Alahmari et al. BMC Oral Health (2025) 25:350
https://doi.org/10.1186/s12903-025-05730-y

RESEARCH

Open Access

Accuracy of artificial intelligence-based segmentation in maxillofacial structures: a systematic review

Manea Alahmari[1], Maram Alahmari[2], Abdulmajeed Almuaddi[1], Hossam Abdelmagyd[3], Kumuda Rao[4], Zaid Hamdoon[5], Mohammed Alsaegh[5], Nallan C. S. K. Chaitanya[6] and Shishir Shetty[5*]

Abstract

Objective The aim of this review was to evaluate the accuracy of artificial intelligence (AI) in the segmentation of teeth, jawbone (maxilla, mandible with temporomandibular joint), and mandibular (inferior alveolar) canal in CBCT and CT scans.

Materials and methods Articles were retrieved from MEDLINE, Cochrane CENTRAL, IEEE Xplore, and Google Scholar. Eligible studies were analyzed thematically, and their quality was appraised using the JBI checklist for diagnostic test accuracy studies. Meta-analysis was conducted for key performance metrics, including Dice Similarity Coefficient (DSC) and Average Surface Distance (ASD).

Results A total of 767 non-duplicate articles were identified, and 30 studies were included in the review. Of these, 27 employed deep-learning models, while 3 utilized classical machine-learning approaches. The pooled DSC for mandible segmentation was 0.94 (95% CI: 0.91–0.98), mandibular canal segmentation was 0.694 (95% CI: 0.551–0.838), maxilla segmentation was 0.907 (95% CI: 0.867–0.948), and teeth segmentation was 0.925 (95% CI: 0.891–0.959). Pooled ASD values were 0.534 mm (95% CI: 0.366–0.703) for the mandibular canal, 0.468 mm (95% CI: 0.295–0.641) for the maxilla, and 0.189 mm (95% CI: 0.043–0.335) for teeth. Other metrics, such as sensitivity and precision, were variably reported, with sensitivity exceeding 90% across studies.

Conclusion AI-based segmentation, particularly using deep-learning models, demonstrates high accuracy in the segmentation of dental and maxillofacial structures, comparable to expert manual segmentation. The integration of AI into clinical workflows offers not only accuracy but also substantial time savings, positioning it as a promising tool for automated dental imaging.

Keywords Artificial intelligence, Segmentation, Dental, Jaws, Dental canal, Temporomandibular joint

*Correspondence:
Shishir Shetty
shishirshettyomr@gmail.com
[1]College of Dentistry, King Khalid University, Abha, Saudi Arabia
[2]Armed Forces Hospital Southern Region, Khamis Mushait, Saudi Arabia
[3]College of Dentistry, Suez Canal University, Ajman, United Arab Emirates
[4]AB Shetty Memorial Institute of Dental Sciences, Nitte (Deemed to be University), Mangalore, India
[5]College of Dental Medicine, University of Sharjah, Sharjah, United Arab Emirates
[6]College of Dental Sciences, RAK Medical and Health Sciences University, Ras-Al-Khaimah, United Arab Emirates

Introduction

Computed Tomography (CT) and Cone Beam Computed Tomography (CBCT) have emerged as indispensable tools in the field of oral and maxillofacial surgery (OMFS) over recent years [1–3]. These imaging modalities provide detailed three-dimensional (3D) representations of the teeth, jaw bones, temporomandibular (TM) joint, and surrounding anatomy, and aiding clinicians in treatment planning and diagnosis [3]. However, manually segmenting these structures from CBCT and CT images is a time-consuming and laborious task, often susceptible to variability between operators, which affects the reliability and reproducibility of the segmentation results [4–6]. This approach is challenging and time-consuming, due to the complex and variable anatomy of the oral and maxillofacial region [7].

Research into the application of Artificial Intelligence (AI) techniques in the segmentation of dental structures, has recently garnered scholarly attention. AI-based segmentation techniques can be broadly categorized into several approaches, which are Active Shape Model-based (ASM-based), Statistical Shape Model-based (SSM-based), Active Appearance Model-based (AAM-based), Atlas-based, Level Set-based, Classical Machine Learning-based, and Deep Learning-based [8]. Classical machine learning includes algorithms, like random forests, support vector machines (SVM), and k-nearest neighbors (k-NN), which are trained on image features to perform segmentation tasks [8].

Deep learning on the other hand is based on artificial neural networks (ANNs), which can be categorized into either convolutional neural networks (CNNs) or recurrent neural networks (RNNs) [9]. CNNs contain convolutional, pooling, and fully connected layers [9]. Compared to classical machine learning, deep learning methods offer greater flexibility and more powerful capabilities. They also require less expert analysis, making it easier to extend them to other segmentation tasks [10]. CNNs are structured to learn basic patterns in the initial layers and progress to more complex patterns in deeper layers. This hierarchical learning approach allows deep neural networks to manage complex tasks without an excessive increase in the number of parameters [11].

In recent years, new AI architectures, such as Segment Anything Models (SAM) and transformer-based networks, have emerged as transformative tools for medical image segmentation. SAMs, for example, offer the ability to handle diverse segmentation tasks with minimal manual input, making them particularly suitable for complex maxillofacial structures like the mandibular canal and teeth [12–14]. Transformer-based architectures, such as Vision Transformers (ViT) and MAMBA, utilize attention mechanisms to capture long-range dependencies, enhancing segmentation accuracy and robustness [15–16]. These models represent a significant leap forward, addressing challenges faced by traditional CNNs in terms of generalization and adaptability.

Deep CNNs have become increasingly adopted in the field of medical image segmentation [17, 18], and have shown exceptional performance [19–22]. According to Singh & Raza [22], the effectiveness of CNNs is primarily attributed to their capacity to learn complex spatial features, representative and predictive, within input images. Several studies have employed CNNs for segmenting dental structures, typically involving binary segmentation tasks, and have demonstrated their capability to achieve precise segmentation results [6, 23, 24].

AI algorithms, particularly those based on deep learning, have shown significant achievement in automating and standardizing the segmentation process, leading to improved efficiency, and accuracy of dental structure segmentation. AI-based segmentation techniques have been shown to be effective in tasks such as planning dental implants, planning orthodontic treatments, and evaluating temporomandibular joint disorders. Therefore, there is a need to determine the performance of AI-based segmentation in maxillofacial structures. This review evaluates AI for the segmentation of teeth, jawbone (maxilla, mandible with TM joint), and mandibular (inferior alveolar) canal in CBCT and CT, with a specific focus on the use of machine learning and deep learning. The focused research question of the present review was as follows "How effective are AI-based segmentation methods in delineating teeth, jaw bones, and mandibular canals in CBCT and CT scans?" Based on the PICO model, the components of the research question were as follows: Population: CT and CBCT scans of the human maxillofacial structure, Intervention: AI-based segmentation methods, Comparison: Ground truth annotations by human experts, Outcome: Performance of segmentation measured through metrics such as DSC, ASD, IoU, and HD.

Materials and methods

Study design

This systematic review and meta-analysis were conducted following the 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure methodological rigor [25]. The primary aim was to evaluate the accuracy and efficacy of AI-based segmentation models for dental structures, including the mandible, mandibular canal, maxilla, and teeth, based on computed tomography (CT) and cone beam computed tomography (CBCT) imaging.

Information sources

Electronic searches were conducted across three databases: PubMed (via MEDLINE), and Cochrane Library.

Additionally, manual searches were performed in Google Scholar, and gray literature was screened in IEEE Xplore. The search strategies for the databases are provided in the supplementary material. The last database search was completed on 1st November 2024.

Reference lists of included studies and similar reviews were screened for additional relevant citations to ensure comprehensive coverage. The search was designed to identify studies published in English with no restriction on publication dates.

Search strategy

The search strategy incorporated a combination of controlled vocabulary terms (MeSH terms for MEDLINE) and free-text keywords to account for variations in terminology. Boolean operators (AND, OR) were used to combine search terms, along with truncation symbols to capture singular and plural variations. Key terms included: Artificial Intelligence Terminology: "artificial intelligence," "machine learning," "deep learning," "neural networks," "CNN," Anatomical Structures: "teeth," "jaw," "mandible," "maxilla," "inferior alveolar canal," "temporomandibular joint," Imaging Modalities: "CBCT," "CT," "cone beam CT," "computed tomography."

Expanded search strategies specific to each database are presented in Table 1. The updated terms for MEDLINE and Cochrane Library included broader anatomical descriptors.

