



Research Paper

Automated Teeth Segmentation in Three-Dimensional Dental Images Using a Hybrid GoogLeNet-Residual Connection Architecture with Attention Mechanism

Thushara Hameed^{*} , S Amala Shanthi^{ID}

Department of Electronics and Communication, Noorul Islam Center for Higher Education, Tamil Nadu, India.

Citation Hameed T, Amala Shanthi S. Automated Teeth Segmentation in Three-Dimensional Dental Images Using a Hybrid GoogLeNet-Residual Connection Architecture with Attention Mechanism. *International Journal of Medical Toxicology and Forensic Medicine*. 2026; 16:E53264.

<https://doi.org/10.22037/ijmtfm.v16.53264>

Article info:

Received: 24 August, 2026

Accepted: 11 September, 2026

Published: 02 October, 2026

Keywords:

Hybrid architecture,
GoogLeNet, Residual
network (ResNet), Attention
mechanism, Pre-processing,
Hierarchical features

ABSTRACT

Background: Automatic tooth segmentation in three-dimensional (3D) dental images is crucial for computer-aided diagnosis, treatment planning, and dental research. However, due to the complexity of dental structures, morphological differences, variations in image quality, background noise, and other factors, segmentation remains challenging. Current deep learning methods may also struggle to preserve fine-grained structural details and identify informative regions in intricate 3D dental images.

Methods: In this study, we propose a hybrid deep learning model that combines an improved GoogLeNet, a Residual Network (ResNet), and an attention mechanism for automatic tooth segmentation in 3D dental images. First, we apply several preprocessing steps to enhance image quality and normalize the data. Then, we use an enhanced GoogLeNet architecture with additional residual connections to capture hierarchical, fine-grained features and mitigate vanishing gradients. Moreover, we design an attention mechanism to adaptively select informative dental regions while suppressing irrelevant background information, making the segmentation results more robust and informative. We test the proposed framework on a large dataset of 3D dental images from patients with various dental conditions.

Results: The proposed hybrid framework outperforms existing approaches in segmentation performance and generalization, achieving 99% Accuracy. The incorporation of residual connections helps extract detailed dental features, while the attention mechanism aids in identifying relevant dental regions and improves the interpretability of the segmentation results.

Conclusion: The proposed framework, GoogLeNet-RSNet-Attention, is an effective model for automatically segmenting teeth in 3D Dental images. The method can be used in computer-aided dental diagnosis, treatment planning, and clinical research. More validation with dentists and clinical data is needed to demonstrate its applicability in clinical practice.

* Corresponding Author:

Thushara Hameed, PhD

Department of Electronics and Communication, Noorul Islam Center for Higher Education, Tamil Nadu, India.

E-mail: thushararesearchscholar@gmail.com



Copyright © 2026 The Author(s);

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0);

<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Three-dimensional (3D) dental imaging is a tool for acquiring detailed images of dental anatomy for diagnosis, treatment planning, and quantification. This section introduces the topic and reviews the literature. Delineating individual teeth, as in other areas of computer-aided dentistry such as diagnosis, treatment planning, orthodontic analysis, and endodontic procedures, is important for automatically segmenting 3D dental images. Manual segmentation is time-consuming, observer-dependent, and difficult when tooth borders are affected by anatomic overlap, restorations, root proximity, and imaging artifacts. Cone-beam computed tomography (CBCT) has been demonstrated to be a useful tool for providing volumetric information. Deep learning has hence become a viable strategy to minimize manual efforts and enhance uniformity. Lahoud et al. [1] demonstrated fast and accurate tooth segmentation in 3D CBCT images, and Shaheen et al. [2] extended this work to multi-class tooth segmentation and classification. Similarly, Jang et al. [3] proposed a fully automatic solution for tooth-level identification and segmentation, demonstrating the feasibility of tooth-level analysis in 3D dental imaging. A few studies have examined specialized architectures to improve segmentation performance. Li et al. [4] combined a U-Net and an RNN to automatically segment tooth roots in CBCT sequences, and Hsu et al. [5] used majority voting and a 3.5D U-Net to enhance the reliability of the segmentation results. Recently, Wen et al. [6] introduced a multi-scale feature fusion method for accurate 3D tooth segmentation, emphasizing the importance of multi-scale feature interactions. Wang et al. [7] used deep multi-task feature learning for simultaneous tooth and root-canal segmentation to assist root-canal treatment planning. Hu et al. [8] focused on challenging samples from oral CBCT and highlighted that the main causes of error are difficult anatomical regions and poorly defined boundaries.

