

SYSTEMATIC REVIEW

Open Access



Evolution of deep learning tooth segmentation from CT/CBCT images: a systematic review and meta-analysis

Wai Ying Kot¹, Sum Yin Au Yeung¹, Yin Yan Leung¹, Pui Hang Leung^{1,2} and Wei-fa Yang^{1,2*}

Abstract

Background Deep learning has been utilized to segment teeth from computed tomography (CT) or cone-beam CT (CBCT). However, the performance of deep learning is unknown due to multiple models and diverse evaluation metrics. This systematic review and meta-analysis aims to evaluate the evolution and performance of deep learning in tooth segmentation.

Methods We systematically searched PubMed, Web of Science, Scopus, IEEE Xplore, arXiv.org, and ACM for studies investigating deep learning in human tooth segmentation from CT/CBCT. Included studies were assessed using the Quality Assessment of Diagnostic Accuracy Study (QUADAS-2) tool. Data were extracted for meta-analyses by random-effects models.

Results A total of 30 studies were included in the systematic review, and 28 of them were included for meta-analyses. Various deep learning algorithms were categorized according to the backbone network, encompassing single-stage convolutional models, convolutional models with U-Net architecture, Transformer models, convolutional models with attention mechanisms, and combinations of multiple models. Convolutional models with U-Net architecture were the most commonly used deep learning algorithms. The integration of attention mechanism within convolutional models has become a new topic. 29 evaluation metrics were identified, with Dice Similarity Coefficient (DSC) being the most popular. The pooled results were 0.93 [0.93, 0.93] for DSC, 0.86 [0.85, 0.87] for Intersection over Union (IoU), 0.22 [0.19, 0.24] for Average Symmetric Surface Distance (ASSD), 0.92 [0.90, 0.94] for sensitivity, 0.71 [0.26, 1.17] for 95% Hausdorff distance, and 0.96 [0.93, 0.98] for precision. No significant difference was observed in the segmentation of single-rooted or multi-rooted teeth. No obvious correlation between sample size and segmentation performance was observed.

Conclusions Multiple deep learning algorithms have been successfully applied to tooth segmentation from CT/CBCT and their evolution has been well summarized and categorized according to their backbone structures. In future, studies are needed with standardized protocols and open labelled datasets.

Keywords Artificial intelligence, Deep learning, Tooth segmentation, Convolutional neural networks, Transformer, CBCT

*Correspondence:
Wei-fa Yang
teddyrun@hku.hk

¹Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China
²Division of Oral & Maxillofacial Surgery, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China



Introduction

As digital dentistry advances as a disruptive technology, 3D imaging techniques such as computed tomography (CT) and cone-beam CT (CBCT) have become essential tools for diagnosis, treatment planning, and outcome verification [1, 2]. CBCT has been recognized as an advanced imaging modality in dental applications due to its ability to offer lower radiation doses and high-resolution images of the craniofacial region, yet CT remains popular among dental professionals due to its superior spatial resolution and detailed anatomical information [3, 4]. While CT/CBCT provides 2D slices for cross-sectional assessment, 3D modelling offers a more intuitive form of visualization [5]. For example, tooth 3D modelling can simulate the dynamic movement and alignment of teeth during orthodontic treatment [6]. The applications of tooth 3D modelling are extensive, encompassing orthodontics [7], restorative dentistry [8], implantology [9], prosthodontics [10], endodontics [11], oral surgery [12], periodontics [13] and more. Ongoing research and development continue to uncover new uses and enhance existing techniques [14, 15].

In practice, 3D modelling involves the crucial process of image segmentation, which extracts groups of voxels with specific intensity and boundaries from 2D slices [16]. However, unlike segmentation of the bone, traditional threshold-based segmentation is unsuitable for tooth segmentation due to the similar density between adjacent teeth and alveolar bone. This similarity makes it difficult to distinguish between bone and teeth, making manual tooth segmentation a cumbersome and time-consuming process [17]. Other methods, including level set-based algorithms, region-growing algorithms, and modifications such as hybrid or hierarchical level sets, have been explored [18–20]. However, these methods often require manual tuning of parameters and lack robustness for the variability of different teeth and imaging features. In recent years, artificial intelligence (AI) has emerged as a promising technology for medical applications, achieving fully automatic segmentation from medical images and improving the effectiveness of segmentation, with deep learning thriving as the most overwhelming tool [15]. Deep learning, a subset of machine learning, utilizes neural networks with many layers to model and understand complex patterns in data. One of its primary advantages is its ability to handle large datasets and automatically extract features, reducing the need for manual intervention. During the training process, deep learning models can learn hierarchical feature representations directly from raw data, which enables them to capture intricate structures and variations within the data. This capability allows them to excel recognizing complex patterns and scales well with increased data and computational resources. Transfer learning

and end-to-end optimization further enhance its versatility and efficiency. Continuous advancements in deep learning algorithms and architectures have consistently pushed the boundaries of AI performance across various domains, making it a leading approach in the field. In the realm of tooth segmentation, deep learning algorithms have been employed to automatically label and segment teeth, demonstrating satisfactory and accurate performance [21, 22].

Previous literature has investigated the application of automatic tooth segmentation in CBCT images [23, 24]. However, there is a lack of focus on deep learning algorithms, and the quantitative analysis of deep learning performance in tooth segmentation remains unknown due to multiple models and diverse evaluation metrics. This systematic review and meta-analysis aims to evaluate the evolution and performance of deep learning algorithms in tooth segmentation and provide updated insights in teeth segmentation for potential clinical applications.

Materials and methods

This systematic review and meta-analysis was developed and reported following the guideline of PRISMA-P (Preferred Items for Reporting Systematic Reviews and Meta-analyses for protocols) [25].

Information source and search strategy

This study emphasizes the implementation of deep learning algorithms in CT/CBCT segmentation, an area driven by interdisciplinary cooperation of dentistry and engineering. To ensure a comprehensive inclusion of relevant studies, the literature searches were performed across medical and engineering databases, including PubMed, Web of Science, Scopus, IEEE Xplore, ArXiv.org, and ACM. An exhaustive search strategy was employed to maximize search results, with a combination of free terms and database thesaurus terms of the following domains: “tooth”, “segmentation”, “CT”, “CBCT”, “computed tomography”, and “cone-beam” (Supplementary Table 1). The search encompassed all fields of relevant studies. The retrieved references were organized in EndNote 20 (Clarivate) for a streamlined screening process.

Eligibility criteria

The inclusion criteria for studies were: (1) the study had to be either retrospective, prospective, or cross-sectional in nature; (2) the deep learning algorithm had to be applied to human tooth segmentation in CT/CBCT imaging using either public or private dataset; (3) the study had to report evaluation metrics for segmentation accuracy. The exclusion criteria were: (1) the study was published in a language other than English; (2) the deep learning algorithm was not applied to human tooth

Table 1 Characteristics of the included studies

Author (Year)	Researcher/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/ slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Cui (2019)	ToothNet	All tooth type	20 (12 for training, 8 for testing)	Not mentioned	Not mentioned	Internal validation	DA: 0.9955 DSC: 0.9237 FA: 0.9685	The proposed fully automatic segmentation (without user annotation and post-processing step) produced superior results by exploiting the novel learned edge map, similarity matrix and spatial relations between teeth. The proposed components make the RPN-based framework suitable for 3D applications with lower GPU memory and less training time requirements.
Ezhov (2019)	V-Net	Not mentioned	815 (20 for testing)	Not mentioned	Not mentioned	Not mentioned	ASSD: 0.17 IoU: 0.94	The study presents the Coarse-to-Fine segmentation pipeline with weakly supervised pretraining, achieving 0.17 mm ASD and 0.94 IoU and showing significant improvements over several typically used baseline setups. The coarse-to-fine framework was effective for handling large volumetric images.
Chen (2020)	V-Net	All tooth type	25 (20 for training, 5 for testing)	376–541 slices per scan	Manual segmentation	Internal validation	ASSD: 0.363 ± 0.145 DSC: 0.936 ± 0.012 IoU: 0.881 ± 0.019 RVD: 0.072 ± 0.027	The study proposed a segmentation method with multi-task 3D FCN and MWT. This proposed method is effective and outperforms other modified FCN methods, with more accurate segmentation results.
Chung (2020)	PATRCNN + TSNet	All tooth type	175 (150 for training, 25 for testing)	402–552 slices per scan	Manual segmentation	Internal validation	AJ: 0.86 ± 0.01 ASSD: 0.20 ± 0.10 F1: 0.93 ± 0.03 HD: 1.59 ± 1.22 Prec: 0.93 ± 0.04 Sen: 0.93 ± 0.07	The study proposed a fully automatic, hierarchical method that performs accurate individual tooth detection followed by a CNN based on single pixel-wise labeling to segment the tooth. It avoids difficult pose-based procedure of separating teeth. It outperforms other state-of-the-art tooth segmentation methods.
Lee (2020)	UDS-Net	All tooth type	102 (69 for training, 1 for validation, 32 for testing)	264–727 slices per scan	Manual segmentation	Internal validation	Sen: 0.952 (validation set); 0.952 (test set) Prec: 0.924 (validation set); 0.904 (test set) DSC: 0.938 (validation set); 0.918 (test set)	The study proposed a fully automated CNN based tooth segmentation method for dental CBCT images. It utilises a multi-phase training strategy. It shows improved performance compared to conventional U-Net architecture.

Table 1 (continued)

Author (Year)	Researched/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Rao (2020)	DBA with DCRF	All tooth type	Not mentioned	110 slices (86 for training, 24 for testing)	Not mentioned	Internal validation	ASSD: 0.25 DSC: 0.9166 MSSD: 1.18 VD: 18.86	The study proposed a symmetric fully convolutional residual network with DCRF. It can outline and refine tooth boundaries, result in a more accurate segmentation of tooth images compared to other existing methods.
Wu (2020)	2-stage U-Net	All tooth type	20 (12 for training, 8 for testing)	543 teeth (324 for training, 219 for testing)	Not mentioned	Internal validation	ASSD: 0.122 DA: 0.995 DSC: 0.962 FA: 0.991	The study proposed a fully automatic segmentation method with two-level hierarchical deep neural network. It acquires accurate and smooth tooth boundaries.
Cui (2021)	3D V-Net	All tooth type	100 (50 for training, 10 for validation, 40 for testing)	Not mentioned	Manual segmentation	Internal validation	ASSD: 0.18 ± 0.02 HD: 1.52 ± 0.28 IoU: 0.891 ± 0.009	The study presents a tooth instance segmentation network guided by hierarchical morphological representations. It achieves high segmentation accuracy especially around tooth boundaries and tooth root apices, outperforms other state-of-the-art methods.
Duan (2021)	2-stage U-Net	Single-rooted tooth (ST) and Multi-rooted tooth (MT)	20	Not mentioned	Manual segmentation	Not mentioned	ASSD: 0.104 ± 0.019 (single-rooted); 0.137 ± 0.019 (multi-rooted) DSC: 0.957 ± 0.005 (single-rooted); 0.962 ± 0.002 (multi-rooted) RVD: 0.049 ± 0.017 (single-rooted); 0.053 ± 0.010 (multi-rooted)	The study proposed a two-phase deep learning solution for accurate tooth and pulp cavity segmentation using U-net model. It outperforms threshold-based method and FCN method. Multi-view data can enhance the accuracy of segmentation of the proposed method.
Jang (2021)	U-Net	All tooth type	11 (7 for training, 4 for testing)	328 teeth (216 for training, 112 for testing)	Not mentioned	Internal validation	ASSD: 0.14 ± 0.04 DSC: 0.9479 ± 0.0134 HD: 1.66 ± 0.72 Prec: 0.9597 ± 0.0200 Sen: 0.9371 ± 0.0208	The study proposed a fully automated segmentation and identification method for individual teeth and jaws from CBCT. It generates panoramic images from the CBCT, then performs 2D tooth segmentation and identification as prior knowledge of 3D segmentation, to improve the accuracy of 3D segmentation. It outperforms other existing automated methods by the fully automated and improved accuracy.

