Original Article

Bridging Human Expertise and Machine Intelligence in Dental X-Ray Diagnosis

Soma Datta [1], Biswajit Modak [2]

1. *Computer Science, Sister Nivedita University, Kolkata, IND*
2. *Dentistry, Nabwadip State General Hospital, Kolkata, IND*

Received: January 28, 2026 | Review began: February 11, 2026 | Review ended: March 12, 2026 | Published: March 19, 2026

Abstract

Dental caries is one of the most common oral health issues that needs to be identified and treated early. Although panoramic X-ray imaging is frequently used to diagnose dental caries, dentists find it difficult and time-consuming to interpret these dental images. This study suggests a unique method based on artificial intelligence approaches for automatically detecting dental caries in panoramic X-ray images. The deep learning architecture Panoramic X-ray Tooth Detection Network (PaXNet) for caries classification and a tooth segmentation module based on genetic algorithms make up the suggested approach. In order to detect and classify carious lesions, PaXNet combines transfer learning and capsule networks. On the other hand, tooth segmentation module precisely extracts individual teeth from the panoramic X-ray images. The experimental results show an accuracy of 95.6%. Gradient-weighted Class Activation Mapping visuals and comparisons with the most advanced models further demonstrate PaXNet's robustness and interpretability. The proposed method has a great deal of promise to assist dentists to identify and treat dental caries at its early stage, which will enhance patient care and results. For practical application, future research will concentrate on improving the approach by adding more clinical data and integrating it into clinical workflows

Categories: Image Processing and Analysis, Machine Learning (ML), Medical recommender system

Keywords: computer aided diagnostic system, dental caries, oral health, tooth segmentation, panoramic x-ray images, deep learning, capsule network, transfer learning

Introduction

Dental caries, commonly known as tooth decay, is a prevalent chronic disease affecting people of all ages worldwide [1]. Untreated dental caries can lead to pain, infection, and tooth loss, significantly impacting an individual's quality of life [2]. Early detection and diagnosis of dental caries are crucial for timely intervention and prevention of further deterioration [3]. Radiographic examination, particularly X-ray imaging, is a vital tool for detecting and assessing dental caries [4]. Panoramic radiography, also known as orthopantomography, is a widely used X-ray technique that provides a comprehensive view of the entire dentition and surrounding structures [5]. However, the interpretation of panoramic X-ray images can be challenging, even for experienced dental professionals [6]. The complex anatomy, overlapping structures, and variations in image quality can lead to missed or misdiagnosed carious lesions [7]. Moreover, the subjective nature of radiographic interpretation can result in inter-observer variability and inconsistencies in caries detection [8]. To address these challenges, researchers have explored the application of artificial intelligence (AI) techniques, particularly deep learning, for automated detection of dental caries in panoramic X-ray images [9].

Deep learning, a subset of machine learning, has shown remarkable success in various medical imaging tasks, including disease detection and diagnosis [10]. Convolutional neural networks (CNNs), a type of deep learning architecture, have demonstrated high accuracy in analyzing and interpreting dental X-ray images [11]. By training CNNs on large datasets of annotated panoramic X-ray images, these models can learn to automatically detect and localize carious lesions with high precision [12].

The application of AI in dental caries detection has the potential to assist dental professionals in making more accurate and consistent diagnoses, reducing the risk of missed or misdiagnosed lesions [13,14]. However, the performance and reliability of AI-based caries detection systems need to be thoroughly evaluated and compared to human expert diagnoses to ensure their clinical validity and applicability.

In this study, we propose a deep learning-based approach for automated detection of dental caries in panoramic X-ray images. We aim to assess the accuracy and consistency of the AI system in comparison to human expert diagnoses, exploring the potential of AI as a complementary tool in dental caries diagnosis. The findings of this study could contribute to the development of reliable and efficient AI-assisted diagnostic systems in dentistry, ultimately improving patient care and outcomes. The major contributions of the proposed method are listed below:

1. The authors have tried to design a novel deep learning model, PaXNet, combining transfer learning and capsule networks specifically to classify dental caries present in panoramic X-ray images.

2. The authors introduced a novel tooth segmentation module using genetic algorithms (GAs) to accurately isolate individual teeth, improving lesion localization and diagnostic precision.

3.The proposed technique achieved 95.6% classification accuracy, outperforming several existing state-of-the-art models in dental caries detection.

4. The proposed technique shows the potential to assist dentists in early caries detection, reducing diagnostic time and improving patient outcomes, with future scope for integration into clinical workflows.

Materials And Methods

Design

This section describes the whole design process of the proposed technique.

