Deep Learning in Cephalometric Analysis: A Scoping Review of Automated Landmark Detection

Idriss Tafala[1], Fatima-Ezzahraa Ben-Bouazza[2], Aymane Edder[3], Oumaima Manchadi[4], Bassma Jioudi[5]
BRET Lab, Mohammed VI University of Health Sciences (UM6SS), Casablanca, Morocco[1,3,4,5]
BRET Lab, Mohammed VI University of Health Sciences (UM6SS), AIRA Lab, FST,
Hassan 1er University, Morocco[2]
LaMSN, La Maison Des Sciences Numériques, France[2]

***Abstract*—Cephalometric landmark identification is fundamental for accurate cephalometric analysis, serving as a cornerstone in orthodontic diagnosis and treatment planning. However, manual tracing is a labor-intensive process prone to inter-observer variability and human error, highlighting the need for automated methods to improve precision and efficiency. Recent advances in Deep Learning have enabled automatic detection of cephalometric landmarks, thereby increasing accuracy and consistency while reducing processing time. This scoping review examines contemporary applications of Deep Learning in cephalometric landmark detection and cephalometric analysis from 2019 to January 2025. We searched IEEEXplore, ScienceDirect, arXiv, Springer, and PubMed databases, identifying 601 articles, of which 76 met inclusion criteria after rigorous screening. Our analysis revealed significant performance improvements with Deep Learning methods achieving Success Detection Rates (SDR) of 75-90% at 2mm thresholds, substantially outperforming traditional methods. Geographical analysis identified China, South Korea, and the United States as leading research centers, with commercial applications like WebCeph and CephX gaining clinical adoption. Deep Learning improves the accuracy and efficiency of cephalometric analysis; however, challenges persist regarding dataset standardization and clinical validation. These technologies show promising potential to support novice clinicians, streamline radiological examinations, and improve landmark identification reliability in routine orthodontic practice.**

***Keywords*—Artificial Intelligence; deep learning; cephalometric analysis; landmark detection**

I. INTRODUCTION

Artificial Intelligence (AI) has significantly reshaped modern healthcare, revolutionizing various domains, including medical imaging, diagnostics, and treatment planning [1]. Orthodontics has experienced significant advancements in AI, attracting considerable interest from both researchers and practitioners. Lateral cephalometry is a commonly employed X-ray imaging technique that is essential for orthodontic diagnosis and treatment assessment, as it quantifies the angles of various characteristic points in relation to a reference plane. The accurate identification of cephalometric landmarks has historically been a labor-intensive process necessitating the expertise of skilled professionals [2].

Despite advancements in computational power and AI algorithms, the localization of cephalometric landmarks continues to pose a considerable challenge in orthodontics and dentistry. Digital cephalometric analysis has been possible for years; however, manual landmark identification was traditionally conducted solely by experts. Researchers have pursued automation of this process due to its complexity, time consumption, and the expertise required, acknowledging the advantages of decreased workload, enhanced efficiency, and reduced inter- and intra-observer variability [3][4].

Deep Learning (DL) has emerged as a transformative approach, exceeding the capabilities of traditional knowledge-based systems in landmark detection. Recent advancements in deep neural networks have resulted in software programs that can perform automated cephalometric landmark identification with reliability comparable to that of human experts. As a result, commercially available DL-powered cephalometric applications have emerged globally, with prominent examples such as CellmatIQ (Hamburg, Germany), ORCA AI (Herzliya, Israel), and WebCeph (Gyeonggi-do, Korea).

We selected a scoping review methodology rather than a systematic review or meta-analysis because the field of Deep Learning in cephalometric analysis is rapidly evolving with heterogeneous methodologies and outcomes. A scoping review allows us to comprehensively map this landscape, identify key concepts, and highlight knowledge gaps without being constrained by the rigid methodological appraisal required in systematic reviews. This approach is particularly appropriate given the diverse technical implementations, varying evaluation metrics, and emerging commercial applications in this domain.

Prior reviews on this topic [5][6][7][8][9] have primarily conducted umbrella reviews of systematic studies that compare the performance of artificial intelligence with manual tracing methods. These reviews did not include comprehensive bibliometric or geographic analyses and failed to explicitly assess the differences in accuracy between AI-generated landmarks and traditional manual landmarking. Additionally, previous meta-analyses have concentrated predominantly on academic performance metrics, with less emphasis on practical clinical implications.

This study is organized as follows: Section II provides essential background on cephalometric analysis fundamentals, the evolution of landmark detection methodologies, and key performance metrics used in the field. Section III details our comprehensive search strategy, study selection process, eligibility criteria, and data extraction methodology following PRISMA-ScR guidelines. Section IV presents our findings through three distinct evolutionary phases of Deep Learning applications (2019-2020, 2021-2022, and 2023-2025), alongside a detailed analysis of commercial software implementations and their clinical validation. Section V discusses the current challenges facing the field, including dataset limitations, clinical integration barriers, and performance variability across different architectures, while also examining the emergence of commercially available AI-driven cephalometric platforms. Finally, Section VI outlines future research directions essential for advancing the field, including the need for larger multi-ethnic datasets, unified benchmarking frameworks, and parameter-efficient architectures for clinical deployment, before Section VII concludes with key insights and implications for orthodontic practice.

A. Contributions and Research Questions

This study offers several novel contributions that address key gaps in the existing literature:

- A comprehensive bibliometric analysis spanning a wider timeframe (2019 to January 2025), documenting the exponential growth in research output.
- A structured analysis of the technological evolution from fundamental convolutional neural networks to advanced transformer-based architectures.
- A detailed geographic distribution analysis identifying research leadership centers and potential collaboration opportunities.
- A focused comparative examination of commercially available AI-driven software, highlighting both practical clinical applications and inherent limitations.

Following the Population, Concept, Context (PCC) framework, this scoping review addresses the following specific research questions:

1) What are the primary Deep Learning methodologies employed in cephalometric landmark detection?

   Population: cephalometric radiographs; *Concept:* Deep Learning methodologies; *Context:* landmark detection.

2) How have these methodologies evolved from 2019 to 2025?

   Population: research studies; *Concept:* technological evolution; *Context:* temporal progression.

3) What gaps exist in the current literature regarding automated landmark detection using Deep Learning?

   Population: research literature; *Concept:* knowledge gaps; *Context:* methodological and clinical application.

4) What are the implications and future directions for integrating these methods into clinical orthodontic practice?

   Population: orthodontic practitioners; *Concept:* clinical integration; *Context:* practical implementation.

Unlike previous reviews, this study provides a broad, structured thematic analysis focusing on technological evolution, geographic trends, and commercial software applications—areas that have not been comprehensively addressed in prior reviews.

II. Background

A. Cephalometric Analysis

Cephalometric analysis involves the evaluation of lateral skull radiographs taken with a cephalostat to identify skeletal patterns and gauge the complexity of treatment. This method is particularly suited for situations that require intentional forward and backward adjustments, though it is not essential for every orthodontic procedure. Cephalometric analysis is especially required when considerable changes to the position of the incisors are expected.

Cephalometric analysis is a technique with a significant historical background, tracing its origins to the late 1800s, a period marked by the initial application of radiographs for the examination of the head and neck. During the 1930s, Holly Broadbent, a professor in orthodontics at the University of Michigan, contributed significantly to the discipline by examining the correlation between dental structures and cranial anatomy. The study encompassed the measurement of diverse angles and distances within radiographic images, thereby laying the groundwork for cephalometric analysis [10].

Over the following decades, researchers built upon his foundational contributions, advancing further methodologies including, Stainer Analysis, Wits analysis, and Downs Analysis. Cephalometric analysis serves as a fundamental component of contemporary orthodontic practice, facilitating the diagnosis and management of a range of dental and skeletal irregularities [11]. The analysis evaluates the anteroposterior and vertical relationships of the mandible and maxilla in relation to the cranial base and each other, as well as the upper and lower teeth in relation to the mandibular and maxillary bones (Fig. 1). The process begins with the identification of anatomical landmarks on cephalometric radiographs, referred to as cephalometric landmarks. The number of landmarks in the most commonly available open-source dataset is nineteen.