Eligibility criteria

This review included original clinical studies, written in English, and published before 1st November 2024. For a study to be eligible it had to include machine learning or deep learning-based segmentation models using CT or CBCT scans. The anatomy of focus in the study was teeth, jaw bone (mandible and maxilla) with TMJ and mandibular canal. Non-journal articles and those written in languages other than English were excluded. Systematic reviews, meta-analyses papers, literature reviews, conference papers, commentaries, letters to editors, and other non-original clinical studies were excluded. Studies that had other forms of AI–based segmentation different from machine and deep learning models were excluded. Studies involving AI segmentation of sinonasal structures, pharyngeal airway, cranial structures, orbit and its contents were excluded from the systematic review.

Screening and data extraction

Two independent reviewers conducted the screening process, resolving any disagreements with the involvement of a third reviewer. Duplicates were removed in Zotero, and full-text articles were assessed for eligibility.

Data were extracted into structured tables, capturing:

1. Study characteristics: author, publication year, sample size, imaging modality, AI model type, and anatomical structures segmented.
2. Quantitative metrics: mean and standard deviation of DSC, ASD, HD.
3. Additional results, such as segmentation time and model comparison metrics.

Where data were incomplete (e.g., missing SDs), they were imputed using values from similar studies. For studies reporting ranges, SDs were estimated using the formula: SD = Range/SD.

Table 1 Search strings used in the study

Database	Search string
MEDLINE	("Artificial Intelligence"[All Fields] OR "Artificial Intelligence"[MeSH Terms] OR "machine learning"[All Fields] OR "Deep Learning"[All Fields] OR "Deep Learning"[MeSH Terms] OR "neural networks"[All Fields] OR "CNN"[All Fields]) AND ("teeth s"[All Fields] OR "teeths"[All Fields] OR "tooth"[MeSH Terms] OR "tooth"[All Fields] OR "teeth"[All Fields] OR "tooth s"[All Fields] OR "tooths"[All Fields] OR ("jaw"[MeSH Terms] OR "jaw"[All Fields]) OR ("maxilla"[MeSH Terms] OR "maxilla"[All Fields] OR "maxillae"[All Fields] OR "maxillas"[All Fields]) OR ("mandible"[MeSH Terms] OR "mandible"[All Fields] OR "mandibles"[All Fields] OR "mandible s"[All Fields]) OR ("mandible"[MeSH Terms] OR "mandible"[All Fields] OR "mandibular"[All Fields] OR "mandibulars"[All Fields]) OR "inferior alveolar canal"[All Fields] OR "temporomandibular joint"[All Fields] OR ("temporomandibular joint"[MeSH Terms] OR ("temporomandibular"[All Fields] AND "joint"[All Fields]) OR "temporomandibular joint"[All Fields] OR "tmj"[All Fields])) AND ("CBCT"[All Fields] OR ("j comput tomogr"[Journal] OR "commun theory"[Journal] OR "child teenagers"[Journal] OR "cancer ther"[Journal] OR "ct"[All Fields]) OR "Computed Tomography"[All Fields] OR "Cone-Beam CT"[All Fields])
Cochrane CENTRAL	("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" OR CNN) AND (teeth OR jaw OR maxilla OR mandible OR mandibular OR "inferior alveolar canal" OR "temporomandibular joint" OR TMJ) AND (CBCT OR CT OR "Computed Tomography" OR "Cone-Beam CT") in Title Abstract Keyword
Google Scholar	('artificial intelligence' OR 'machine learning' OR 'deep learning' OR 'neural networks' OR CNN) AND (segmentation OR delineation) AND (teeth OR jaw OR maxilla OR mandible OR mandibular OR 'inferior alveolar canal' OR 'temporomandibular joint' OR TMJ) AND (CBCT OR CT OR 'Computed Tomography' OR 'Cone-Beam CT')
IEEE Xplore	("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" OR "convolutional neural networks" OR CNN OR "U-Net") AND (segmentation OR delineation OR identification OR detection) AND (teeth OR tooth OR "dental structures" OR jaw OR maxilla OR mandible OR "mandibular canal" OR "temporomandibular joint" OR TMJ) AND ("computed tomography" OR CT OR CBCT OR "cone-beam CT" OR "cone beam computed tomography")

Assessment of methodological quality

The JBI checklist for diagnostic test accuracy studies was used for quality appraisal [26]. There were slight modifications in the wording of Item 1 (Modified to focus on the appropriateness of segmentation techniques rather than diagnostic tests) and Item 4 (Adapted to assess the accuracy of ground truth comparisons in segmentation), to fit the type of studies included in this review. This was done because the included studies evaluated the performance of a technological model, not a diagnostic test as is the case with 'traditional' diagnostic test accuracy studies. The overall quality of the study was determined based on the number of 'Yes' scores received, "Good" (8–10 Yes scores), "Fair" (4–7), and "Poor" (0–3).

Statistical analysis

The meta-analysis was conducted using R software with the *meta* and *metafor* packages. Metrics analyzed included DSC and ASD, with standard errors calculated as the standard deviation (SD) divided by the root of the sample (n).

A random-effects model was applied using the Restricted Maximum Likelihood (REML) method to account for between-study heterogeneity. Heterogeneity was assessed using the I^2 statistic, with values interpreted as: Low (< 25%), Moderate (25–75%), High (> 75%).

Forest plots were generated for each anatomical region (mandible, mandibular canal, maxilla, and teeth) to visualize pooled estimates and confidence intervals. Studies not meeting pooling criteria were described narratively.

Results

Database search yielded 807 articles; 25 clinical trials from CENTRAL, 488 studies from PubMed, 94 articles from IEEE Xplore, and 300 studies from Google Scholar. 767 non-duplicate articles were subjected to title and abstract screening, and 722 were excluded due to topic irrelevancy. 23 articles were excluded during the full-text review because they reported on other segmentation approaches, which were not machine learning or deep learning. The level-set-based method was the most reported among the excluded studies. 18 articles from the study selection and 12 articles from citation searching were included in this review. Figure 1 below shows the selection process.

Study characteristics

After filtration of the articles based on the review criteria, 30 studies qualified for the systematic review. Among these, 27 studies employed deep-learning models, whereas the remaining 3 studies utilized classical machine-learning approaches [4, 48, 49]. The most commonly used imaging modality was Cone Beam Computed Tomography (CBCT), reported in 24 studies, while the remainder used conventional CT imaging.

The studies were distributed across different anatomical targets, with 9 studies focusing on mandibular bone segmentation and another 9 studies specifically targeting individual teeth segmentation. Ground truth/reference for model evaluation was predominantly derived from manual segmentation performed by radiologists, dentists, or surgeons. However, semi-automatic segmentation was used as the reference in the studies by Hsu et al., Lahoud et al., Pankert et al., and Verhelst et al. [31, 35, 40, 46].

Table 2 provides a comprehensive overview of the study characteristics, including the imaging modalities, AI model types, anatomical targets, and evaluation metrics.

Results of quality assessment

Among the selected studies, 23 of the studies scored "Good", 7 scored "Fair" and none scored "Poor", in terms of overall quality Table 3.

Results of included studies

Mandible segmentation

The mandible was the most studied anatomical site, with 18 studies included in this review. Performance was evaluated using a range of metrics, with DSC and ASD being the most frequently reported. The pooled DSC for mandible segmentation was 0.94 (95% CI: 0.91–0.98), reflecting a high level of accuracy. Individual studies reported DSC values ranging from 0.89 [37] to 0.972 ± 0.01 [37]. The high-performing studies utilized deep learning methods, specifically CNNs, often enhanced with advanced preprocessing techniques such as artifact reduction [40, 46]. Variability in DSC may stem from differences in dataset size, imaging modalities, and annotation quality. ASD values ranged from 0.217 mm [43] to 1.2827 ± 0.2780 mm [8]. The variability likely reflects anatomical complexity and the differing objectives of studies (e.g., cortical bone vs. marrow bone segmentation). Studies employing multi-scale or hierarchical architectures demonstrated lower ASD values, indicating higher accuracy (Figure 2).

Intersection over Union (IoU) values were reported in three studies, ranging from 0.734 by Kim et al., cortical bone [34] to 0.954 [36]. Hausdorff Distance (HD) varied significantly, with values of 2.656 mm reported by [42] and 4.1583 ± 2.8549 mm [46]. Precision ranged from 0.68 to 0.99, with sensitivity values above 0.93 in most studies. Studies utilizing deep learning models, particularly CNNs and U-Nets, consistently outperformed classical machine learning methods. The use of larger and more diverse datasets resulted in improved segmentation performance, highlighting the importance of training data. Variability in reported metrics highlights the need for standardization in evaluation and reporting practices.