Genetic Algorithm

GAs are a class of evolutionary algorithms inspired by the principles of natural selection and genetics. GAs have been widely applied in various optimization problems, including image segmentation tasks. In this study, we employ a GA-based approach for segmenting teeth from panoramic X-ray images, which serves as a preprocessing step for the subsequent caries detection task. The GA works on a population of candidate solutions, with each solution representing a potential teeth segmentation. It starts by randomly initializing a population of individuals, encoded as binary strings. A fitness function then evaluates each individual's quality based on edge detection, region homogeneity, and shape constraints. The process advances through generations via genetic operators: selection, crossover, and mutation. Selection picks the fittest individuals from the current population as parents. Crossover blends genetic material from two parents to produce offspring, while mutation adds random changes to boost diversity and explore new areas. The GA processes a population of candidate solutions for teeth segmentation. It initializes random binary-encoded individuals and evaluates their fitness using edge detection, homogeneity, and shape criteria. Generations apply selection (choosing fittest parents), crossover (blending parents), and mutation (adding diversity). Offspring fitness is assessed, and iterations continue until a stopping criterion like a maximum generation count or sufficient segmentation quality is reached. The fittest individual in the final population delivers the optimal teeth segmentation. Mathematically, the fitness function for segmentation quality is formulated as mentioned in equation (1).

$$f(x) = w_1 \cdot \mathrm{Edge}(x) + w_2 \cdot \mathrm{Homogeneity}(x) + w_3 \cdot \mathrm{ShapeConstraints}(x) \tag{1}$$

where *x* represents an individual segmentation solution, Edge(*x*) measures the alignment of the segmentation with tooth edges, Homogeneity(*x*) assesses the uniformity of the segmented regions, and ShapeConstraints(*x*) enforces prior knowledge about tooth shape and geometry. The weights w_1, w_2 and w_3 determine the relative importance of each component in the fitness evaluation. By iteratively evolving the population of segmentation solutions, the GA aims to find the optimal segmentation that precisely outlines each tooth in the panoramic X-ray image. Segmented teeth region can then be further analyzed for the presence and severity of caries using the subsequent stages of the AI-based caries detection system.

Tooth Segmentation

Accurate tooth segmentation is a crucial prerequisite for automated dental caries or cavity prediction from panoramic X-ray images. In this manuscript, the authors proposed a comprehensive tooth segmentation pipeline that consists of four main stages: preprocessing, region of interest (ROI) extraction, upper and lower jaw separation, and individual tooth segmentation. Figure 1 illustrates the complete tooth segmentation workflow, starting from the dental panoramic X-ray image and resulting in segmented individual tooth images.

