Explainable AI for Panoramic Dental Radiographs Using Contrastive Learning and U-Net Based Segmentation

Aadyoth Sreeram[1], Balamurugan R.[2], Aditya MN.[3]

[1,3]Deperatment of Electronics and Communication Engineering, S.R.M Institute of Science and Technology, Kattankulathur, Chennai, India.

[2]Department of Mechatronics Engineering, S.R.M Institute of Science and Technology, Kattankulathur, Chennai, India.

E-mail: [1]as2235@srmist.edu.in, [2]br9471@srmist.edu.in, [3]an8594@srmist.edu.in

Abstract

Self-supervised segmentation of panoramic dental radiographs often lacks interpretability, leading clinicians to be hesitant in relying on these predictions. The practical application of such segmentation is limited, as the results must be comprehensible for effective diagnosis and treatment planning. To address this issue, we implement Explainable AI (XAI) techniques, such as Grad-CAM, which generate heat maps that highlight significant areas within the radiographs, thereby enhancing the model's transparency. This approach mitigates the black-box nature of self-supervised models and provides a reliable AI-driven solution. We propose a hybrid strategy that integrates attention-based segmentation with both supervised and self-supervised learning. Self-supervised learning is utilized to analyse unlabelled radiographs to extract features, which are subsequently fine-tuned through supervised learning to achieve accurate segmentation. This model aids in the detection of anomalies such as cavities, fractures, and periodontal diseases, thereby improving diagnostic precision. Additionally, it facilitates treatment planning and bridges the gap between clinical trust and AI automation, rendering AI-based dental imaging more practical and acceptable.

Keywords: Contraceptive Learning, U-Net, Grad-CAM, Medical Imaging.

DOI: https://doi.org/10.36548/jscp.2025.2.004
Received: 16.04.2025, received in revised form: 15.05.2025, accepted: 25.05.2025, published: 03.06.2025

1. Introduction

Artificial intelligence has significantly transformed medical imaging, enhancing diagnostic accuracy and efficiency within the healthcare industry. Panoramic radiographs are frequently utilized in dentistry to detect abnormalities such as cavities, fractures, and periodontal disease. However, deep learning models operate as "black boxes," presenting a challenge to clinical utility, as practitioners require interpretability to validate AI-based diagnoses. The lack of transparency impedes trust in automated segmentation, complicating the integration of AI into clinical practice.

To address this issue, we employ Explainable AI (XAI) techniques to improve the interpretability of dental radiograph segmentation. Specifically, we utilize Gradient-weighted Class Activation Mapping (Grad-CAM) to generate visual heatmaps that highlight key areas influencing model predictions. This approach bridges the gap between clinical trust and AI automation, ensuring that segmentation outputs are both precise and interpretable. Additionally, we adopt a hybrid methodology that incorporates contrastive learning and segmentation through attention mechanisms. Self-supervised learning is used to develop robust features, which are subsequently refined through supervised learning for accurate segmentation.

We compare models such as U-Net and attention-based architectures to assess their performance in segmenting dental structures. Our research aims to enhance the reliability of AI-assisted diagnoses and promote the adoption of interpretable deep learning models in dentistry.

2. Related Work

Artificial intelligence has also had a tremendous influence on clinical practice and dental research, improving diagnostics, image analysis, and treatment planning. [1] discusses AI-related dental research from bibliometric and altimetric standpoints. [2] synthesizes the last decade of the use of explainable AI in the healthcare field. [3] compares deep learning models for dental disease prediction based on X-ray imaging. [4] introduces IDD-Net, a deep learning model for the early detection of dental disease through X-ray imaging. [5] introduces an interpretable deep learning architecture for mandibular canal segmentation of CBCT volumes. [6] discusses machine learning applications for dental image analysis. [7] proposes SimCLR, a contrastive learning framework for self-supervised visual representation learning. [8]

formulates an AI-driven system for automatic dental cavity detection using explainable AI. [9] critically reviews AI-based dental image analysis methods and algorithms. [10] performs a scoping review of machine learning in dentistry. [11] employs machine learning for the automatic detection of dental caries and calculus through hyperspectral fluorescence imaging. [12] surveys developments in machine learning for dental, oral, and craniofacial imaging. [13] reports on advances in dental imaging technology and in clinical applications. [14] employs DeepLabv3+ to segment apical lesions in panoramic radiography. [15] systematically tracks a decade's worth of research on deep learning in dentistry. [16] presents basic AI concepts and dental imaging applications such as diagnostics and forensics.[17] examines CNN-based tooth segmentation and numbering of panoramic dental X-rays. The initial CAM method was proposed in [18], where the network structure needed to be modified by replacing fully connected layers with convolutional layers and global average pooling. Although revolutionary, this architectural limitation restricted its general applicability. Grad-CAM, introduced in [19], overcame this restriction by employing gradient information entering the terminal convolutional layer. Deep learning achieved significant leaps: FCNs [20] eliminated hand-designed feature extraction but suffered from loss of detail. U-Net [21] enhanced organ and tumor segmentation with skip connections.