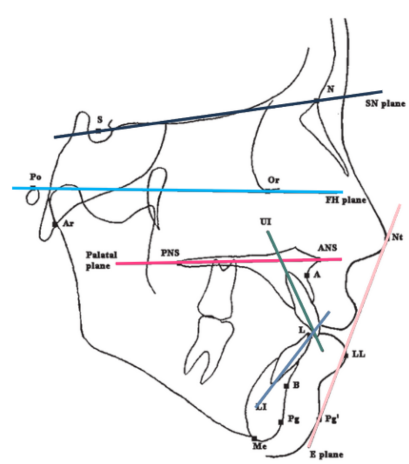

Fig. 1. Common cephalometric points and planes.

Lateral skull radiographs (Fig. 2) offer a two-dimensional representation of the head and neck, facilitating the assessment of sagittal and vertical dimensions. Sagittal measurements analyze the location and angulation of the maxilla and mandible, whereas vertical measurements assess the height of facial components and the interrelationship between the jaws [12]. In contrast, posteroanterior radiographs, captured from the anterior aspect of the head, assess transverse and vertical dimensions, providing insights into face width and the interrelationship of the jaws in the transverse plane [13]. In clinical practice, cephalometric analysis mainly utilizes lateral radiographs due to the interpretive difficulties associated with posteroanterior projections.

Fig. 2. Lateral skull radiograph (cephalogram).

Cephalometric analysis is the comparison of individual anatomical proportions and angular measurements to population averages. Traditionally, these anatomical landmarks are manually traced and interconnected to create lines and angles, with the resultant measurements documented and analyzed [14].

Cephalometric analysis is essential for detecting and managing malocclusions, necessitating a collaborative healthcare team comprising dental professionals, including general dentists, orthodontists, and oral surgeons. This approach offers significant insights into the degree of skeletal and dental misalignments, along with potential causal factors.

B. Landmark Detection Evolution

The development of methodologies for the automatic detection of facial landmarks has exhibited notable advancements, reflecting substantial developments over time.

Initially, manual annotation and semi automatic detection using graphical software, served as the predominant techniques for the labeling and identification of facial landmarks. Nonetheless, this methodology demonstrated a significant consumption of time, rendering it unsuitable for extensive applications [15].

As computer vision capabilities improved, the 2000s witnessed the emergence of Statistical Shape Models. Active Shape Models (ASMs) were developed as a method to represent object shape using a set of landmark points and an associated shape model.in other words, it aligns the model to new images using edge detection. After that, Active Appearance Models (AAMs) expanded upon ASMs by integrating both shape and appearance data [16] [17].

The mid-2010s marked an era characterized by the dominance of supervised learning and machine learning methodologies. The detection techniques focused on feature extraction methods, including Haar features, Histogram of Oriented Gradients (HOG), and Scale-Invariant Feature Transform (SIFT), as well as learning-based regression models such as decision trees, support vector machines (SVM), and boosted regression forests [18][19].

Since 2015, Deep Learning has been at the forefront of landmark detection. Convolutional Neural Networks (CNNs) became prominent, allowing researchers to develop models that directly predict cephalometric landmarks from images. Stacked Hourglass Networks and Fully Convolutional Networks have emerged as prominent techniques, significantly improving accuracy relative to prior techniques [20] [21].

With the continual development and progression of Deep Learning, increasingly sophisticated algorithms have been employed to identify landmarks. Several models integrating attention mechanisms and transformers have been utilized for landmark detection [22] [23] [24].

C. Performance Metrics

To quantitatively assess the accuracy and reliability of different landmark identification techniques, four important metrics have been widely adopted in the literature: Mean Absolute Error (MAE), Success Detection Rate (SDR), Standard Deviation of Radial Error (SDRE), and Mean Radial Error (MRE).

1) Mean Absolute Error (MAE): MAE [Eq. (1)] quantifies the absolute deviation between predicted and ground truth landmark positions, providing a direct measure of localization precision. It is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |RE_i| \quad (1)$$

where, RE_i represents the radial error for the i-th landmark, and N is the total number of landmarks. A lower MAE value indicates higher accuracy in landmark localization.

2) Success Detection Rate (SDR): SDR [Eq. (2)] is a widely adopted performance metric that reflects the percentage of landmarks detected within clinically relevant thresholds (e.g., 2.0 mm, 2.5 mm, 3.0 mm, and 4.0 mm). It is calculated as:

$$SDR_x = \frac{\text{Number of landmarks with error} \leq x \text{ mm}}{\text{Total number of landmarks}} \times 100 \quad (2)$$

where, x represents the error threshold. Higher SDR values indicate greater robustness in landmark detection.

3) Standard Deviation of Radial Error (SDRE): SDRE [Eq. (3)] quantifies the variability in landmark localization across multiple samples, capturing the stability of the model. It is expressed as:

$$SDRE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(RE_i - MRE)^2} \tag{3}$$

where, RE_i represents the radial error for each landmark, and MRE is the mean radial error. A lower SDRE value suggests more consistent landmark localization.

4) Mean Radial Error (MRE): MRE [Eq. (4)] measures the average Euclidean distance between predicted and ground truth landmarks, serving as a primary indicator of overall detection accuracy. It is given by:

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \tag{4}$$

where, (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ represent the ground truth and predicted coordinates of the i-th landmark, respectively. A lower MRE value signifies higher landmark localization accuracy.

III. Methods

A. Search Strategy

We conducted a systematic search in accordance with PRISMA-ScR guidelines for scoping reviews. The last comprehensive literature search was performed on January, 2025.

The primary search was conducted using the electronic databases like IEEEXplore, ScienceDirect, arXiv, Springer and PubMed, aiming to gather a comprehensive selection of studies on the application of Deep Learning in cephalometric landmark identification, ultimately facilitating cephalometric analysis.

The search process employed specific keywords and queries such as:

Query 1

"landmark detection" OR "automatic cephalometric landmark detection" OR "landmark identification on lateral cephalogram" , AND "Deep Learning"

Query 2

(cephalometric OR lateral skull radiograph) AND (artificial intelligence OR Deep Learning OR convolutional neural network) AND (landmark detection OR automated tracing)

Query 3

(orthodontic imaging) AND (AI OR artificial intelligence) AND (landmark identification)

Query 4

"Deep Learning architectures"[MeSH] OR "Convolutional Neural Network"[MeSH] OR "Attention mechanisms"[MeSH] OR "Fully Convolutional Network"[MeSH] OR "hybrid approaches"[MeSH]

Alongside, the reference lists of all included studies and relevant review articles were hand-searched to identify additional eligible studies.

A total of 601 articles were found within the timeframe of 2019 to 2024. The year-wise distribution is depicted in Fig. 3, and the geographic distribution is in Fig. 4.

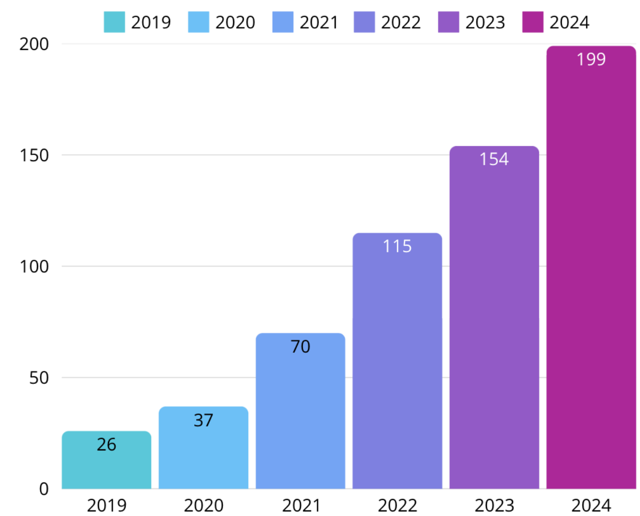

Fig. 3. Annual distribution of published studies included in the review.

B. Study Selection Process

The selection process followed the four-stage PRISMA flow diagram (Fig. 5). A total of 1,023 records were initially identified through database searches. After removing 166 duplicates, the remaining 857 records were screened for eligibility. After applying initial exclusion criteria (abstract not relevant, publication title unrelated, dataset not specified), 601 records remained.

Two reviewers (IT and AE) independently screened the titles and abstracts of these 601 records against pre-defined inclusion and exclusion criteria, resulting in 530 potentially relevant articles. Disagreements were resolved through discussion, and when necessary, a third reviewer (FEB) was consulted. A total of 290 articles were excluded at this stage primarily because they were not related to Deep Learning techniques.