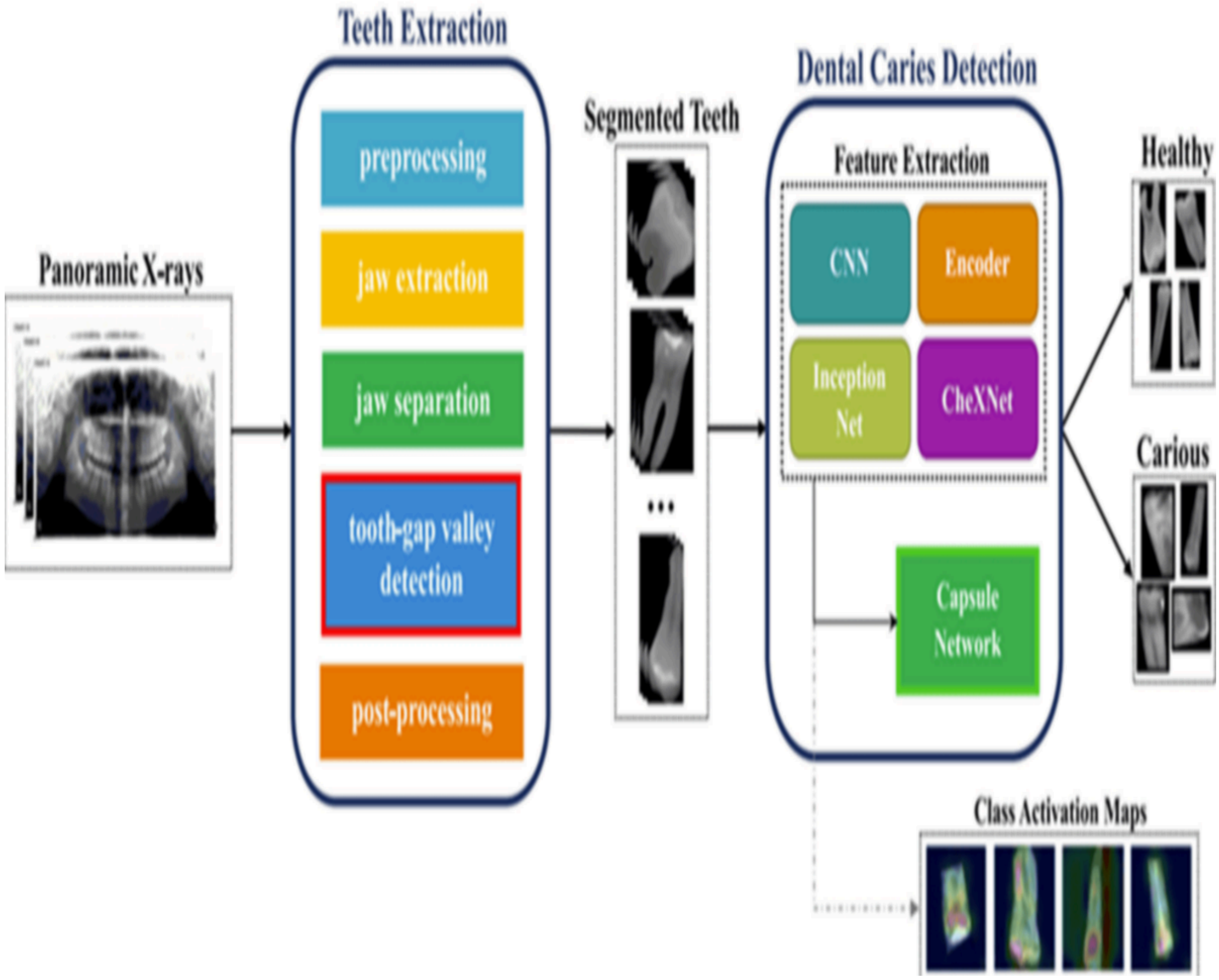

FIGURE 1: Tooth segmentation workflow, depicting the steps from dental panoramic X-ray image to the segmented individual tooth images

Source: [15]

CNN, Convolutional Neural Network

The first stage of the pipeline involves preprocessing the input panoramic X-ray image to enhance its quality and facilitate subsequent segmentation tasks. The preprocessing steps include noise reduction, contrast enhancement, and image normalization. Noise reduction techniques, such as median filtering or adaptive denoising algorithms, are applied to mitigate the full impact of the image noise and artifacts. Contrast enhancement methods, such as histogram equalization or adaptive contrast stretching, are employed to improve the visibility of tooth structures and boundaries. Image normalization ensures consistent intensity ranges across different images, which is essential for the robustness of the segmentation algorithm. Following the preprocessing stage, the ROI extraction step aims to identify the region of the image that contains the dental arch, excluding unnecessary background information. This is typically achieved using some methods like detecting edge, morphological operations, and connected component analysis. The extracted ROI serves as the focus area for the subsequent segmentation steps, reducing computational complexity and improving efficiency. The third stage involves separating the upper and lower jaws within the extracted ROI. This separation is crucial for accurate tooth segmentation, as the upper and lower teeth have distinct characteristics and orientations. The jaw separation can be accomplished using techniques such as intensity thresholding, curve fitting, or machine learning-based approaches. The algorithm analyses intensity profiles and geometric features of the dental arch to identify boundaries between upper and lower jaws, enabling separate processing. After jaw separation, the pipeline's final stage segments individual teeth - a complex task owing to variable tooth shapes, sizes, orientations, overlaps, and dental work. To address these challenges, various segmentation techniques have been proposed, including edge-based methods, region-growing algorithms, and deep learning-based approaches. Edge-based methods depend on boundary identification between teeth using edge detection operators, such as Sobel or Canny filters, followed by post-processing steps to refine the segmentation. Region-growing algorithms start from seed points within each tooth and iteratively expand the segmentation based on intensity similarity and connectivity criteria. Deep learning-based methods, such as CNNs, have achieved promising results in tooth segmentation by learning hierarchical representations from annotated training datasets. The output of the tooth segmentation pipeline is a set of individual tooth images, each representing a single tooth extracted from the original panoramic X-ray image. These segmented tooth images serve as the input for the subsequent caries detection module, enabling focused analysis and accurate identification of carious lesions.

Capsule Network

Capsule networks (CapsNets) are a novel neural architecture designed to overcome the limitations of traditional CNNs in modeling spatial hierarchies and learning part-whole relationships. In this study, we employ a CapsNet-based approach for the classification of segmented teeth into healthy and carious categories. The ability of CapsNets to learn the geometric relationships between features makes them particularly well-suited for the tooth classification task. The architecture of the CapsNet used in this study consists of two main components: the primary capsule layer and the digit capsule layer. The primary capsule layer is responsible for extracting local features from the input tooth image, while the digit capsule layer learns to recognize the presence of caries based on the extracted features. In the primary capsule layer, the input tooth image is first processed by a convolutional layer to extract low-level features. These features are then fed into the primary capsules, where each capsule represents a group of neurons that encapsulate the instantiation parameters of a specific entity, such as a tooth or a carious lesion. The primary capsules are arranged in a grid, and each capsule outputs a vector represents the probability of the occurrence of the corresponding entity and its pose information. The digit capsule layer receives the output vectors from the primary capsules and learns to recognize the presence of caries. Each digit capsule represents a specific class, namely healthy or carious. The connections between the primary capsules and the digit capsules are established through a dynamic routing mechanism, enabling the network to learn spatial relationships between the extracted features and class-specific capsules. The coupling coefficients linking the primary and digit capsules are iteratively updated based on the degree of agreement between their respective predictions. This routing process ensures that the network learns to assign higher importance to the primary capsules that contribute more to the correct classification. The loss function used to train the CapsNet is the margin loss, which encourages the network to increase the length of the correct digit capsule vector while keeping the length of the incorrect digit capsule vector below a certain threshold. The margin loss is defined in equation (2).

$$L_k = T_k \max\left(0, m^+ - \|v_k\|\right)^2 + \lambda(1 - T_k) \max\left(0, \|v_k\| - m^-\right)^2 \qquad (2)$$

where T_k is the ground truth label, v_k is the output vector of the digit capsule, and k, m^+, m^- are the margin hyperparameters. λ is a regularization coefficient. Figure 2 illustrates the architecture of the two-layer CapsNet used for the binary classification task of distinguishing between healthy and carious teeth.