3. Proposed Work

3.1 Contrastive Learning for Enhanced Tooth Segmentation

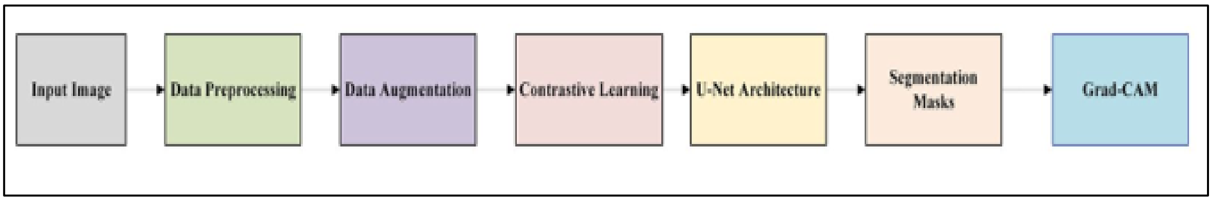

Figure 1. Flow Map

As per Figure 1, the workflow offers a cutting-edge tooth segmentation algorithm for dental panoramic radiographs utilizing contrastive learning as pretraining. The proposed two-stage model includes self-supervised pretraining and supervised fine-tuning. Contrastive learning trains an encoder network with augmentation, producing views of the same image. The NT-Xent loss function (Equation 4) is designed to push embedded views toward each other while pushing away from all other images. This minimizes reliance on small annotation sets and maximizes robustness to image alterations. Domain-specific augmentations like synthetic exposure variations are added to the pipeline. Visualization of the embedding space monitors pretraining performance to guarantee that dental anatomy clusters appropriately.

3.2 U-Net Architecture

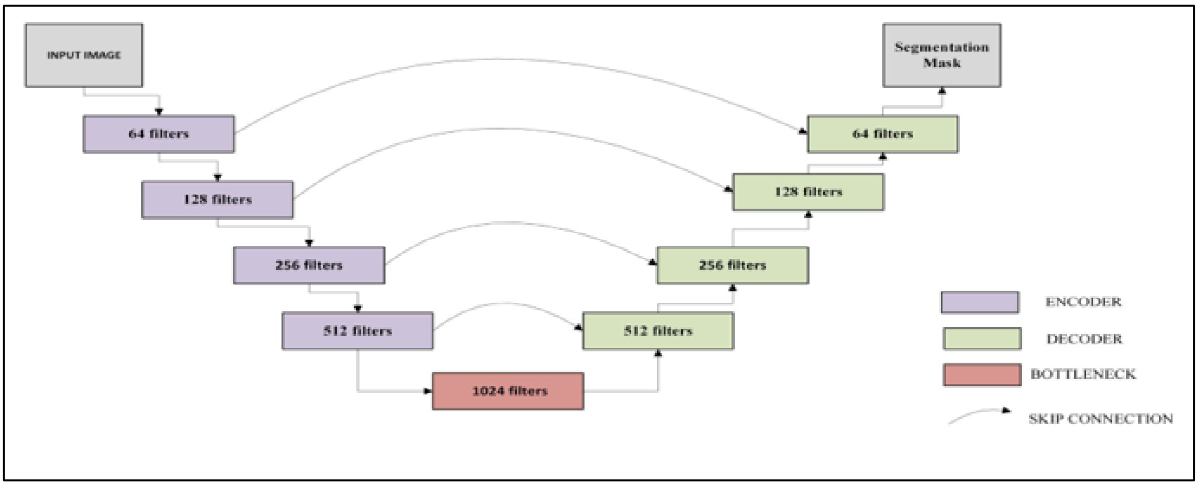

Figure 2. U-Net Architecture with Skip Connections

After pretraining, a U-Net-like architecture as portrayed in Figure 2 is employed, with the pretrained encoder as the contracting path combined with a decoder network. Skip connections maintain spatial information to accurately delineate teeth. The model is trained with the Dice loss function (Equation 1) to address class imbalance through pixel-wise overlap weighting. The encoder consists of feature dimension augmentation, creating a hierarchical dental structure representation, and the decoder iteratively reconstructs spatial resolution through the respective encoder features.