Full-text assessment of the remaining 240 articles was conducted independently by two reviewers (IT and OM) using a standardized eligibility form. Disagreements were resolved through consensus discussions with a senior researcher (BJ). This process excluded 164 additional articles due to datasets

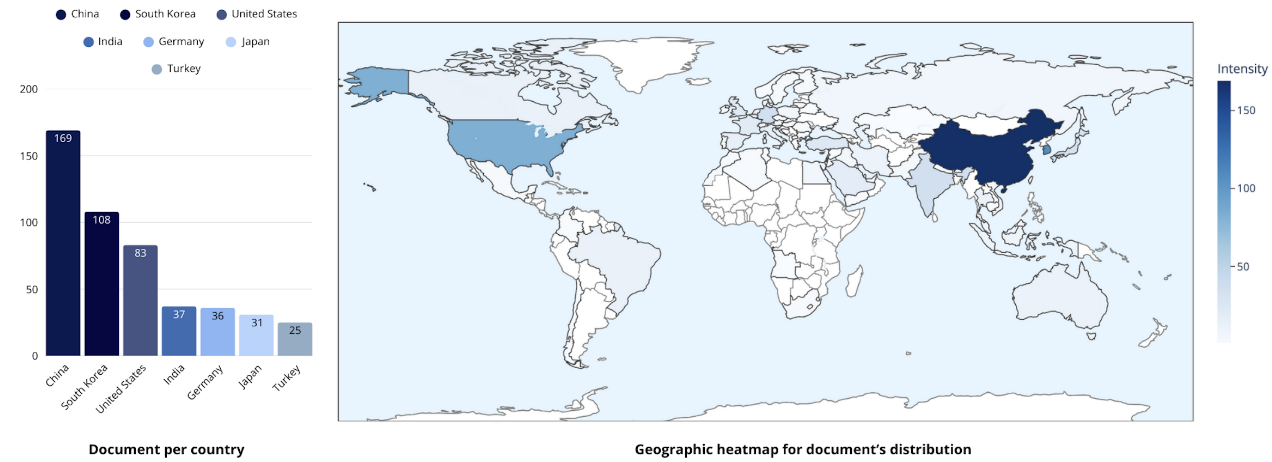

Fig. 4. Geographic distribution of the included studies across countries.

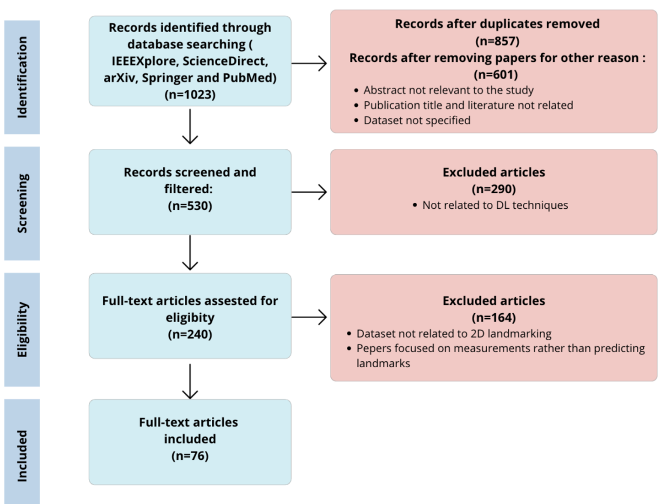

Fig. 5. PRISMA-ScR flow diagram of the study selection process.

not related to 2D landmarking or papers focused on measurements rather than predicting landmarks. Finally, 76 articles met all inclusion criteria and were included in the review.

C. Eligibility Criteria

1) Inclusion criteria:

- Original research studies utilizing Deep Learning for cephalometric landmark detection
- Studies published between January 2019 and January 2025
- Studies with clearly described methodology and evaluation metrics
- Papers reporting on 2D lateral cephalometric radiographs
- Studies comparing commercial software performance with manual tracing

2) Exclusion criteria:

- Review articles, case reports, editorials, and letters
- Studies without clear methodology descriptions
- Studies without quantitative performance metrics

D. Data Extraction

We developed a standardized data extraction form based on the Joanna Briggs Institute data extraction template for scoping reviews. Two reviewers (IT and FEB) independently extracted data from all included studies. For each study, the following information was recorded:

- Publication details (author, year, country, journal)
- Study characteristics (sample size, dataset origin, number of landmarks)
- Methodological details (Deep Learning architecture, implementation approach)
- Performance metrics (Success Detection Rate at different thresholds, Mean Radial Error)
- Clinical validation methods (if any)
- Commercial application details (for relevant studies)

IV. Results

Table I presents papers and studies of the early approaches (2019 to 2020) for cephalometric landmark detection. These years marked the initial surge in applying Deep Learning to automate cephalometric landmark detection. Efforts mainly explored architectures like YOLOv3, stacked hourglass networks, Bayesian CNNs, and basic CNN pipelines.

Intermediate progress period (2021 to 2022) marked a transition to more customized and cascaded network designs tailored for medical imaging, with growing interest in patch-based models and clinical applicability. Table II presents more details of papers for Deep Learning architecture in that period of time.

TABLE I. SUMMARY OF CEPHALOMETRIC STUDIES BETWEEN 2019 AND 2020

Paper	Year	Country	Dataset	Nb of Points	Methods	Results
[25]	2019	South Korea	1,028 cephalograms	80	YOLOv3	SDR: 2 mm: 80.4%, 2.5 mm: 87.4%, 3 mm: 92.0%, 4 mm: 96.0%
[26]	2020	Germany	1,792 cephalograms	18	Customized convolutional neural network	Absolute mean differences: Angular parameters ¡ 0.37°, Metric ¡ 0.20 mm, Proportional ¡ 0.25%
[27]	2020	South Korea	1,311 cephalograms	46	YOLOv3	Mean detection error: 1.46 ± 2.97 mm
[28]	2020	South Korea	2,075 cephalograms	23	Stacked hourglass network	SDR: 2 mm: 84.53%, 2.5 mm: 90.11%, 3 mm: 93.21%, 4 mm: 96.79%
[29]	2020	South Korea	ISBI 2015 dataset	19	Bayesian convolutional neural network (BCNN)	SDR: 2 mm: 82.11%, 3 mm: 92.28%, 4 mm: 95.95%
[30]	2020	Netherlands	ISBI 2015 dataset	19	Fully convolutional neural network (ResNet34)	Median Euclidean error: Test1: 0.46–2.12 mm, Test2: 0.42–4.32 mm, SDR (4 mm): 95%
[31]	2020	Japan	ISBI 2015 dataset	19	Two-step method: ROI extraction + ResNet50 network	SDR: Test1: 2 mm: 86.4%, 4 mm: 97.8%; Test2: 2 mm: 74%, 4 mm: 94.3%
[32]	2020	China	ISBI 2015 dataset	19	CephaNN Network	SDR: Test1 - 2 mm: 87.61%, 4 mm: 94.63%; Test2 - 2 mm: 76.30%, 4 mm: 94.63%