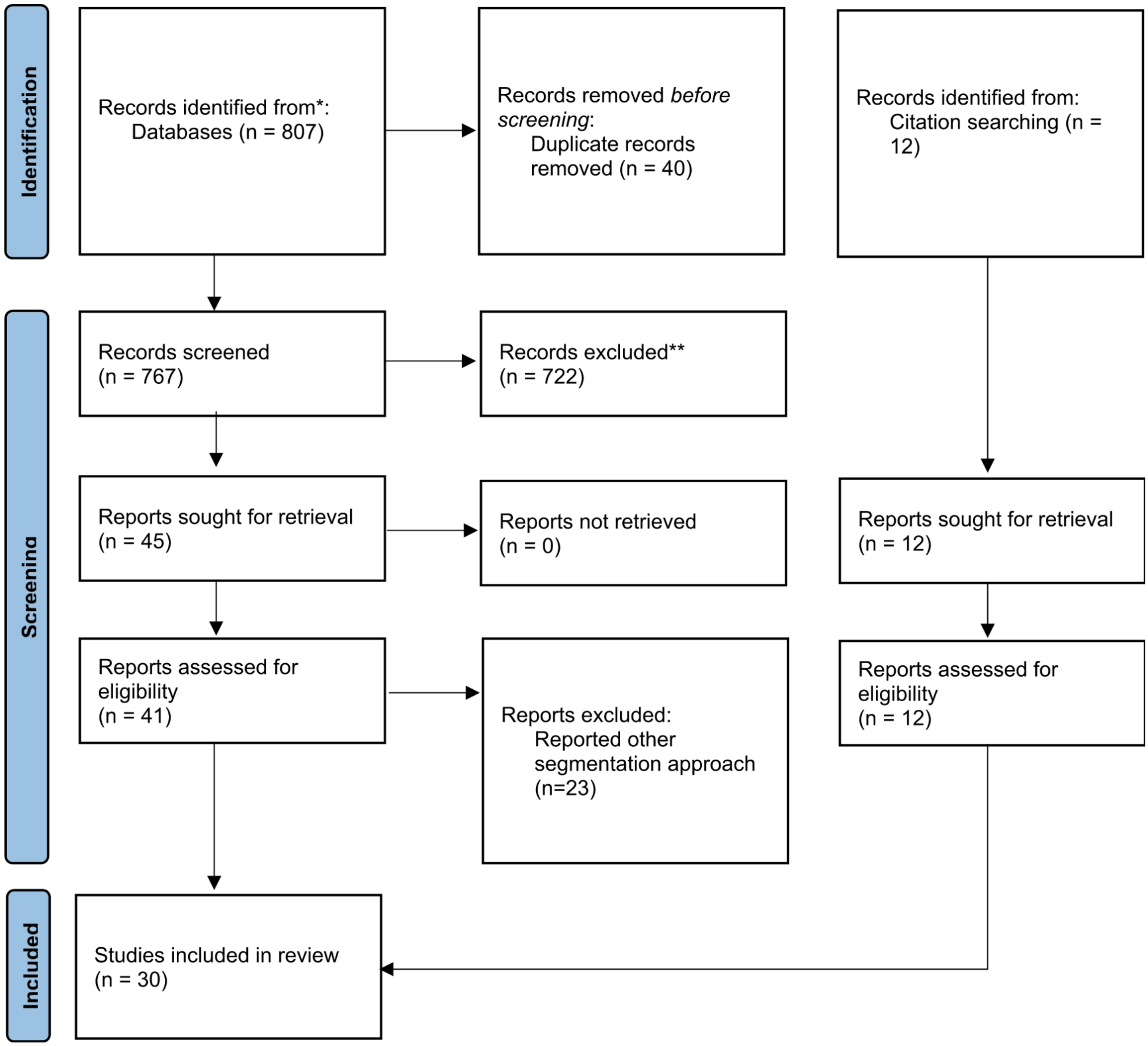

Fig. 1 PRISMA flowchart of the systematic review

Mandibular canal segmentation

Mandibular canal segmentation was analyzed in five studies, with performance evaluated using Dice Similarity Coefficient (DSC), Average Surface Distance (ASD), and other metrics such as Hausdorff Distance (HD) and Intersection over Union (IoU). This anatomical site presented unique challenges due to its small and complex structure, requiring models with high precision for accurate delineation.

The meta-analysis of DSC for mandibular canal segmentation included 4 studies, yielding a pooled DSC of 0.694 (95% CI: 0.551–0.838) under the random-effects model. High heterogeneity ($I^2 = 100\%$) was observed, likely due to differences in segmentation methodologies, dataset size, and reference standards.

Individual study results for DSC varied significantly, ranging from 0.570 ± 0.008 (Jaskari et al. [33], left canal) to 0.875 ± 0.0045 [36]. Studies employing advanced deep learning architectures, such as two-stage U-Net pipelines, achieved higher performance compared to simpler CNN models.

The meta-analysis of ASD included 4 studies, resulting in a pooled ASD of 0.534 mm (95% CI: 0.366–0.703) under the random-effects model. Heterogeneity was high (I2 = 100%), reflecting significant variability in anatomical complexity and segmentation protocols across studies.

ASD values ranged from 0.442 mm (± 0.0571) [36] to 0.785 mm (± 0.0009) [45]. Notably, studies using attention-based or hierarchical segmentation architectures

Table 2 Study descriptors including the anatomical site, imaging modality, segmentation approach and evaluation matrix

Study	Subjects	Part	Training and evaluation data	Image Modality	Segmentation approach	Evaluation metrics	Results
Chen, Wang, et al. [4]	30 subjects, 20 females	maxilla	30 scans for training, 6 for testing	CBCT	ML, Learning-based multi-source IntegratioN frameworK for Segmentation (LINKS)	DSC	The average Dice ratio of the maxilla was 0.800 ± 0.029.
Chen, Du, et al. [27]	25 subjects, 11 females, age range of 10 to 49 years	teeth	20 scans for training, 5 for testing	CBCT	DL, multi-task 3D fully convolutional network (FCN), and marker-controlled watershed transform (MWT)	DSC, Jaccard index	The average Dice similarity coefficient, and Jaccard index were 0.936 (± 0.012) and 0.881 (± 0.019).
Chung et al. [28]	NR	teeth	150 scans for training, 25 for testing	CBCT	DL, single CNN, base architecture of the 3D U-net	Aggregate Jaccard index (AJI), ASSD	AJI score of 0.86 ± 0.01, and ASSD [mm] of 0.20 ± 0.10
Cui et al. [29]	4,215 subjects, 2312 females, mean age of 38.4 years	teeth, alveolar bone	4531 images for training, 407 for testing	CBCT	DL	Dice ratio, ASD	Average Dice scores of 94.1%, ASD (mm) of 0.17 for tooth. Average Dice scores of 94.1 (76.9–96.9) %, ASD (mm) of 0.35 (0.18–0.84) for maxillary bone. Average Dice scores of 94.8 (80.3–97.3) %, ASD (mm) of 0.29 (0.13–0.77) for mandible bone.
Duan et al. [30]	NR	teeth	20 sets of images for training and testing	CBCT	DL, Region Proposal Network (RPN) with Feature Pyramid Network (FPN) method	DSC, ASD, RVD	They achieved an average dice of 95.7% (0.957 ± 0.005) for single-rooted tooth ST, 96.2% (0.962 ± 0.002) for multirooted tooth MT. The ASD was 0.104 ± 0.019 for ST and 0.137 ± 0.019 for MT. The RVD was 0.049 ± 0.017 for ST and 0.053 ± 0.010 for MT.
Hsu et al. [31]	24 subjects, 9 females, mean age of 29.1 ± 14.7 years	teeth	24 sets of images for training and testing	CBCT	DL, 3.5D U-Net	DSC	The 3.5 Dv5 U-Net achieved highest DSC among all U-Nets.
Ileșan et al. [32]	NR	mandible	120 scans for training, 40 for validation	CBCT and CT	DL, CNN, a 3D U-Net	DSC	Mean DSC was 0.884 for teeth segmentation and 0.894 for mandible segmentation
Jaskari et al. [33]	594 subjects	mandibular canal	457 scans for training, 52 for validation, 128 for testing	CBCT	DL, fully CNN	DSC, MCD	DSC was 0.57 for the left canal and 0.58 for the right canal, MCD was 0.61 mm for the left canal and 0.50 mm for the right canal.
Kim et al. [34]	25 subjects	mandibular condyle	18 subject cases for training, 5 for validation, 2 for training	CBCT	DL, modified U-Net, and a CNN	IoU, HD	Intersection over union (IoU) was 0.870 ± 0.023 for marrow bone and 0.734 ± 0.032 for cortical bone. The Hausdorff distance was 0.928 ± 0.166 mm for marrow bone and 1.247 ± 0.430 mm for cortical bone.
Kwak et al. [5]	102 subjects, age range of 18–90 years	mandibular canal	29, 456 images for training, 9818 for validation, 9818 for testing	CBCT	DL, deep CNN	Global classification accuracy	Global accuracy of 0.99 using a 3D U-Net CNN model for segmentation of mandibular canal
Lahoud et al. [35]	46 subjects	teeth	2095 samples for training, 501 for optimization, 328 for validation	CBCT	DL, deep CNN	IoU,	The mean intersection over union IOU for full-tooth segmentation was 0.87 (60.03) and 0.88 (60.03) for semiautomated (SA)