$$\text{Dice Loss} = 1 - \frac{2 \cdot \sum_i y_{\text{true},i} \cdot y_{\text{pred},i} + \epsilon}{\sum_i y_{\text{true},i} + \sum_i y_{\text{pred},i} + \epsilon} \quad (1)$$

$$\text{Dice} = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} \quad (2)$$

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \quad (3)$$

Performance is evaluated using the Dice coefficient (Equation 2) and IoU (Equation 3), preserving high overlap between regions and boundary accuracy.

3.3 Grad-CAM Integration for Model Interpretability

An important addition is Gradient-weighted Class Activation Mapping (Grad-CAM) for interpretability. The computation (Equation 5) produces localization maps based on gradient flows from the last convolutional layer.

$$\mathcal{L}i, j = -\log \quad \frac{\exp\left(\text{sim}\left(z_i, z_j\right)/\tau\right)}{\sum k{=}1^{2N} 1_{[k \neq i]} \exp\left(\text{sim}\left(z_i, z_k\right)/\tau\right)} c \tag{4}$$

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right) \tag{5}$$

In dental segmentation, this facilitates the visualization of tooth boundary detection influences. A dedicated overlay technique uses heatmaps on raw radiographs to emphasize network attention on anatomical features. A dedicated colormap increases contrast against grayscale backgrounds. Comparative evaluation of Grad-CAM outputs identifies errors or biases, such as demographic-based variations. For clinical application, a real-time visualization interface offers immediate feedback during dental treatment, making the model a transparent, explainable AI system.

4. Results and Discussion

The integration of Explainable AI, particularly Grad-CAM, significantly enhances the interpretability of the model by highlighting regions within dental panoramic radiographs that influence AI decision-making. This section presents both qualitative and quantitative assessments of model performance across various visualization techniques and comparative benchmarks.

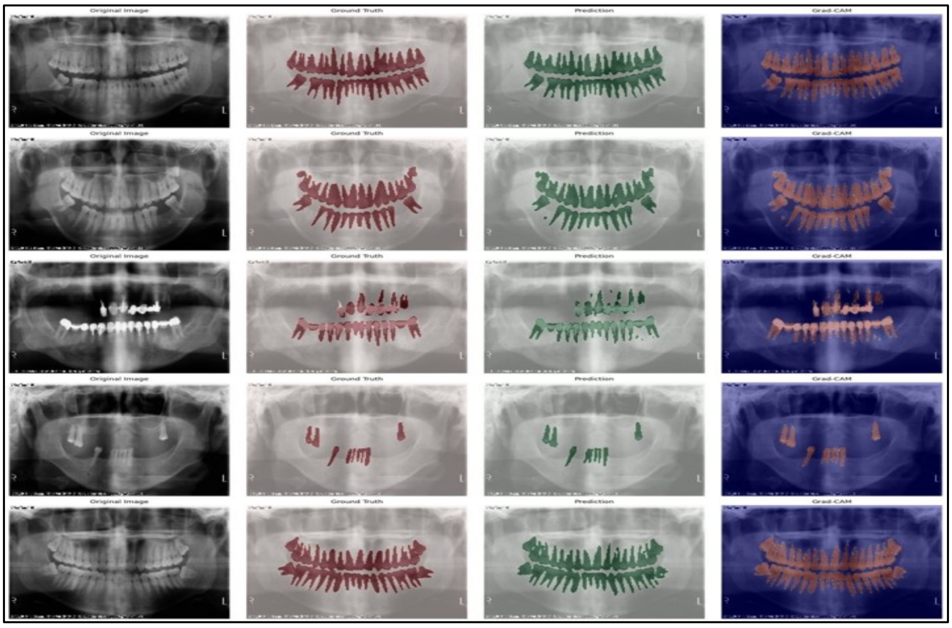

Figure 3. Comparative Visualization of Dental Panoramic Radiographs

4.1 Interpretability via Grad-CAM

Grad-CAM was employed to visualize the regions of attention during model inference. As illustrated in Figure 3, each row represents a dental X-ray evaluated using four distinct methods:

- **Original Image**: The first column presents the raw grayscale panoramic radiographs, depicting diverse dental anatomies.
- **Ground Truth**: The second column shows the manually segmented teeth (highlighted in red) annotated by expert dentists, serving as the reference for accuracy.
- **Model Prediction**: The third column highlights the AI model's automated segmentation outputs in green, illustrating its capability to delineate dental structures.
- **Grad-CAM Visualization**: The fourth column provides attention heatmaps using Grad-CAM, where red/orange regions indicate high relevance to the model's predictions, while blue areas have minimal impact.