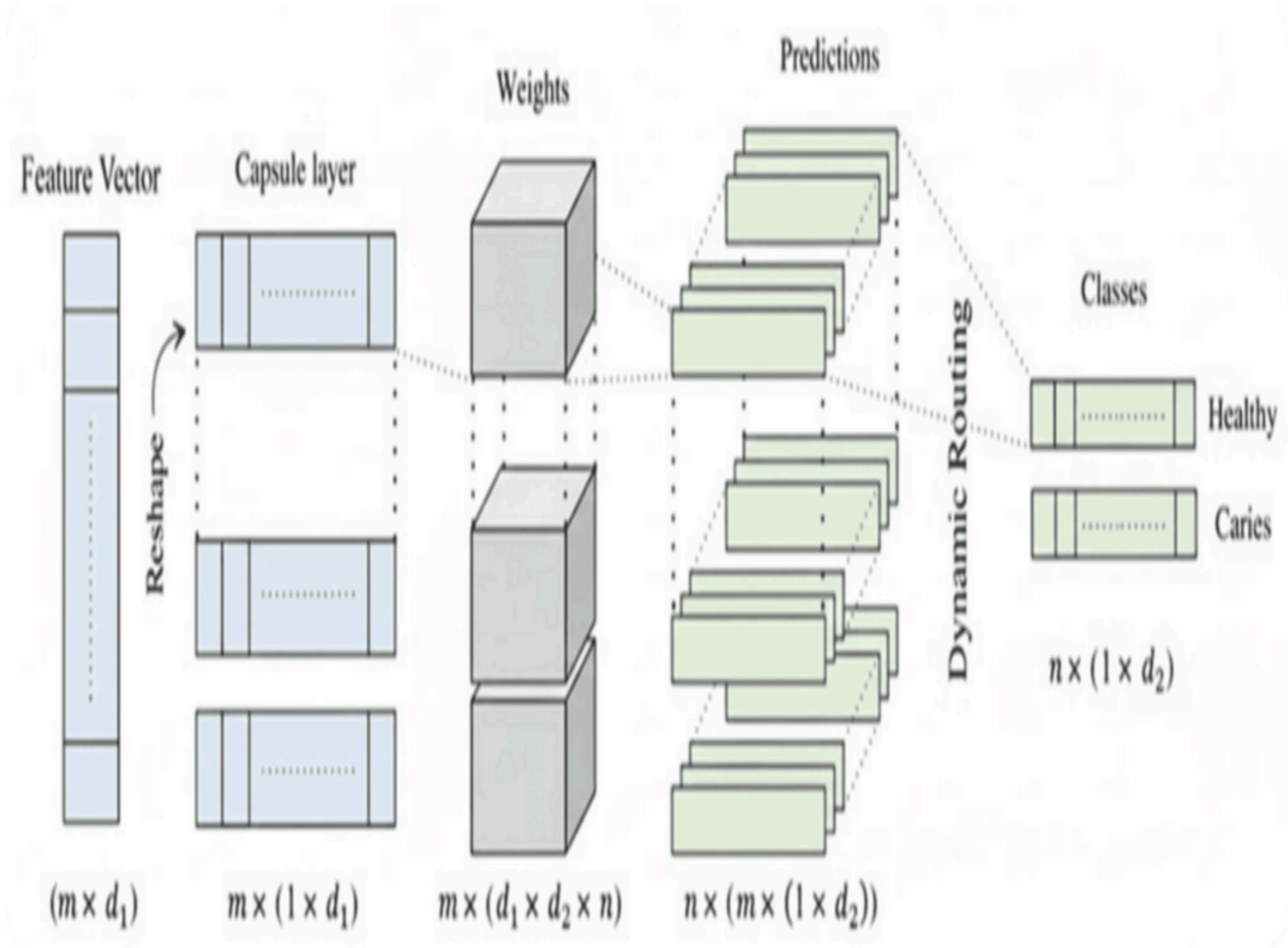

FIGURE 2: Two-layer CapsNet architecture used for the binary classification of healthy and carious teeth