TABLE II. SUMMARY OF CEPHALOMETRIC STUDIES BETWEEN 2021 AND 2022

Paper	Year	Country	Dataset	Nb of Points	Methods	Results
[33]	2021	South Korea	ISBI 2015 dataset	19	DACFL	EDR: Test1+Test2 - 2 mm: 17.92%, 4 mm: 3.08%
[34]	2021	China	ISBI 2015 dataset	19	Cascaded three-stage CNN	SDR: 2 mm: 76.82%, 4 mm: 95.58%
[33]	2021	South Korea	ISBI 2015 dataset	19	YOLOv3	SDR: 2 mm: 75.45%, 4 mm: 94.24%
[35]	2021	South Korea	3,150 lateral cephalograms	20	Cascade CNN	Mean detection error: 1.36 ± 0.98 mm
[36]	2021	Japan	1,785 cephalograms	26	CNN-PC + CNN-PE	Mean detection error: 1.32 mm - 1.50 mm
[37]	2021	South Korea	950 cephalograms	13	Two-step method (ROI + detection)	SDR: 2 mm: 64.3%, 4 mm: 95.1%
[38]	2021	China	ISBI 2015 dataset + Huaxi-Analysis dataset	19/37	GDN + LRN	SDR: Dataset1 - 2 mm: 75.47%, 4 mm: 94.63%; Dataset2 - 2 mm: 90.30%, 4 mm: 98.61%
[39]	2021	South Korea	ISBI 2015 grand challenge dataset	19	Multistage probabilistic approach (global, local, and refinement stages)	SDR: Test1 - 2 mm: 77.16%, 2.5 mm: 84.74%, 3 mm: 89.55%, 4 mm: 95.12%; Test2 - 2 mm: 77.16%, 2.5 mm: 84.74%, 3 mm: 89.25%, 4 mm: 95.12%
[40]	2021	China	ISBI 2015 grand challenge dataset	19	Four-step DCNN-based framework	SDR: 2 mm: 87.51%, 2.5 mm: 91.83%, 3 mm: 94.74%, 4 mm: 98.01%
[41]	2022	China	512 cephalograms	37	Neural network using two modules (global detection module and locally modified module)	SDR: 1 mm: 54.05%, 1.5 mm: 91.89%, 2 mm: 97.30%, 2.5 mm: 100%, 3 mm: 100%, 4 mm: 100%
[42]	2022	South Korea	JBNU dataset	41	Deep anatomical context feature learning (DACFL) model	SDR: 2 mm: 73.17%, 2.5 mm: 80.39%, 3 mm: 85.61%, 4 mm: 91.68%
[43]	2022	South Korea	ISBI 2015 grand challenge dataset	19	Network architecture composed of global and local encoders and patch-wise attentions	SDR: Test1 - 2 mm: 86.42%, 4 mm: 98.46%; Test2 - 2 mm: 74.58%, 4 mm: 94.26%
[44]	2022	South Korea	600 cephalograms	16	Cascaded convolutional neural network (CNN)	SDR Comparison: Skeletal landmarks - 2 mm: 93.43%, 4 mm: 98.71%; Upper airway - 2 mm: 72.22%, 4 mm: 92.78%
[45]	2022	South Korea	2,798 cephalograms	16	Cascaded convolutional neural network (CNN)	SDR: 1 mm: 47.9%, 2 mm: 83.3%, 4 mm: 98.0%
[46]	2022	Slovenia	ISBI 2015 grand challenge dataset + AUDAX Private Dataset	19/72	SpatialConfiguration-Net	MRE: ISBI: 1.13 mm; AUDAX: 11.26 ± 17.51 pixels
[47]	2022	China	ISBI 2015 grand challenge dataset	19	Shooting Reward Learning Network (SRLN)	MRE: 1.09 mm

Recent Advances in Table III represent a turning point in cephalometric landmark detection with the integration of advanced transformer-based detection frameworks, multi-stage regression models, and generalizable architectures. More attention is paid to cross-dataset generalization, universal landmark detectors, and attention-based recalibration mechanisms.

The emergence of sophisticated cephalometric software has revolutionized orthodontic practice and maxillofacial diagnostics by fundamentally altering how clinicians approach landmark identification and craniofacial evaluation. This digital transformation has redefined traditional methodologies, with specialists increasingly relying on computational tools for precision and efficiency. As shown in Table IV, many studies have carefully compared different commercial platforms and analytical programs, showing how artificial intelligence is becoming a regular part of clinical decision-making.

V. DISCUSSION

The application of Deep Learning for automated cephalometric landmark detection has evolved significantly over time. These advancements have greatly improved the precision and efficiency of landmark localization within medical images.

TABLE III. SUMMARY OF CEPHALOMETRIC STUDIES BETWEEN 2023 AND JANUARY 2025

Paper	Year	Country	Dataset	Nb of Points	Methods	Results
[48]	2023	China	ISBI 2015 grand challenge dataset	19	Feature Aggregation and Refinement Network (FARNet)	SDR: Test1 - 2.0 mm: 88.03%, 2.5 mm: 92.73%, 3.0 mm: 95.96%, 4.0 mm: 98.48%; Test2 - 2.0 mm: 77.00%, 2.5 mm: 84.42%, 3.0 mm: 89.47%, 4.0 mm: 95.21%
[49]	2023	China	9,870 cephalograms	30	CephNet	SDR: 1.0 mm: 66.15%, 2.0 mm: 91.73%, 3.0 mm: 97.99%
[50]	2023	Germany	30 CT scans	60	Deep Neural Patchworks (DNPs)	SDR: 2.0 mm: 66.4%, 4.0 mm: 91.9%
[51]	2023	South Korea	1,286 cephalograms	19	Ceph-Net	SDR: 1.0 mm: 41.35%, 2.0 mm: 73.14%, 3.0 mm: 85.22%, 4.0 mm: 94.65%
[52]	2023	India	ISBI 2015 grand challenge dataset + 100 cephalograms from Solanki Dental Care Clinic	19	CephXNet	Classification Accuracy: Training Set: 98.51%, Testing Set: 97.72%; MRE: Test1: 0.92 mm, Test2: 1.41 mm; SDR within 2mm: Test1: 88.06%, Test2: 78.72%
[53]	2023	India	ISBI 2015 grand challenge dataset	19	Coarse localization through region of interest (ROI) extraction and fine localization utilizing histogram-oriented gradient (HOG) feature	SDR: 2.0 mm: 77.11%
[54]	2023	China	ISBI 2015 grand challenge dataset + 163 cephalograms	19	MS-YOLOv3	SDR: 2 mm: 80.84%, 4 mm: 98.14%
[55]	2023	China	481 cephalograms	61	Point-locating model CephaNET	SDR: 2.0 mm: 85.4%, 2.5 mm: 90.2%, 3.0 mm: 93.5%, 4.0 mm: 97%
[56]	2023	Japan	2,000 facial profile images	23	HRNetv2	SDR: 2.0 mm: 98.20%, 2.5 mm: 99.55%, 3.0 mm: 99.79%, 4.0 mm: 99.93%
[57]	2024	China	ISBI 2015 grand challenge dataset	19	Mask Region-based Convolutional Neural Network	SDR: 1.0 mm: 43.44%, 2.0 mm: 68.27%, 4.0 mm: 85.74%
[58]	2024	China	ISBI 2015 grand challenge dataset	19	Feature Decouple and Gated Recalibration Network (FDGR-Net)	SDR: 2.0 mm: 75.16%, 2.5 mm: 82.53%, 3.0 mm: 88.58%, 4.0 mm: 94.79%
[59]	2024	Taiwan	1,002 cephalograms	14	CNN architecture enhanced with U-Net and MobileNetV2	SDR: 2.0 mm: 83.14%, 2.5 mm: 89.62%, 3.0 mm: 93.97%, 4.0 mm: 98.23%
[60]	2024	China	ISBI 2015 grand challenge dataset	19	UniverDetect	SDR: 2 mm: 75.87%, 3 mm: 88.35%, 4 mm: 94.59%
[61]	2024	Pakistan	ISBI 2015 grand challenge dataset	19	Two-stage regression framework (LDM for coarse localization, LRM for fine-tuning)	SDR: Test1 - 2.0 mm: 87.17%, 3.0 mm: 95.18%, 4.0 mm: 98.12%; Test2 - 2.0 mm: 75.79%, 3.0 mm: 89.0%, 4.0 mm: 94.53%
[53]	2024	India	ISBI 2015 grand challenge dataset + Diverse-CEPH19	19	Detectron2, and YOLOv8	Detectron2: Best Model: rcnn_R_101_FPN_3x; MRE: 1.2 mm (±0.96 mm); SDR (2 mm): 85.89%; SDR (4 mm): 98.45%; YOLOv8: Best Model: YOLOv8m-pose; MRE: 1.62 mm (±1.03 mm); SDR (2 mm): 72.92%; SDR (4 mm): 96.80%
[62]	2024	Morocco	ISBI 2015 grand challenge dataset	19	YOLOv8	SDR : 2 mm: 86.31%; 2.5 mm: 87.69%; 3 mm: 90.84%

Our analysis of 76 studies reveals distinct developmental phases and several key insights that warrant critical examination. Cephalometric landmark detection has progressed through three distinct evolutionary phases. The first phase, from 2019 to 2020, focused on basic abilities using general models like YOLOv3, which reached an SDR value of 80.4% at 2mm thresholds. The intermediate refinement phase (2021-2022) implemented architectural innovations aimed at resolving challenges in medical imaging. Multi-stage frameworks, cascaded approaches, and attention mechanisms have become leading strategies. The advanced optimization phase (2023-2025) includes the use of complex transformer-based architectures [57], feature recalibration mechanisms [58], and multi-scale approaches [54], which have pushed performance limits even further. Research from this period indicates that SDR values consistently surpass 87% at 2mm thresholds in controlled settings, with certain methodologies [56] attaining notable accuracy (SDR of 98.2% at 2mm) on facial profile images.

Our statistical analysis of performance metrics across architectural categories provides crucial insights into effective design approaches (Table V).