Table 1 (continued)

Author (Year)	Researched/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Shahen (2021)	3D U-Net	All tooth type	186 (Training: 140, validation: 35, test: 11)	832 teeth (Training: 400, validation: 100, test: 332)	Previously validated AI tool	Internal validation	95HD: 0.56 ± 0.38 IoU: 0.82 ± 0.05 Prec: 0.98 ± 0.02 Sen: 0.83 ± 0.05 Time: 13.7 ± 1.2 s	The study proposed a new cloud-based deep learning system for automatic tooth segmentation and classification of teeth. It is accurate and time-efficient.
Wang (2021)	MSD-CNN	All tooth type	30 (23 for training, 7 for testing)	9507 slices	Manual segmentation	Internal validation	DSC: 0.945 ± 0.021 (multiclass segmentation); 0.948 ± 0.021 (binary segmentation) MAD: 0.204 ± 0.061 (multiclass segmentation); 0.163 ± 0.051 (binary segmentation)	The study applied a novel MS-D network to segment CBCT scans into jaw, teeth and background. Multiclass segmentation achieved comparable segmentation accuracy as binary segmentation. It has high accuracy and the potential to segment both jaw and teeth simultaneously in a shorter time.
Yang (2021)	FCN	All tooth type	10	512 slices per scan	Manual segmentation	Internal validation	BF: 0.9824 ± 0.0324 DA: 0.9733 ± 0.0169 DSC: 0.9791 ± 0.0145 IoU: 0.9595 ± 0.0271	The study presents an improved active contour model with level set formulation. The proposed model has satisfactory result in segmentation, and is more accurate than other AI models.

Table 1 (continued)

Author (Year)	Researched/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Cui (a) (2022)	Multiple models included: SkipDenseNet, DenseVoxelNet, 3D U-Net, V-Net, Voxresnet, nnU-Net, Dense U-Net, AttU-Net, MT, CPS, DCT, CTCT, ENT, MAR, CEAL	All tooth type	168 (22 annotated, 146 unlabeled)	31,380 slices (5504 annotated, 25876 unlabeled)	Manual segmentation	Internal validation	ASSD: 1.08 (SkipDenseNet); 0.62 (DenseVoxelNet); 1.01 (3D U-Net); 0.29 (V-Net); 0.45 (Voxresnet); 0.27 (nnU-Net); 0.39 (Dense U-Net); 0.27 (AttU-Net) DSC: 0.6499 (SkipDenseNet); 0.7645 (DenseVoxelNet); 0.7951 (3D U-Net); 0.8121 (V-Net); 0.8507 (Voxresnet); 0.8548 (nnU-Net); 0.8627 (Dense U-Net); 0.8660 (AttU-Net) HD: 7.61 (SkipDenseNet); 5.10 (DenseVoxelNet); 8.02 (3D U-Net); 1.61 (V-Net); 5.14 (Voxresnet); 1.29 (nnU-Net); 2.08 (Dense U-Net); 1.72 (AttU-Net) IoU: 0.4916 (SkipDenseNet); 0.6222 (DenseVoxelNet); 0.6640 (3D U-Net); 0.6858 (V-Net); 0.7425 (Voxresnet); 0.7483 (nnU-Net); 0.7611 (Dense U-Net); 0.7645 (AttU-Net) PPV: 0.6949 (SkipDenseNet); 0.7536 (DenseVoxelNet); 0.8278 (3D U-Net); 0.8327 (V-Net); 0.8429 (Voxresnet); 0.8722 (nnU-Net); 0.8323 (Dense U-Net); 0.8779 (AttU-Net) SD: 0.7640 (SkipDenseNet); 0.8876 (DenseVoxelNet); 0.8876 (3D U-Net); 0.9290 (V-Net); 0.9404 (Voxresnet); 0.9503 (nnU-Net); 0.9591 (Dense U-Net); 0.9520 (AttU-Net) Sen: 0.7354 (SkipDenseNet); 0.8316 (DenseVoxelNet); 0.7821 (3D U-Net); 0.8088 (V-Net); 0.8658 (Voxresnet); 0.8456 (nnU-Net); 0.9080 (Dense U-Net); 0.8611 (AttU-Net) SO: 0.8017 (SkipDenseNet); 0.8954 (DenseVoxelNet); 0.8922 (3D U-Net); 0.9311 (V-Net); 0.9411 (Voxresnet); 0.9509 (nnU-Net); 0.9598 (Dense U-Net); 0.9525 (AttU-Net)	The study presents a 3D dental dataset CTooth + consisting extensive number of images with manual annotation of 3D structures. It evaluates tooth volume segmentation on fully-supervised learning, semi-supervised learning, and active learning methods as benchmarks for tooth segmentation.
Cui (b) (2022)	AttU-Net	All tooth type	22	Not mentioned	Manual segmentation	Internal validation	DSC: 0.8804 IoU: 0.7871 PPV: 0.8230 Sen: 0.9471 wDSC: 0.9514	The study presents an open-source 3D dental CT dataset with full tooth annotations. It proposed an attention-based benchmark for tooth segmentation, outperforms other 3D segmentation methods.

Table 1 (continued)

Author (Year)	Researched/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Cui (c) (2022)	Multi-stage V-Net	All tooth type	5345 (4938 for internal set, 407 for external set)	Not mentioned	Manual segmentation	Internal validation and External validation	<p>ASSD: 0.17 (internal validation); 0.21 (external validation)</p> <p>DSC: 0.941 (internal validation); 0.925 (external validation)</p> <p>Sen: 0.939 (internal validation); 0.921 (external validation)</p> <p>Time: 0.23 min</p>	The study proposed a deep-learning-based AI system for clinically stable and accurate fully automatic tooth and alveolar bone segmentation from dental CBCT images. It can segment individual teeth, maxillary and mandibular alveolar bone separately. It is fully automatic, accurate, robust and clinically applicable.
Dot (2022) ^Δ	nnU-Net	Upper teeth and Lower teeth	453 (300 for training and validation, 153 for testing)	Not mentioned	Semi-automatic segmentation	External validation	<p>95HD: 0.5909 ± 0.3977 (validation for upper teeth dataset); 0.8066 ± 2.7852 (validation for lower teeth dataset); 0.5403 ± 0.2273 (test for upper teeth dataset); 0.589 ± 0.5276 (test for lower teeth dataset)</p> <p>ASSD: 0.1021 ± 0.0825 (validation for upper teeth dataset, GT to prediction); 0.1049 ± 0.0985 (validation for lower teeth dataset, GT to prediction); 0.0909 ± 0.2332 (validation for upper teeth dataset, prediction to GT); 0.1161 ± 0.4811 (validation for lower teeth dataset, prediction to GT); 0.0912 ± 0.0537 (test for upper teeth dataset, GT to prediction); 0.102 ± 0.1119 (test for lower teeth dataset, GT to prediction); 0.0972 ± 0.3492 (test for upper teeth dataset, prediction to GT); 0.0779 ± 0.0697 (test for lower teeth dataset, prediction to GT)</p> <p>HD: 3.9237 ± 7.6032 (validation for upper teeth dataset); 3.8274 ± 7.1537 (validation for lower teeth dataset); 4.6572 ± 18.8037 (test for upper teeth dataset); 3.5852 ± 2.7675 (test for lower teeth dataset)</p> <p>IoU: 0.8976 ± 0.0407 (validation for upper teeth dataset); 0.8907 ± 0.0495 (validation for lower teeth dataset); 0.9022 ± 0.0316 (test for upper teeth dataset); 0.8944 ± 0.0389 (test for lower teeth dataset)</p> <p>sDSC: 0.9860 ± 0.0176 (validation for upper teeth dataset); 0.9832 ± 0.0262 (validation for lower teeth dataset); 0.9887 ± 0.0118 (test for upper teeth dataset); 0.9853 ± 0.02 (test for lower teeth dataset)</p> <p>vDSC: 0.9455 ± 0.024 (validation for upper teeth dataset); 0.9414 ± 0.0309 (validation for lower teeth dataset); 0.9483 ± 0.0181 (test for upper teeth dataset); 0.9438 ± 0.0232 (test for lower teeth dataset)</p> <p>VS: -0.0194 ± 0.0583 (validation for upper teeth dataset); -0.0188 ± 0.0683 (validation for lower teeth dataset); -0.0183 ± 0.0527 (test for upper teeth dataset); -0.0177 ± 0.0606 (test for lower teeth dataset)</p>	The study evaluated the performance of nnU-Net framework for automatic segmentation of craniomaxillofacial structures from CT scan. It segments the upper skull, mandible, upper teeth, lower teeth and mandibular canal separately and evaluates the accuracy. The results are comparable or superior to previously published studies.