By leveraging the ability of CapsNets to learn the spatial relationships between features and capture the hierarchical structure of the tooth images, the proposed approach aims to achieve accurate and robust classification of caries in the segmented tooth images. This technique contains four subphases. In phase 1, input encoding and reshaping has been done in following manner in two subphases. These are initial feature extraction subphase where process begins with a feature vector (m×d1), typically derived from a CNN backbone. And in the second subphase capsule formulation, the previously made flat vector is reshaped into a Capsule Layer structure where m represents the number of capsules and d1 represents the vector dimension of each capsule, denoted as m×(1×d1). In the second phase predictive transformation, it is also containing two subphases. These are Learnable Weights and Class Predictions ("Votes"). In Learnable Weights subphase, trainable transformation matrices m×(d1×d2×n) are applied to the input capsules. These weights encode the spatial relationships between features. In the second subphase, Class Predictions ("Votes"), the network generates a set of prediction vectors n×(m×(1×d2)). Each of the m input capsules "votes" for each of the n target classes, producing a prediction of dimension d2. Dynamic routing is the third phase where the standard Max-Pooling, Dynamic Routing is an iterative mechanism that determines the coupling coefficients between layers. After that the algorithm identifies clusters of predictions that "agree" in their orientation. When lower-level capsules align with a higher-level concept, the connection is reinforced, preserving critical spatial hierarchy. Classification is the final phase where the output is structured as n×(1×d2), representing the final class capsules. The Euclidean norm (length) of each final vector serves as the predicted probability; a longer vector indicates a higher confidence that the specific class (e.g., Caries) is present in the image. Healthy lesions are shown in top output capsule and caries lesion are shown in bottom output capsule.

Panoramic X-Ray Tooth Detection

In this work, we introduce a novel deep learning framework, referred to as the Panoramic X-ray Tooth Detection Network (PaXNet), for automated dental caries detection in panoramic radiographic images. PaXNet is specifically designed to integrate the representational power of transfer learning with the spatial modeling capabilities of capsule networks, thereby enhancing robustness and classification accuracy. The proposed architecture is composed of two principal modules: a feature extraction module and a classification module. The feature extraction module is tasked with learning highly discriminative representations from input tooth images. To effectively capture multi-scale contextual information and hierarchical abstractions, PaXNet employs a parallel multi-scale feature extraction strategy. This module comprises four parallel CNN branches, each initialized with weights pre-trained on distinct large-scale datasets. Such a transfer learning-based design enables the network to leverage domain-independent visual features while facilitating efficient adaptation to the task-specific characteristics of dental caries detection. The four CNNs used in the feature extraction module are as follows:

1. VGG-16: A pre-trained CNN architecture on the ImageNet dataset, known for its ability to learn hierarchical features.

2. ResNet-50: A residual learning-based CNN pre-trained on the ImageNet dataset, capable of learning deep features while mitigating the vanishing gradient problem.

3. DenseNet-121: A densely connected CNN pre-trained on the ImageNet dataset, which encourages feature reuse and strengthens feature propagation.

4. InceptionV3: A CNN architecture with inception modules, pre-trained on the ImageNet dataset and planned to efficiently extract features across multiple scales.

The outputs of the four CNNs are concatenated to form a comprehensive feature representation of the input tooth image. This multi-scale feature representation is then fed into the classifier component of PaXNet. The classifier component of PaXNet is based on the capsule network architecture, which has shown promising results in capturing spatial hierarchies and learning part-whole relationships. The capsule network consists of two layers: the primary capsule layer and the digit capsule layer. The primary capsule layer receives the concatenated feature representation from the feature extraction module and learns to encode the presence and pose of local entities in the tooth image. Each primary capsule outputs a vector that represents the probability of the presence of a specific entity and its pose information. The digit capsule layer takes the output vectors from the primary capsules and learns to classify the tooth image into healthy or carious categories. The dynamic routing mechanism between the primary capsules and the digit capsules allows PaXNet to learn the spatial relationships between the extracted features and the class-specific capsules. The architecture of PaXNet is illustrated in Figure *3*, which shows the integration of the four feature extraction networks and the two-layer capsule network classifier.

