, these tasks include regression calculations of continuous variables (e.g., SNA and SNB angles), the automated detection and localization of anatomical landmarks, and the segmentation of structures such as craniofacial bones, teeth, and airways. They also include multi-task joint modelling.

Deep learning excels at processing high-dimensional images and automatically learning hierarchical features. It generally outperforms traditional machine learning in both accuracy and efficiency. The choice of model architecture significantly impacts task performance. Convolutional neural networks (CNNs) and their variants, such as ResNets and UNets, are widely used for 2D and 3D tasks. For 2D lateral radiographs, coarse-to-fine strategies based on heatmap regression are commonly used to ensure landmark localization stability. For 3D CT/CBCT, approaches such as 3D U-Net, Mask R-CNN, and two-stage strategies combining initial detection with fine-grained refinement are frequently used. These approaches explicitly model local dependencies to enhance robustness. A representative study achieved localization of 105 craniofacial landmarks on CBCT with an average error of approximately 1.38 ± 0.95 mm. This result demonstrates the feasibility of 3D methods while highlighting challenges in internal and external generalization [8].

The model training process typically revolves around three phases: data governance, model construction, and validation. During data preparation, tasks include de-identification, uniform positioning and scaling, resampling, and intensity normalization, as well as establishing multi-expert annotation and quality control. In the modelling and validation phase, the training, validation, and test sets must be partitioned appropriately and supplemented with independent external data for verification. This systematic workflow minimizes data leakage and result bias, ensuring model robustness and transferability in clinical settings.

4 Applications of AI in Cephalometry

4.1 Automated Landmark Detection

Automated landmark detection is the most critical AI task in cephalometric analysis. It aims to precisely localize anatomical key points using deep learning models. The outputs are typically landmark coordinates, and the common evaluation metrics are Mean Radial Error (MRE) and Success Detection Rate (SDR) at various thresholds. Studies indicate that the accuracy of automatic landmark recognition on two-dimensional (2D) lateral cephalometric radiographs is approaching clinical thresholds. The SDR is approximately 79% within a 2 mm range and 90% within a 3 mm range; however, high inter-study heterogeneity and low certainty of evidence suggest cautious interpretation and the need for robust external validation [9].

Beyond 2D imaging, AI is also extensively applied in three-dimensional (3D) cephalometric analysis. Using 3D Mask R-CNN combined with local-to-global dependency modelling on CBCT data achieved an MRE of approximately 1.38 ± 0.95 mm across 105 landmarks.

This demonstrates the effectiveness of transitioning from global candidates to local refinement while incorporating topological prior strategies. However, it also reveals sensitivity to sample size and protocol. Another approach, heatmap regression and configuration networks based on coarse-to-fine structure, reported MREs near 1 mm and high SDRs. However, this approach exhibited greater errors for tooth-specific or contour-type points, indicating location-dependent challenges [8]. Automatic landmark detection establishes the technical foundation for subsequent "point-to-line-to-angle" quantitative analysis. However, the heterogeneity of the current evidence and the sensitivity of 3D tasks necessitate manual verification prior to critical decisions [10].

4.2 Structural Segmentation and Facial Feature Recognition

Facial structure segmentation provides clear geometric boundaries and voxel-level structural priors for bones, teeth, airways, and soft tissues. This stabilizes subsequent point annotation, enhances the repeatability of angle and distance measurements, and supports preoperative orthognathic and airway assessments alongside 3D simulation [11]. Within 3D pipelines, a cascading approach involving detection, segmentation, reconstruction, and local refinement is commonly employed. Within this framework, incorporating local dependencies or topological relationship modelling significantly improves point stability and overall consistency while reducing biases caused by anatomical complexity [12].

In limited field-of-view (FOV) or CBCT scout scenarios, partial anatomical structures and reference planes may be missing, which poses a challenge to measurement integrity. To address this issue, the open-source tool CEPHOSS can estimate and complete critical measurements (e.g., SNA, SNB, and ANB) when structures or points are missing. CEPHOOSS performs consistency checks using RMSE/MAE, ICC, and Bland–Altman plots. This reduces the need for retakes and radiation exposure while maintaining reading reliability [13]. This approach is grounded in structural priors and visibility modelling, leading to robust measurements and providing a viable engineering pathway to address real-world imaging constraints.