Table 1 (continued)

Author (Year)	Researched/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Dou (2022)	TSDNet	All tooth type	40 (30 for training, 5 for validation, 5 for testing)	Not mentioned	Not mentioned	Internal validation	ASSD: 0.15 DA: 0.996 DSC: 0.952 HD: 2.12 IoU: 0.902	The study proposed a fully automated tooth instance segmentation deep learning network. It is accurate and outperforms other state-of-the-art deep learning-based segmentation methods.
Fon-tenele (2022)	3D U-Net	All tooth type	249 (140 for training, 35 for validation, 74 for testing)	726 teeth (400 for training, 100 for validation, 226 for test)	Manual segmentation	Internal validation	95HD: 0.03 ± 0.16 (anterior without fillings); 0.01 ± 0.09 (premolar without fillings); 0.08 ± 0.39 (molar without fillings); 0.25 ± 0.34 (anterior with fillings); 0.17 ± 0.38 (premolar with fillings); 0.19 ± 0.43 (molar with fillings) Acc: 0.9997 ± 0.0017 (anterior without fillings); 0.9998 ± 0.0016 (premolar without fillings); 0.9973 ± 0.0045 (molar without fillings); 0.9952 ± 0.0051 (anterior with fillings); 0.9968 ± 0.0047 (premolar with fillings); 0.9940 ± 0.0050 (molar with fillings) DSC: 0.99 ± 0.02 (anterior without fillings); 0.99 ± 0.02 (premolar without fillings); 0.98 ± 0.02 (molar without fillings); 0.95 ± 0.03 (anterior with fillings); 0.97 ± 0.03 (premolar with fillings); 0.97 ± 0.03 (molar with fillings) ICC: 0.97 IoU: 0.98 ± 0.04 (anterior without fillings); 0.99 ± 0.03 (premolar without fillings); 0.97 ± 0.04 (molar without fillings); 0.91 ± 0.05 (anterior with fillings); 0.94 ± 0.06 (premolar with fillings); 0.95 ± 0.04 (molar with fillings) Prec: 1.00 ± 0 (anterior without fillings); 1.00 ± 0 (premolar without fillings); 1.00 ± 0 (molar without fillings); 1.00 ± 0 (anterior with fillings); 1.00 ± 0 (premolar with fillings); 1.00 ± 0 (molar with fillings) Sen: 0.99 ± 0.04 (anterior without fillings); 0.99 ± 0.03 (premolar without fillings); 0.97 ± 0.04 (molar without fillings); 0.91 ± 0.05 (anterior with fillings); 0.94 ± 0.05 (premolar with fillings); 0.94 ± 0.04 (molar with fillings) Time: 29.8 s	The study proposed a multiple 3D U-net network model, and the proposed method demonstrated high accuracy metrics regardless of tooth type and presence of artifacts generated by high-density materials.
Gerhardt (2022)	3D U-Net	All tooth type	216 (130 for training, 46 for validation, 40 for testing)	1472 teeth for clinical validation	Manually refined AI segmentation	Internal validation	95HD: 0.33 (fully dentate); 0.15 (partially dentate) IoU: 0.96 (fully dentate); 0.97 (partially dentate) Time: 1.5 (1.2–2.9) seconds	The study proposed a 3D U-Net architecture based algorithm, and the proposed method demonstrated accurate tooth detection, labelling and segmentation.
Khan (2022)	DRNet	All tooth type	70 (60% for training, 20% for validation, 10% for testing)	Not mentioned	Semi-automatic segmentation	Not mentioned	Acc: 0.9554 DSC: 0.90 IoU: 0.70	The study proposed a deep convolutional neural network algorithm, and the proposed method had an accuracy comparable to 3D U-Net model, while requiring less time to train.

Table 1 (continued)

Author (Year)	Researched/developed algorithm	Tooth type	Number of scans/individuals	Number of teeth/slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Tao (2022) [^]	AttU-Net	All tooth type	Not mentioned	1500 slices (1200 for training, 300 for validation)	Not mentioned	Internal validation	IoU: 0.8373 PA: 0.8591 Time: 0.126 s	The study proposed a tooth CT image segmentation method using attention module integrated into the U-Net network that has better segmentation performance, segmentation efficiency and clearer contours to assist diagnosis.
Wu (2022)	2D CNN ** FCN	All tooth type	11 (9 for training, 2 for testing)	1660 slices (1360 for training, 300 for testing)	Not mentioned	Internal validation	IoU: 0.8474	The study proposed a practical local enhancement module for tooth segmentation, and it demonstrated superior performance.
Xie (2022)	FCN	All tooth type	78 (39 for training, 14 for validation, 25 for testing)	38,082 slices (19,416 for training, 6,820 for validation, 11,846 for testing)	Manual segmentation	Internal validation	ASSD: 0.53 ± 0.34 DSC: 0.88 ± 0.03 Prec: 0.98 ± 0.03 Sen: 0.93 ± 0.05	The study utilised CNNs and watershed transform method for automated individual tooth segmentation, and it achieved excellent agreement with manual tooth segmentation results
Alqahani (2023)	U-Net	All tooth type	215 (140 for training, 35 for validation, 40 for testing)	1780 slices (400 for training, 100 for validation, 1280 for testing)	Virtual Patient Creator	Internal validation	95HD: 0.12 ± 0.15 Acc: 0.99 ± 0.01 DSC: 0.99 ± 0.06 IoU: 0.99 ± 0.02 Prec: 0.99 ± 0.01 Sen: 0.99 ± 0.01 Time: 43.56 ± 20.31 s	The study proposed a multiclass CNN model, showing excellent performance with high accuracy and efficiency for segmentation.
Chen (2023)	CTA U-Net	All tooth type	45 (27 for training, 9 for validation, 9 for testing)	440 slices	Not mentioned	External validation	95HD: 0.536 ASSD: 0.211 IoU: 0.7812 DSC: 0.8650 (after training); 0.8715 (external validation)	The study proposed a CTA-U-Net model pre-trained with CTAMIM, and it outperformed traditional automated segmentation methods.
Chun # (2023)	Dense U-Net	Mandibular third molars	50 (32 for training and validation, 18 for testing)	100 teeth (64 for training and validation, 36 for testing); 5,350 slices (3,546 for training and validation, 1,804 for testing)	Not mentioned	Internal validation	DSC: 0.920 ± 0.131 IoU: 0.872 ± 0.161 Prec: 0.946 ± 0.091 RVD: 0.038 ± 0.025 Sen: 0.918 ± 0.148 VOE: 0.088 ± 0.024	The study proposed a Dense121 U-Net algorithm, and it achieved the highest average precision among the tested methods.

Table 1 (continued)

Author (Year)	Researched/ developed algorithm	Tooth type	Number of scans/ individuals	Number of teeth/ slices	Standard	Performance evaluation	Evaluation metrics	Main outcome
Kim (2023)	DHU-Net	All tooth type	70 (42 for training, 14 for validation, 14 for testing)	Not mentioned	Not mentioned	Not mentioned	95HD: 1.32 ± 0.30 DSC: 0.9391 ± 0.0034 IoU: 0.8867 ± 0.0068	The study proposed DHU-Net, a segmentation network capable of effectively learning hierarchical features of tooth structures, and it demonstrated improved segmentation performance for both the whole tooth and internal structures.
Li (a) (2023)	Swin-Transformer	All tooth type	151	Not mentioned	Not mentioned	Not mentioned	95HD: 1.50 ± 0.27 ASSD: 0.23 ± 0.06 DSC: 0.9423 ± 0.0061 IoU: 0.8964 ± 0.0086	The study proposed a swin-transformer-based network, showing a high accuracy for instance tooth segmentation and detection for pulp calcifications.
Li (b) (2023)	ZXYformer	All tooth type	157	Not mentioned	Not mentioned	Not mentioned	95HD: 1.47 ± 0.26 ASSD: 0.21 ± 0.05 DSC: 0.9447 ± 0.0045 IoU: 0.9035 ± 0.0073 Sen: 0.9383 ± 0.0131	The study proposed a ZXYformer with uncertainly guidance and macro weight transfer, and it outperformed other tooth segmentation methods.

*Notes: For study type: retrospective studies are indicated with #, other studies are cross-sectional studies. Imaging modality: studies using CT are indicated with Δ ; other studies are using CBCT. Evaluation metrics: 95HD (95% Hausdorff Distance, mm); Acc (Accuracy); AJI (Aggregated Jaccard Index); ASSD (Average Symmetric Surface Distance, mm); BF (Boundary F1 Score); DA (Detection Accuracy); DSC (Dice Similarity Coefficient); F1 (F1 Score); FA (Identification Accuracy); HD (Hausdorff Distance, mm); ICC (Intra-Class Correlation Coefficient); IoU (Jaccard Index / Intersection over Union); MAD (Mean Absolute Deviation, mm); MSSD (Maximum Symmetric Surface Distance, mm); PA (Pixel Accuracy); PPV (Positive Predictive Values); Prec (Precision); RVD (Relative Volume Difference); sDSC (Surface Dice Similarity Coefficient at 1 mm); Sen (Sensitivity/ Recall); SO (Surface Overlap); Time (Time spent, seconds); VD (Volume Difference, mm3); vDSC (Volumetric Dice Similarity Coefficient); VOE (Volumetric Overlap Error); VS (Volume Similarity); wDSC (Weighted Dice Similarity Coefficient)

segmentation; (3) the study used image modalities other than CT/CBCT; (4) the segmentation algorithm was manual or semi-automatic segmentation without deep learning algorithms; (5) the segmentation was applied to tooth pulp or only a part of the tooth; (6) the study did not report evaluation metrics for segmentation performance; (7) the studies were animal studies, forensic studies, literature reviews, questionnaire-based studies, or other non-original studies.

Study selection

The selection was conducted in two phases by three authors. In phase 1, the titles and abstracts of all relevant references were independently reviewed. In phase 2, the full texts were retrieved and independently reviewed, ensuring that all eligibility criteria were met. Any discrepancies among reviewers were resolved through consensus. The final decision was always made based on the full text of the publication.

Data collection

The following data were extracted from each included study: (1) year and type of study; (2) structure and features of the deep learning algorithm; (3) characteristics of samples, including sample size, tooth type, and image modality; (4) characteristics of evaluation metrics, including evaluation methods, type and result of the metric, and validation methods.

Risk of bias assessment

The risk of bias in the included studies was evaluated using the Quality Assessment of Diagnostic Accuracy Study (QUADAS-2) tool [26]. This scoring system assesses four domains: (1) patient selection; (2) index test; (3) reference standard; and (4) flow of patients through the study and timing of the index test(s) and reference standard. Signalling questions were employed in each domain to evaluate the risk of bias, with the first three domains also addressing applicability concerns. Based on the responses to these signalling questions, the risk of bias was determined to be either 'low', 'high', or 'unclear'. Any differences of opinion were resolved through discussion and mutual agreement between the authors.

Data analysis

The primary outcome of the meta-analysis was the segmentation accuracy of teeth with deep learning algorithms, as indicated by evaluation metrics. The number of teeth was used as the sample size; if the number of individuals or scans was reported instead of the number of teeth, an estimate was made assuming that each individual had 28 teeth unless the number of teeth per individual was specified. The mean values and standard deviations of the included studies were extracted;

for studies that did not report a standard deviation, the pooled standard deviation of the remaining studies was used as an estimate. Z test was performed to compare synthesized results that incorporated estimated standard deviations and those that did not. Furthermore, a comparison was conducted between the segmentation accuracy of single-rooted teeth and multi-rooted teeth. Meta-analyses were performed to calculate the pooled means of included evaluation metrics. The I^2 value of the Cochran Q test was used to evaluate statistical heterogeneity, and the random inverse-variance model was applied to provide the best pooled outcome estimates. The meta-analysis was conducted using Stata Statistical Software (Release 16; StataCorp LLC, College Station, TX, USA).