This comparative analysis demonstrates a clear progression in accuracy metrics, with transformer-based and attention-enhanced architectures consistently outperforming earlier approaches. Notably, cascaded networks exhibit a favorable balance between accuracy and computational efficiency, an important consideration for clinical deployment.

Despite the significant progress in Deep Learning applications for cephalometric landmark detection, several persistent challenges and limitations remain that impede widespread clinical adoption. The reliance on limited datasets represents perhaps the most significant barrier to the clinical translation of Deep Learning for cephalometric analysis. Our comprehensive analysis of dataset characteristics across the literature (Table VI) reveals several critical concerns.

The median dataset size of 723 images falls dramatically short compared to medical imaging datasets in other domains, such as ChestX-ray14 (112,120 images) or MIMIC-CXR (377,110 images). This limitation is particularly concerning given the high dimensional complexity of craniofacial anatomy

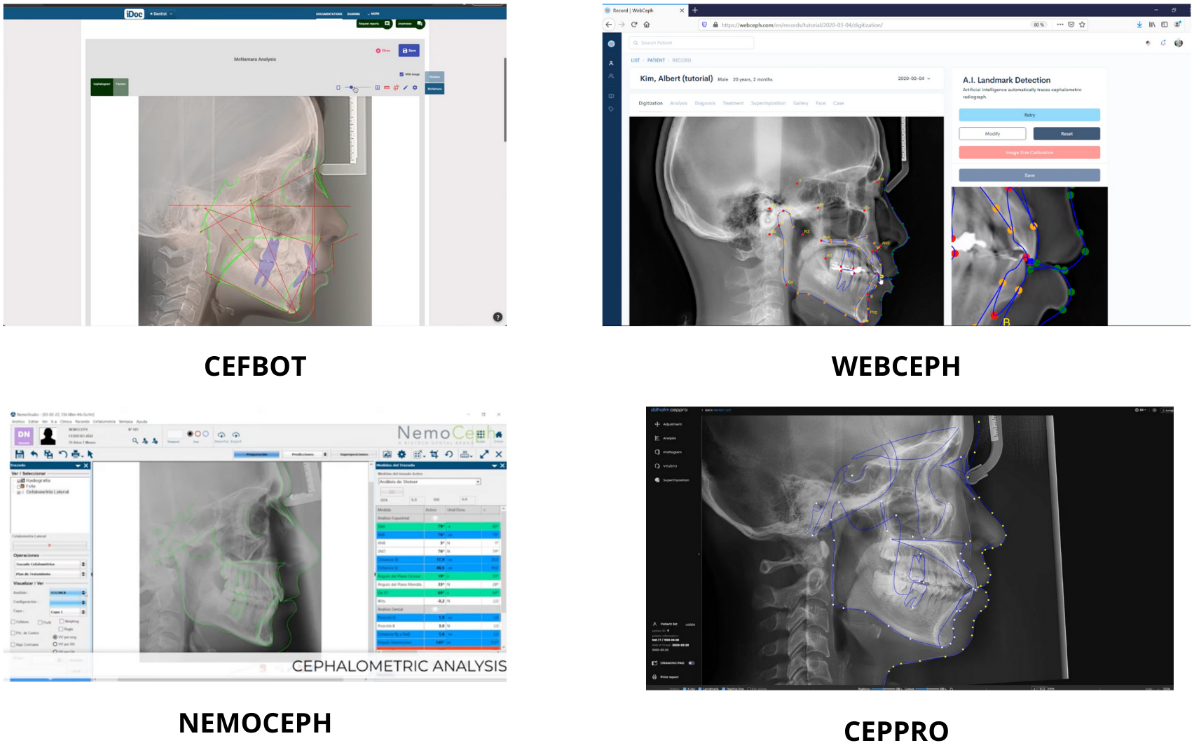

Fig. 6. Localization of oral cancer process.

TABLE IV. SUMMARY OF STUDIES EVALUATING COMMERCIAL SOFTWARE FOR AUTOMATED CEPHALOMETRIC ANALYSIS

Paper	Year	Software included
[63]	2021	Ceppro software by DDH Inc
[64]	2022	WebCeph
[65]	2022	WebCeph
[66]	2022	CEFBOT
[67]	2022	AudaxCeph
[68]	2022	Dolphin Imaging softwar, and CS Imaging V8 software
[69]	2022	WebCeph
[70]	2022	CephX
[71]	2023	Planmeca Romexis software
[72]	2023	MyOrthoX, Angelalign, Digident
[73]	2023	CEFBOT
[74]	2023	DentaliQ.ortho, WebCeph, AudaxCeph, and CephX
[75]	2023	CellmatIQ, CephX, AudaxCeph, and WebCeph
[76]	2023	CEFBOT
[77]	2024	CefBot, and WebCeph
[78]	2024	OneCeph app and WebCeph
[79]	2024	WebCeph, WeDoCeph, and CephX
[80]	2024	WebCeph, and CephX
[81]	2024	NemoCeph, and WebCeph
[82]	2024	WebCeph, Cephio, and Ceppro software

and the substantial anatomical variation across human populations.

For our analysis of dataset characteristics across the literature. It reveals that 55.3% of studies (42/76) relied exclusively on the ISBI 2015 dataset [4], which contains only 400 images primarily from a specific demographic population.

Studies that performed cross-dataset validation [38], [41], [48] consistently reported performance degradation when models trained on one dataset were applied to another. He et al. [38] demonstrated performance degradation of 15.4% when evaluating models trained on the ISBI dataset against their institutional dataset, highlighting the challenge of domain shift in clinical applications. Similarly, Šavc et al. [48] reported MRE values of 1.13mm on the ISBI dataset but substantially higher errors (11.26 ± 17.51 pixels) on their private AUDAX dataset.

A positive development is the emergence of larger datasets, such as that used by Jiang et al. [49], which included 9,870 cephalograms and demonstrated improved generalization capabilities. However, even this represents a fraction of the dataset sizes commonly used in other medical imaging domains, as highlighted by Zeng et al. [37] and Wang et al. [40].

The emergence of commercially available AI-driven cephalometric software (Fig. 6) represents a pivotal step toward clinical integration.

Our analysis of twenty studies evaluating commercial systems [83]-[81] reveals a complex picture of real-world application challenges. While commercial systems demonstrate impressive automation capabilities, validation studies consistently identify three critical limitations:

1) Inconsistent performance across landmark types: Commercial systems exhibit variable accuracy depending on landmark anatomical characteristics, with mean differences of 0.8 to 1.7 mm between predicted and ground truth positions for well-defined skeletal landmarks, but significantly higher deviations (1.9 to 3.2 mm) for soft tissue landmarks.

TABLE V. COMPARATIVE PERFORMANCE ANALYSIS BY ARCHITECTURE TYPE

Architecture Category	Mean SDR (2mm)	Mean MRE (mm)	Mean Detection Time (s)
Standard CNN-based	77.3% ± 3.8%	1.86 ± 0.27	0.78 ± 0.15
YOLO-based	79.5% ± 4.2%	1.53 ± 0.18	0.43 ± 0.07
Cascaded Networks	83.7% ± 3.1%	1.38 ± 0.15	0.95 ± 0.23
Attention-enhanced	85.9% ± 2.9%	1.29 ± 0.11	0.83 ± 0.19
Transformer-based	87.4% ± 2.4%	1.12 ± 0.09	1.21 ± 0.27

TABLE VI. DATASET SIZE AND DIVERSITY ANALYSIS; *DEMOGRAPHIC DIVERSITY INDEX (1-5 SCALE): 1=SINGLE POPULATION, 5=GLOBALLY REPRESENTATIVE

Dataset Type	Median Sample Size	Demographic Diversity*	Usage Frequency
ISBI 2015	400	1.2	55.30%
Private Institutional	723 (range: 163–9,870)	1.7	34.20%
Multi-center	1,872 (range: 950–3,150)	2.3	10.50%

2) Limited interchangeability with human experts: No commercial system achieved the clinically required reliability threshold (ICC > 0.95) across all cephalometric measurements necessary for complete diagnostic interchangeability with human experts.

3) Insufficient transparency: Commercial systems often lack methodological transparency regarding their underlying architectures, training protocols, and dataset characteristics, complicating clinical validation efforts.

Anatomical variability adds to the problems, especially when it comes to soft tissue landmarks like the soft tissue pogonion, which have many different structures and make predictions less accurate. The closeness of landmarks inside the skull makes automation harder because there aren't clear lines between structures that are next to each other. This can lead to unclear predictions, which lowers the accuracy of medical care.

