Keywords
Deep learning; Explainable artificial intelligence (XAI); Grad-CAM; Instance segmentation; Panoramic dental radiographs.

Authors
Dennis Dennis[1*],
* Department of Conservative Dentistry, Faculty of Dentistry, Universitas Sumatera Utara, Medan, Indonesia. Email: dennis@usu.ac.id, https://orcid.org/0000-0003-1718-4363

Siriwan Suebnukarn[2]
[2]Faculty of Dentistry, Thammasat University, Pathum Thani, Thailand. ssiriwan@tu.ac.th
https://orcid.org/0000-0003-1237-1274

*Correspondence to:
Dennis Dennis
Department of Conservative Dentistry, Faculty of Dentistry, Universitas Sumatera
Utara, Jl, Alumni No.2, Kampus USU, Medan 20155, Sumatera Utara, Indonesia
https://orcid.org/0000-0003-1718-4363
E-mail: dennis@usu.ac.id

Received: 11.12.2025
Accepted: 15.02.2026
Doi: 10.1922/ejprd.v34i2.1321

Deep Learning-Based Dental Image Analysis Using Grad Cam Convolutional Network

Abstract
Panoramic radiographs have been regularly applied in the field of prosthodontic and restorative dentistry to aid in treatment planning, tooth morphology, edentulous space evaluation, and preliminary screening of the implant site. Nevertheless, panoramic images may not be easily interpreted because of anatomy overlap, distortion, and varying quality of images. This paper provides a deep learning architecture that can be explained and used to aid the radiographic evaluation of prosthodontics by means of automated localization of teeth and interpretable visualisation. It applied a multi-stage pipeline, which involved dataset validation, COCO-based annotation auditing, instance segmentation with a Mask R-CNN backbone based on ResNet-FPN, and incorporation of explainable artificial intelligence (XAI) methods. Grad-CAM, occlusion sensitivity mapping, and mask confidence visualisation were automatically used to segment and analyse tooth regions to get transparent decision-support outputs. Qualitative data showed that there was anatomical localization and activation of teeth in patterns that were consistent with morphologically relevant structures. Despite the fact that quantitative metrics of segmentation were affected by rigid confidence levels, explainability analysis showed that model attention was mainly focused on tooth anatomy instead of background artefact. The suggested framework offers a reproducible and interpretable base of AI-assisted panoramic radiograph interpretation, which has possible applications in the field of prosthodontic planning, restorative assessment, and implant-oriented screening procedures.

1.Introduction

Panoramic dental radiography has become a common tool in the field of prosthodontics and restorative dentistry, to assess the condition of the dentition, alveolar bone support, tooth morphology, edentulous spaces, and the condition of the maxillofacial region before the application of the restorative interventions. It is also a vital part of treatment planning of fixed prostheses, removable partial dentures, full-mouth rehabilitation, and implant-supported restorations. Nevertheless, panoramic radiographs are difficult to interpret because of structural overlap, distortion artifacts, differences in contrast, and changes in patient positioning. Such factors can decrease diagnostic consistency especially when several teeth and anatomical landmarks should be evaluated at the same time. Therefore, automated image analysis systems are also becoming a more studied decision-support tool to make workflows in restorative activities more efficient and reliable.

Deep learning, especially convolutional neural networks (CNNs) have shown great potential in dental image detection as they make it possible to automatically detect, segment and classify anatomical structures and pathological conditions. CNN-based has been effectively used in detecting dental caries and interpreting panoramic images, and Grad-CAM visualization has been used to facilitate clear clinical validation [1]. Deep learning has also emerged as a powerful approach to dental and maxillofacial imaging studies, with positive outcomes in segmentation, detection, and classification tasks [2]. These advancements demonstrate the practicability of using deep learning to panoramic radiographs to localize and analyze the structure of the teeth automatically. Although such deep learning models have high predictive performance, most of them are used as black-box systems, producing results without the radiographic features that determine decisions. This Uninterpretability has restricted clinical acceptability, especially in the area of prosthodontics and restorative dentistry, where the treatment planning involves

traceable and clinically justifiable reasoning. Deep learning models that can be interpreted to classify dental diseases into multiple categories have demonstrated that explainability modules enhance reliability by drawing attention to clinically relevant parts of the image [3]. On the same note, segmentation-based methods show that the isolation of significant anatomical structures improves diagnostic attention in complicated dental radiographs [4]. Deep learning pipelines (Multi-stage deep learning) have also been used to improve the performance of diagnostic screening [5], and explainable models used on panoramic radiographs ensure that decision-making processes can be more transparent through the use of visualization techniques like Grad-CAM [6]. Grad-CAM in conjunction with transfer learning has also enhanced interpretability in caries detection tasks [7].

Explainable models have also been used to identify anatomy structure in panorama radiographs such as classifying the complex canal morphologies, which further supports the significance of interpretable outputs in dental decision-support systems [8]. Automated tooth localization is of special interest in the context of prosthodontic and restorative applications because tooth-level segmentation can be used to analyse the quality of crown preparation, the presence of missing teeth, and the screening of implant sites. Explainability methods also allow validation showing that predictions are based on meaningful morphology of the tooth and not imaging artifacts [9]. Despite the significant advances in dental imaging by deep learning, panoramic radiographs are technically difficult, as they are distorted, structurally overlapping, and have unpredictable quality. In addition, most systems do not have interpretable justification that limits clinical trust. Thus, the necessity to design an automated framework that will correctly localize the teeth and will also be able to explain the way the decision-making process works in a transparent manner exists [10].