Results

Study selection

A total of 3361 studies were identified from databases. After removing duplicates, 2607 references remained for title and abstract screening. This process led to the selection of 181 articles for full-text evaluation. A full-text assessment was conducted, and 151 studies were excluded according to pre-defined eligibility criteria. Thereafter, 30 studies were selected for the systematic review (Supplementary Table 2). 2 studies were excluded from the meta-analysis due to insufficient data, specifically lacking standard deviation and the number of teeth, leaving 28 studies for the meta-analyses. A flowchart describing the process is illustrated in Supplementary Fig. 1.

Characteristics of included studies

The characteristics of the included articles are listed in Table 1. Overall, almost all studies were cross-sectional studies (96.7%), except for one (3.3%) [27], which was a retrospective study. The earliest paper included was from 2019. The most common imaging modality studied was CBCT (93.3%), followed by CT (6.7%). Manual segmentation was commonly used (40%) as the reference standard, while other studies used semi-automatic segmentation as the standard (16.7%), or did not mention their standard (43.3%). To evaluate training performance, 21 studies (70%) used internal validation, 2 (6.67%) used external validation, 1 (3.33%) combined external and internal validation, and the remaining studies (20%) did not mention their evaluation methods. The majority of studies (86.7%) did not specify the segmented tooth type or included all types of teeth, whereas studies (13.3%) classified the segmented teeth based on the tooth type or as single-rooted and multi-rooted teeth.

Evolution of deep learning algorithms

The study identified multiple deep learning algorithms for tooth segmentation and depicted their relationships. Figure 1 illustrates the evolution and classification of these algorithms according to the neural networks, which are further explained in Table 2. The deep learning algorithms involved in tooth segmentation were classified into single-stage convolutional model, convolutional model with U-Net architecture, Transformer model, convolutional model with attention mechanism, and combination of multiple models. For the evolution of deep learning, the single-stage convolutional model, represented by FCN, underwent modifications to incorporate VoxResNet and DenseVoxelNet. This evolution progressed with the introduction of a U-shaped architecture, culminating in the creation of the U-Net. U-Net and its variants, such as UDS-Net and Dense U-Net, proved to be the most frequently utilized algorithms. The advancement of deep learning was further boosted by the emergence of the multi-head attention mechanism, embodied by the Transformer and its modifications. This attention mechanism was seamlessly integrated with U-Net-based algorithms, resulting in the formulation of the TDS-Net. The combinations of multiple models were also introduced to achieve robustness and offset the limitations of single-stage models.

Evaluation metrics of segmentation accuracy

A total of 29 evaluation metrics for segmentation accuracy were identified, which were categorized into three groups: overlap-based metrics expressed as percentages (%), distance-based metrics measured in millimetres (mm), and volume-based metrics represented in millimetres cubed (mm³) or as percentages (%). These metrics are listed in Table 1. The overlap-based metrics encompassed various measurements, such as Jaccard Index/Intersection over Union (IoU), Dice Similarity (DSC), Aggregated Jaccard Index (AJI), Accuracy (Acc), Boundary F1 Score (BF), Detection Accuracy (DA), F1 Score (F1), Identification Accuracy (FA), Pixel Accuracy (PA), Positive Predictive Values (PPV), Precision, Surface Dice (SD), Surface Dice Similarity Coefficient at 1 mm (sDSC), Sensitivity/Recall (Sen), Surface Overlap (SO), Volumetric Dice Similarity Coefficient (vDSC), Volumetric Overlap Error (VOE), and Weighted Dice Similarity Coefficient (wDSC). The distance-based metrics included 95% Hausdorff Distance (95HD), Average Symmetric Surface Distance (ASSD), Hausdorff Distance (HD), Mean Absolute Deviation (MAD), and Maximum Symmetric Surface Distance (MSSD). Lastly, the volume-based metrics consisted of Relative Volume Difference (RVD), Volume Difference (VD), and Volume Similarity (VS). Among all evaluation metrics, DSC was the most popular one [6, 27–47], followed by IoU [27, 29, 35–37,

39–41, 43–54], ASSD [29, 31–34, 36, 38, 39, 42, 44, 46–49, 51, 55], sensitivity [27, 30, 36–38, 40, 42, 43, 47, 50, 55, 56], 95HD [40, 43–47, 50–52], and precision [27, 30, 40, 42, 43, 50, 55, 56]. Meta-analyses were performed for these top five evaluation metrics in below.

Risk of bias and applicability concerns

The results of the QUADAS-2 tool are provided in Supplementary Tables 3 and Supplementary Fig. 2. Within the patient selection domain, only 9.1% studies were found to have a low risk of bias, while the others 90.1% were found to have an unclear risk of bias; all studies presented low applicability concerns. Within the index test domain, all studies exhibited a low risk of bias and applicability concerns. Within the reference standard domain, 51.5% of studies were found to have a high risk of bias, 30.3% low risk, and 18.2% unclear risk; all studies presented low applicability concerns. In the flow and timing domain, all studies had a low risk of bias.

Pooled accuracy of tooth segmentation

The synthesized results were 0.93 [0.93, 0.93] for DSC (Fig. 2), 0.86 [0.85, 0.87] for IoU (Supplementary Fig. 3), 0.22 [0.19, 0.24] for ASSD (Supplementary Fig. 4), 0.92 [0.90, 0.94] for sensitivity (Supplementary Fig. 5), 0.71 [0.26, 1.17] for 95% Hausdorff distance (Supplementary Fig. 6), and 0.96 [0.93, 0.98] for precision (Supplementary Fig. 7). The synthesized results without estimation of standard deviations were shown in Supplementary Figs. 8–13; no statistically significant differences were observed between with and without SD estimations ($p > 0.05$) (Supplementary Table 4). High I^2 values ($> 90\%$) were shown in all forest plots, demonstrating high heterogeneity of the synthesized results. No significant difference was observed between the pooled results of single-rooted and multi-rooted teeth in all available evaluation metrics ($p > 0.05$) (Supplementary Figs. 14–19). The exploration of the potential correlation between segmentation models, sample size, publication year, and DSC is demonstrated in the bubble plot in Fig. 3.

Discussion

This systematic review and meta-analysis is the first to comprehensively analyse the evolution of deep learning algorithms specifically in the context of tooth segmentation and provides an updated insight into their performance within the field. Multiple deep learning algorithms were identified and categorized, including single-stage convolutional models, convolutional models with U-Net architecture, Transformer models, convolutional models with attention mechanism, and combination of multiple models. Convolutional models with U-Net architecture have been the dominant structure for tooth segmentation, while the integration of attention mechanism is

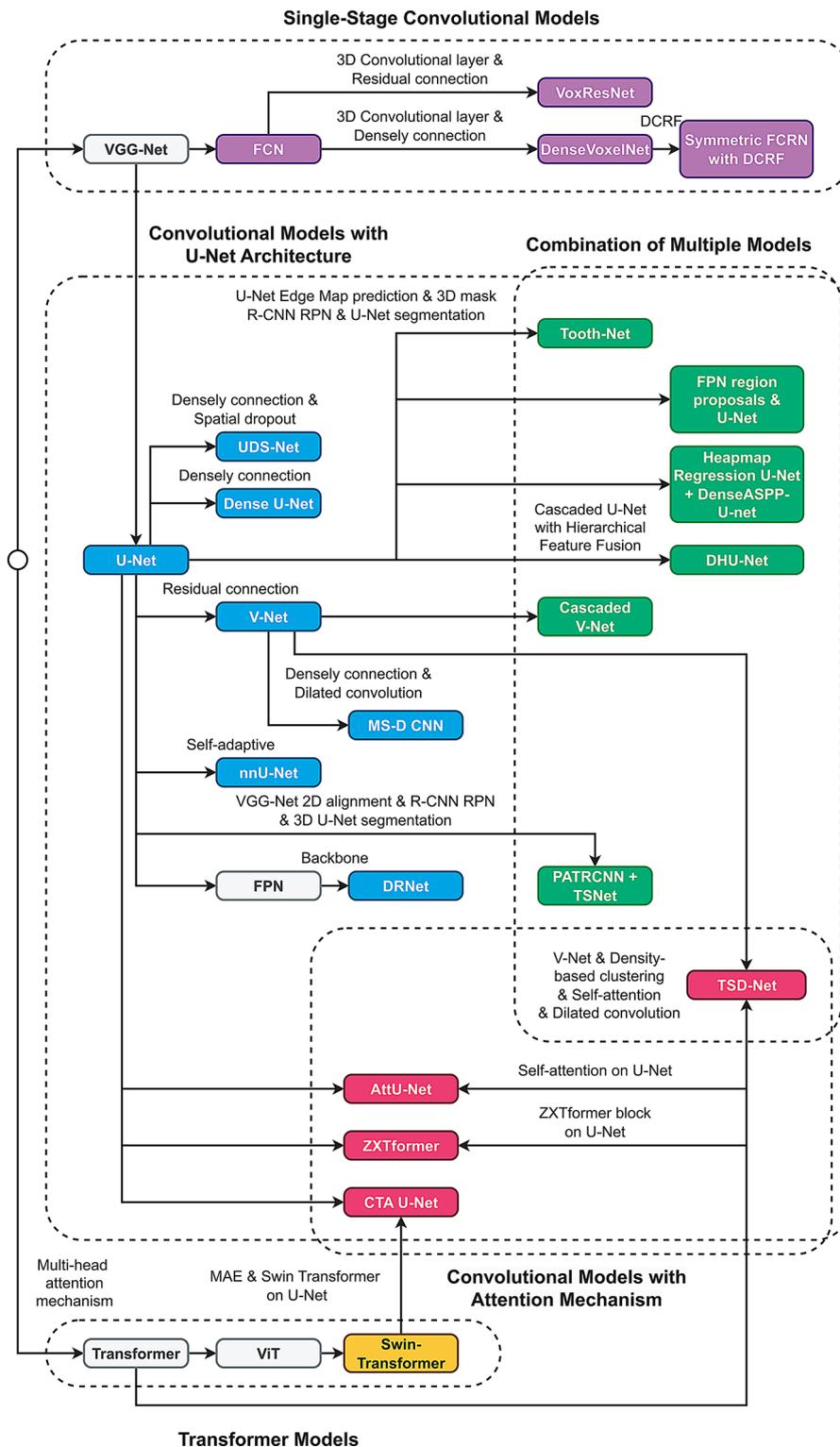


Fig. 1 The evolution of deep learning models used for tooth segmentation *Notes: The colours of the models indicate their respective groups of deep learning architectures. Models marked in grey indicate the absence of included studies. Layer modifications for each model are detailed along the arrows. AttU-Net: Attention U-Net; CTA U-Net: CNN-transformer architecture U-Net; DenseVoxelNet: Densely-Connected Volumetric Convolutional Neural Network; DHU-Net: Dual-Hierarchy U-Net; DRNet: end-to-end Decomposition and Reasoning Network; FPN: Feature pyramid network; MS-D CNN: Mixed-scale dense (MS-D) Convolutional Neural Network; nnU-Net: no new Net; PATRCNN + TSNet: Pose-aware Track R-CNN; Swin-Transformer: Shifted windows Transformer; Symmetric FCRN with DCRF: Symmetric Fully Convolutional Residual Network (FCRN) with DCRF (Dense Conditional Random Field); TSDNet: Tooth Segmentation Deeplearning Network; UDS-Net: Combination of U-Net, Dense Block and Spatial Dropout; VGG-Net: Visual Geometry Group Net (Very Deep Convolutional Networks); ViT: Vision Transformer; Voxresnet: Deep Voxelwise Residual Network