Besides basic landmark detection, effective clinical integration needs more features, like measuring angles and distances, and should work well with common orthodontic software. Even though it works very well in controlled environments, how widely it is used in busy clinical settings depends a lot on things like how easy it is to use, how fast the computer can process data, and how well it works with other imaging systems. The computational resource demands for model training and deployment hinder the widespread adoption of AI-based approaches in clinical settings, despite their demonstrated promise.

VI. FUTURE DIRECTIONS

Given the significant challenges identified, various research directions require focus. The creation of extensive, multi-ethnic cephalometric datasets that include pathological diversity and are annotated with standardized protocols is crucial for enhancing model generalizability and fairness .

The establishment of unified benchmarking frameworks that include diverse test cohorts, clinically relevant evaluation metrics, and rigorous statistical analyses would greatly enhance the field's cross-study comparability.

Third, the design of parameter-efficient Deep Learning architectures specifically for clinical hardware deployment is essential for enabling real-world integration. Fourth, enhancing human–AI collaboration through interactive systems that use the complementary strengths of expert clinicians and AI models may facilitate adoption in standard practice.

Finally, the clinical utility of AI-driven cephalometric tools must be substantiated through prospective trials evaluating their impact on diagnostic accuracy, treatment planning, and patient outcomes.

VII. CONCLUSION

Artificial intelligence (AI) is progressively incorporated into orthodontic treatment, serving as a valuable tool for automating cephalometric landmark tracing in routine clinical practice. AI-powered systems are transforming contemporary orthodontic workflows by aiding clinicians in orthodontic treatment planning and decreasing the time needed for radiological diagnoses. Cephalometric analysis is an important way to assess the bones and structures of the face, making it a great option for using AI to automate processes in orthodontics, maxillofacial surgery, and craniofacial treatment.

However, the trustworthiness of AI in cephalometry varies, affected by different things like how accurately landmarks are identified by hand, the skill level of the person using it, the quality and number of X-rays, and the kind of algorithm used. These factors must be meticulously considered when analyzing AI-generated outcomes.

This scoping review indicates that although AI-powered cephalometric analysis has considerable potential, notable limitations persist, including issues related to accuracy, accessibility, expertise, ethical considerations, costs, and regulatory challenges.

Future research and technology should focus on improving accuracy, increasing the variety of data used, and tackling clinical validation issues to make sure that AI effectively helps and supports human skills in orthodontic and craniofacial diagnostics.

REFERENCES

[1] J. Liu, C. Zhang, and Z. Shan, "Application of artificial intelligence in orthodontics: current state and future perspectives," in *Healthcare*, vol. 11, p. 2760, MDPI, 2023.

[2] H. L. D. da Silveira and H. E. D. Silveira, "Reproducibility of cephalometric measurements made by three radiology clinics," *The Angle Orthodontist*, vol. 76, no. 3, pp. 394–399, 2006.

[3] S. Ö. Arık, B. Ibragimov, and L. Xing, "Fully automated quantitative cephalometry using convolutional neural networks," *Journal of Medical Imaging*, vol. 4, no. 1, pp. 014501–014501, 2017.

[4] C.-W. Wang, C.-T. Huang, J.-H. Lee, C.-H. Li, S.-W. Chang, M.-J. Siao, T.-M. Lai, B. Ibragimov, T. Vrtovec, O. Ronneberger, *et al.*, "A benchmark for comparison of dental radiography analysis algorithms," *Medical image analysis*, vol. 31, pp. 63–76, 2016.

[5] A. Polizzi and R. Leonardi, "Automatic cephalometric landmark identification with artificial intelligence: An umbrella review of systematic reviews," *Journal of Dentistry*, p. 105056, 2024.

[6] S. Narkhede, P. Rao, V. Sawant, S. S. Sachdev, S. Arora, A. M. Pawar, R. Reda, and L. Testarelli, "Digital versus manual tracing in cephalometric analysis: a systematic review and meta-analysis," *Journal of Personalized Medicine*, vol. 14, no. 6, p. 566, 2024.

[7] P. Smołka, K. Nelke, N. Struzik, K. Wiśniewska, S. Kiryk, J. Kensy, W. Dobrzyński, J. Kiryk, J. Matys, and M. Dobrzyński, "Discrepancies in cephalometric analysis results between orthodontists and radiologists and artificial intelligence: A systematic review," *Applied Sciences*, vol. 14, no. 12, p. 4972, 2024.

[8] M. Kiełczykowski, K. Kamiński, K. Perkowski, M. Zadurska, and E. Czochrowska, "Application of artificial intelligence (ai) in a cephalometric analysis: a narrative review," *Diagnostics*, vol. 13, no. 16, p. 2640, 2023.

[9] J. Londono, S. Ghasemi, A. H. Shah, A. Fahimipour, N. Ghadimi, S. Hashemi, Z. Khurshid, and M. Dashti, "Evaluation of deep learning and convolutional neural network algorithms accuracy for detecting and predicting anatomical landmarks on 2d lateral cephalometric images: a systematic review and meta-analysis," *The Saudi Dental Journal*, vol. 35, no. 5, pp. 487–497, 2023.

[10] M. S. Jha, "Cephalometric evaluation based on steiner's analysis on adults of bihar," *Journal of Pharmacy and Bioallied Sciences*, vol. 13, no. Suppl 2, pp. S1360–S1364, 2021.

[11] M. G. Hans, J. M. Palomo, and M. Valiathan, "History of imaging in orthodontics from broadbent to cone-beam computed tomography," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 148, no. 6, pp. 914–921, 2015.

[12] A. M. Aldrees, "Lateral cephalometric norms for saudi adults: A meta-analysis," *The Saudi dental journal*, vol. 23, no. 1, pp. 3–7, 2011.

[13] A. Dinesh, S. Mutalik, J. Feldman, and A. Tadinada, "Value-addition of lateral cephalometric radiographs in orthodontic diagnosis and treatment planning," *The Angle Orthodontist*, vol. 90, no. 5, pp. 665–671, 2020.

[14] G. Tsorovas and A. Linder-Aronson Karsten, "A comparison of hand-tracing and cephalometric analysis computer programs with and without advanced features—accuracy and time demands," *The European Journal of Orthodontics*, vol. 32, no. 6, pp. 721–728, 2010.

[15] T. Kanade, "Identification of human faces," in *Computer recognition of human faces*, pp. 47–63, Springer, 1977.

[16] J. Liu and J. K. Udupa, "Oriented active shape models," *IEEE Transactions on medical Imaging*, vol. 28, no. 4, pp. 571–584, 2008.

[17] T. F. Cootes and P. Kittipanya-ngam, "Comparing variations on the active appearance model algorithm.," in *BMVC*, pp. 1–10, Citeseer, 2002.

[18] C. Rahmad, R. A. Asmara, D. Putra, I. Dharma, H. Darmono, and I. Muhiqqin, "Comparison of viola-jones haar cascade classifier and histogram of oriented gradients (hog) for face detection," in *IOP conference series: materials science and engineering*, vol. 732, p. 012038, IOP Publishing, 2020.

[19] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1867–1874, 2014.

[20] A. Bulat, *Deep learning for real world face alignment*. PhD thesis, University of Nottingham, 2019.

[21] X. Lan, Q. Hu, Q. Chen, J. Xue, and J. Cheng, "Hih: Towards more accurate face alignment via heatmap in heatmap," *arXiv preprint arXiv:2104.03100*, 2021.

[22] A. Yueyuan and W. Hong, "Swin transformer combined with convolutional encoder for cephalometric landmarks detection," in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 184–187, IEEE, 2021.

[23] A. Tamhane, T. Mida, E. Posner, and M. Bouhnik, "Colonoscopy landmark detection using vision transformers," in *MICCAI Workshop on Imaging Systems for GI Endoscopy*, pp. 24–34, Springer, 2022.

[24] M. J. Rajabi, M. Mirzai, and A. Nickabadi, "Enhancing landmark detection in cluttered real-world scenarios with vision transformers," *arXiv preprint arXiv:2308.13671*, 2023.

[25] J.-H. Park, H.-W. Hwang, J.-H. Moon, Y. Yu, H. Kim, S.-B. Her, G. Srinivasan, M. N. A. Aljanabi, R. E. Donatelli, and S.-J. Lee, "Automated identification of cephalometric landmarks: Part 1—comparisons between the latest deep-learning methods yolov3 and ssd," *The Angle Orthodontist*, vol. 89, no. 6, pp. 903–909, 2019.