4.3 Intelligent Measurement and Diagnostic Assistance

The intelligent measurement pathway comprises two modes: marker-based quantification and end-to-end prediction. The former automatically marks landmarks and calculates angles or linear metrics, such as SNA, SNB, ANB, FMA, IMPA, U1-NA, and Co-A. It then evaluates these metrics against clinical reference ranges using MAE, ICC, and Bland–Altman plots to make diagnoses, thereby mirroring traditional orthodontic workflows. The end-to-end approach directly regresses or classifies key measurements against malocclusion categories and can be applied to extraction and orthognathic surgery decisions and facial prediction [14]. Compared to point-to-quantity approaches, end-to-end methods avoid pointwise error propagation, but they face challenges in interpretability and out-of-distribution generalization. Current evidence suggests that models incorporating multi-source information, such as cephalometric measurements, can predict orthognathic surgery indications and extraction plans with 78–97% accuracy, indicating their potential value in borderline cases. However, sample coverage and generalization capabilities require enhancement. Concurrently, image-to-image generative methods (e.g., cGAN) and virtual dental setup tools integrate quantification, aesthetics, and patient experience to predict the impact of cephalometric and facial changes. This provides quantitative support for treatment planning and doctor-patient communication [11].

Multi-platform comparisons reveal systematic discrepancies in measurements such as ANB, FMA, IMPA, U1-NA, and Co-A between commercial AI systems (WebCeph, WeDoCeph, and CephX) and manual digital tracing (NemoCeph). Sources of discrepancy include training data distribution, point definition, and post-processing strategies. This indicates that centralized calibration is essential prior to deployment across platforms, institutions, and populations. Physician final review must be retained for critical conclusions [6]. Systematic reviews indicate that automated workflows can reduce processing time per case from 15 minutes to a few seconds or minutes, making them suitable for routine follow-ups and large-scale screening; however, there are unsuitable scenarios. For example, critical measurements such as ANB, FMA, and IMPA require secondary verification and physician final review before high-risk decisions like tooth extraction or orthognathic surgery. This creates a collaborative, closed-loop system where AI provides preliminary information combined with physician final review [7].

5 Challenges and Future Development Directions

The most critical current issue is inconsistent standards and criteria. There is a lack of unified norms in defining landmarks, measuring angles or linear dimensions, and setting evaluation thresholds. For instance, SDR values fluctuate across studies (e.g., 2, 2.5, or 3 mm). Many studies use different point sets and interpretation criteria, which leads to poor comparability across studies and high heterogeneity in meta-analyses. Both systematic and umbrella reviews identify inconsistent standards as a primary source of bias and caution against equating a single threshold (e.g., 2 mm) with "clinical equivalence" [10]. Additionally, there are ongoing issues regarding insufficient external validation and inadequate reporting. Many studies only report results on in-house test sets, lacking cross-center external testing and out-of-time validation. This leaves the models' generalization capabilities in real clinical settings unclear. The literature also indicates high risks of bias in subject selection and reference standards, leaving generalization in the real world uncertain. Recommendations include standardizing training, validation, and testing phases, along with external validation, in study design and publication. There should also be transparent reporting of data sources, annotation processes, and failed samples. Platform variability and reproducibility issues pose significant barriers to clinical implementation. Multi-platform comparisons reveal systematic discrepancies between commercial AI and manual digital tracings across key metrics, including ANB, FMA, IMPA, U1-NA, and Co-A. These discrepancies stem from variations in training data distribution, point definitions, post-processing, and calibration strategies, implying that readings across different centers or software cannot be directly compared. This necessitates centralized calibration and ongoing monitoring for clinical deployment [6].

Future research must establish unified consensus standards and open benchmarks to standardize marker point definitions, measurement checklists, and evaluation systems, including MRE, SDR at different thresholds, MAE for angular and linear measurements, ICC, and Bland–Altman plots, to reconstruct a comparable evidence framework. For 2D tasks, traditional benchmarks such as ISBI should be maintained while advancing 3D head phantom benchmarks and point cloud standardization. Submission and registration requirements for cross-center external validation should also be implemented. Simultaneously, cross-domain image techniques can be leveraged to bridge gaps in head shadow measurements. When dealing with disease imbalance, small sample sizes, or variable imaging conditions, one can adapt approaches from radiology and fundus imaging. Diffusion models can synthesize challenging locations, uncommon cranial shapes, and lateral views or CBCT slices under varying exposure and noise conditions to mitigate data scarcity and domain shift. Existing cranial studies demonstrate that controllable, diffusion-based data generation significantly improves landmark detection. Leveraging structural priors and point-to-point topological relationships can stabilize training; this approach can also be extended to dataset construction and augmentation pipelines.