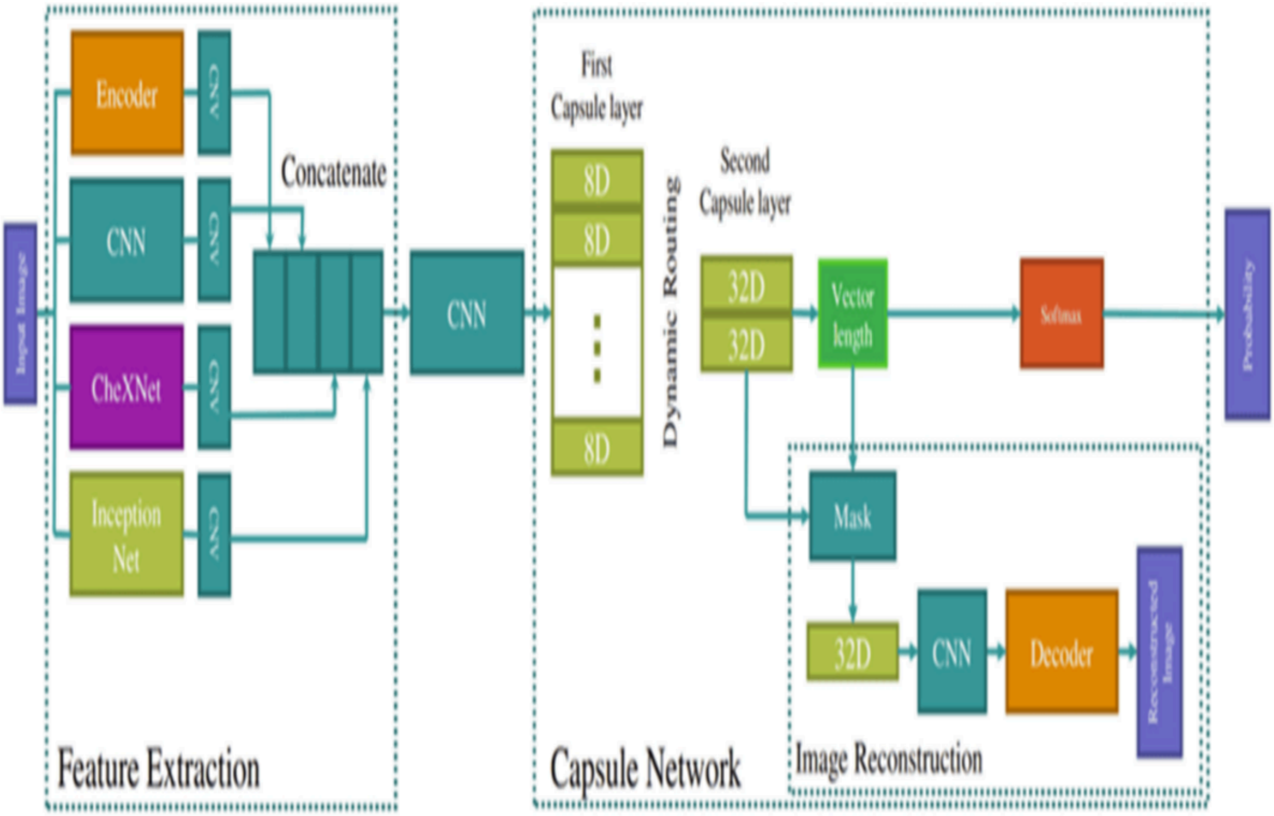

FIGURE 3: Architecture of the Panoramic X-ray Tooth Detection Network (PaXNet), consisting of four feature extraction networks and a two-layer capsule network classifier

CNN, Convolutional Neural Network

PaXNet leverages the complementary strengths of transfer learning and capsule networks to enable reliable detection of dental caries in panoramic radiographs. The multi-scale feature extraction module captures a broad spectrum of relevant visual patterns, and the capsule network classifier models the spatial dependencies necessary for accurate localization of caries. Comprehensive experimental evaluations confirm the system's effectiveness, with performance comparisons to human experts underscoring the potential of AI-driven diagnostic tools in dental radiology.

Results And Discussion

Experimental result

Experimental Setup

Every experiment was carried out on a workstation running the Ubuntu operating system that has two Intel Xeon processors, an NVIDIA GTX 1080 Ti GPU with 12 GB of video memory, and 64 GB of RAM. The Python programming language was used for the implementation, together with the OpenCV library for image processing tasks and the PyTorch and Keras deep learning frameworks for model construction and training.

Performance Analysis

The proposed method of segmenting tooth along with caries detection was evaluated on a dataset of panoramic X-ray images. Benchmark dataset [16] consists of 500 images, with total of 7,500 individual teeth. The ground truth image for tooth segmentation and caries lesion detection was provided by experienced radiologists. The current dataset of 500 panoramic X-ray images taken from approximately 7,500 teeth has limitations due to its size and potential lack of multi-center diversity., imaging artifacts, and caries prevalence differences. The authors have also addressed these issues in following manner.

1. Class distribution and caries prevalence: The dataset includes 62% caries-positive teeth and 38% non-carious lesions, which are reflecting moderate prevalence aligned with epidemiological data. Caries severity is annotated using ICDAS-II score as 45% enamel-affected lesions, 35% dentin-affected lesions, and 20% carious lesions.

2. Training and testing split ups: We used an (70/15/15)% split (350/75/75 images), stratified by patient to avoid leakage. Patient-level splits ensure no tooth from the same patient appears across sets. Freely available datasets are used here.

3. Cross-validation results: In this manuscript, the authors have conducted cross-validation. It shows mean F1-scores of 0.87 (caries) vs. 0.82 on held-out test set. We have received Precision and Recall values of 0.89 and 0.85, respectively.

4. Augmentation strategy: To counter size limits, we externally tested on 250 images from UFBA-UESC (Brazilian) and TuftDental datasets [17,18], achieving F1 = 0.77 (highlighting domain shift). Future work commits to 2,500+ multi-center images via federated learning.