[26] F. Kunz, A. Stellzig-Eisenhauer, F. Zeman, and J. Boldt, "Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network," *J Orofac Orthop*, vol. 81, no. 1, pp. 52–68, 2020.

[27] H.-W. Hwang, J.-H. Park, J.-H. Moon, Y. Yu, H. Kim, S.-B. Her, G. Srinivasan, M. N. A. Aljanabi, R. E. Donatelli, and S.-J. Lee, "Automated identification of cephalometric landmarks: Part 2-might it be better than human?," *The Angle Orthodontist*, vol. 90, no. 1, pp. 69–76, 2020.

[28] H. Kim, E. Shim, J. Park, Y.-J. Kim, U. Lee, and Y. Kim, "Web-based fully automated cephalometric analysis by deep learning," *Computer methods and programs in biomedicine*, vol. 194, p. 105513, 2020.

[29] J.-H. Lee, H.-J. Yu, M.-j. Kim, J.-W. Kim, and J. Choi, "Automated cephalometric landmark detection with confidence regions using bayesian convolutional neural networks," *BMC oral health*, vol. 20, pp. 1–10, 2020.

[30] J. M. Noothout, B. D. De Vos, J. M. Wolterink, E. M. Postma, P. A. Smeets, R. A. Takx, T. Leiner, M. A. Viergever, and I. Išgum, "Deep learning-based regression and classification for automatic landmark localization in medical images," *IEEE transactions on medical imaging*, vol. 39, no. 12, pp. 4011–4022, 2020.

[31] Y. Song, X. Qiao, Y. Iwamoto, and Y.-w. Chen, "Automatic cephalometric landmark detection on x-ray images using a deep-learning method," *Applied Sciences*, vol. 10, no. 7, p. 2547, 2020.

[32] J. Qian, W. Luo, M. Cheng, Y. Tao, J. Lin, and H. Lin, "Cephann: a multi-head attention network for cephalometric landmark detection," *Ieee Access*, vol. 8, pp. 112633–112641, 2020.

[33] K. Oh, I.-S. Oh, D.-W. Lee, *et al.*, "Deep anatomical context feature learning for cephalometric landmark detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 806–817, 2020.

[34] M. Zeng, Z. Yan, S. Liu, Y. Zhou, and L. Qiu, "Cascaded convolutional networks for automatic cephalometric landmark detection," *Medical Image Analysis*, vol. 68, p. 101904, 2021.

[35] J. Kim, I. Kim, Y.-J. Kim, M. Kim, J.-H. Cho, M. Hong, K.-H. Kang, S.-H. Lim, S.-J. Kim, Y. H. Kim, *et al.*, "Accuracy of automated identification of lateral cephalometric landmarks using cascade convolutional neural networks on lateral cephalograms from nationwide multi-centres," *Orthodontics & Craniofacial Research*, vol. 24, pp. 59–67, 2021.

[36] C. Tanikawa, C. Lee, J. Lim, A. Oka, and T. Yamashiro, "Clinical applicability of automated cephalometric landmark identification: Part i—patient-related identification errors," *Orthodontics & Craniofacial Research*, vol. 24, pp. 43–52, 2021.

[37] Y. H. Kim, C. Lee, E.-G. Ha, Y. J. Choi, and S.-S. Han, "A fully deep learning model for the automatic identification of cephalometric landmarks," *Imaging Science in Dentistry*, vol. 51, no. 3, p. 299, 2021.

[38] T. He, J. Yao, W. Tian, Z. Yi, W. Tang, and J. Guo, "Cephalometric landmark detection by considering translational invariance in the two-stage framework," *Neurocomputing*, vol. 464, pp. 15–26, 2021.

[39] H. J. Kwon, H. I. Koo, J. Park, and N. I. Cho, "Multistage probabilistic approach for the localization of cephalometric landmarks," *IEEE Access*, vol. 9, pp. 21306–21314, 2021.

[40] L. Wang, L. Ma, Y. Li, K. Niu, and Z. He, "A dcnn system based on an iterative method for automatic landmark detection in cephalometric x-ray images," *Biomedical Signal Processing and Control*, vol. 68, p. 102757, 2021.

[41] J. Yao, W. Zeng, T. He, S. Zhou, Y. Zhang, J. Guo, and W. Tang, "Automatic localization of cephalometric landmarks based on convolutional neural network," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 161, no. 3, pp. e250–e259, 2022.

[42] V. N. T. Le, J. Kang, I.-S. Oh, J.-G. Kim, Y.-M. Yang, and D.-W. Lee, "Effectiveness of human–artificial intelligence collaboration in cephalometric landmark detection," *Journal of Personalized Medicine*, vol. 12, no. 3, p. 387, 2022.

[43] M. Lee, M. Chung, and Y.-G. Shin, "Cephalometric landmark detection via global and local encoders and patch-wise attentions," *Neurocomputing*, vol. 470, pp. 182–189, 2022.

[44] H.-J. Yoon, D.-R. Kim, E. Gwon, N. Kim, S.-H. Baek, H.-W. Ahn, K.-A. Kim, and S.-J. Kim, "Fully automated identification of cephalometric landmarks for upper airway assessment using cascaded convolutional neural networks," *European Journal of Orthodontics*, vol. 44, no. 1, pp. 66–77, 2022.

[45] S.-M. Gil, I. Kim, J.-H. Cho, M. Hong, M. Kim, S.-J. Kim, Y.-J. Kim, Y. H. Kim, S.-H. Lim, S. J. Sung, *et al.*, "Accuracy of auto-identification of the posteroanterior cephalometric landmarks using cascade convolution neural network algorithm and cephalometric images of different quality from nationwide multiple centers," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 161, no. 4, pp. e361–e371, 2022.

[46] M. Šavc, G. Sedej, and B. Potočnik, "Cephalometric landmark detection in lateral skull x-ray images by using improved spatialconfiguration-net," *Applied Sciences*, vol. 12, no. 9, p. 4644, 2022.

[47] K. Huang and F. Feng, "An intelligent shooting reward learning network scheme for medical image landmark detection," *Applied Sciences*, vol. 12, no. 20, p. 10190, 2022.

[48] Y. Ao and H. Wu, "Feature aggregation and refinement network for 2d anatomical landmark detection," *Journal of Digital Imaging*, vol. 36, no. 2, pp. 547–561, 2023.

[49] F. Jiang, Y. Guo, C. Yang, Y. Zhou, Y. Lin, F. Cheng, S. Quan, Q. Feng, and J. Li, "Artificial intelligence system for automated landmark localization and analysis of cephalometry," *Dentomaxillofacial Radiology*, vol. 52, no. 1, p. 20220081, 2023.

[50] J. V. Weingart, S. Schlager, M. C. Metzger, L. S. Brandenburg, A. Hein, R. Schmelzeisen, F. Bamberg, S. Kim, E. Kellner, M. Reisert, *et al.*, "Automated detection of cephalometric landmarks using deep neural patchworks," *Dentomaxillofacial Radiology*, vol. 52, no. 6, p. 20230059, 2023.

[51] S. Yang, E. S. Song, E. S. Lee, S.-R. Kang, W.-J. Yi, and S.-P. Lee, "Correction: 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*, vol. 24, 2024.

[52] R. Neeraja and L. J. Anbarasi, "Cephxnet: A deep convolutional squeeze-and-excitation model for landmark prediction on lateral cephalograms," *IEEE Access*, 2023.

[53] S. Rashmi, S. Srinath, K. Patil, P. S. Murthy, and S. Deshmukh, "Lateral cephalometric landmark annotation using histogram oriented gradients extracted from region of interest patches," *Journal of Maxillofacial and Oral Surgery*, vol. 22, no. 4, pp. 806–812, 2023.

[54] C. Zhao, Z. Yuan, S. Luo, W. Wang, Z. Ren, X. Yao, and T. Wu, "Automatic recognition of cephalometric landmarks via multi-scale sampling strategy," *Heliyon*, vol. 9, no. 6, 2023.

[55] B. Gong, S. Chang, F. Zuo, X. Xie, S. Wang, Y. Wang, Y. Sun, X. Guan, and Y. Bai, "Automated cephalometric landmark identification and location based on convolutional neural network," *Zhonghua kou Qiang yi xue za zhi= Zhonghua Kouqiang Yixue Zazhi= Chinese Journal of Stomatology*, vol. 58, no. 12, pp. 1249–1256, 2023.