Table 2 Deep learning model descriptions

Group	Models	Full Name of Model	Descriptions	Paper: Author (Year)
Single-stage convolutional models	VGG-Net	Visual Geometry Group Net (Very Deep Convolutional Networks)	VGG-Net, a classical deep CNN architecture with 16 (VGG-16) or 19 (VGG-19) layers, is known for its simplicity and depth. It comprises convolutional layers with small filters, max-pooling layers, and fully connected layers for classification. While effective at capturing hierarchical patterns, its high computational cost and large parameter count make it less efficient than modern architectures like ResNet.	N/A
Single-stage convolutional models	FCN	Fully Convolutional Network	FCN is a neural network architecture that replaces traditional CNN's last fully connected layers with convolutional layers, allowing for dense pixel-wise predictions suitable for semantic segmentation.	Yang (2021), Wu (2022), Xie (2022)
3D convolutional design				
Single-stage convolutional models	DenseVoxelNet	Densely-Connected Volumetric Convolutional Neural Network	DenseVoxNet is a 3D convolutional neural network designed for volumetric medical image segmentation, incorporating dense blocks from DenseNet to enable feature reuse and improved gradient flow. (DenseNet is a CNN architecture that densely connects each layer to every other layer in a feed-forward fashion, improving gradient flow, encouraging feature reuse, and reducing the number of parameters. This efficient design leads to better performance in computer vision tasks such as image classification and object detection.)	Cui a (2022)
Single-stage convolutional models	Symmetric FCRN with DCRF	Symmetric Fully Convolutional Residual Network (FCRN) with DCRF (Dense Conditional Random Field)	A Symmetric FCRN with DCRF is a deep learning architecture that combines the strengths of both FCNs and ResNets within the U-Net-inspired symmetric encoder-decoder structure, with the powerful pixel-level post-processing capabilities of DCRFs to improve segmentation accuracy.	Rao (2020)
Single-stage convolutional models	VoxResNet	Deep Voxelwise Residual Network	VoxResNet is a 3D convolutional neural network designed for volumetric medical image segmentation, extending ResNet's residual connections to handle 3D data. (ResNet is a deep convolutional neural network known for its ability to train extremely deep models using residual or skip connections, which help avoid the vanishing gradient problem.)	Cui a (2022)
U-shaped design				
Convolutional models with U-Net architecture	U-Net		U-Net features a contracting path (encoder) that captures context and a symmetric expanding path (decoder) that enables precise localization, making it the most common CNN for segmenting medical images with complex structures.	Jang (2021), Shaheen (2021), Cui a (2022), Fontenele (2022), Gerhardt (2022), Alqahtani (2023)
Convolutional models with U-Net architecture	UDS-Net	U-Net + Dense Block + Spatial Dropout	Variant of U-Net; dense block and spatial dropout layer were added after each convolution encoding level; spatial dropout layers were added to skip connections between the contractive and extractive paths.	Lee (2020)
Convolutional models with U-Net architecture	Dense U-Net		Dense U-Net incorporates dense connections (inspired by DenseNet) within the U-Net structure to improve the information flow and gradient propagation throughout the network.	Cui a (2022), Chun (2023)

Table 2 (continued)

Group	Models	Full Name of Model	Descriptions	Paper: Author (Year)
Convolutional models with U-Net architecture	V-Net		V-Net is a 3D convolutional neural network architecture designed for volumetric medical image segmentation, inspired by the 2D U-Net architecture. It operates on 3D data using 3D convolutional, pooling, and upsampling layers, and features an encoder-decoder structure with skip connections to effectively capture spatial relationships and contextual information in medical images.	Ezhov (2019), Chen (2020), Cui (2021), Cui a (2022)
Convolutional models with U-Net architecture	MS-D CNN	Mixed-scale dense Convolutional Neural Network	MS-D CNN is a deep learning architecture that combines mixed-scale feature learning, dilated convolutional layers, and dense connections to efficiently capture multi-scale features and enhance information flow within the network.	Wang (2021)
Convolutional models with U-Net architecture	nnU-Net	no new Net	nnU-Net is a self-adaptive medical image segmentation framework based on the U-Net architecture, which automatically configures itself to different tasks and datasets to achieve the high performance.	Cui a (2022), Dot (2022)
Convolutional models with U-Net architecture	FPN	Feature pyramid network	Feature Pyramid Network (FPN) is a neural network architecture designed to extract feature maps at different scales by combining low-resolution, semantically strong features with high-resolution, semantically weak features through the use of lateral connections and top-down pathways.	N/A
Convolutional models with U-Net architecture	DRNet	end-to-end Decomposition and Reasoning Network	VGG-16 and FPN backbone for image feature maps calculation and embedding; mathematical methods were applied for head descriptor extraction and pedestrian inflow reasoning	Khan (2022)
Attention mechanism	Transformer	Transformer	Neural network with multi-head attention gate (query, key, value) for encode and decode on image input; better performance and efficiency on processing sequential data compared with recurrent neural network; usually used for NLP (neural language processing) and have been successfully applied to computer vision tasks.	N/A
Adaptation of vision task	Transformer	Vision Transformer	ViT (Vision Transformer) processes non-overlapping fixed-size image patches as input. These patches are linearly embedded into flat vectors and fed into the Transformer encoder, with positional embeddings added to retain spatial information. ViT has achieved state-of-the-art performance in image classification tasks, proving the viability of Transformers in computer vision.	N/A
Transformer	Swin-Transformer	Shifted windows Transformer	The Swin-Transformer integrates a hierarchical structure similar to convolutional neural networks (CNNs) and employs non-overlapping multi-head self-attention modules to facilitate cross-window connections, thereby enhancing the efficiency and scalability of Vision Transformer in computer vision tasks.	Li a (2023)
Combination of U-shaped design model	Convolutional model with attention mechanism	Tooth Segmentation Deep learning Network	The Tooth Segmentation Deep Learning Network (TSDNet) is a two-phase framework for tooth instance segmentation in CBCT data. In the first phase, the centroid prediction network obtains tooth centroid offsets using the V-Net framework and a density-based fast search clustering algorithm. In the second phase, the tooth instance segmentation network incorporates a 3D self-attention-based mechanism for guiding the tooth geometric structure information and a dilated convolution-based multiscale fusion for the tooth feature integration.	Dou (2022)
Convolutional model with attention mechanism	AttU-Net	Attention U-Net	U-Net with attention gate in each encoding and decoding level; it takes the advantage of attention network design and improves the performance of encoding and decoding and hence the features representativeness	Cui a (2022), Cui b (2022), Tao (2022)

Table 2 (continued)

Group	Models	Full Name of Model	Descriptions	Paper: Author (Year)
Convolutional model with attention mechanism	CTA U-Net	CNN-transformer architecture U-Net	By adding the swin-transformer block for each encoding layer, it can achieve a better performance on feature extraction and segmentation; MAE (masked autoencoder) design was used for self-supervised learning which can improve the encoder performance	Chen (2023)
Convolutional model with attention mechanism	ZXYformer		Adding ZXYformer block for each shortcut connection in U-Net; ZXYformer block contains deformable convolution and multi-head cross attention layer in order to take the advantage of attention mechanism and perform segmentation task	Li b (2023)
Multiple-stage model	Tooth-Net		ToothNet is a two-stage deep neural network designed for tooth segmentation in CBCT images. The first stage focuses on edge map extraction to enhance boundary information, while the second stage uses a 3D Mask R-CNN architecture with a 3D RPN module, tooth identification branch, similarity matrix, and spatial relationship component to segment and identify individual teeth.	Cui (2019)
Multiple-stage model	FPN region proposals & U-Net	Region Proposal Network (RPN) with Feature Pyramid Network (FPN) & U-Net	The proposed deep learning solution employs a two-phase approach, combining RPN with FPN for bounding box extraction and U-Net for refined tooth and pulp segmentation, while incorporating a smoothness penalty loss function and multi-view data augmentation to address data and morphological challenges.	Duan (2021)
Multiple-stage model		Heatmap Regression U-Net + DenseASPP-U-Net	Heatmap Regression U-Net in the Global stage to guide the localization of tooth ROIs and ROI-based DenseASPP-U-Net in the Local stage for tooth segmentation and classification.	Wu (2020)
Multiple-stage model	DHU-Net	Dual-Hierarchy U-Net	Dual-Hierarchy U-Net (DHU-Net), consists of two cascaded U-Net networks: P-Net and C-Net. P-Net learns higher-level features, while C-Net learns lower-level features. DHU-Net also includes a Hierarchical Feature Fusion (HFF) module, which combines features from both hierarchies using channel attention.	Kim (2023)
Multiple-stage model	Cascaded V-Net		Refining the segmentation results by connecting the V-Net with another V-Net in order to improve the accuracy	Cui c (2022)
Multiple-stage model	PATRCNN + TSNet	pose-aware track R-CNN	In the first phase, the pose-aware TRCNN is a VGG-16 model which was used to extract and realign the position of teeth; In the second phase, TSNet is a 3D U-Net with SkipBlock design which reduce the parameters and increase the performance	Chung (2020)