In the broader literature, there has also been a rise in studies on automated dental segmentation. Polizzi et al. [9] conducted a thorough survey of automatic tooth segmentation in CBCT scans and noted the growing use of deep learning in this area. Zheng et al. [10] compared semi-automatic and fully automatic tooth-segmentation methods, while Chen et al. [11] summarized deep learning-based tooth-segmentation methods used in medical imaging. Zhou et al. [12] suggested an open deep learning method for the instance segmentation of dental CBCT. To further document the evolution of deep learning for tooth segmentation in CT and CBCT images, Kot et al. [13] conducted a systematic review

and meta-analysis of deep learning for tooth segmentation in CT and CBCT images, respectively. Beyond segmentation, Zhou et al. [14] evaluated an automated dental segmentation system's ability to generate diagnostic reports, suggesting that segmentation systems could be integrated into a broader clinical decision-support system.

Medical imaging also provides insights into architecture. Zakareya et al. [15] demonstrated the effectiveness of deep learning in medical image diagnosis, and Balraj et al. [16] showed that a multi-level attention mechanism and dilated residual learning can enhance feature representation in medical image segmentation. Nevertheless, accurate tooth segmentation remains challenging due to the diversity of 3D dental images in tooth morphology, image quality, background structures, and local boundary contrast. The current methods can also fail to preserve fine details during deep feature extraction and may not always focus on important anatomical areas. Therefore, this work proposes a hybrid approach consisting of GoogLeNet, residual connections, and an attention mechanism. GoogLeNet captures multi-scale hierarchical features; residual connections help ensure stable deep feature learning and preserve structural information; the attention mechanism focuses the network on tooth-related regions and suppresses responses from irrelevant background areas. The framework is designed to improve segmentation accuracy, robustness, and interpretability in 3D dental image analysis and to aid clinicians' interpretation across various imaging scenarios. The proposed hybrid GoogLeNet-ResNet model for automated tooth segmentation therefore fuses the multi-scale features of GoogLeNet, ResNet-inspired residual connections, and a channel attention mechanism. Residual connections help pass gradients through the network, and the attention mechanism adaptively weights informative channels. Measurement criteria for 3D dental images include usability for routine dental image analysis and treatment planning. The generalizability and actual impact of the integrated architecture on clinical practice should be demonstrated through more comparative assessments and independent clinical validation.

Materials and Methods

Dataset Description

The given methodology uses a large set of 3D images of dental features, denoted as D , comprising volumetric scans from multiple imaging modalities, including cone-beam computed tomography (CBCT) and magnetic resonance imaging (MRI). The 3D dental image in the dataset is denoted by X_i , where i ranges

from 1 to N. Ground-truth segmentation masks indicating individual teeth, denoted Y_i , are annotated in the dataset and used for training and evaluation supervision. Figure 1 depicts the Proposed Architecture of the work.

Pre-processing Techniques:

Intensity Normalization: Intensity normalization is applied to 3D dental images to improve image quality and equalize intensities across scans. This measure reduces the effects of variations in imaging parameters and ensures a consistent intensity distribution in the input data. Intensity normalization is done to match the intensity values of various scans:

Mathematical Formulation of Pre-processing

Let the input 3D dental image volume be represented as

$$X \in \mathbb{R}^{H \times W \times D} \quad (1)$$

where H, W, and D denote the height, width, and depth of the volumetric scan, respectively. The objective of preprocessing is to transform X into an enhanced and standardized volume $\tilde{X}$ suitable for deep feature extraction.

Intensity Normalization

Due to variations in acquisition devices and

imaging conditions, voxel intensity values vary across scans. To reduce inter-scan variability, z-score normalization is applied:

$$X_n(i, j, k) = \frac{X(i, j, k) - \mu_X}{\sigma_X} \quad (2)$$

where $X(i, j, k)$ denotes the voxel intensity at spatial location (i, j, k) , μ_X is the mean intensity of the volume, and σ_X is the corresponding standard deviation. This normalization ensures a zero-mean and unit-variance intensity distribution.