The main aim of the research is to create and test an interpretable deep learning model of panoramic radiograph analysis to aid in the making of prosthodontic and restorative decisions based on a transparent tooth-level interpretation.

Specific Objectives:

- To design a deep learning pipeline integrating instance segmentation and explainable AI for tooth-level analysis in panoramic radiographs.
- To preprocess and validate the dataset to ensure annotation accuracy and data reliability.
- To develop a Mask R-CNN–based instance segmentation model for automated tooth detection and segmentation.
- To extract tooth-level regions of interest (ROIs) from segmentation outputs for localized structural assessment.
- To integrate explainability techniques, including Grad-CAM heatmap visualization, occlusion sensitivity mapping, and mask confidence overlays, for interpretation of model predictions.

2. Literature Review

Deep learning has made a major breakthrough in the field of dental radiographic analysis especially with regard to automated detection and classification of dental caries, periodontal disease, and other oral abnormalities. It has been made possible by the growing access to digital imaging data and the creation of convolutional neural networks (CNNs) to develop computer-aided diagnostic systems that enhance efficiency and minimize inter-observer variability. Although there has been an increase in predictive performance, lack of interpretability is a significant obstacle to clinical use. In turn, recent studies have focused on explainable artificial intelligence (XAI) approaches, in particular, gradient-based visualization methods, including Gradient-weighted Class Activation Mapping (Grad-CAM) that can maximize transparency and trust in clinicians.

One of the most important developments in AI in the field of dentistry is the combination of transfer learning and interpretability frameworks. Transfer learning enables CNN models that have been trained on large datasets to be adapted successfully to dental radiographs even in the situation when labeled data are scarce. A caries detection system that uses transfer learning with Grad-CAM showed that trained models have the potential to produce robust radiographic features and attention heatmap can give clinically useful lesion localization [11]. This type of visualization enhances the validation of the diagnosis and promotes the clinical applicability. The interest in panoramic radiographs has been particularly high because of its overall presentation of dentition and maxillofacial structures. Nevertheless, panoramic imaging presents the issues of structural overlap, distortion and noise. A caries prediction deep learning model was designed using panoramic radiographs, which was interpretable and focused on the use of visualization tools to verify predictions in intricate anatomical settings [12]. These results indicate that explainable models are of particular significance in panoramic imaging, where the wide-range of anatomy makes interpretation of regions difficult.

Multi-condition recognition systems have also been extended to deep learning. The generalized learning of radiographic features across several diseases was shown as possible by a CNN-based framework that was able to identify periodontitis and dental caries [13]. Although multi-class systems allow making the systems more relevant to the clinical context, they lead to higher classification complexity and misinterpretation, which supports the importance of XAI tools to make sure that clinically significant patterns are considered. Explainable deep learning has also been used in structural and demographic analysis. Gender-characterization model based on Grad-CAM visualisation identified discriminative anatomical features learnt by the network [14]. This does not focus on radiographs, but this is indicative of the wider applicability of gradient-based interpretability to dental imaging. Multi-input CNN frameworks are innovative architectures that have enhanced the detection

performance because of incorporating multiple feature representations [15]. Nevertheless, there is a tendency of higher interpretability to be less as architectural complexity increases, which points to the necessity to balance predictive power and transparency.

Explainable frameworks that are segmentation-based have drawn interest in offering pixel-level localization. U-Net segmentation and contrastive learning approach showed better feature discrimination and localization of tooth structure in panoramic radiographs [16]. Segmentation-based approaches make the results more interpretable by emphasizing anatomical boundaries as opposed to simply using global classification results. Ensemble learning techniques have also enhanced diagnostic strength but also make predictions more difficult to interpret since they are made by many models [17]. Systematic reviews verify that CNN-based methods prevail in dental imaging studies and tend to be more effective than conventional machine learning methods [18]. Nonetheless, these reviews note that there are consistent limitations, such as inconsistent evaluation protocols, lack of dataset transparency, and lack of explainability integration. The deep learning applications have also been extended to panoramic radiograph tasks like dental age estimation where the rich morphological information of panoramic images is evident [19].

Altogether, despite significant achievements in the transfer learning field, the multi-disease classification, segmentation-based models, and the ensemble frameworks, the lack of interpretability is one of the key issues. It is evident that explainable systems that combine segmentation, classification, and visualization methods including Grad-CAM and occlusion sensitivity mapping are necessary. These insights form the basis of the proposed study

3. Methodology

3.1 Research Design

The research design adopted in this study was a quantitative and experimental one to create and test a deep learning-based system of automated dental images analysis using explainable artificial intelligence (XAI). The general paradigm of the methodology was a supervised learning model, according to which annotated panoramic dental radiographs were employed to train convolutional neural network (CNN) models. The workflow was built as a multi-step computational pipeline, that is, data preparation, instance segmentation, tooth-region extraction, classification modeling, and interpretability analysis with the help of Grad-CAM and other explainability tools. The primary aim of using this research design was to guarantee high diagnostic performance and model transparency. Compared to the conventional deep learning workflow, which emphasises prediction accuracy, the proposed framework has modules of explainability to emphasise the areas of panoramic X-rays that affect the decision-making of the model. This research design consequently guarantees clinical interpretability, which is a crucial requirement of medical imaging use and publication criteria of high-impact (Q1) journals. The experiment process was broken down into five key sequential steps: (1) auditing and preprocessing of a dataset, (2) training a model of instance segmentation, (3) obtaining tooth-level cropped images, (4) training a CNN classification model, and (5) explainability analysis by visualising mask confidence, the sensitivity of occlusions, and Grad-CAM heatmap. The implementation of each stage was done in a programmable manner to allow reproducibility and traceability of results.