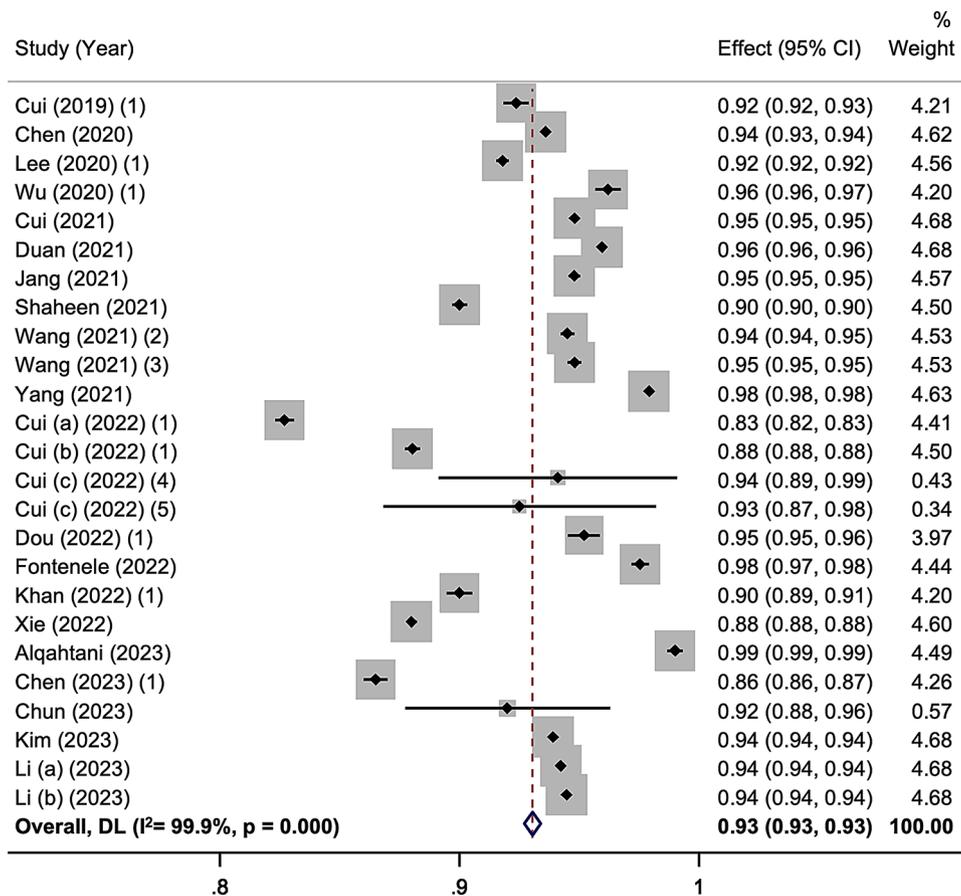


Fig. 2 Forest plot of pooled mean of Dice Similarity Coefficient. *Notes: (1) The missing standard deviation (SD) values were imputed by pooling all given SD values that were ≤ 1 . (2) Multiclass segmentation subgroup. (3) Binary segmentation subgroup. (4) Internal validation dataset. (5) External validation dataset

becoming a new topic of exploration. Based on the synthesis of the most updated evidence, deep learning algorithms can achieve a DSC accuracy as high as 0.93 for tooth segmentation.

The continuous evolution of deep learning algorithms is driven by the primary goal of achieving accurate and efficient tooth segmentation [23]. The most popular deep learning algorithm in tooth segmentation is the U-Net architecture, which consists of the convolutional encoding stage and the deconvolutional decoding stage. U-Net can utilize skip connections, directly connecting the appropriate layers between encoder and decoder, combining information from different levels of abstraction, and assisting the model in comprehending complex patterns in the data, which is particularly beneficial for tooth segmentation [57, 58]. This results in higher segmentation accuracy compared to other convolutional neural networks (CNNs). To address the volumetric nature of CT/CBCT images, researchers have also explored 3D U-Net architectures, leading to the development of VoxResNet and DenseVoxelNet [59]. These networks are designed to capture 3D information from volumetric

data, enabling the prediction of fine-grained labels for objects within a 3D scene [60]. By leveraging 3D convolutions, these models capture spatial relationships across all three dimensions resulting in enhanced segmentation performance compared to 2D networks [61].

Another approach in tooth segmentation is the incorporation of multi-scale information, exemplified by the use of feature pyramid networks (FPN). This enables capturing fine-grained details and global context, which are crucial for achieving accurate tooth segmentation [62, 63]. Furthermore, data augmentation has become a valuable technique for improving the performance of deep learning in tooth segmentation, particularly when dealing with limited annotated data. Data augmentation contributes to the efficient training of deep learning models, as demonstrated in the application of Region Proposal Network (RPN) combined with FPN and U-Net [64].

Aside from the mainstream convolutional models, Transformer-based architectures have also been explored for tooth segmentation. Transformer models are multi-head self-attention networks that overcome memory constraints and capture long-range dependencies, originally

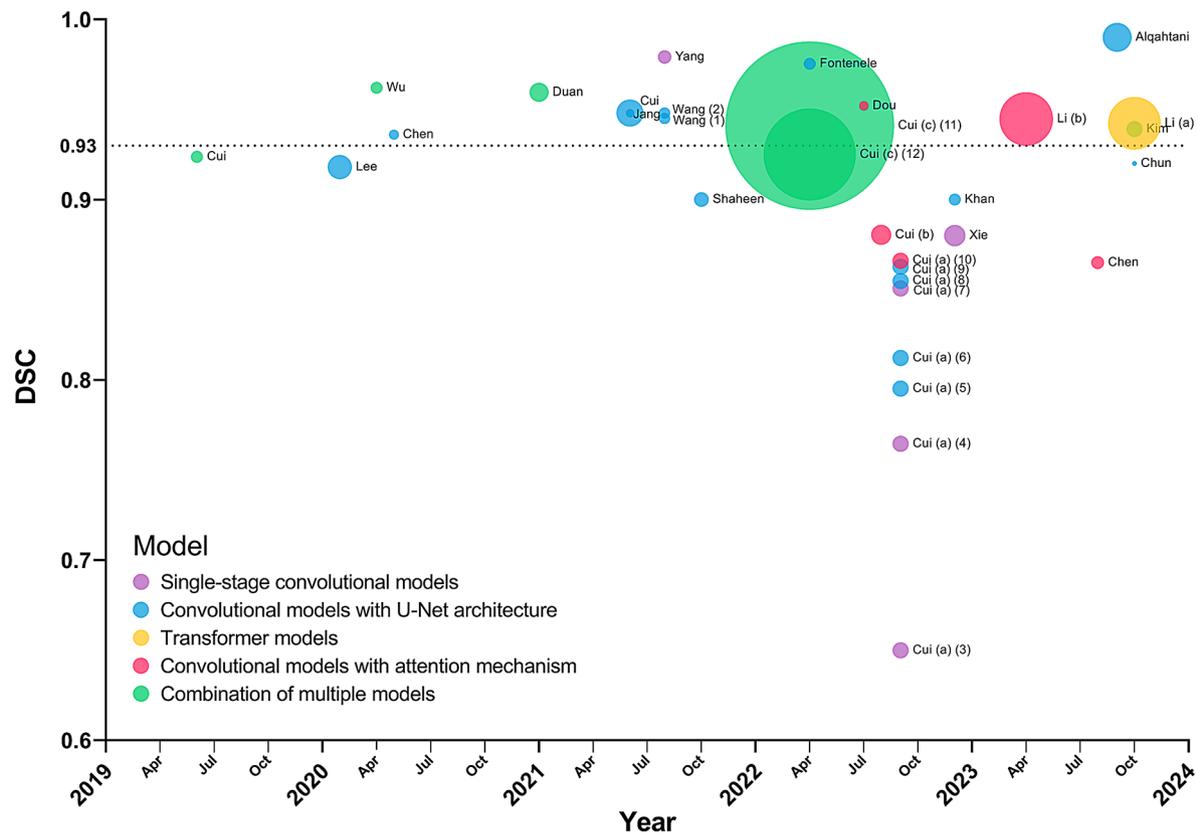


Fig. 3 Bubble plot of Dice Similarity Coefficient of different models used in different studies over the years. *Notes: Circle colours indicate different groups of deep learning architectures. Circle sizes are proportional to the sample size (number of test scans) used in validation of the models. The dotted line represents the pooled mean of Dice Similarity Coefficient. (1) Multiclass segmentation subgroup. (2) Binary segmentation subgroup. (3) SkipDenseNet model. (4) DenseVoxelNet model. (5) 3D U-Net model. (6) V-Net model. (7) Voxresnet model. (8) nnU-Net model. (9) Dense U-Net model. (10) AttU-Net model. 11. Internal validation dataset. 12. External validation dataset

developed for natural language processing [65]. These networks have revolutionized sequence modeling by global dependency modeling and parallelization, leading to enhanced computational efficiency [66–68]. With the advent of the Vision Transformer (ViT), the application of Transformer-based architectures in image segmentation has expanded significantly. These networks process a sequence of image patches and exhibit more consistent prediction errors compared to human beings than CNNs [67]. Transformer-based models are adept at encoding long-range dependencies and learning highly effective feature representations compared to CNNs [69]. While Transformer-based algorithms excel at capturing global relationships within data, a significant challenge remains in the necessity for pre-training on large datasets, limiting its real-world feasibility [67]. Further modifications of deep learning mainly incorporate the combination of different algorithms, such as combining convolutional models with the attention mechanism from Transformer models [44], aiming to integrate the strengths of various models while mitigating their weaknesses. The future progression of deep learning algorithms is likely to involve the continued integration of convolutional

models with Transformer attention mechanisms, with the aim of enhancing segmentation accuracy while simultaneously reducing computational time.

The sample sizes varied across different studies, and no clear relationship between sample size and segmentation accuracy has been observed. In traditional clinical studies, researchers need to calculate the target sample size using precision or power analysis [70]. However, in the AI era, sample size calculation aims to determine the number of images required for a machine learning algorithm to reach a specific performance threshold or maintain a sufficiently low generalization error [71]. A common method for optimizing sample size is the use of a post-hoc (curve fitting) approach, which involves fitting the learning curve of the AI model on varying sizes of training datasets [72]. As the sample size increases, the training loss tends to stabilize, allowing the sample size to be determined to avoid undertraining or overtraining of AI models. Therefore, an excessive sample size may not contribute to further improvement in the model's performance, explaining our findings.

Various evaluation metrics have been employed to assess the performance of tooth segmentation, with

DSC being the most prevalent, as observed in previous systematic reviews [23]. These diverse metrics can be classified into three categories: overlap-based metrics, distance-based metrics, and volume-based metrics [23]. For example, DSC is an overlap-based metric that measures the spatial overlap between two sets of binary segmentation results [23, 73]. The performance of deep learning algorithms often varies depending on the metrics used. The prevalent use of DSC during model training usually contributes to its high score in the evaluation phase compared to other metrics [74]. Therefore, although DSC is frequently used as the primary metric to evaluate an algorithm, it is crucial to consider other metrics to obtain a comprehensive assessment of the overall performance.