Table 2 (continued)

Study	Subjects	Part	Training and evaluation data	Image Modality	Segmentation approach	Evaluation metrics	Results
Lee et al. [24]	102 subjects	teeth	1066 images for training, 400 for validation, 151 for testing	CBCT	DL, CNN	Dice ratio	Dice value of 0.918.
Lin et al. [36]	220 subjects, 136 females, mean age of 36.93 ± 13.77 years	mandibular canal	132 images for training, 44 for validation, 44 for testing	CBCT	DL, CNN, two-stage 3D-Unet	DSC, 95% HD	The mean DSC was 0.875 ± 0.045 and the mean 95% HD was 0.442 ± 0.379.
Lo Giudice et al. [37]	40 subjects, 20 females, mean age of 23.37 ± 3.34 years	mandible	20 scans for training, 20 for testing	CBCT	DL, CNN	DSC, surface-to-surface matching	DSC of 0.972.
Macho et al. [38]	40 subjects	teeth	36 images for training, 4 for validation	CT	DL, CNN employing 3D volumetric convolutions	dice ratio	Dice value between 0.88 and 0.94.
Miki et al. [39]	52 subjects	teeth	42 subject cases for training, 10 for testing	CBCT	DL, deep CNN	classification accuracy	Classification accuracy 88.8%
Minnema et al. [20]	NR	teeth	20 scans for training,2 for validation, 18 for testing	CBCT	DL, mixed-scale dense (MS-D) CNN	DSC	DSC of 0.87 ± 0.06
Pankert et al. [40]	307 subjects, 112 females, mean age of 63 years	mandible	248 images for testing, 30 for validation, 29 for testing	CT	DL, two-stepped CNN	DSC, ASD, HD	Dice coefficient of 94.824% and an average surface distance of 0.31 mm.
Park et al. [41]	171 subjects	mandible, maxilla	146 sets for training, 10 for validation, 15 for testing	CT	DL, hierarchical, parallel, and multi-scale residual block to the U-Net (HPMR-U-Net)	DSC, ASD, 95HD	DC of 90.2 ± 19.5 for maxilla segmentation and 97.4 ± 0.4 for mandible segmentation
Qiu et al. [6]	11 subjects	mandible	8 scans for training, 2 for validation, 1 for testing	CT	DL, CNN	DSC	Average dice coefficient of 0.89
Qiu et al. [42]	48 subjects	mandible	52 scans for training, 8 for validation, 49 for testing	CT	DL, Single-planar CNN	DSC, 3D surface error	Average dice score of0.9328(± 0.0144), 95HD (mm) of 1.4333(± 0.5564)
Qiu, Guo, et al. [43]	48 subjects	mandible	90 scans for training, 2 for validation, 17 for testing	CT	DL Recurrent Convolutional Neural Networks for Mandible Segmentation (RCNNSeg)	DSC, ASD, 95HD	Average DSC of 97.48%, ASD of 0.2170 mm, and 95HD of 2.6562 mm
Qiu, van der Wel, et al. [44]	59 subjects	mandible	38 subject cases for training, 1 for validation, 20 for testing	CBCT	DL, SASeg	DSC, ASD, 95HD	Average DSC of 95.35 (± 1.54)%, ASD of0.9908 (± 0.4128) mm, and 95HD of 2.5723 (± 4.1192) mm
Qiu, van der Wel, et al. [8]	59 subjects	mandible	38 scans for training, 1 for validation, 20 for testing	CBCT	DL, 3D CNN and recurrent SegUnet	DSC, ASD, 95HD	Average DSC of 95.31 (± 1.11)%, ASD of 1.2827 (± 0.2780) mm, and 95HD of 3.1258 (± 3.2311) mm
Usman et al. [45]	NR	mandibular canal	400 scans for training, 500 for testing	CBCT	DL, Attention-Based CNN model with Multi-Scale input Residual UNet (MSiR-UNet)	Dice ratio, mean IoU,	A dice score of 0.751, mean IoU of 0.795
Verhelst et al. [46]	NR	mandible	160 scans for training, 30 for testing	CBCT	DL, 2-stageed 3D U-Net	DSC, IoU, HD,	IoU of 94.6 ± 1.17%, DSC of 0.9722 ± 0.0062, HD (mm) of 4.1583 ± 2.8549,

Table 2 (continued)

Study	Subjects	Part	Training and evaluation data	Image Modality	Segmentation approach	Evaluation metrics	Results
Vinayahalingam et al. [47]	81 subjects	mandibular condyles and glenoid f	130 scans for training, 24 for validation, 8 for testing	CBCT	DL, 3D U-Net	Dice ratio, IoU,	The dice ratio was 0.976 (0.01) and IoU was 0.954 (0.02) for mandibular condyle segmentation. The dice ratio was 0.966 (0.03) and IoU was 0.936 (0.04) for Glenoid fossa segmentation.
Wang et al. [7]	30 subjects, 19 females, mean age of 14.2 ± 3.4 years	mandible, maxilla, teeth	21 scans for training, 7 for testing	CBCT	DL, mixed-scale dense (MS-D) CNN	DSC, surface deviation	Dice similarity coefficient: 0.934 ± 0.019, jaw; 0.945 ± 0.021, teeth
Wang et al. [48]	15 subjects, 9 females, mean age of 26 ± 10 years	mandible, maxilla	30 scans for training, 15 scans for validation	CBCT	ML, patch-based sparse representation, and convex optimization	DSC	Dice ratio of 0.92 ± 0.02 for mandible segmentation and 0.87 ± 0.02 for maxilla segmentation.
Wang et al. [49]	30 subjects, 18 females, mean age of 24 ± 10 years	mandible, maxilla	30 scans for training, 30 CBCT scans, and 60 spiral multislice CT (MSCT) scans for validation	CBCT	ML, prior-guided sequential random forests	DSC	Dice ratio of 0.94 ± 0.02 for mandible segmentation and 0.91 ± 0.03 for maxilla segmentation

DL, Deep learning, ML, Machine learning, CNN, convolutional neural network; DSC, Dice similarity Coefficient; HD, Hausdorff distance; 95HD, 95% Hausdorff distance; ASD, average surface distance; RVD, Relative Volume Difference; ASSD, Average Symmetric Surface Distance; JSC, Jaccard similarity coefficient, IoU, intersection over union; RMS, root mean square; MCD, mean curve distance; RVD, relative volume difference

demonstrated better performance, with lower ASD values (Figs. 3 and 4).

Regarding Hausdorff Distance (HD), Jaskari et al. [33] reported HD values of 1.40 ± 0.63 mm (left canal) and 1.38 ± 0.47 mm (right canal), reflecting moderate discrepancies between predicted and reference segmentations. IoU values ranged from 0.734 ± 0.032 [34] to 0.895 ± 0.06 [47]. Fully automated models demonstrated faster segmentation times compared to semi-automated approaches.

Mandibular canal segmentation results indicate the utility of advanced deep learning architectures (multi-stage U-Net and attention-based CNNs) in achieving higher DSC and lower ASD values. Variability in imaging modalities, model architectures, and training datasets likely contributed to heterogeneity. Studies employing CBCT imaging tended to report slightly higher segmentation accuracy compared to those using CT, although small sample sizes limit definitive conclusions.

Maxilla segmentation

Maxilla segmentation was evaluated in 3 studies, with metrics including Dice Similarity Coefficient (DSC) and Average Surface Distance (ASD). This anatomical site posed moderate challenges due to its complex structure and variable anatomical landmarks.

The meta-analysis of DSC for maxilla segmentation included 3 studies, yielding a pooled DSC of 0.907 (95% CI: 0.867–0.948) under the random-effects model. High heterogeneity ($I^2 = 99\%$) was observed, reflecting differences in model architectures, dataset size, and segmentation objectives.

Reported DSC values ranged from 0.870 ± 0.0052 [7] to 0.941 ± 0.0025 [29]. The high-performing study utilized a deep learning model with advanced preprocessing techniques, enabling accurate segmentation of maxillary structures.