Noise Reduction

To suppress acquisition noise while preserving anatomical details, Gaussian filtering is applied:

$$X_d = X_n * G_\sigma \quad (3)$$

where $*$ denotes the 3D convolution operator and G_σ represents a Gaussian kernel with standard deviation σ . This step improves structural continuity in dental regions.

Voxel Resampling

To ensure spatial consistency across datasets, voxel resampling is performed to obtain isotropic resolution:

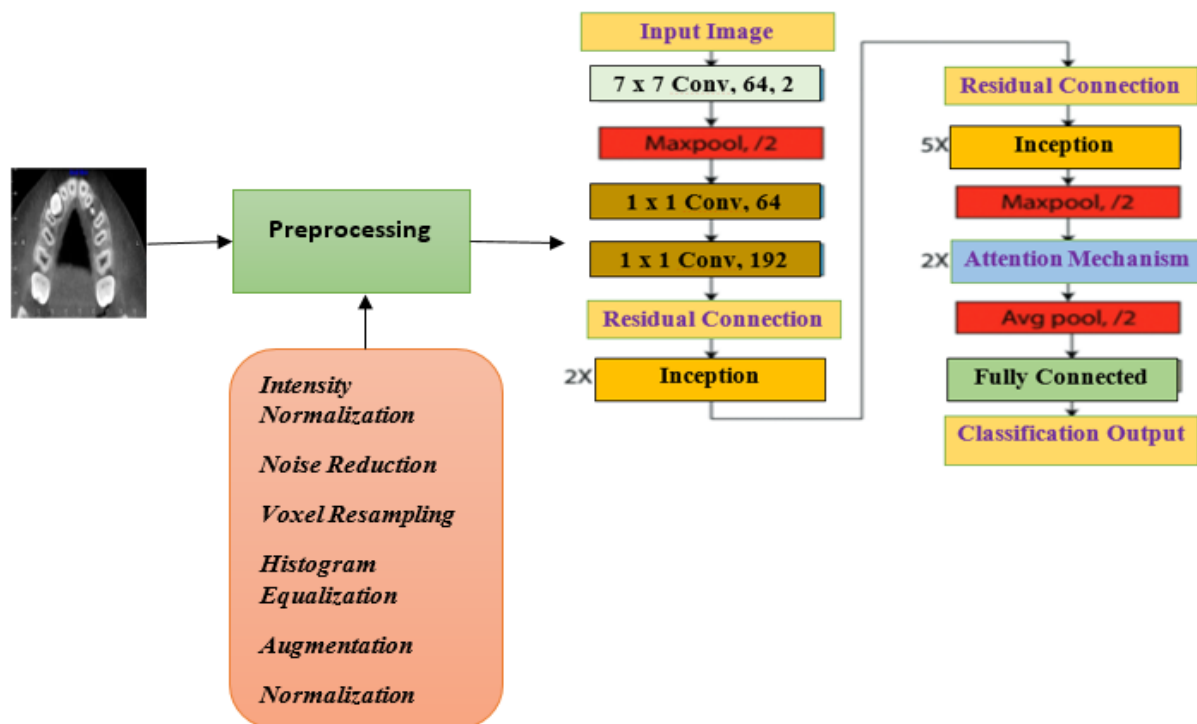


Figure 1. Proposed Architecture.

$$X_r = \mathcal{R}(X_d; s_x, s_y, s_z) \quad (4)$$

where $\mathcal{R}(\cdot)$ denotes the resampling operator and (s_x, s_y, s_z) are the target voxel spacings along each spatial axis.

Histogram Equalization

Contrast enhancement is achieved using histogram equalization:

$$X_h = \mathcal{H}(X_r) \quad (5)$$

where $\mathcal{H}(\cdot)$ redistributes voxel intensities to improve contrast between teeth and surrounding tissues, enhancing segmentation discriminability.

Data Augmentation

To improve generalization and mitigate overfitting, data augmentation is applied during training:

$$X_a = \mathcal{A}(X_h) \quad (6)$$

where $\mathcal{A}(\cdot)$ includes random 3D rotations, translations, scaling, and flipping operations.

Intensity Scaling

For numerical stability and faster convergence, voxel intensities are scaled to a fixed range:

$$X_s = \frac{X_a - \min(X_a)}{\max(X_a) - \min(X_a)} \quad (7)$$

ensuring $X_s \in [0,1]$.