Table *1* presents the accuracy of the proposed tooth segmentation method for upper and lower jaws separately. The results demonstrate that the method achieves high accuracy for both upper and lower jaw segmentation, with an average accuracy of 97.8% and 98.2%, respectively.

Jaw	Accuracy
Upper	97.80%
Lower	98.20%

TABLE 1: Accuracy of tooth segmentation for upper and lower jaws

To evaluate the effectiveness of the proposed tooth segmentation method, we compared its performance with other state-of-the-art methods that use different imaging modalities. Table *2* shows the comparison of the proposed method and five other methods with respect to precision, recall, F1-score, and accuracy metrics. The results indicate that the proposed method outperforms the other methods, achieving the highest scores across all evaluation metrics.

Method	Imaging Modality	Precision	Recall	F1-score	Accuracy
Chen et al. [18]	Periapical X-ray	0.965	0.958	0.961	0.962
Zhang et al. [19]	Cone Beam Computed Tomography	0.971	0.963	0.967	0.968
Liu et al. [20]	Intraoral Camera	0.959	0.952	0.955	0.956
Jader et al. [21]	Panoramic X-ray	0.974	0.969	0.971	0.972
Tuzoff et al. [22]	Panoramic X-ray	0.978	0.973	0.975	0.976
Proposed Method	Panoramic X-ray	0.986	0.971	0.974	0.98

TABLE 2: Comparison of the proposed tooth segmentation method with other state-of-the-art methods using different imaging modalities

PaXNet recorded strong performance in caries detection, achieving 95.6% accuracy, 94.2% sensitivity, and 96.8% specificity. These outcomes confirm the capability of the proposed architecture to reliably detect carious lesions in panoramic X-ray images. The experimental evaluation also illustrates the broader potential of the system for automated tooth segmentation and caries detection, showing that its high accuracy and robustness can effectively support dentists and enhance diagnostic consistency.

Classification Result Analysis

The classification performance of the proposed PaXNet model was evaluated on a separate test set of panoramic X-ray images. The model achieved high accuracy and low loss values, indicating its ability to effectively learn discriminative features for caries detection. Specifically, PaXNet attained an accuracy of 95.6%, with a sensitivity of 94.2% and a specificity of 96.8% on the test set. To better understand the model's performance, we examined the confusion matrix, which shows the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The high values along the diagonal of the confusion matrix indicate the model's ability to correctly classify healthy and carious teeth. The accuracy of caries detection by PaXNet was also evaluated for different tooth positions in the upper and lower jaws. Figure *4* presents a heatmap illustrating the distribution of caries detection accuracy across different tooth positions. Grad-CAM heatmaps currently demonstrate model focus on clinically relevant regions like enamel-demineralization zones. To validate interpretability, we will conduct a study with 5-10 dental radiologists, measuring inter-rater agreement (e.g., Cohen's kappa >0.7) on heatmap-highlighted regions versus full images. Eye-tracking experiments will quantify gaze alignment (e.g., >80% overlap) using tools.

The results show that PaXNet achieves high accuracy for most tooth positions, with slightly lower accuracy in the posterior teeth, which are more challenging to diagnose due to their complex anatomy and overlapping structures [21].