The meta-analysis of ASD included the same 3 studies, resulting in a pooled ASD of 0.468 mm (95% CI: 0.295–0.641) under the random-effects model. Heterogeneity was high ($I^2 = 96\%$), likely due to variability in imaging modalities and segmentation protocols.

Reported ASD values ranged from 0.350 ± 0.0023 mm [29] to 0.650 ± 0.0491 mm [49]. Lower ASD values were observed in studies employing multi-scale or hierarchical models (Figs. 5 and 6).

Reported sensitivity for maxilla segmentation was high, with values exceeding 90% in Cui et al. [29]. Precision metrics were rarely reported but were consistent with high DSC values. Cui et al. [29] reported an IoU of 0.92, indicating strong alignment between predicted and reference segmentations. Fully automated segmentation was faster compared to manual methods, though time metrics were inconsistently reported across studies.

Studies employing deep learning architectures, especially CNN-based pipelines, demonstrated robust segmentation performance with high DSC and low ASD values. Variations in imaging modality, dataset quality, and evaluation methods were major contributors to heterogeneity. Multi-scale architectures and preprocessing

Table 3 Results of quality assessment of the studies

Study	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Overall quality
Chen, Wang, et al. [4]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Chen, Du, et al. [27]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Chung et al. [28]	Unclear	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Cui et al. [29]	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Good
Duan et al. [30]	Unclear	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Hsu et al. [31]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Ileșan et al. [32]	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Jaskari et al. [33]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Kim et al. [34]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Kwak et al. [5]	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Good
Lahoud et al. [35]	Yes	Yes	Unclear	Unclear	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Lee et al. [24]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Lin et al. [36]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Lo Giudice et al. [37]	Yes	Yes	Unclear	Yes	No	Yes	Yes	Yes	Yes	Yes	Good
Macho et al. [38]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Miki et al. [39]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Good
Minnema et al. [20]	Unclear	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Pankert et al. [40]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Park et al. [41]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Good
Qiu et al. [6]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Qiu et al. [42]	Yes	Yes	Yes	Unclear	No	Yes	Unclear	Yes	Yes	Yes	Good
Qiu, Guo, et al. [43]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Qiu, van der Wel, et al. [44]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Qiu, van der Wel, et al. [8]	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Usman et al. [45]	Unclear	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Verhelst et al. [46]	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Vinayahalingam et al. [47]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good
Wang et al. [7]	Yes	Yes	Yes	Unclear	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Wang et al. [48]	Yes	Yes	Yes	Unclear	Unclear	Yes	Unclear	Yes	Yes	Yes	Fair
Wang et al. [49]	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Good

Item (1) Was a consecutive or random dataset used? Item (2) Was a case control design avoided? Item (3) Did the study avoid inappropriate exclusions? Item (4) Were the index method results interpreted without knowledge of the results of the reference standard? Item (5) If a threshold was used, was it pre-specified? Item (6) Is the reference standard likely to correctly classify the target condition? Item (7) Were the reference standard results interpreted without knowledge of the results of the index test? Item (8) Was there an appropriate interval between index test and reference standard? Item (9) Did all patients receive the same reference standard? Item (10) Were all patients included in the analysis?