Dataset Partitioning

The processed dataset is divided into training, validation, and testing subsets:

$$\{X_{\text{train}}, X_{\text{val}}, X_{\text{test}}\} = \mathcal{S}(X_s) \quad (8)$$

where $\mathcal{S}(\cdot)$ represents the dataset splitting function.

Overall Pre-processing Pipeline

The complete preprocessing transformation is summarized as:

$$\tilde{X} = \mathcal{S} \left(\mathcal{N} \left(\mathcal{A} \left(\mathcal{H} \left(\mathcal{R} \left(\mathcal{D} \left(\mathcal{Z}(X) \right) \right) \right) \right) \right) \right) \quad (9)$$

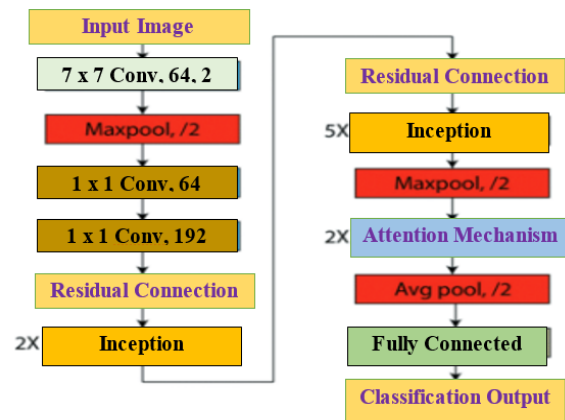
where $\mathcal{Z}$, $\mathcal{D}$, $\mathcal{R}$, $\mathcal{H}$, $\mathcal{A}$, $\mathcal{N}$, and $\mathcal{S}$ represent normalization, denoising, resampling, histogram equalization, augmentation, scaling, and dataset splitting, respectively. The suggested hybrid GoogLeNet-ResNet structure is combined with an inbuilt attention mechanism, which enables intensive hierarchical feature learning and precise localization of individual teeth. The close integration of rigorous mathematical preprocessing with the proposed deep learning system significantly improves segmentation precision, robustness, and extrapolation capability in complex 3D dental imaging settings.

Hybrid Architecture with Attention Mechanism

The suggested architecture comprises 3 major parts: a modified GoogLeNet with residual connections, an attention mechanism, and a final segmentation module. The architecture will be designed to effectively extract hierarchical features, dynamically accentuate regions of interest, and accurately segment teeth. The Hybrid Architecture with Attention Mechanism is shown in Figure 2.

Modified GoogLeNet with Residual Connections

The GoogLeNet architecture is modified and used as the feature extractor for the input 3D dental images.



International Journal of
Medical Toxicology & Forensic Medicine

Figure 2. Hybrid architecture with attention mechanism.

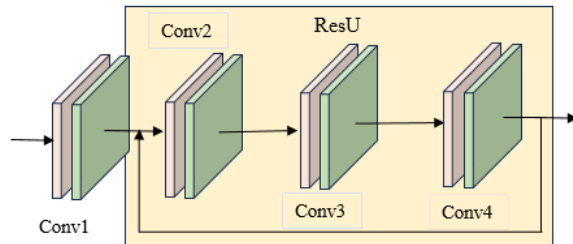
It has several convolutional layers arranged in parallel routes, thereby extracting multi-scale features. The residual connections, introduced in ResNet, are implemented in every convolutional block to mitigate vanishing gradients and enable effective gradient propagation throughout training. These residual connections enable information to pass through multiple layers, leading to improved feature learning

and segmentation, as shown in Figures 3 and 4.



International Journal of
Medical Toxicology & Forensic Medicine

Figure 3. Residual connections on multiple layers.



International Journal of
Medical Toxicology & Forensic Medicine

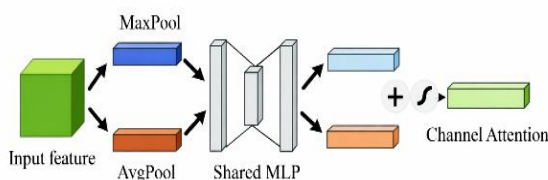
Figure 4. Residual network framework for prediction.