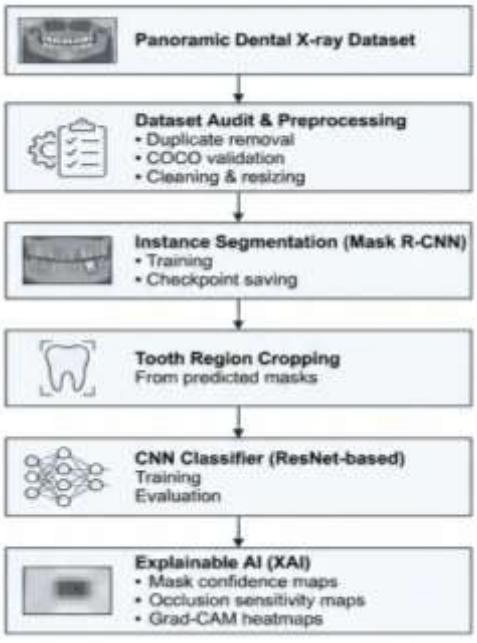

Figure 1. Overall Research Design Flowchart

This flowchart above in figure 1 represents the structured research design applied in this work. The design guarantees methodological rigour in that it isolates segmentation and classification tasks and incorporates explainability methods on various levels.

3.2 Data Collection Methods

3.2.1 Dataset Source

The data of this research was a publicly accessible panoramic dental X-ray dataset of high-resolution panoramic radiographs [20]. The data was also provided with annotated files in COCO (Common Objects in Context) format. The COCO format has been chosen due to its popularity in object detection and segmentation studies and offering standardised format of image metadata, category names, polygon coordinates, and bounding box data [21]. The panoramic radiographs are real clinical imaging situations, such as differences in illumination, tooth position, noise, and complexity of anatomy. The dataset is thus a good reference point when assessing the performance of deep learning models in dental image analysis tasks.

3.2.2 Data Structure

The data was divided into three sets, namely training, validation, and testing. Such a predetermined separation was necessary to make sure that model training and evaluation adhered to the typical machine learning experimental procedures [22]. Model learning was done on the training set, hyperparameter monitoring and checkpoint selection was done on the validation set, and final evaluation was done on the test set. Both subsets included two main data elements, namely, panoramic dental radiographs in image format (e.g., JPG/PNG) and annotation files in JSON format. The annotation files contained critical parameters of COCO like image identifiers, class labels, segmentation polygons and bounding box coordinates. This design allowed direct compatibility with Torchvision and PyTorch-based training pipelines.

3.2.3 Data Preprocessing and Validation

An extensive preprocessing pipeline was adopted in order to provide data integrity and reproducibility. The preprocessing stage entailed the detection of duplicate images, COCO annotation validation, standardisation of file names, and resolution of inconsistencies. Duplicate detection was used to prevent bias in training and did not allow the same or similar images to be found in different split of the dataset. This measure minimised the possibility of data leaking, which can artificially increase model accuracy. Also, validation of annotation was conducted to verify accuracy of polygon segmentation masks and bounding box values. Bounding boxes were found to be invalid, segmentation polygons were absent and annotations were corrected. Resizing and normalisation of images were also carried out to ensure that the input data is consistent to be trained on the deep learning. Cleaned COCO annotation files were produced after the preprocessing and stored so that they are compatible with the segmentation model training pipeline.

Lastly, preview overlays were created through the overlaying of masks and bounding boxes on sample pictures. This enabled visual examination of correctness of annotations and guaranteed that the training dataset was prepared. The dataset is moderate but can be compared to various studies investigating the use of deep learning in exploratory research in the field of dental radiographs analysis where quality annotation is valued over quantity. Polygon annotation of panoramic radiographs is a tedious task that requires manual annotation of panoramic radiographs, and thus the current size of the dataset is realistic in terms of clinical annotation of dental imaging studies..

3.3 Population and Sampling

3.3.1 Target Population

This research had a target population of panoramic dental radiographs of the structures of adult dentition. The images find extensive application in dental diagnostics to determine the location of teeth, the structure of the jaw, periodontal disorders, and other anatomical disorders. The dataset represents the real radiographic variability that is usually present in dental clinical settings.

3.3.2 Sampling Strategy

The implicitly used stratified sampling method was the predefined division into training, validation, and testing sets of the dataset. The stratification was done to make sure that the subsets had representative samples of the annotated dental structures. This strategy enhances better generalisation because the model is not trained on a biassed sample of images. In order to enhance the sampling validity, cross-split leakage analysis and duplicate detection were conducted. Duplicates were detected and the data set was checked so that the training set did not have the same images as the validation or the test set. This made sure that the model evaluation was carried out on hidden samples giving a realistic measure of generalisation performance.

3.3.3 Inclusion and Exclusion Criteria

The inclusion criteria of the study were that the images had to be high-resolution panoramic radiographs, and the annotation files of the COCO format had to include complete polygon and bounding box data. The missing metadata of images, corrupted files, invalid annotations, or duplicates were not analysed. This made the training and evaluation processes reliable.