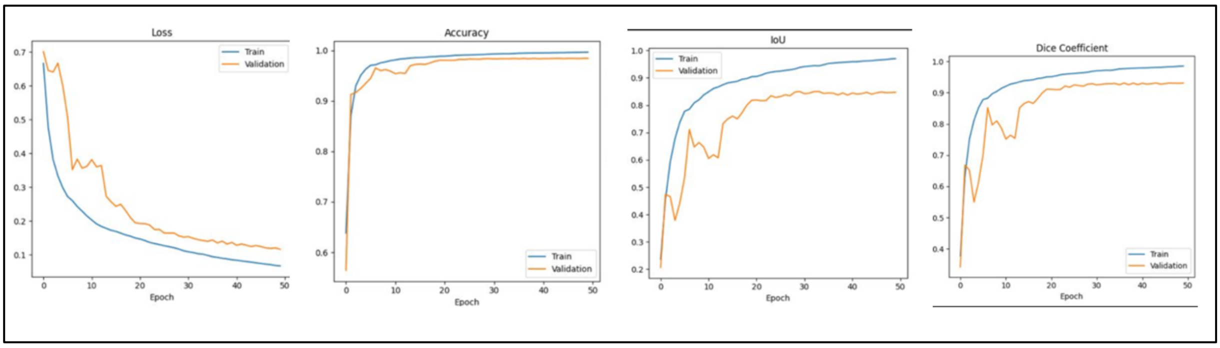

Figure 4. Model Performance Analysis

Figure 4 presents training and validation metrics over 50 epochs.

4.2 Performance Across Training Epochs

- **Loss Curve**: A steady decrease in both training and validation loss suggests effective learning without signs of overfitting.
- **Accuracy**: Training accuracy quickly converges toward 100%, while validation accuracy stabilizes around 97–98%, reflecting strong generalization.
- **Intersection over Union (IoU)**: The model achieves 98% IoU on training data and 85% on validation, indicating slightly reduced boundary precision on unseen samples.
- **Dice Coefficient**: This metric mirrors the IoU trend, with a high Dice score affirming precise segmentation despite some variability early in training (epochs 0–15).

Table 1. SOTA Analysis (All Values are in %)

	IOU	Dice Coefficient	Accuracy	Recall	References
U-NET	73.9	84.2	95.1	79.4	[22]
Deeplab V3+	54.6	-	-	24.8	[16]
Attention U-NET	79.5	88.5	95.6	85.7	[22]
Seg Net	77.1	87.1	95.1	83.3	[22]
Ours	85.5	92.7	98.2	92.8	-

As observed in Table 1 the proposed model demonstrates a significant performance in comparison to the existing architectures.

5. Conclusion

The integration of explainable artificial intelligence (XAI) into panoramic dental radiograph segmentation enhances the interpretability and clinical utility of deep learning models. The application of gradient-weighted class activation mapping (Grad-CAM) to generate visual heatmaps improves diagnostic transparency and fosters clinician trust in the segmentation results. The combination of contrastive learning with an attention-based segmentation module enhances feature representation and segmentation accuracy, resulting in superior performance compared to conventional methods. However, potential biases in the training dataset and dependence on pre-annotated data may impact generalizability. Future research will focus on enhancing model robustness and generalizability by incorporating more diverse datasets and mitigating annotation bias. Additionally, real-time visualization will be integrated into clinical practice, and the framework will be expanded to encompass other dental imaging modalities.

References

[1] Lu, Wei, Xueqian Yu, Yueyang Li, Yi Cao, Yanning Chen, and Fang Hua. "Artificial intelligence–related dental research: bibliometric and altmetric analysis." International Dental Journal 75, no. 1 (2025): 166-175.

[2] Loh, Hui Wen, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, and U. Rajendra Acharya. "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)." Computer methods and programs in biomedicine 226 (2022): 107161.