Attention Mechanism

The architecture uses an attention mechanism to dynamically highlight informative regions relevant to tooth segmentation. The attention system dynamically scales the contribution of the various regions in the input images, enabling the model to attend to salient aspects and ignore irrelevant background noise:

$$\text{Attention}(X) = \sigma(W \cdot \text{ReLU}(V \cdot X)) \quad (8)$$

where X represents the input feature maps, W and V are learnable weight matrices, $\text{ReLU}(\cdot)$ denotes the rectified linear unit activation function, and $\sigma(\cdot)$ represents the sigmoid activation function, as shown in Figure 5.



International Journal of
Medical Toxicology & Forensic Medicine

Figure 5. Attention mechanism.

Results

The pseudocode outlines the steps for configuring and training the proposed automated tooth segmentation model. The InitializeModel process sets the adjusted GoogLeNet architecture parameters, such

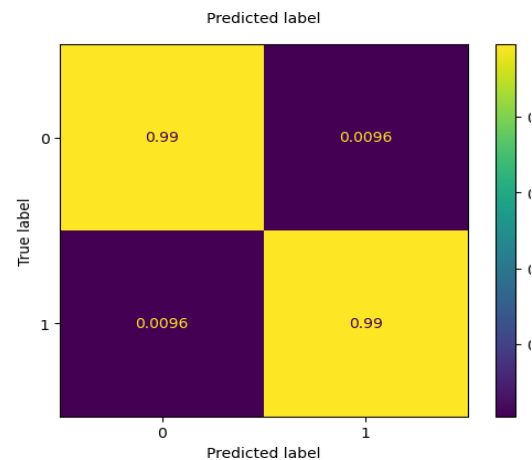
as the learning rate, kernel size, and stride, which are essential for training the model. After this initialization, the TrainModel process coordinates training. It repeats each training-set epoch, updating the model parameters during forward and backward passes. Also, the model is evaluated on the validation set after each epoch to track its progress. In general, the pseudocode offers a systematic method for configuring and training the automated tooth segmentation model, thereby preparing it for further evaluation and analysis.

The proposed model has specific parameters as demonstrated in Table 1. A confusion matrix is a table, shown in Figure 6, commonly used to assess the accuracy of a classification model on a set of test data with known correct values. The first two lines of the context show the predicted and true labels, while the figures below show the confidence scores for the predicted labels. The first row shows that the model predicted label 0 (negative class), whereas the actual label was 1 (positive class). The confidence for the

Table 1. Specific parameters for the proposed model.

Parameters	Accuracy
Sensitivity	98.95%
Specificity	98.51%
F-score	99.23%
Accuracy	99%

International Journal of
Medical Toxicology & Forensic Medicine



International Journal of
Medical Toxicology & Forensic Medicine

Figure 6. Confusion Matrix for the proposed model.
predicted label is 0.0096.

Discussion

The graphic appears to show the training and test accuracy of a machine learning model. Its x-axis represents epochs (i.e., passes through the training set) and ranges from 0 to 55 in steps of 5. The y-axis indicates accuracy, the number of correct predictions

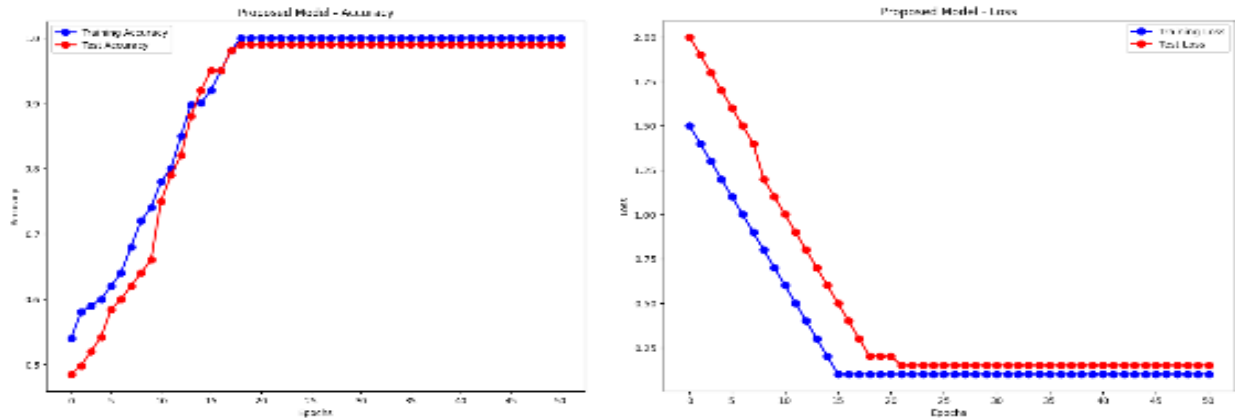


Figure 7. (a) and (b): Shows the Training and Testing Accuracy and Loss for the Proposed Model.