3.4 Data Analysis Techniques

The analysis of data in this paper involved three calculational phases, which were instance segmentation, classification, and explainability analysis. A Mask R-CNN model was used to detect teeth at the tooth level with Torchvision. The choice of mask R-CNN was based on the fact that it can be used to detect objects and perform pixel-wise segmentation simultaneously, which allows extracting individual tooth structures in panoramic radiographs with high precision. In the architecture, a ResNet backbone was used with a Feature Pyramid Network (FPN) to note both the low level texture features and as well as the high level semantic features. The model was trained through supervised learning through ground-truth masks, and stochastic gradient descent was used to optimize the model. To be able to reproduce and conduct further explainability analysis, model checkpoints were stored at specified epochs. The generated trained network produced a prediction of masks, bounding boxes, and confidence scores of the detected tooth instances.

After the segmentation, individual tooth regions were extracted in panoramic images by post processing the predicted masks. These tooth images were cropped and stored separately and the labels in a structured CSV file. The step allowed local structural analysis and training a secondary convolutional neural network classifier. The extracted tooth crops were trained on a ResNet-based CNN classifier through supervised classification. The performance of the models was measured based on accuracy, precision, recall, and F1-score, and confusion matrices were obtained to study the performance of the models in classes and to determine the patterns of misclassification. The stage of classification offered quantitative performance measurements and feature representations that can be analyzed with the help of interpretability.

Explainability was one of the key elements of the methodology. The first generation of mask confidence overlays was produced by visualizing the scores of segmentation probability to produce high-confidence areas. Subsequently, occlusion sensitivity analysis was performed, which involved masking patches of the image in a systematic way and quantifying variation in confidence of the prediction yielding heatmaps that retrieved spatially significant regions. Lastly, the CNN classifier was used with Grad-CAM to derive gradient-based activation maps on the last convolutional layers. These class-discriminative heatmaps were overlaid on tooth crop images to see areas that make the most contribution to classification decisions. Grad-CAM results were divided into correct and incorrect predictions so that the analysis of errors can be structured and to determine whether the attention of the model was focused on clinically relevant anatomical structures.

3.5 Ethical Considerations

This research was conducted in an ethical manner. The data used was publicly available and anonymized, i.e. there was no patient-identifying data (names, clinical record, demographic identifiers, etc.) provided. Consequently, the study adhered to the normal privacy and confidentiality requirements on medical imaging research. Moreover, this study is supposed to be purely scholarly and experimental. The suggested model is not to substitute professional clinical diagnosis but to assist and complement the decision-making process based on automated analysis and interpretable results. This restriction is also significant to avoid the abuse of AI-based predictions in the actual clinical setting without professional confirmation. To reduce bias and increase fairness, duplicate detection and train-test separation tests were conducted to decrease the possibility of data leakage and inflated performance reporting. Furthermore, explainability methods, including Grad-CAM, occlusion mapping, and confidence overlay, are used to guarantee the transparency and responsibility of the AI system. Lastly, reproducibility was also considered an ethical scientific duty. All the preprocessing logs, training checkpoints, evaluation reports, and explainability outputs were stored in a systematic manner, which means that other researchers can replicate and validate the experimental workflow.

4. Results

4.1 Dataset Preparation and Preprocessing Outcomes

Preparation of the dataset was done to guarantee that panoramic dental radiographs were clean, consistent and devoid of cross-split leakage prior to training the deep learning models. The COCO annotation files were checked in terms of structural integrity, keys absence, annotation mismatch, and duplicate data contamination. The audit indicated that all the necessary COCO keys (images, annotations and categories) were included, and no missing image files were found in the training, validation and test sets. In particular, the dataset was composed of 42 training, 12 validation and 6 test images, having 1255, 378, and 178 annotations respectively. Polygon segmentation annotations and bounding boxes were also checked against correctness and no invalid polygon structure or bounding box errors were detected in any split, which confirms the high annotation reliability. Duplicate leakage between train and validation and test partitions was checked with SHA hashing and no duplicate groups were detected across the three splits which is methodologically rigorous and will not cause inflated performance of the model because of repeated samples. Moreover, aspect-ratio-preserving resizing (letterbox) of standardized image resolution to a fixed size of 1024 × 512 pixels, to provide the same input representation. There were no missing outputs, and all the images and annotations were transformed successfully without any writing failures, which proved the consistency of preprocessing execution. This measure enhanced reproducibility and made sure that the Mask R-CNN segmentation model was provided with the same input geometry.

4.2 Tooth Category Distribution and Dataset Balance

The most important dataset feature was the distribution of the classes of the tooth categories. Eight clinically significant tooth classes were included in the dataset, and one category, named dental (category id=0), was not used and was not trained. The last categories of teeth were: canine, central incisor, first molar, first premolar, lateral incisor, second molar, second premolar and third molar. The number of classes across the training split was between 123 instances (third molar) and 171 instances (canine) with a comparatively equal distribution but with slight underrepresentation of third molars. The same pattern was found in validation and test split where third molars had the lowest frequency at all times. This imbalance is important because it may contribute to reduced classification confidence for rare categories, especially in later-stage CNN classification.

Table 1. Tooth **Class Distribution Across Dataset Splits**

Tooth Class	Train	Validation	Test
Canine	171	49	23
Central incisor	170	48	24

First molar	149	48	24
First premolar	167	48	23
Lateral incisor	166	51	24
Second molar	154	49	25
Second premolar	155	48	19
Third molar	123	37	16

Table 1 shows that tooth classes are generally balanced across the train, validation, and test splits, but third molars remain consistently underrepresented compared to other categories.