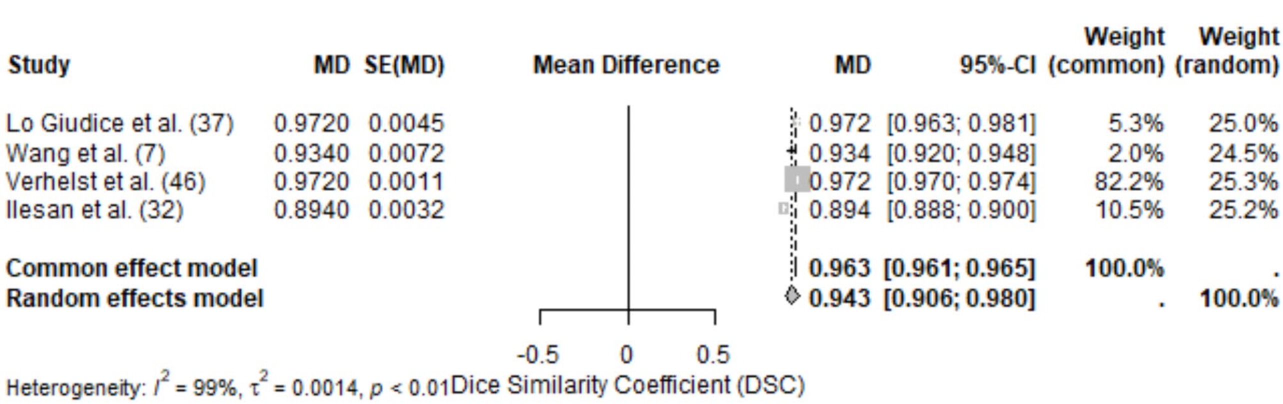

Fig. 2 Forest plot of DSC values for mandible segmentation

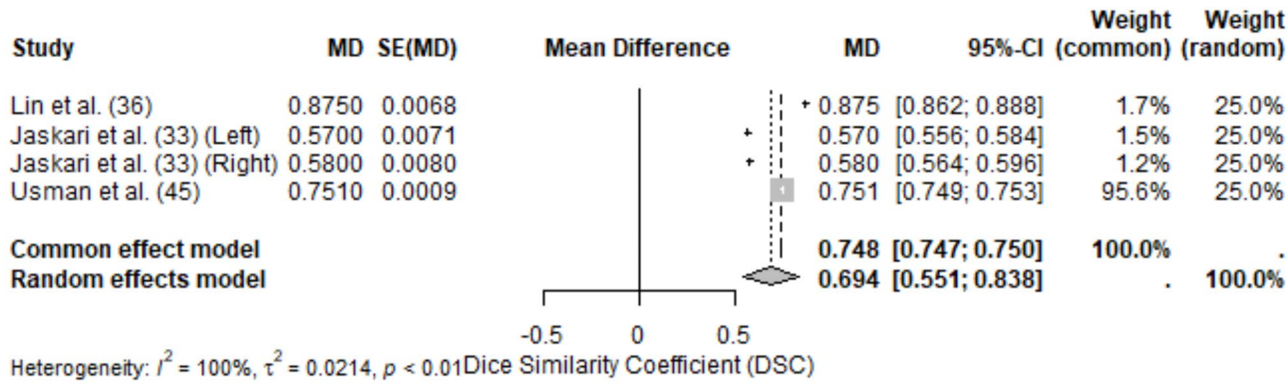

Fig. 3 Forest plot of DSC values for mandibular canal segmentation

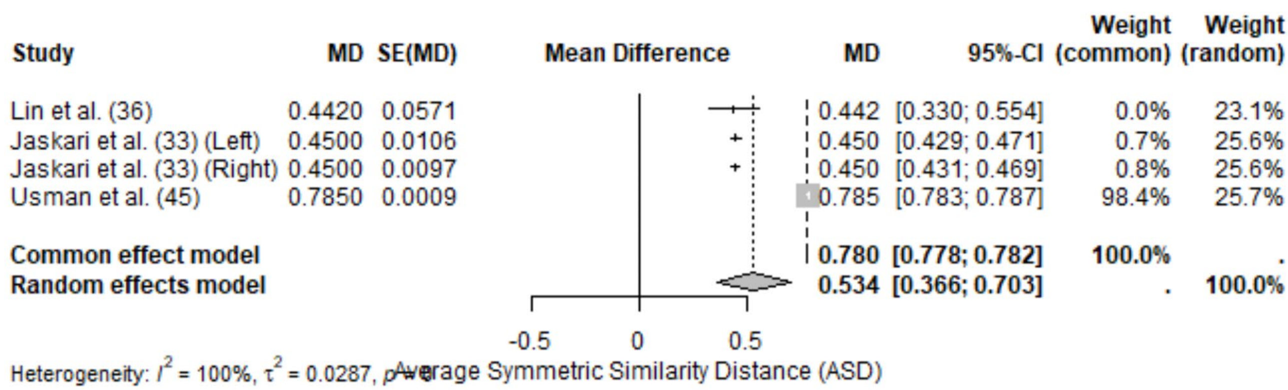

Fig. 4 Forest plot of ASD values for mandibular canal segmentation

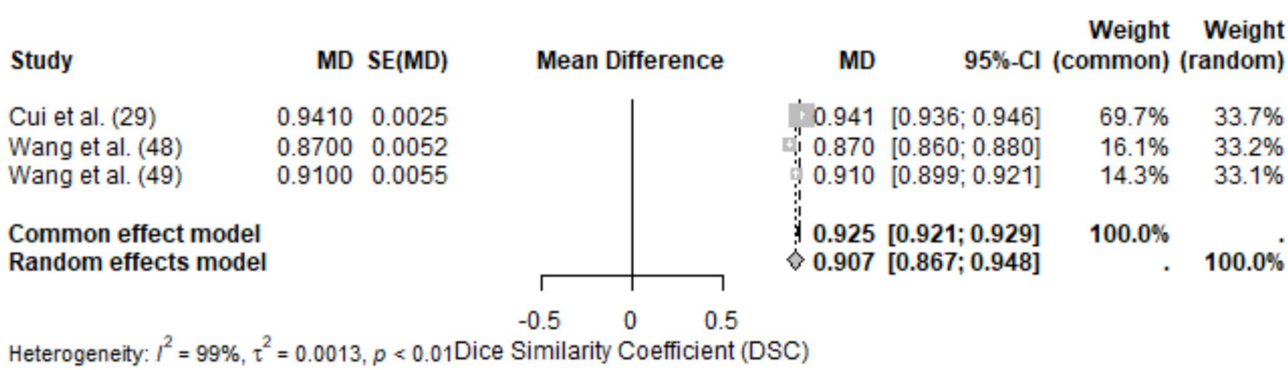

Fig. 5 Forest plot of DSC values for maxilla segmentation

enhancements were associated with improved segmentation accuracy and precision.

Teeth segmentation

Teeth segmentation was evaluated in 5 studies, with performance metrics including Dice Similarity Coefficient (DSC) and Average Surface Distance (ASD). The segmentation of teeth presented challenges due to varying anatomical complexity, especially for multi-rooted teeth.

The meta-analysis of DSC for teeth segmentation included 5 studies, yielding a pooled DSC of 0.925 (95% CI: 0.891–0.959) under the random-effects model. High heterogeneity ($I^2 = 100\%$) was observed, likely due to differences in model types and the anatomical features of single-rooted vs. multi-rooted teeth.

Reported DSC values ranged from 0.880 ± 0.0017 [35] to 0.962 ± 0.0004 (Duan et al. [30], multi-rooted teeth). Studies utilizing CNNs with advanced preprocessing demonstrated higher DSC values.

The meta-analysis of ASD included 3 studies, resulting in a pooled ASD of 0.189 mm (95% CI: 0.043–0.335) under the random-effects model. Moderate heterogeneity ($I^2 = 95\%$) was observed, reflecting variability in evaluation methods and segmentation protocols.

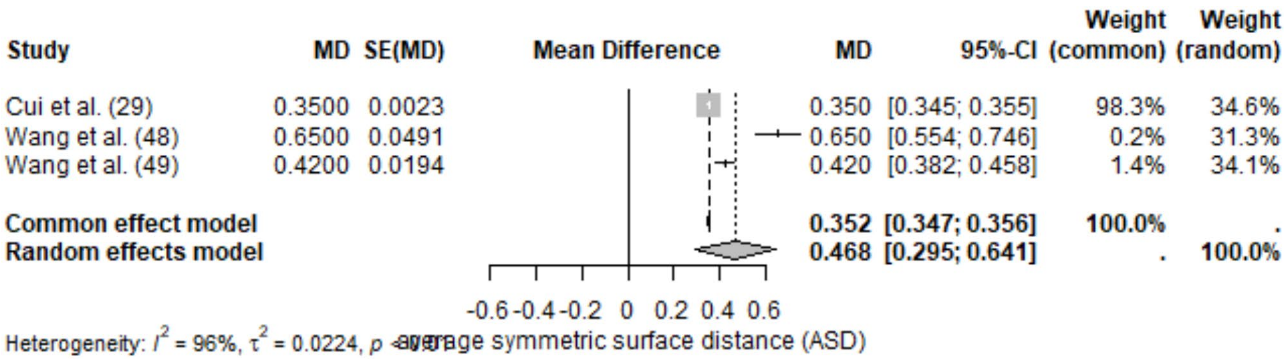

Fig. 6 Forest plot of ASD values for maxilla segmentation

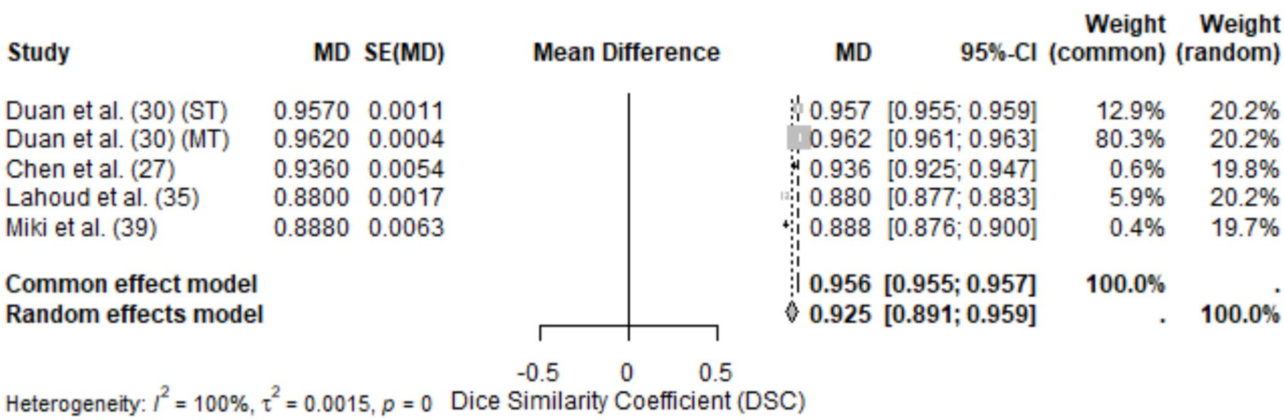

Fig. 7 Forest plot of DSC values for teeth segmentation

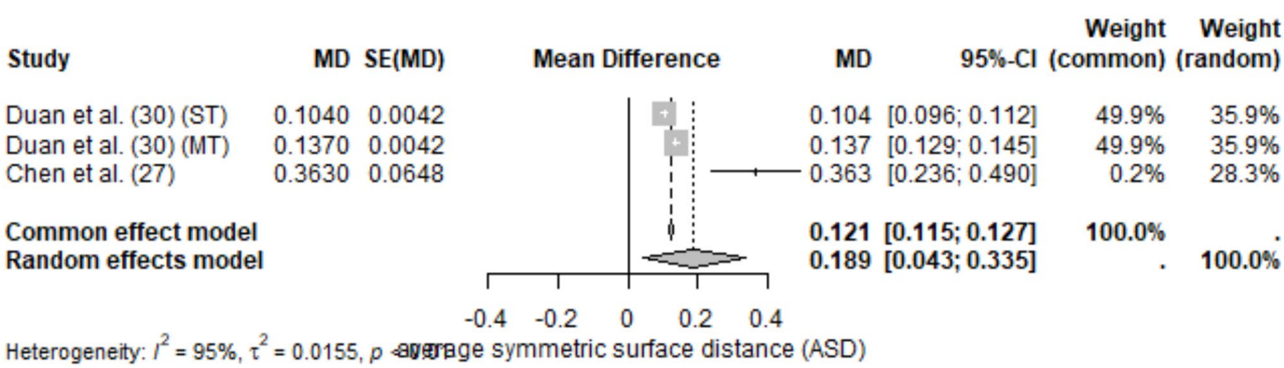

Fig. 8 Forest plot of ASD values for teeth segmentation

Reported ASD values ranged from 0.104 ± 0.0042 mm (Duan et al. [30], single-rooted teeth) to 0.363 ± 0.0648 mm [27]. Lower ASD values were associated with studies that used highly specialized CNN architectures (Figs. 7 and 8).

Sensitivity exceeded 90% in studies by Duan et al. [30] and Chen et al. [27], reflecting the models' ability to correctly identify teeth structures. Lahoud et al. [35] reported segmentation times for fully automated methods ranging from 1.2 min ± 33 s to 6.6 min ± 76.2 s depending on the algorithm used.

Deep learning models demonstrated strong segmentation performance for teeth, with DSC values consistently exceeding 0.88. Multi-rooted teeth presented greater challenges but benefited from specialized architectures like U-Net variants. Differences in imaging modalities, dataset diversity, and evaluation metrics contributed to high heterogeneity in the pooled results. Studies using CBCT imaging and CNN-based pipelines generally outperformed older machine learning methods, emphasizing the advantages of deep learning in handling complex tooth structures.