[56] K. Takahashi, Y. Shimamura, C. Tachiki, Y. Nishii, and M. Hagiwara, "Cephalometric landmark detection without x-rays combining coordinate regression and heatmap regression," *Scientific reports*, vol. 13, no. 1, p. 20011, 2023.

[57] Z. Jiao, Z. Liang, Q. Liao, S. Chen, H. Yang, G. Hong, and H. Gui, "Deep learning for automatic detection of cephalometric landmarks on lateral cephalometric radiographs using the mask region-based convolutional neural network: a pilot study," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 137, no. 5, pp. 554–562, 2024.

[58] X. Li, S. Lv, J. Zhang, M. Li, J. J. Rodriguez-Andina, Y. Qin, S. Yin, and H. Luo, "Fdgr-net: Feature decouple and gated recalibration network for medical image landmark detection," *Expert Systems with Applications*, vol. 238, p. 121746, 2024.

[59] H.-T. Lee, P.-Y. Chiu, C.-W. Yen, S.-T. Chou, and Y.-C. Tseng, "Application of artificial intelligence in lateral cephalometric analysis," *Journal of Dental Sciences*, vol. 19, no. 2, pp. 1157–1164, 2024.

[60] C. Lu, G. Yang, X. Qiao, W. Chen, and Q. Zeng, "Univerdetect: Universal landmark detection method for multidomain x-ray images," *Neurocomputing*, vol. 600, p. 128157, 2024.

[61] M. A. Khalid, A. Khurshid, K. Zulfiqar, U. Bashir, and M. M. Fraz, "A two-stage regression framework for automated cephalometric landmark detection incorporating semantically fused anatomical features and multi-head refinement loss," *Expert Systems with Applications*, vol. 255, p. 124840, 2024.

[62] I. Tafala, F.-E. Ben-Bouazza, A. Edder, O. Manchadi, M. Et-Taoussi, and B. Jioudi, "Cephalometric landmarks identification through an object detection-based deep learning model.," *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 2, 2024.

[63] G. Bulatova, B. Kusnoto, V. Grace, T. P. Tsay, D. M. Avenetti, and F. J. C. Sanchez, "Assessment of automatic cephalometric landmark identification using artificial intelligence," *Orthodontics & Craniofacial Research*, vol. 24, pp. 37–42, 2021.

[64] R. K. Mahto, D. Kafle, A. Giri, S. Luintel, and A. Karki, "Evaluation of fully automated cephalometric measurements obtained from web-based artificial intelligence driven platform," *BMC Oral Health*, vol. 22, no. 1, p. 132, 2022.

[65] Y. A. Yassir, A. R. Salman, and S. A. Nabbat, "The accuracy and reliability of webceph for cephalometric analysis," *Journal of Taibah University Medical Sciences*, vol. 17, no. 1, pp. 57–66, 2022.

[66] T. P. Silva, M. M. Hughes, L. d. S. Menezes, M. d. F. B. de Melo, P. H. L. d. Freitas, and W. M. Takeshita, "Artificial intelligence-based cephalometric landmark annotation and measurements according to arnett's analysis: can we trust a bot to do that?," *Dentomaxillofacial Radiology*, vol. 51, no. 6, p. 20200548, 2022.

[67] B. Ristau, M. Coreil, A. Chapple, P. Armbruster, and R. Ballard, "Comparison of audaxceph®'s fully automated cephalometric tracing technology to a semi-automated approach by human examiners," *International orthodontics*, vol. 20, no. 4, p. 100691, 2022.

[68] I. A. Tsolakis, A. I. Tsolakis, T. Elshebiny, S. Matthaios, and J. M. Palomo, "Comparing a fully automated cephalometric tracing method to a manual tracing method for orthodontic diagnosis," *Journal of Clinical Medicine*, vol. 11, no. 22, p. 6854, 2022.

[69] G. Çoban, T. Öztürk, N. Hashimli, and A. Yağci, "Comparison between cephalometric measurements using digital manual and web-based artificial intelligence cephalometric tracing software," *Dental Press Journal of Orthodontics*, vol. 27, no. 4, p. e222112, 2022.

[70] M. Davidovitch, T. Sella-Tunis, L. Abramovicz, S. Reiter, S. Matalon, and N. Shpack, "Verification of convolutional neural network cephalometric landmark identification," *Applied Sciences*, vol. 12, no. 24, p. 12784, 2022.

[71] H. Bao, K. Zhang, C. Yu, H. Li, D. Cao, H. Shu, L. Liu, and B. Yan, "Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence," *BMC Oral Health*, vol. 23, no. 1, p. 191, 2023.

[72] H. Ye, Z. Cheng, N. Ungvijanpunya, W. Chen, L. Cao, and Y. Gou, "Is automatic cephalometric software using artificial intelligence better than orthodontist experts in landmark identification?," *BMC Oral Health*, vol. 23, no. 1, p. 467, 2023.

[73] S. Panesar, A. Zhao, E. Hollensbe, A. Wong, S. S. Bhamidipalli, G. Eckert, V. Dutra, and H. Turkkahraman, "Precision and accuracy assessment of cephalometric analyses performed by deep learning artificial intelligence with and without human augmentation," *Applied Sciences*, vol. 13, no. 12, p. 6921, 2023.

[74] F. Kunz, A. Stellzig-Eisenhauer, L. M. Widmaier, F. Zeman, and J. Boldt, "Assessment of the quality of different commercial providers using artificial intelligence for automated cephalometric analysis compared to human orthodontic experts," *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie*, pp. 1–16, 2023.

[75] J. Lee, S.-R. Bae, and H.-K. Noh, "Commercial artificial intelligence lateral cephalometric analysis: part 1–the possibility of replacing manual landmarking with artificial intelligence service.," *Journal of Clinical Pediatric Dentistry*, vol. 47, no. 6, 2023.

[76] L. d. S. Menezes, T. P. Silva, M. A. Lima dos Santos, M. M. Hughes, S. d. R. Mariano Souza, P. M. Leite Ribeiro, P. H. L. d. Freitas, and W. M. Takeshita, "Assessment of landmark detection in cephalometric radiographs with different conditions of brightness and contrast using the an artificial intelligence software," *Dentomaxillofacial Radiology*, vol. 52, no. 8, p. 20230065, 2023.

[77] T. P. Silva, M. C. R. Pinheiro, D. Q. Freitas, H. Gaêta-Araujo, and C. Oliveira-Santos, "Assessment of accuracy and reproducibility of cephalometric identification performed by 2 artificial intelligence-driven tracing applications and human examiners," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 137, no. 4, pp. 431–440, 2024.

[78] S. Gupta, S. Shetty, S. Natarajan, S. Nambiar, M. Ashith, and S. Agarwal, "A comparative evaluation of concordance and speed between smartphone app-based and artificial intelligence web-based cephalometric tracing software with the manual tracing method: A cross-sectional study," *Journal of clinical and experimental dentistry*, vol. 16, no. 1, p. e11, 2024.

[79] S. Bor, S. Ç. Ciğerim, and S. Kotan, "Comparison of ai-assisted cephalometric analysis and orthodontist-performed digital tracing analysis," *Progress in Orthodontics*, vol. 25, no. 1, p. 41, 2024.

[80] R. Zaheer, H. Z. Shafique, Z. Khalid, R. Shahid, A. Jan, T. Zahoor, R. Nawaz, and M. ul Hassan, "Comparison of semi and fully automated artificial intelligence driven softwares and manual system for cephalometric analysis," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 271, 2024.

[81] J.-P. Mercier, C. Rossi, I. N. Sanchez, I. D. Renovales, P. M.-P. Sahagún, and L. Templier, "Reliability and accuracy of artificial intelligence-based software for cephalometric diagnosis. a diagnostic study," *BMC Oral Health*, vol. 24, no. 1, pp. 1–19, 2024.

[82] N. Baig, K. S. Gyasudeen, T. Bhattacharjee, J. Chaudhry, and S. Prasad, "Comparative evaluation of commercially available ai-based cephalometric tracing programs," *BMC Oral Health*, vol. 24, no. 1, p. 1241, 2024.

[83] M. Hori, M. Jincho, T. Hori, H. Sekine, A. Kato, K. Miyazawa, and T. Kawai, "Automatic point detection on cephalograms using convolutional neural networks: A two-step method," *Dental Materials Journal*, vol. 43, no. 5, pp. 701–710, 2024.