4.3 Mask R-CNN Instance Segmentation Performance

The instance segmentation model of the Mask R-CNN was trained on a supervised approach to learning based on Torchvision code and the ResNet50-FPN backbone. The training was successfully performed under the acceleration of the graphics processor, which proves the possibility to use deep learning-based segmentation to analyse panoramic radiographs. Qualitative assessment of the model results indicated that the model could localize tooth structures and produce relevant bounding boxes and segmentation masks with anatomically significant spatial relationship. Quantitative analysis was done with a detection confidence score threshold of 0.05 to keep low-confidence predictions and a mask probability threshold of 0.5 to binarize soft mask predictions to compute and visualise overlaps. In this assessment setup, the average IoU and Dice scores were near to zero in each of the tooth categories, which means that the mask boundary accuracy of the current training configuration is low.

This finding indicates that the complexity of panoramic radiograph such as overlapping dental anatomy and indistinct structural borders have a significant impact on segmentation confidence and accuracy of overlap. However, probability mask overlays and visualizations of confidence showed consistent activation in tooth-shaped areas, which proved that the network acquired meaningful localization information in spite of the low quantitative overlap values. These results show that additional optimization in terms of longer training, threshold optimization, and augmentation techniques are needed to enhance the accuracy of segmentation without compromising the clinically interpretable localization performance. It is worth mentioning that the main goal of segmentation module in this paper was tooth-region localization to aid in the process of providing prosthodontic radiographic evaluation as opposed to pixel-level boundary optimization. Thus, in this framework of exploration, qualitative anatomical correspondence and region correspondences are regarded as more clinically relevant measures compared to rigid mask overlap measures.

4.4 Segmentation Visualization and Prediction Patterns

The predicted segmentation outputs were visually inspected and the model was found to have learned how to approximate the location of the molar regions, although with low confidence in certain samples. An illustrative case is presented in Figure 2 whereby the predicted mask probability map identified an area of teeth and the final overlay indicated a bounding box containing the identified second molar.

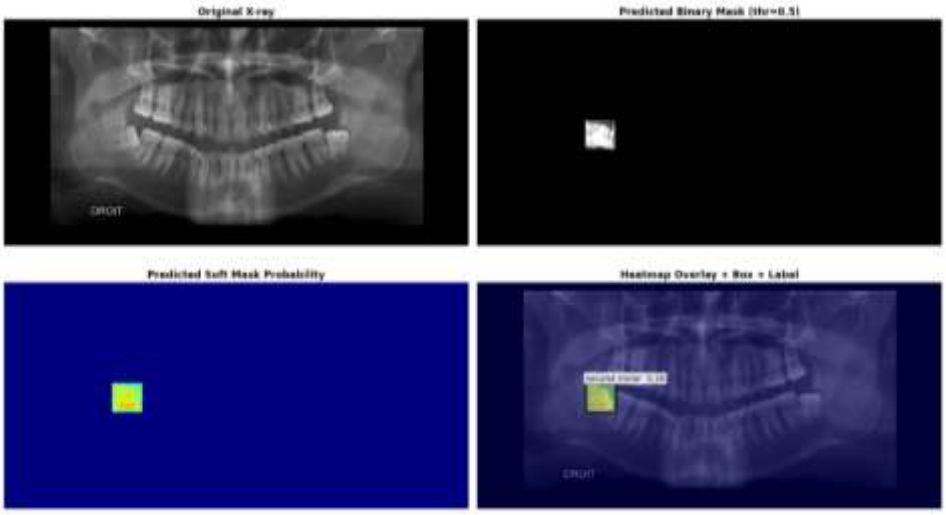

Figure 2. Mask R-CNN Segmentation XAI Output (Mask Probability and Overlay)

Figure 2 illustrates a four-panel segmentation output: (a) original panoramic X-ray, (b) predicted binary mask at threshold 0.5, (c) predicted soft mask probability heatmap, and (d) final overlay with bounding box and predicted label ("second molar"). The result demonstrates that at low confidence (score=0.20), the network is activated in the correct areas with an anatomical shape of teeth. This points to the new localization ability of the model but also to better segmentation confidence calibration.

4.5 Tooth Cropping and CNN Classification Outcomes

After the stage of instance segmentation, tooth regions were automatically localised based on predicted bounding boxes and mask-based localization. The patches of cropped teeth were mainly produced so that they could be analysed in terms of localised explainability with Grad-CAM, which is most effective with CNN-based classification structures. This middle data set enabled visualisation of discriminative regions of the anatomy at the tooth level and enabled the assessment of interpretability of correct and incorrect predictions. The produced cropped outputs were then utilised to create Grad-CAM heatmaps to confirm that the attention of the model was related to the clinically significant tooth morphology.

4.6 Grad-CAM Explainability Results

One of the strongest points of this study is the use of explainable AI (XAI) techniques to prove clinical interpretability. The predictions of CNN classifiers were subjected to grad-CAM, which produced heatmaps indicating the most discriminative anatomical parts that were utilised to classify teeth.

The Grad-CAM results were classified as correct and incorrect predictions. An example is given in Figure 3, where Grad-CAM placed a lot of attention on the crown of the tooth crop, and a correct prediction was made, which was the third molar. The pattern of attention of the model matched clinically significant morphology, which favoured interpretability.