Although various evaluation metrics have been employed in the research [23], there is a lack of clinically relevant metrics that evaluate the clinical applicability of tooth segmentation. Tooth segmentation in CT/CBCT images can be used to examine tooth morphology and positioning and can be applied in different clinical scenarios. For instance, in orthodontic treatment, metal artefacts caused by brackets and wires can influence the segmentation accuracy of crowns, potentially affecting the simulation of tooth movements during alignment [75]. The impact of these artefacts, however, remains unverified. In dentoalveolar surgeries, accurately segmenting the root apex is vital for proper apical surgery guidance and prevention of root remnants, especially in curved roots. Overlooking small root tips might not noticeably affect segmentation accuracy, but it can significantly impact clinical outcomes [24, 27]. While our study pooled existing data to present a reference DSC value of 0.93, it is important to approach clinical interpretation with caution. The collaboration between AI researchers, mathematicians, and clinicians is crucial for developing evaluation metrics specifically tailored for clinical applications, taking into account factors such as the type of tooth and the location of teeth in relation to adjacent vital structures, to comprehensively assess the performance of tooth segmentation.

In our meta-analyses, all pooled metrics exhibited significant data heterogeneity ($I^2 > 99\%$). Statistically, I^2 describes the percentage of variability in effect estimates due to heterogeneity [76], and statistical heterogeneity may also arise from clinical and methodological heterogeneity. Each included study had unique datasets and deep learning models, leading to clinical heterogeneity. Furthermore, the included studies demonstrated variability in their designs, including the use of manual or semi-automatic segmentation as the reference standard, the implementation of internal and/or external validation for model development, and the utilization of different imaging modalities such as CBCT and CT.

For methodological heterogeneity, it is worth noting that some studies reported the number of teeth, while others used the number of scans or individuals as the sample size. Therefore, an estimate of the number of teeth was made, assuming that each participant had 28 teeth. This potential overestimation of the number of teeth may distort the weighting of studies and contribute to the heterogeneity of the pooled result. Additionally, our findings indicated a high prevalence of unclear risk of bias in the patient selection domain, with more than half of the included studies exhibiting a high risk of bias in the reference standard domain. Although most studies provided clear inclusion and exclusion criteria for their populations, they often failed to specify whether the sample was selected through random sampling, resulting in an unclear risk of bias. In the reference standard domain, while the methods for obtaining ground truth were generally specified, the absence of cross-checking for the reference standard led to a high risk of bias.

These concerns regarding the risk of bias in the included studies should be taken into account when interpreting the results. To improve comparability between future studies, it is recommended that research be conducted with standardized protocols and open-labelled datasets. Additionally, it is essential to specify the sampling methods, cross-check the reference standard, and report the number of teeth used for training and validation of the dataset.

To the best of our knowledge, this systematic review and meta-analysis is the first and most comprehensive effort to investigate the application of deep learning algorithms in tooth segmentation and characterize the evolution of deep learning algorithms in this field. The DSC value of 0.93 represents the updated accuracy of tooth segmentation utilizing deep learning that could provide a reference for similar future studies. However, it should be acknowledged that the included studies exhibited significant heterogeneity and high risk of bias, which impacted the pooled results. Furthermore, further clinical studies are warranted to confirm the clinical applicability of deep learning tooth segmentation from CT/CBCT. Future research could explore advanced applications such as the recognition of supernumerary teeth, the automatic detection of periapical lesions, and the integration of deep learning tools into dental education.

Conclusion

The application of deep learning algorithms in tooth segmentation has significantly advanced the segmentation process. Notably, the popularity of U-Net and U-Net-based algorithms, as well as the emergence of the Transformer model and the combination of multiple models, should be highlighted as potential future trends in the development of deep learning algorithms for

tooth segmentation. This study provides a reference DSC value of 0.93, which can serve as a benchmark for future research in this field. However, it is crucial to conduct studies with standardized protocols, evaluation metrics, and open-labelled datasets to gain a better comparison of different deep learning algorithms employed in tooth segmentation. Additionally, more appropriate evaluation metrics should be developed for use in the clinical setting.

Abbreviations

CT	Computed tomography
CBCT	Cone-beam computed tomography
AI	Artificial intelligence
PRISMA-P	Preferred Items for Reporting Systematic Reviews and Meta-analyses for protocols
QUADAS-2	Quality Assessment of Diagnostic Accuracy Study tool
IoU	Intersection over Union
DSC	Dice similarity
AJI	Aggregated JACCARD Index
Acc	Accuracy
BF	Boundary F1 Score
DA	Detection Accuracy
F1	F1 Score
FA	Identification Accuracy
PA	Pixel Accuracy
PPV	Positive Predictive Values
SD	Surface Dice
sDSC	surface Dice Similarity Coefficient at 1 mm
Sen	Sensitivity/Recall
SO	Surface Overlap
vDSC	volumetric Dice Similarity Coefficient
VOE	Volumetric Overlap Error
wDSC	weighted Dice Similarity Coefficient
95HD	95% Hausdorff Distance
ASSD	Average Symmetric Surface Distance
HD	Hausdorff Distance
MAD	Mean Absolute Deviation
MSSD	Maximum Symmetric Surface Distance
RVD	Relative Volume Difference
VD	Volume Difference
VS	Volume Similarity

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12903-025-05984-6>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors would like to express their gratitude to Kar Yan Li from the Data Center of the Faculty of Dentistry at the University of Hong Kong for her invaluable assistance in conducting the statistical analysis for the meta-analysis.

Author contributions

W.Y.K: Contributed to conceptualization, methodology, data curation, formal analysis, and writing – original draft, review and editing. S.Y.A.Y: Contributed to conceptualization, data curation, formal analysis, and writing – review and editing. Y.Y.L: Contributed to conceptualization, data curation, formal analysis, and writing – review and editing. P.H.L: Contributed to data curation, software, and writing – review and editing. W-f.Y: Contributed to conceptualization, supervision, and writing – review and editing. All authors gave their final approval and agreed to be accountable for all aspects of the work.

Funding

This study is partially supported by: (1) Undergraduate Research Programme 2023, Faculty of Dentistry, University of Hong Kong; (2) Health and Medical Research Fund (11221966), Food and Health Bureau, Hong Kong.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 13 February 2025 / Accepted: 10 April 2025

Published online: 26 May 2025

References

- Camps-Perepérez I, Guijarro-Martínez R, Peiró-Guijarro MA, Hernández-Alfaro F. The value of cone beam computed tomography imaging in surgically assisted rapid palatal expansion: a systematic review of the literature. *Int J Oral Maxillofac Surg*. 2017;46(7):827–38.
- Alkhayer A, Piffko J, Lippold C, Segatto E. Accuracy of virtual planning in orthognathic surgery: a systematic review. *Head Face Med*. 2020;16(1):34.
- Borges CC, Estrela C, Decurcio DA, PÉcora JD, Sousa-Neto MD, Rossi-Fedele G. Cone-beam and micro-computed tomography for the assessment of root Canal morphology: a systematic review. *Braz Oral Res*. 2020;34:e056.
- Andriola FO, Haas Junior OL, Guijarro-Martínez R, Hernández-Alfaro F, Oliveira RB, Pagnoncelli RM, et al. Computed tomography imaging superimposition protocols to assess outcomes in orthognathic surgery: a systematic review with comprehensive recommendations. *Dentomaxillofac Radiol*. 2022;51(3):20210340.
- Peralta-Mamani M, Rubira CM, Lopez-Lopez J, Honorio HM, Rubira-Bullen IR. CBCT vs panoramic radiography in assessment of impacted upper canine and root resorption of the adjacent teeth: A systematic review and meta-analysis. *J Clin Exp Dent*. 2024;16(2):e198–222.
- Wang H, Minnema J, Batenburg KJ, Forouzanfar T, Hu FJ, Wu G. Multiclass CBCT image segmentation for orthodontics with deep learning. *J Dent Res*. 2021;100(9):943–9.
- 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*. 2020;90(1):77–84.
- Ahmed S, Saifuddin KM, Ahmed AS, Hossain ABMA, Iqbal MT. Identification and volume Estimation of dental caries using CT image. *2nd IEEE Int Conf Telecommunications Photonics ICTP 2017*. 2017;2017–December:48–51.
- Elgarba BM, Van Aelst S, Swaity A, Morgan N, Shujaat S, Jacobs R. Deep learning-based segmentation of dental implants on cone-beam computed tomography images: A validation study. *J Dent*. 2023;137.
- Dumbryte I, Narbutis D, Vailionis A, Juodkazis S, Malinauskas M. Revelation of microcracks as tooth structural element by X-ray tomography and machine learning. *Sci Rep*. 2022;12(1).
- Albitar L, Zhao T, Huang C, Mahdian M. Artificial intelligence (AI) for detection and localization of unobturated second mesial buccal (MB2) canals in Cone-Beam computed tomography (CBCT). *Diagnostics*. 2022;12(12).
- Ahmad R, Abu-Hassan MI, Li Q, Swain MV. Three dimensional quantification of mandibular bone remodeling using standard tessellation Language registration based superimposition. *Clin Oral Implants Res*. 2013;24(11):1273–9.
- Xu P, Gholamalizadeh T, Moshfeghifar F, Darkner S, Erleben K. Deep-Learning-Based segmentation of individual tooth and bone with periodontal ligament interface details for simulation purposes. *IEEE Access*. 2023;11:102460–70.
- De Grauwe A, Ayaz I, Shujaat S, Dimitrov S, Gbadegbegnon L, Vande Vannet B, et al. CBCT in orthodontics: a systematic review on justification of CBCT