Table 4 Segmentation time of the AI models

Study	Segmentation approach	Segmentation time
Chen, Wang, et al. [4]	ML, LINKS	15 min
Cui et al. [29]	DL	0.23 min
Ileşan et al. [32]	DL, CNN, a 3D U-Net	2 min 03 s
Kim et al. [34]	DL, modified U-Net, and a CNN	10 s
Lahoud et al. [35]	DL, deep CNN	68.64 s
Lo Giudice et al. [37]	DL, CNN	50 s
Qiu et al. [42]	DL, Single-planar CNN	2.5 min
Verhelst et al. [46]	DL, 2-stageed 3D U-Net	17 s
Vinayahalingam et al. [47]	DL, 3D U-Net	3.6 s
Wang et al. [7]	DL, mixed-scale dense (MS-D) CNN	25 s
Wang et al. [49]	ML, prior-guided sequential random forests	20 min

DL, deep learning, ML, machine learning, CNN, convolutional neural network; mins, minutes; Secs, seconds

Segmentation time

The lowest segmentation time was 3.6 s in Vinayahalingam et al. which used a deep learning, 3D U-Net model [47]. Chen et al. used a machine learning model, had the longest time of 15 min [4]. A detailed description of the segmentation time in different models is presented in Table 4.

Segmentation model performance

The performance of segmentation models varied significantly across anatomical structures and AI architectures. CNNs were widely used and achieved high DSC values across most anatomical structures, particularly for mandibular bone and teeth [30, 37]. U-Net variants demonstrated robust performance in studies targeting the mandibular canal and maxilla [29, 36]. Emerging models, including transformer-based architectures, were noted for their potential but lacked sufficient data for inclusion in this review.

Discussion

This systematic review and meta-analysis evaluated the performance of AI-based segmentation models for four anatomical regions: the mandible, mandibular canal, maxilla, and teeth. Across all regions, the meta-analysis revealed high Dice Similarity Coefficient (DSC) values, indicating the strong performance of AI models in delineating dental and maxillofacial structures.

For mandible segmentation, the pooled DSC was 0.94 (95% CI: 0.91–0.98), confirming the high accuracy of AI models. Similarly, studies reported robust Average Surface Distance (ASD) values, indicating precise delineation. The heterogeneity in results ($I^2 = 99\%$) highlights the influence of methodological and dataset variability, particularly in model architecture and preprocessing techniques.

Mandibular canal segmentation posed additional challenges due to its small and complex structure. The pooled DSC for this region was 0.694 (95% CI: 0.551–0.838), with ASD values of 0.534 mm (95% CI: 0.366–0.703). Advanced deep learning models such as multi-stage U-Nets and attention-based architectures consistently outperformed classical machine learning models.

For maxilla segmentation, the pooled DSC was 0.907 (95% CI: 0.867–0.948), with ASD values of 0.468 mm (95% CI: 0.295–0.641). These results reflect the complexity of segmenting maxillary structures, where multi-scale architectures demonstrated superior performance.

Finally, for teeth segmentation, the pooled DSC was 0.925 (95% CI: 0.891–0.959), with ASD values of 0.189 mm (95% CI: 0.043–0.335). Single-rooted teeth were easier to segment compared to multi-rooted teeth, where specialized CNN pipelines demonstrated higher accuracy.

Across all anatomical regions, studies consistently showed that deep learning models outperformed classical machine learning methods. The adoption of advanced preprocessing steps, hierarchical architectures, and larger datasets further improved segmentation accuracy and precision.

Variation in segmentation outcomes

The results from this systematic review and meta-analysis demonstrated significant variability across studies, as evidenced by the high heterogeneity indices ($I^2 > 95\%$) observed in most pooled analyses. This variability can be attributed to several factors:

The majority of studies employed Cone Beam Computed Tomography (CBCT), with a few utilizing Computed Tomography (CT). While CBCT provides high-resolution imaging for dental structures, CT scans are often used for broader craniofacial assessments, potentially influencing segmentation accuracy. The variation in imaging protocols, voxel sizes, and contrast settings likely contributed to differences in reported metrics. Also, deep learning models, particularly U-Net variants and attention-based architectures, consistently outperformed classical machine learning methods. Multi-stage approaches, such as two-stage 3D U-Nets, showed superior performance in handling complex anatomical structures like the mandibular canal. However, studies that employed simpler models (single-layer CNNs) reported lower DSC and higher ASD values, emphasizing the importance of architectural sophistication.

The size and diversity of training datasets played a critical role in segmentation outcomes. Studies with larger, more diverse datasets generally reported higher DSC and lower ASD values, reflecting the importance of robust training data. For instance, Duan et al. [30] and Cui et al. [29], which used extensive datasets, reported some of the highest performance metrics across teeth and maxilla segmentation tasks.

Again, most studies relied on manual segmentation by experts (radiologists, dentists, or surgeons) as the reference standard. However, a few studies utilized semi-automated segmentation as the reference, which may introduce bias due to differences in the accuracy of semi-automated methods compared to manual delineations. While DSC and ASD were the most frequently reported metrics, there was considerable variability in how results were presented. Some studies reported DSC as percentages, others as decimal values, and a few omitted standard deviations altogether, complicating meta-analysis efforts. Standardized reporting guidelines are needed to improve the comparability of results.

Variations in segmentation accuracy also reflected the inherent complexity of the anatomical structures being segmented. For example: Mandible segmentation achieved the highest overall accuracy, likely due to its relatively large and well-defined structure, Mandibular canal segmentation posed greater challenges due to its small size and proximity to surrounding anatomical features, leading to lower pooled DSC values and Teeth segmentation results varied between single-rooted and multi-rooted teeth, with the latter presenting more difficulties for segmentation models.

Emerging models: SAM vs. traditional AI approaches

This review highlights the dominance of CNNs and U-Nets in dental segmentation research but also recognizes the potential of emerging models like Segment Anything Models (SAMs). SAMs are designed to generalize across diverse segmentation tasks with minimal manual input, offering a transformative approach to medical image segmentation. However, their application to dental imaging remains largely unexplored. Future studies should prioritize comparisons between SAMs and traditional models like CNNs, focusing on challenging anatomical regions such as the mandibular canal and multi-rooted teeth. Evaluating these models across standardized benchmarks, including DSC, ASD, and IoU, could provide valuable insights into their relative strengths and weaknesses.

Conclusion

High speed segmentation along with accuracy of AI in replication of the gold standard (manual segmentation), makes the integration of AI-based segmentation into the dental workflow a viable undertaking.

Acknowledgements
Not applicable.

Author contributions
MNA, MAA, ABA, HA, KR, ZH, MA, NC and SS collected the articles and wrote the manuscript. All authors reviewed the manuscript.

Funding
Not applicable.

Data availability
The data can also be accessed at figshare using the link https://doi.org/10.6084/m9.figshare.25512052.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 1 August 2024 / Accepted: 26 February 2025
Published online: 07 March 2025

References

1. Ciocca L, Mazzoni S, Fantini M, Persiani F, Marchetti C, Scotti R. CAD/CAM guided secondary mandibular reconstruction of a discontinuity defect after ablative cancer surgery. J Cranio-Maxillofacial Surg [Internet]. 2012;40(8):e511–5.
2. Kedar Kawsankar V, Tile V, Ambulgekar, Deshpande A. Application of CBCT in oral & maxillofacial surgery. J Dent Specialities. 2023;11(2):88–91.
3. Weiss R, Read-Fuller A. Cone beam computed tomography in oral and maxillofacial surgery: an Evidence-Based review. Dentistry J [Internet]. 2019;7(2):52.
4. Chen S, Wang L, Li G, Wu TH, Diachina S, Tejera B, et al. Machine learning in orthodontics: introducing a 3D auto-segmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients. Angle Orthod [Internet]. 2020;90(1):77–84.
5. Kwak GH, Kwak EJ, Song JM, Park HR, Jung YH, Cho BH et al. Automatic mandibular Canal detection using a deep convolutional neural network. Sci Rep [Internet]. 2020;10(1).
6. Qiu B, Guo J, Kraeima J, Borra RJH, Witjes MJH, van Ooijen PMA. 3D segmentation of mandible from multisectional CT scans by convolutional [internet] eural [internet]etworks [Internet]. ArXiv Org. 2018.
7. Wang H, Minnema J, Batenburg KJ, Forouzanfar T, Hu FJ, Wu G. Multiclass CBCT image segmentation for orthodontics with deep learning. J Dent Res [Internet]. 2021;002203452110053.
8. Qiu B, van der Wel H, Kraeima J, Glas HH, Guo J, Borra RJH et al. Automatic Segmentation of Mandible from Conventional Methods to Deep Learning—A Review. Journal of Personalized Medicine [Internet]. 2021cn Jul 1;11(7):629.
9. Vaz JM, Balaji S. Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. Mol Divers [Internet]. 2021;25(3):1569–84.
10. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, et al. Deep learning vs. Traditional Comput Vis Adv Intell Syst Comput [Internet]. 2019;943:128–44.
11. Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: introduction, application, and perspective in the big data era. Methods [Internet]. 2019;166:4–21.
12. Zhang Y, Shen Z, Jiao R. Segment anything model for medical image segmentation: current applications and future directions. Comput Biol Med. 2024;171:108238–8.
13. Huang Y, Yang X, Liu L, Han Z, Ao C, Zhou X, et al. Segment anything model for medical images? Med Image Anal. 2024;92:103061–1.
14. Zhang K, Liu D. Customized Segment Anything Model for Medical Image Segmentation [Internet]. arXiv.org. 2023. Available from: https://arxiv.org/abs/2304.13785
15. Khan A, Rauf Z, Sohail A, Khan AR, Asif H, Asif A et al. A survey of the vision transformers and their CNN-transformer based variants. Artificial Intelligence