6 Conclusion

Deep learning has demonstrated remarkable potential in cephalometric analysis by enabling automatic landmark detection, structural segmentation, and intelligent measurement with diagnostic support. These AI-driven approaches have shown accuracy approaching that of human experts, while offering substantial advantages in efficiency, reproducibility, and workflow standardization. Automated systems can shorten analysis time from minutes to seconds, making them suitable for large-scale screening and routine follow-up, and they also enhance consistency across operators and clinical settings. However, several limitations must be addressed before widespread clinical adoption can be realized. Current research suffers from heterogeneity in landmark definitions, measurement protocols, and evaluation thresholds, which compromises comparability across studies. Furthermore, external validation, multi-center datasets, and standardized benchmarks are still lacking. Future progress requires unified consensus standards, integration of advanced generative models for data augmentation, and cross-domain imaging transfer to improve robustness in real-world clinical scenarios. With these advancements, AI-based cephalometry is expected to evolve into a reliable, standardized, and indispensable tool in orthodontics and craniofacial care.

References

1. Yang S, Song ES, Lee ES, et al. Ceph-Net: automatic detection of cephalometric landmarks on scanned lateral cephalograms from children and adolescents using an attention-based stacked regression network. BMC Oral Health. **23**, 803 (2023). https://doi.org/10.1186/s12903-023-03452-7
2. Chen S, Sedghi Gamechi Z, Dubost F, et al. An end-to-end approach to segmentation in medical images with CNN and posterior-CRF. Med Image Anal. **76**, 102311 (2022). https://doi.org/10.1016/j.media.2021.102311
3. Ye H, Cheng Z, Ungvijanpunya N, et al. Is automatic cephalometric software using artificial intelligence better than orthodontist experts in landmark identification? BMC Oral Health. **23**, 467 (2023). https://doi.org/10.1186/s12903-023-03188-4
4. Wu Y, Yu Q, Xia Y, et al. Does mandibular advancement with clear aligners have the same skeletal and dentoalveolar effects as traditional functional appliances? BMC Oral Health. **23**, 65 (2023). https://doi.org/10.1186/s12903-023-02709-5
5. Dot G, Schouman T, Chang S, et al. Automatic 3-dimensional cephalometric landmarking via deep learning. J Dent Res. **101**, 1380 (2022). https://doi.org/10.1177/00220345221112333
6. Bor S, Ciğerim S, Kotan S. Comparison of AI-assisted cephalometric analysis and orthodontist-performed digital tracing analysis. Prog Orthod. **25**, 41 (2024). https://doi.org/10.1186/s40510-024-00539-x
7. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: an umbrella review of systematic reviews. J Dent. **146**, 105056 (2024). https://doi.org/10.1016/j.jdent.2024.105056
8. Lang Y, Lian C, Xiao D, et al. Localization of craniomaxillofacial landmarks on CBCT images using 3D Mask R-CNN and local dependency learning. IEEE Trans Med Imaging. **41**, 2856 (2022). https://doi.org/10.1109/tmi.2022.3174513
9. de Queiroz Tavares Borges Mesquita G, Vieira WA, Vidigal MTC, et al. Artificial intelligence for detecting cephalometric landmarks: a systematic

review and meta-analysis. J Digit Imaging. **36**, 1158 (2023). https://doi.org/10.1007/s10278-022-00766-w

10. Hendrickx J, Gracea RS, Vanheers M, et al. Can artificial intelligence-driven cephalometric analysis replace manual tracing? A systematic review and meta-analysis. Eur J Orthod. **46**, cjae029 (2024). https://doi.org/10.1093/ejo/cjae029
11. Liu J, Zhang C, Shan Z. Application of artificial intelligence in orthodontics: current state and future perspectives. Healthcare. **11**, 2760 (2023). https://doi.org/10.3390/healthcare11202760
12. Guo D, Han W, Lyu P, et al. Towards better cephalometric landmark detection with diffusion data generation. IEEE Trans Med Imaging. **44**, 2784 (2025). https://doi.org/10.1109/tmi.2025.3557430
13. Zecca PA, Caccia M, Levrini L, et al. AI-based open-source software for cephalometric analysis from limited FOV radiographs. J Dent. **151**, 105412 (2024). https://doi.org/10.1016/j.jdent.2024.105412
14. Bao H, Zhang K, Yu C, et al. Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. BMC Oral Health. **23**, 191 (2023). https://doi.org/10.1186/s12903-023-02881-8