the model made. The scale is 0.5- 1.0 with a step size of 0.1. The graph has two lines: one for training accuracy and the other for test accuracy. The training accuracy line, which will begin at about 0.6 and increase at the same rate to approximately a little more or a little less than 1 after 50 epochs, appears to shake and then drift away in a different direction at a slightly lower or the same time, until it again reaches about 0.95 after the 55th epoch. Starting at approximately 0.5 and rising to approximately 0.9 at 40 epochs, the test accuracy curve then oscillates more erratically before leveling off at approximately 0.85 by 55 epochs (Figure 7a). It is a line graph illustrating the accuracy of a few of the proposed machine learning models versus the number of epoch updates. The epoch count is shown on the x-axis, ranging from 5 to 55 in 5-step increments. The model accuracy (0.5-1.0 with step size = 0.1) is plotted on the y-axis, with the step size on the x-axis. The graph has two line plots: one for training accuracy and the other for test accuracy. For training accuracy, the value in the first epoch is approximately 0.6 and slowly approaches 1.0 as the number of epochs increases to 50. The test accuracy line also starts at 0.5; it increases, but not as much as the training accuracy, which peaks around epoch 40 at approximately 0.9 on both the training and test sets. Since the beginning of the 40th epoch, the test accuracy curve has been slowly decreasing, as shown in Figure 7b.

The figure is a graph plot of the machine learning model training and test accuracy with epoch(s). The x-axis is the number of epochs, i.e., complete traversals through the entire training set. The y-axis shows the accuracy, i.e., the percentage of correct predictions. The graph shows two lines: the training accuracy line and the testing accuracy line. The Training Accuracy line indicates how well the model is learning the training data, while the Testing Accuracy line indicates how well the model generalizes to new, unseen data.

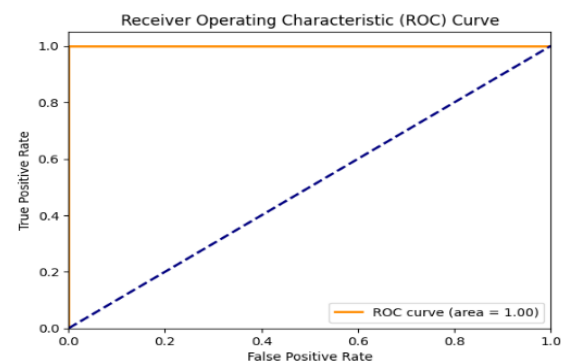


Figure 8. ROC curve for the proposed model.

During the initial stages of training, the training and testing accuracies are low, at approximately 0.6 and 0.5, respectively. However, as the number of epochs increases, the training accuracy increases and approaches 1.0. This implies that the model was adequately trained, as shown in Figure 8. In contrast, the testing accuracy increases more slowly and reaches approximately 0.9 after 40 epochs. Accuracy begins to decline beyond this point, indicating that the model is overfitting to the training data.

Conclusion

In conclusion, this study proposes an automated tooth segmentation framework for 3D dental images based on a hybrid deep learning architecture that integrates GoogLeNet, ResNet, and an attention mechanism. The proposed model effectively captures hierarchical features and highlights informative regions, improving segmentation accuracy and robustness compared with existing approaches. Incorporating residual connections into the modified GoogLeNet architecture further enhances feature learning and overall segmentation performance,

supporting more reliable clinical interpretation and diagnosis. Although the results demonstrate the effectiveness of the proposed method, certain limitations remain, including the dependence on accurately annotated datasets, potential variations in imaging protocols, and computational complexity associated with deep hybrid architectures. Future research will focus on reducing annotation requirements through semi-supervised or weakly supervised learning, incorporating multimodal dental imaging data such as panoramic radiographs and intraoral scans, and applying domain adaptation techniques to improve generalization across diverse clinical environments. Furthermore, large-scale clinical validation and integration with clinical decision support systems will be explored to enhance the practical applicability of the proposed approach in routine dental diagnosis, treatment planning, and dental research.