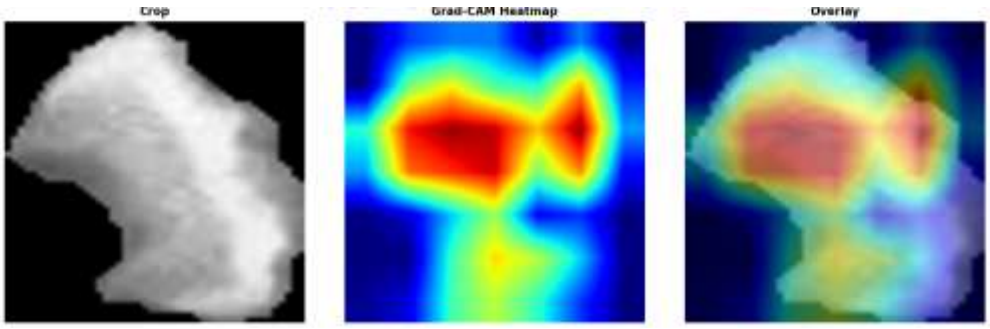

Figure 3. Grad-CAM Visualization for Correct Tooth Classification

Figure 3 demonstrates a three-panel Grad-CAM visualization: (a) cropped tooth image, (b) Grad-CAM heatmap, and (c) overlay of heatmap onto the tooth structure. The highest activation is observed on the upper crown region, which shows that the classifier was trained to recognise the discriminative patterns of tooth shapes and not the background artefacts. This promotes the transparency and clinical plausibility of the model.

On the contrary, false classification Grad-CAM outputs showed attention shift to irrelevant areas, including blank edges or non-tooth edges.

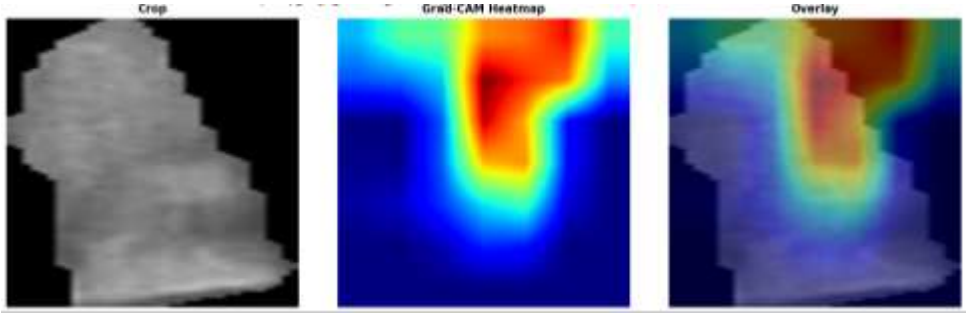

Figure 4. Grad-CAM Visualization for Incorrect Tooth Classification

Figure 4 depicts a false prediction in which the model had wrongly identified a tooth crop as a lateral incisor when the actual label was the second molar. The Grad-CAM heatmap shows diffuse activation that is localised to one side of the crop, as opposed to localising to important molar anatomical landmarks. This implies that the misclassification was affected by missing tooth boundaries and lack of structural contrast.

4.7 Occlusion Sensitivity and Perturbation Analysis

Occlusion sensitivity mapping, which was used to complement Grad-CAM, was done by moving a patch over the image and measuring the loss of confidence. The occlusion setup had a patch size of 32 and stride of 32 with three sample images being analysed. The resulting occlusion maps indicated regions of high impact of the teeth and ensured that the confidence of the classifier reduced in the presence of important tooth regions that were blocked.

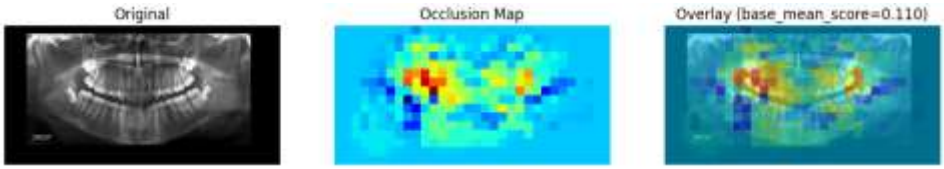

Figure 5. Occlusion Sensitivity Heatmap Overlay

Figure 5 represents results of occlusion sensitivity in which blocked patches in the molar area produced the greatest decrease in prediction confidence. This confirms the Grad-CAM results and reinforces the interpretability evidence, which proves that the model is based on actual morphology of teeth as opposed to irrelevant background morphology.

In order to have a unified view of the explainability outputs produced in this study, the following table summarises the number of visualisations and interpretability artefacts produced in various XAI modules.

Table 2. Summary of Explainability Outputs Generated in the Proposed Framework

Output Component	Result
Total tooth crops generated	178
Mask confidence visualizations	6 images
Occlusion sensitivity analyses performed	3 images
Grad-CAM correct prediction visualizations	7
Grad-CAM incorrect prediction visualizations	20
Occlusion patch size	32 × 32
Occlusion stride	32
Score threshold (segmentation)	0.05
Mask threshold	0.5

The interpretability artifacts produced during the course of the experiment are presented in a tabularized manner in Table 2. Segmentation outputs were used to extract 178 tooth-level crops that were used to support localized analysis and Grad-CAM visualization. To test segmentation confidence, mask confidence overlays were produced, whereas to test occlusion sensitivity, three representative panoramic images were tested using a 32 × 32 sliding patch set up. The visualisations of Grad-CAM were split into two groups, namely correct (n = 7) and incorrect (n = 20) predictions, which allowed the analysis of errors in detail. The two sets of correct and misclassified examples are beneficial to the assessment of interpretability because they enable the evaluation of the alignment of attention to anatomical structures.