- in a paediatric population prior to orthodontic treatment. *Eur J Orthod.* 2019;41(4):381–9.
15. Muresanu S, Almasan O, Hedesiu M, Diosan L, Dinu C, Jacobs R. Artificial intelligence models for clinical usage in dentistry with a focus on dentomaxillofacial CBCT: a systematic review. *Oral Radiol.* 2023;39(1):18–40.
 16. Naumovich SS, Naumovich SA, Goncharenko VG. Three-dimensional reconstruction of teeth and jaws based on segmentation of CT images using watershed transformation. *Dentomaxillofacial Radiol.* 2015;44(4).
 17. Zhang L, Li W, Lv J, Xu J, Zhou H, Li G et al. Advancements in oral and maxillofacial surgery medical images segmentation techniques: an overview. *J Dent.* 2023;138.
 18. Hosntalab M, Zoroofi RA, Tehrani-Fard AA, Shirani G, editors. Automated dental recognition in MSCT images for human identification. IIH-MSP 2009–2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing; 2009.
 19. Jiang B, Zhang S, Shi M, Liu HL, Shi H. Alternate level set evolutions with controlled switch for tooth segmentation. *IEEE Access.* 2022;10:76563–72.
 20. Mortaheb P, Rezaeian M, Soltanian-Zadeh H, editors. Automatic dental CT image segmentation using mean shift algorithm. Iranian Conference on Machine Vision and Image Processing, MVIP; 2013.
 21. Yang H, Wang X, Li G. Tooth and pulp chamber automatic segmentation with artificial intelligence network and morphometry method in Cone-beam CT. *Int J Morphology.* 2022;40(2):407–13.
 22. Alqahtani KA, Jacobs R, Smolders A, Van Gerven A, Willems H, Shujaat S, et al. Deep convolutional neural network-based automated segmentation and classification of teeth with orthodontic brackets on cone-beam computed-tomographic images: a validation study. *Eur J Orthod.* 2023;45(2):169–74.
 23. Polizzi A, Quinzi V, Ronsivalve V, Venezia P, Santonocito S, Lo Giudice A, et al. Tooth automatic segmentation from CBCT images: a systematic review. *Clin Oral Invest.* 2023;27(7):3363–78.
 24. Badr FF, Jadu FM. Performance of artificial intelligence using oral and maxillofacial CBCT images: A systematic review and meta-analysis. *Niger J Clin Pract.* 2022;25(11):1918–27.
 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. *Rev Esp Cardiol (Engl Ed).* 2021;74(9):790–9.
 26. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529–36.
 27. Chun SY, Kang YH, Yang S, Kang SR, Lee SJ, Kim JM et al. Automatic classification of 3D positional relationship between mandibular third molar and inferior alveolar Canal using a distance-aware network. *BMC Oral Health.* 2023;23(1).
 28. Cui Z, Li C, Wang W, editors. ToothNet: Automatic Tooth Instance Segmentation and Identification From Cone Beam CT Images. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019 15–20 June 2019.
 29. 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.* 2020;8:97296–309.
 30. 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.* 2020;8:50507–18.
 31. Rao Y, Wang Y, Meng F, Pu J, Sun J, Wang Q. A symmetric fully convolutional residual network with DCRF for accurate tooth segmentation. *IEEE Access.* 2020;8:92028–38.
 32. Wu XY, Chen H, Huang YJ, Guo HY, Qiu TT, Wang LS editors. CENTER-SENSITIVE AND BOUNDARY-AWARE TOOTH INSTANCE SEGMENTATION AND CLASSIFICATION FROM CONE-BEAM CT. IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020 Apr 03–07; Iowa, IA2020.
 33. 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).
 34. Jang TJ, Kim KC, Cho HC, Seo JK. A fully automated method for 3D individual tooth identification and segmentation in dental CBCT. *IEEE Trans Pattern Anal Mach Intell.* 2022;44(10):6562–8. <https://doi.org/10.1109/TPAMI.2021.3086072>
 35. Yang Y, Xie R, Jia W, Chen Z, Yang Y, Xie L, et al. Accurate and automatic tooth image segmentation model with deep convolutional neural networks and level set method. *Neurocomputing.* 2021;419:108–25.
 36. Cui WW, Wang YQ, Li YL, Song D, Zuo XY, Wang JJ, et al. editors. CTooth plus: A Large-Scale Dental Cone Beam Computed Tomography Dataset and Benchmark for Tooth Volume Segmentation. 2nd MICCAI International Workshop on Data Augmentation, Labeling, and Imperfections (DALI); 2022; Singapore, SINGAPORE2022.
 37. Cui WW, Wang YQ, Zhang QN, Zhou HY, Song D, Zuo XY, et al. editors. CTooth: A Fully Annotated 3D Dataset and Benchmark for Tooth Volume Segmentation on Cone Beam Computed Tomography Images. 15th International Conference on Intelligent Robotics and Applications (ICIRA) - Smart Robotics for Society; 2022 Aug 01–03; Harbin, PEOPLES R CHINA2022.
 38. 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.* 2022;13(1).
 39. Dou W, Gao S, Mao D, Dai H, Zhang C, Zhou Y, editors. Tooth instance segmentation based on capturing dependencies and receptive field adjustment in cone beam computed tomography. *Computer Animation and Virtual Worlds;* 2022.
 40. Fontenele RC, Gerhardt MD, Pinto JC, Van Gerven A, Willems H, Jacobs R et al. Influence of dental fillings and tooth type on the performance of a novel artificial intelligence-driven tool for automatic tooth segmentation on CBCT images-A validation study. *J Dent.* 2022;119.
 41. Khan S, Mukati A, Rizvi SSH, Yazdanie N, TOOTH, SEGMENTATION IN 3D CONE-BEAM CT IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORK. *Neural Netw World.* 2022;32(6):301–18.
 42. Xie L, Liu B, Cao Y, Yang C, editors. Automatic Individual Tooth Segmentation in Cone-Beam Computed Tomography Based on Multi-Task CNN and Watershed Transform. Proceedings–24th IEEE International Conference on High Performance Computing and Communications, 8th IEEE International Conference on Data Science and Systems. 20th IEEE International Conference on Smart City and 8th IEEE International Conference on Dependability in Sensor, Cloud and Big Data Systems and Application, HPCC/DSS/SmartCity/DependSys 2022; 2022.
 43. Alqahtani KA, Jacobs R, Shujaat S, Politis C, Shaheen E. Automated three-dimensional quantification of external root resorption following combined orthodontic-orthognathic surgical treatment. A validation study. *J Stomatol-ogy Oral Maxillofacial Surg.* 2023;124(1).
 44. Chen Z, Chen S, Hu F. CTA-UNet: CNN-transformer architecture UNet for dental CBCT images segmentation. *Phys Med Biol.* 2023;68(17).
 45. Kim SH, Song IS, Baek SJ, editors. Automatic Segmentation of Internal Tooth Structure from CBCT Images Using Hierarchical Deep Learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics);* 2023.
 46. Li S, Li C, Du Y, Ye L, Fang Y, Wang C, et al. editors. Transformer-Based Tooth Segmentation, Identification and Pulp Calcification Recognition in CBCT. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics);* 2023.
 47. Li S, Du Y, Ye L, Li C, Fang Y, Wang C, et al. editors. Teeth and Root Canals Segmentation using Zxyformer with Uncertainty Guidance and Weight Transfer. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI); 2023 18–21 April 2023.
 48. Ezhov M, Zakirov A, Gusarev M, editors. Coarse-to-fine volumetric segmentation of teeth in cone-beam ct. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019); 2019 8–11 April 2019.
 49. Cui Z, Zhang B, Lian C, Li C, Yang L, Wang W, et al. editors. Hierarchical Morphology-Guided Tooth Instance Segmentation from CBCT Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics);* 2021.
 50. Shaheen E, Leite A, Alqahtani KA, Smolders A, Van Gerven A, Willems H, et al. A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study. *J Dent.* 2021;115:103865.
 51. Dot G, Schouman T, Dubois G, Rouch P, Gajny L. Fully automatic segmentation of craniomaxillofacial CT scans for computer-assisted orthognathic surgery planning using the nnU-Net framework. *Eur Radiol.* 2022;32(6):3639–48.
 52. Gerhardt MDN, Fontenele RC, Leite AF, Lahoud P, Van Gerven A, Willems H et al. Automated detection and labelling of teeth and small edentulous regions on cone-beam computed tomography using convolutional neural networks. *J Dent.* 2022;122.
 53. Tao S, Wang Z. Tooth CT image segmentation method based on the U-Net network and attention module. *Comput Math Methods Med.* 2022;2022:3289663.
 54. Wu J, Zhang M, Yang D, Wei F, Xiao N, Shi L, et al. Clinical tooth segmentation based on local enhancement. *Front Mol Biosci.* 2022;9:932348.

55. 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. *Comput Biol Med.* 2020;120.
56. Jang TJ, Kim KC, Cho HC, Seo JK. A fully automated method for 3D individual tooth identification and segmentation in dental CBCT. *IEEE Trans Pattern Anal Mach Intell.* 2022;44(10):6562–8.
57. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015 May 01, 2015:[arXiv:1505.04597]. Available from: <https://ui.adsabs.harvard.edu/abs/2015arXiv150504597R>
58. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol.* 2017;10(3):257–73.
59. Chen H, Dou Q, Yu L, Heng P-A, VoxResNet. Deep Voxelwise Residual Networks for Volumetric Brain Segmentation 2016 August 01, 2016:[arXiv:1608.05895 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv160805895C>
60. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation 2016 June 01, 2016:[arXiv:1606.06650 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv160606650C>
61. He Y, Yu H, Liu X, Yang Z, Sun W, Anwar S et al. Deep Learning Based 3D Segmentation: A Survey 2021 March 01, 2021:[arXiv:2103.05423 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2021arXiv210305423H>
62. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection 2016 December 01, 2016:[arXiv:1612.03144 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv161203144L>
63. Elizar E, Zulkifley MA, Muharrar R, Zaman MHM, Mustaza SM. A review on Multiscale-Deep-Learning applications. *Sens (Basel).* 2022;22(19).
64. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. *J Med Imaging Radiat Oncol.* 2021;65(5):545–63.
65. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al. Attention Is All You Need 2017 June 01, 2017:[arXiv:1706.03762 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170603762V>
66. Khan RF, Lee BD, Lee MS. Transformers in medical image segmentation: a narrative review. *Quant Imaging Med Surg.* 2023;13(12):8747–67.
67. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: A survey. *Med Image Anal.* 2023;88:102802.
68. Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y, et al. Advances in medical image analysis with vision Transformers: A comprehensive review. *Med Image Anal.* 2024;91:103000.
69. Chaudhari S, Mithal V, Polatkan G, Ramanath R. An Attentive Survey of Attention Models 2019 April 01, 2019:[arXiv:1904.02874 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2019arXiv190402874C>
70. Chow S-C, Shao J, Wang H, Lokhnygina Y. Sample size calculations in clinical research. 3rd ed. New York: Chapman and Hall/CRC; 2017. 24 August 2017.
71. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-Size determination methodologies for machine learning in medical imaging research: A systematic review. *Can Assoc Radiol J.* 2019;70(4):344–53.
72. Zhang X, Oymak S, Chen J. Post-hoc Models for Performance Estimation of Machine Learning Inference 2021 October 01, 2021:[arXiv:2110.02459 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2021arXiv211002459Z>
73. Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a Spatial overlap index. *Acad Radiol.* 2004;11(2):178–89.
74. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support* (2017). 2017;2017:240-8.
75. Kim TH, Park JJ. Teeth segmentation for orthodontics based on deep learning. *Trans Korean Inst Electr Eng.* 2023;72(3):440–6.
76. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023): Cochrane; 2023. Available from: www.training.cochrane.org/handbook

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.