Ethical Approval

Not applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report there are no competing interests to declare.

References

- [1] 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 Endod*. 2021;47(5):827-35. [DOI: 10.1016/j.joen.2020.12.020]
- [2] 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. [DOI: 10.1016/j.jdent.2021.103865]
- [3] 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.

[DOI:

10.1109/TPAMI.2021.3086072]

- [4] Li Q, Han L, Zhuang Y, Li J, Lin J. Automatic tooth root segmentation of cone beam computed tomography image sequences using U-Net and RNN. *J Xray Sci Technol*. 2020;28(5):905-922. [DOI: 10.3233/XST-200678]
- [5] Hsu K, Yuh D, Lin S, Lyu P, Pan G, Zhuang Y, 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;12:19809. [DOI: 10.1038/s41598-022-23901-7]
- [6] Wen X, Liu Z, Chu Y, Le M, Li L. MRCM-UCTransNet: automatic and accurate 3D tooth segmentation network from cone-beam CT images. *Int J Imaging Syst Technol*. 2024;34:e23139. [DOI: 10.1002/ima.23139]
- [7] Poorghobadi S, Baesi K, Gharibzadeh S, Shirzad R, Khosravy MS, Fazeli M, et al. Autophagy and unfolded protein response induction: a crosstalk between street rabies virus and the host. *Cell Stress Chaperones*. 2023 Jul. [DOI: 10.1007/s12192-023-01335-y]
- [8] Hu F, Chen Z, Wu F. A novel difficult-to-segment samples focusing network for oral CBCT image segmentation. *Sci Rep*. 2024;14:5068. [DOI: 10.1038/s41598-024-55522-7]
- [9] Polizzi A, Quinzi V, Ronsivalle V, Venezia P, Santonocito S, Lo Giudice A, et al. Tooth automatic segmentation from CBCT images: a systematic review. *Clin Oral Investig*. 2023;27(7):3363-78. [DOI: 10.1007/s00784-023-05048-5]
- [10] Zheng Q, Gao Y, Zhou M, Li H, Lin J, Zhang W, et al. Semi- or fully automatic tooth segmentation in CBCT images: a review. *PeerJ Comput Sci*. 2024;10:e1994. [DOI: 10.7717/peerj-cs.1994]
- [11] Chen X, Ma N, Xu T, Xu C. Deep learning-based tooth segmentation methods in medical imaging: a review. *Proc Inst Mech Eng H*. 2024;238(2):115-131. [DOI: 10.1177/09544119231217603]
- [12] Zhou Y, Xu Y, Khalil B, Nalley A, Tarce M. An open deep learning-based framework and model for tooth instance segmentation in dental CBCT. *Clin Oral Investig*. 2025;29. [DOI:

[10.1007/s00784-025-06578-w](https://doi.org/10.1007/s00784-025-06578-w)]

025-00555-0]

[13] Kot WY, Au Yeung SY, Leung Y, Leung P, Yang WF. Evolution of deep learning tooth segmentation from CT/CBCT images: a systematic review and meta-analysis. BMC Oral Health. 2025;25. [DOI: [10.1186/s12903-025-05984-6](https://doi.org/10.1186/s12903-025-05984-6)]

[15] Zakareya S, Izadkhah H, Karimpour J. A new deep-learning-based model for breast cancer diagnosis from medical images. Diagnostics (Basel). 2023;13(11):1944. [DOI: [10.3390/diagnostics13111944](https://doi.org/10.3390/diagnostics13111944)]

[14] Zhou W, Zhao D, Zhao B, Shen M, Liu X. A deep learning framework for automated dental segmentation and diagnostic report generation from cone-beam computed tomography. Head Face Med. 2025;21:80. [DOI: [10.1186/s13005-](https://doi.org/10.1186/s13005-025-00555-0)

[16] Balraj K, Ramteke M, Mittal S, Bhargava R, Rathore AS. MADR-Net: multi-level attention dilated residual neural network for segmentation of medical images. Sci Rep. 2024;14. [DOI: [10.1038/s41598-024-63538-2](https://doi.org/10.1038/s41598-024-63538-2)]