All in all, the explainability pipeline generated various artefacts that complement each other, which proves that the suggested framework is a transparent and interpretable framework and not a black-box model. This is a summative overview that strengthens the focus of the study on the explainable integration of AI in panoramic dental image analysis.

5. Discussion

This paper suggested a deep learning-based panoramic dental radiograph analysis framework by combining instance segmentation and explainable artificial intelligence (XAI). The pipeline ensured the use of structured preprocessing and tooth-level localization based on Mask R-CNN and interpretation based on the use of Grad-CAM, occlusion sensitivity mapping, and mask confidence overlay. The COCO annotation structures were validated and no duplication of cross-split was detected, which guarantees experiment reliability. Despite the comparably equal distribution of data between eight types of teeth, the number of third molars was relatively low, and this might have affected the confidence of the models, indicating the necessity of class-wise training methods.

The results of qualitative segmentation proved that the Mask R-CNN model could identify tooth-shaped regions in panoramic radiographs and generate anatomically meaningful masks and bounding boxes. Nevertheless, quantitative assessment showed that the IoU and Dice scores were low at a low confidence threshold (0.05), which suggests the inability to refine the boundaries and calibrate the confidence. Such difficulties can be partly explained by the panoramic radiograph features of structural overlap and indistinct edges, which decrease the segmentation certainty. In spite of these shortcomings, the explainability analysis yielded clinically meaningful results. Grad-CAM visualization revealed that correct predictions were primarily determined by the areas of interest on the tooth including the crown and the boundaries of the enamel, and incorrect predictions had diffuse or disoriented attention. These results were further supported by occlusion sensitivity mapping which demonstrated a decrease in confidence when the important parts of the teeth were covered, suggesting that the decision making of the model was mainly informed by anatomically important regions.

Grad-CAM integration is consistent with the existing literature that shows that gradient-based visualization enhances transparency in dental diagnostic systems [1], and that explainable analysis is especially relevant in panoramic imaging because of the complexity of the anatomy [6]. The systematic reviews have also pointed out that the dental deep learning studies do not have standardized interpretability reporting, and frameworks that combine XAI methods are needed [2]. Although the

performance of the segmentation was lower than what could be obtained by optimized or ensemble-based methods, which can be more accurate due to lower variance [17], the main input of the work is the creation of an interpretable and modular proof-of-concept model. The areas that should be improved in the future include threshold optimization, longer training, more robust augmentation, and class imbalance. Before clinical deployment, external validation on bigger and more diverse datasets is required as well. In general, the paper has shown that explainable deep learning can offer clear and anatomically guided information to analyse panoramic dental radiographs.

6. Conclusion

The paper created and tested a deep learning-based panoramic dental radiograph analysis framework through the combination of instance segmentation and explainable artificial intelligence (XAI) methods. The suggested pipeline included dataset auditing, preprocessing, tooth localization with the help of Mask R-CNN, and interpretability-driven visualization with the help of Grad-CAM, occlusion sensitivity mapping, and mask confidence analysis. The results confirm that the framework is capable of producing anatomically significant localization results and produce clear visual descriptions that make models more interpretable. The findings showed that preprocessing was successfully used to guarantee reliability of the dataset by authenticating the COCO annotations and removing the duplication of cross-splits. The results of qualitative segmentation revealed that the Mask R-CNN model had the capacity to recognize tooth-shaped areas, but quantitative IoU and Dice scores were constrained at high confidence levels. Notably, Grad-CAM and occlusion-based explainability findings demonstrated that accurate predictions were always centered on clinically significant areas of the tooth whereas inaccurate predictions were characterized by attention drift, which suggests the utility of XAI in the interpretation of errors and clinical confidence. The implications of this work indicate the necessity to incorporate interpretability mechanisms into the dental AI systems to mitigate the black-box constraint and facilitate acceptance by the clinicians. The framework has been developed as a proof of concept system that can be replicated to provide a base to bigger prosthodontic imaging data sets as well as multi-center validation experiments. The suggested framework can be used as a basis to create effective decision-support instruments in dental diagnostics and research. The current research should be developed in the future by working on segmentation performance by optimizing thresholds, training, augmentation, and class balancing techniques. Also, it is advisable to test the framework on bigger and more varied datasets in order to enhance generalization.

Acknowledgements

We would like to thank Faculty of Dentistry, Thammasat University for the Research Fund, Contract No.6/2568

Authors' contributions

D.D., and S.S. contributed to conceptualization, methodology, supervision, and writing the original draft. D.D., and S.S. contributed to data collection, project administration and statistical analysis. All authors wrote, revised and edited the manuscript prior to submission.

Funding

This work was supported by Faculty of Dentistry Thammasat University Research Fund (6/2568).