Review [Internet]. 2023; Available from: https://arxiv.org/ftp/arxiv/papers/2305/2305.09880.pdf
16. Zhou R, Wang J, Xia G, Xing J, Shen H, Shen X. Cascade Residual Multiscale Convolution and Mamba-Structured UNet for Advanced Brain Tumor Image Segmentation. Entropy [Internet]. 2024 Apr 30 [cited 2024 Nov 19];26(5):385–5. Available from: https://www.mdpi.com/1099-4300/26/5/385
17. Altaf F, Islam SMS, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges, and future directions. IEEE Access. 2019;7:99540–72.
18. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M et al. A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis [Internet]. 2017; 42:60–88. Available from: https://pubmed.ncbi.nlm.nih.gov/28778026/
19. Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, et al. Caries detection with Near-Infrared transillumination using deep learning. J Dent Res [Internet]. 2019;98(11):1227–33.
20. Minnema J, van Eijnatten M, Hendriksen AA, Liberton N, Pelt DM, Batenburg KJ, et al. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. Med Phys [Internet]. 2019;46(11):5027–35.
21. Nguyen KCT, Duong DQ, Almeida FT, Major PW, Kaipatur NR, Pham TT, et al. Alveolar bone segmentation in intraoral ultrasonographs with machine learning. J Dent Res [Internet]. 2020;99(9):1054–61.
22. Singh NK, Raza K. Progress in deep learning-based dental and maxillofacial image analysis: A systematic review. Expert Syst Appl. 2022;199:116968.
23. Cui Z, Li C, Wang W, ToothNet. Automatic Tooth Instance Segmentation and Identification From Cone Beam CT Images. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019 Jun.
24. Lee S, Woo S, Yu J, Seo J, Lee J, Lee C. Automated CNN-Based tooth segmentation in Cone-Beam CT for dental implant planning. IEEE Access [Internet]. 2020;8:50507–18.
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Br Med J [Internet]. 2021;372(71).
26. Joanna Briggs Institute. Critical appraisal tools [Internet]. JBI. 2020. Available from: https://jbi.global/critical-appraisal-tools
27. Chen Y, Du H, Yun Z, Yang S, Dai Z, Zhong L, et al. Automatic segmentation of individual tooth in dental CBCT images from tooth surface map by a Multi-Task FCN. IEEE Access [Internet]. 2020;8:97296–309.
28. Chung M, Lee M, Hong J, Park S, Lee J, Lee J, et al. Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation. Computers Biology Med [Internet]. 2020;120:103720.
29. Cui Z, Fang Y, Mei L, Zhang B, Yu B, Liu J, et al. A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat Commun [Internet]. 2022;13(1):2096.
30. Duan W, Chen Y, Zhang Q, Lin X, Yang X. Refined tooth and pulp segmentation using U-Net in CBCT image. Dentomaxillofacial Radiol. 2021;50(6):20200251.
31. Hsu K, Yuh DY, Lin SC, Lyu PS, Pan GX, Zhuang YC, et al. Improving performance of deep learning models using 3.5D U-Net via majority voting for tooth segmentation on cone beam computed tomography. Sci Rep. 2022;12(1):19809.
32. Ileşan RR, Beyer M, Kunz C, Thieringer FM. Comparison of Artificial Intelligence-Based Applications for Mandible Segmentation: From Established Platforms to In-House-Developed Software. Bioengineering [Internet]. 2023 May 1 [cited 2024 Mar 7];10(5):604.
33. Jaskari J, Sahlsten J, Järnstedt J, Mehtonen H, Karhu K, Sundqvist O et al. Deep Learning Method for Mandibular Canal Segmentation in Dental Cone Beam Computed Tomography Volumes. Scientific Reports [Internet]. 2020 Apr 3 [cited 2020 Nov 25];10(1):5842.
34. Kim YH, Shin JY, Lee A, Park S, Han SS, Hwang HJ. Automated cortical thickness measurement of the mandibular condyle head on CBCT images using a deep learning method. Scientific Reports [Internet]. 2021;11(1).
35. Lahoud P, EzEldeen M, Beznik T, Willems H, Leite A, Van Gerven A, et al. Artificial intelligence for fast and accurate 3-Dimensional tooth segmentation on Cone-beam computed tomography. J Endodontics [Internet]. 2021;47(5):827–35.
36. Lin X, Xin W, Huang J, Jing Y, Liu P, Han J et al. Accurate mandibular Canal segmentation of dental CBCT using a two-stage 3D-UNet based segmentation framework. BMC Oral Health [Internet]. 2023;23(1).
37. Lo Giudice A, Ronsivalle V, Spampinato C, Leonardi R. Fully Automatic Segmentation of The Mandible Based On Convolutional Neural Networks (CNNs). Orthodontics & Craniofacial Research [Internet]. 2021 Sep 23.
38. Macho PM, Kurz N, Ulges A, Brylka R, Gietzen T, Ulrich Schwanecke. Segmenting teeth from volumetric CT data with a hierarchical CNN-based approach. 2018;109–13.
39. Miki Y, Muramatsu C, Hayashi T, Zhou X, Hara T, Katsumata A, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. Computers Biology Med [Internet]. 2017;80:24–9.
40. Pankert T, Lee H, Peters F, Hölzle F, Modabber A, Raith S. Mandible segmentation from CT data for virtual surgical planning using an augmented two-stepped convolutional neural network. Int J Comput Assist Radiol Surg. 2023;18(8):1479–88.
41. Park S, Kim H, Shim E, Hwang BY, Kim Y, Lee JW, et al. Deep Learning-Based automatic segmentation of mandible and maxilla in Multi-Center CT images. Appl Sci [Internet]. 2022;12(3):1358.
42. Qiu B, Guo J, Kraeima J, Glas HH, Borra RJH, Witjes MJH, et al. Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network. Phys Med Biol. 2019;64(17):175020.
43. Qiu B, Guo J, Kraeima J, Glas HH, Zhang W, Borra RJH, et al. Recurrent convolutional neural networks for 3D mandible segmentation in computed tomography. J Personalized Med [Internet]. 2021;11(6):492.
44. Qiu B, van der Wel H, Kraeima J, Glas HH, Guo J, Borra RJH, et al. Robust and accurate mandible segmentation on dental CBCT scans affected by metal artifacts using a prior shape model. J Personalized Med [Internet]. 2021;11(5):364.
45. Usman M, Rehman A, Saleem AM, Jawaid R, Byon SS, Kim SH, et al. Dual-Stage deeply supervised Attention-Based convolutional neural networks for mandibular Canal segmentation in CBCT scans. Sens [Internet]. 2022;22(24):9877.
46. Verhelst PJ, Smolders A, Beznik T, Meewis J, Vandemeulebroucke A, Shaheen E, et al. Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography. J Dent. 2021;114:103786.
47. Vinayahalingam S, Berends B, Baan F, Moin DA, van Luijn R, Bergé S et al. Deep learning for automated segmentation of the temporomandibular joint. J Dent. 2023;104475.
48. Wang L, Chen KC, Gao Y, Shi F, Liao S, Li G et al. Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. Med Phys. 2014;41(4).
49. Wang L, Gao Y, Shi F, Li G, Chen KC, Tang Z, et al. Automated segmentation of dental CBCT image with prior-guided sequential random forests. Med Phys [Internet]. 2015;43(1):336–46.
50. Qiu B, van der Wel H, Kraeima J, Glas HH, Guo J, Borra RJH, et al. Mandible segmentation of dental CBCT scans affected by metal artifacts using Coarse-to-Fine learning model. J Personalized Med [Internet]. 2021;11(6):560.

Publisher's note