[3] Hasnain, Muhammad Adnan, Hassaan Malik, Muhammad Mujtaba Asad, and Fahad Sherwani. "Deep learning architectures in dental diagnostics: a systematic comparison of techniques for accurate prediction of dental disease through x-ray imaging." International Journal of Intelligent Computing and Cybernetics 17, no. 1 (2024): 161-180.

[4] Walasai, Gordhan Das, Stefano Landini, Muhammad Nawaz Chandio, Abdul Azeem Kumbhar, Shoaib Ali Bhatti, and Sheeraz Ahmed. "Reducing carbon emissions and costs of electricity with solar PV systems in QUEST Nawabshah, Pakistan Administration Building." Journal of Computing & Biomedical Informatics 7, no. 02 (2024).

[5] Barzas, Konstantinos, Shereen Fouad, Gainer Jasa, and Gabriel Landini. "An explainable deep learning framework for mandibular canal segmentation from cone beam computed tomography volumes." In International Conference on Computational Advances in Bio and Medical Sciences, pp. 1-13. Cham: Springer Nature Switzerland, 2023.

[6] Yu, Young-Jun. "Machine learning for dental image analysis." arXiv preprint arXiv:1611.09958 (2016).

[7] Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. "A simple framework for contrastive learning of visual representations." In International conference on machine learning, PmLR, 2020, 1597-1607.

[8] Bhattacharjee, Niharika. "Automated dental cavity detection system using deep learning and explainable AI." AMIA Summits on Translational Science Proceedings 2022 (2022): 140.

[9] Bonny, Talal, Wafaa Al Nassan, Khaled Obaideen, Tamer Rabie, Maryam Nooman AlMallahi, and Swati Gupta. "Primary Methods and Algorithms in Artificial-Intelligence-Based Dental Image Analysis: A Systematic Review." Algorithms 17, no. 12 (2024): 567.

[10] Ding, Hao, Jiamin Wu, Wuyuan Zhao, Jukka P. Matinlinna, Michael F. Burrow, and James KH Tsoi. "Artificial intelligence in dentistry—A review." Frontiers in Dental Medicine 4 (2023): 1085251.

[11] Arsiwala-Scheppach, Lubaina T., Akhilanand Chaurasia, Anne Mueller, Joachim Krois, and Falk Schwendicke. "Machine learning in dentistry: a scoping review." Journal of Clinical Medicine 12, no. 3 (2023): 937.

[12] Talpur, Sarena, Fahad Azim, Munaf Rashid, Sidra Abid Syed, Baby Alisha Talpur, and Saad Jawaid Khan. "Uses of different machine learning algorithms for diagnosis of dental caries." Journal of Healthcare Engineering 2022, no. 1 (2022): 5032435.

[13] Wang, Cheng, Rongjun Zhang, Xiaoling Wei, Le Wang, Wenqian Xu, and Qi Yao. "Machine learning-based automatic identification and diagnosis of dental caries and calculus using hyperspectral fluorescence imaging." Photodiagnosis and Photodynamic Therapy 41 (2023): 103217.

[14] Ren, Ruiyang, Haozhe Luo, Chongying Su, Yang Yao, and Wen Liao. "Machine learning in dental, oral and craniofacial imaging: a review of recent progress." PeerJ 9 (2021): e11451.

[15] Vandenberghe, Bart, Reinhilde Jacobs, and Hilde Bosmans. "Modern dental imaging: a review of the current technology and clinical applications in dental practice." European radiology 20 (2010): 2637-2655.

[16] Ketenci Çay, Fatmanur, Çağrı Yeşil, Oktay Çay, Büşra Gül Yılmaz, Fatma Hasene Özçini, and Dilhan İlgüy. "DeepLabv3+ method for detecting and segmenting apical lesions on panoramic radiography." Clinical Oral Investigations 29, no. 2 (2025): 101.

[17] Pauwels, Ruben. "A brief introduction to concepts and applications of artificial intelligence in dental imaging." Oral radiology 37, no. 1 (2021): 153-160.

[18] Zhou, Bolei, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. "Learning deep features for discriminative localization." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 2921-2929.

[19] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: visual explanations from deep networks via gradient-based localization." International journal of computer vision 128 (2020): 336-359.

[20] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, 3431-3440.

[21] Zhong, Wenjin, XiaoXiao Ren, and HanWen Zhang. "Automatic X-ray teeth segmentation with grouped attention." Scientific Reports 15, no. 1 (2025): 64.