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

This study was approved by the Human Research Ethics Committee of the Thammasat University (review board number COA 047/2567) and was performed in accordance with the tenets of the Declaration of Helsinki. Informed consent was waived from all patients because of the retrospective nature of the fully anonymized radiographic images.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Kim, D.; Kim, J.; Choi, S.G. CNN-based remote dental diagnosis model for caries detection with grad-CAM. Sci. Rep. 2025, 15, 26555.
2. Singh, N.K.; Raza, K. Progress in deep learning-based dental and maxillofacial image analysis: A systematic review. Expert Syst. Appl. 2022, 199, 116968.
3. Ali, D.A.; Sadeeq, H.T. An interpretable deep learning framework for multi-class dental disease classification from intraoral RGB images. Stat. Optim. Inf. Comput. 2025, 14, 3380–3397.
4. Wen, C.; Bai, X.; Yang, J.; Li, S.; Wang, X.; Yang, D. Deep learning based approach: automated gingival inflammation grading model using gingival removal strategy. Sci. Rep. 2024, 14, 19780.
5. Parkhi, P.; Harjal, S.; Sahu, A.; Agrawal, P.; Shingne, H.; Bobde, Y.; Padole, A. A comprehensive deep learning framework for dental disease classification. J. Eur. Syst. Autom. 2025, 58, (3).
6. Oztekin, F.; Katar, O.; Sadak, F.; Yildirim, M.; Cakar, H.; Aydogan, M.; Acharya, U.R. An explainable deep learning model to prediction dental caries using panoramic radiograph images. Diagnostics 2023, 13, 226.
7. Asghar, S.; Rashid, J.; Masood, A. CariesXplainer: Enhancing dental caries detection using gradient-weighted class activation mapping and transfer learning. Multimed. Tools Appl. 2025, 84, 1–26.
8. Yang, S.; Lee, H.; Jang, B.; Kim, K.D.; Kim, J.; Kim, H.; Park, W. Development and validation of a

visually explainable deep learning model for classification of C-shaped canals of the mandibular second molars in periapical and panoramic dental radiographs. J. Endod. 2022, 48, 914–921.
9. Pishghadam, N.; Esmaeilyfard, R.; Paknahad, M. Explainable deep learning for age and gender estimation in dental CBCT scans using attention mechanisms and multi task learning. Sci. Rep. 2025, 15, 18070.
10. Ong, S.H.; Kim, H.; Song, J.S.; Shin, T.J.; Hyun, H.K.; Jang, K.T.; Kim, Y.J. Fully automated deep learning approach to dental development assessment in panoramic radiographs. BMC Oral Health 2024, 24, 426.
11. Inani, H.; Mehta, V.; Bhavsar, D.; Gupta, R.K.; Jain, A.; Akhtar, Z. AI-enabled dental caries detection using transfer learning and gradient-based class activation mapping. J. Ambient Intell. Humaniz. Comput. 2024, 15, 3009–3033.
12. Mani, P.; Meenatchi, K.; Gowrishankar, C.; Nallakumar, R.; Kanikha, M.; Kavin, B.; Rohith, G.K. A model employing interpretable deep learning techniques to forecast dental caries through the analysis of panoramic radiograph images. In Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Chennai, India, June 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
13. Chen, I.D.S.; Yang, C.M.; Chen, M.J.; Chen, M.C.; Weng, R.M.; Yeh, C.H. Deep learning-based recognition of periodontitis and dental caries in dental x-ray images. Bioengineering 2023, 10, 911.
14. Zhou, Y.; Jiang, F.; Cheng, F.; Li, J. Detecting representative characteristics of different genders using intraoral photographs: a deep learning model with interpretation of gradient-weighted class activation mapping. BMC Oral Health 2023, 23, 327.
15. Imak, A.; Celebi, A.; Siddique, K.; Turkoglu, M.; Sengur, A.; Salam, I. Dental caries detection using score-based multi-input deep convolutional neural network. IEEE Access 2022, 10, 18320–18329.
16. Sreeram, A.; Balamurugan, R.; Aditya, M.N. Explainable AI for panoramic dental radiographs using contrastive learning and U-Net based segmentation. J. Soft Comput. Paradigm 2025, 7, 114–123.
17. Kang, S.; Shon, B.; Park, E.Y.; Jeong, S.; Kim, E.K. Diagnostic accuracy of dental caries detection using ensemble techniques in deep learning with intraoral camera images. PLoS ONE 2024, 19, e0310004.
18. Forouzeshfar, P.; Safaei, A.A.; Ghaderi, F.; Hashemi Kamangar, S.; Kaviani, H.; Haghi, S. Dental caries diagnosis using neural networks and deep learning: A systematic review. Multimed. Tools Appl. 2024, 83, 30423–30466.
19. Salehizeinabadi, M.; Ameli, N.; Kouchehbaghi, K.; Arastoo, S.; Neghab, S.; Kornerup, I.M.; Pacheco-Pereira, C. Dental age prediction from panoramic radiographs using machine learning techniques. PLOS Digit. Health 2025, 4, e0001077.
20. Brahmi, W.; Jdey, I.; Drira, F. Panoramic Dental Xray Dataset. Mendeley Data 2025, 3. https://doi.org/10.17632/73n3kz2k4k.3
21. Brahmi, W.; Jdey, I.; Drira, F. Exploring the role of convolutional neural networks (CNN) in dental radiography segmentation: A comprehensive systematic literature review. Eng. Appl. Artif. Intell. 2024, 133, 108510.
22. Brahmi, W.; Jdey, I. Automatic tooth instance segmentation and identification from panoramic X-ray images using deep CNN. Multimed. Tools Appl. 2024, 83, 55565–55585.
23. Dennis D, Suebnukarn S, Vicharueang S, Limprasert W. Development and evaluation of a deep learning segmentation model for assessing non-surgical endodontic treatment outcomes on periapical radiographs: a retrospective study. PLoS One. 2024;19(12):e0310925.
24. Dennis D, Suebnukarn S, Heo MS, Abidin T, Nurliza C, Yanti N, et al. Artificial intelligence application in endodontics: a narrative review. Imaging Sci Dent. 2024;54(4):305–312.