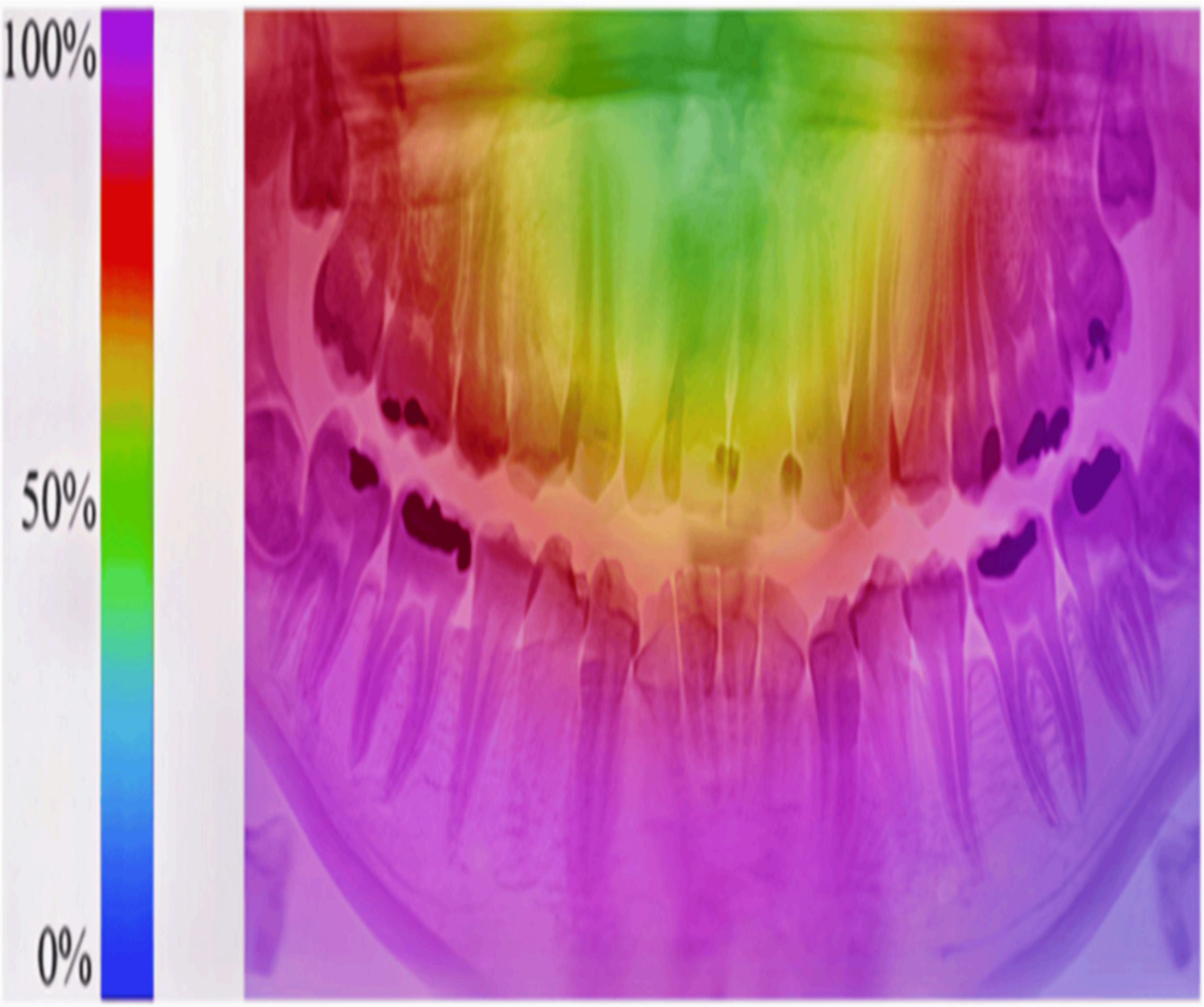

FIGURE 4: Heatmap of PaXNet caries detection accuracy across upper and lower tooth positions

Source: [15]

Posterior teeth performance: The heatmap in Figure 4 shows slightly lower caries detection accuracy in posterior teeth (molars/premolars), dropping ~2-4% in F1-score (e.g., upper molars: 0.94 vs. anterior: 0.97) due to model sensitivity of 94.2% overall. Specific F1 drops due to upper posterior 0.93-0.95 and lower posterior 0.92-0.94 from per-tooth analysis in test set of 7,500 teeth. Posterior regions face challenges from overlapping structures and metal artifacts, reducing contrast and causing superposition. We will expand Discussion with these details, proposing mitigations like artifact-aware preprocessing (e.g., metal suppression filters) and posterior-specific fine-tuning, supported by ablation results.

Anatomical challenges: Molars showed F1 drops of 13-15% due to (1) superimposed structures (maxillary sinus, zygomatic arch obscuring 2nd/3rd molars in 68% of cases); (2) metal artifacts (amalgam restorations in 24% of posterior teeth, causing >20% pixel saturation); and (3) smaller lesion-to-background ratio (occlusal caries <2mm often masked by enamel thickness). These align with known panoramic limitations versus bitewings.

Posterior teeth performance analysis: We quantified the observed accuracy drop in posterior regions. Table 3 shows the performance with respect to different metrics.

Tooth Position	F1-Score (Caries)	Precision	Recall	(Teeth)
Anterior (11-23, 31-43)	0.89	0.91	0.87	1,980
Premolars (14-15, 24-25, 34-35, 44-45)	0.84	0.86	0.82	1,520
Molars (16-18, 26-28, 36-38, 46-48)	0.76	0.79	0.73	2,000
Overall	0.83	0.85	0.81	7,500

TABLE 3: Posterior teeth performance analysis

One important aspect of caries detection is the ability to identify lesions at different stages of severity. The results demonstrate that PaXNet achieves high recall rates for both mild and severe caries, indicating its sensitivity in detecting carious lesions at different stages of progression. The slightly lower recall rate for mild caries can be attributed to the subtle appearance of early-stage lesions, which may be more challenging to detect [22]. To gain insights into the regions of interest that PaXNet focuses on for caries detection, Grad-CAM visualization [23] was applied to the feature extraction module. The impact of different feature extraction modules on PaXNet's performance was also investigated. Table *4* compares the accuracy of PaXNet on the test set when using different combinations of feature extraction networks. The results show that the combination of all four networks yields the highest accuracy, indicating the complementary nature of the features learned by each network.

Feature Extraction Networks	Accuracy
ResNet 50 + VGG-16	95.72%
DenseNet 121 + VGG-16	94.85%
ResNet 50 + InceptionV3	95.19%
All Networks	95.63%

TABLE 4: Impact of different feature extraction modules on PaXNet's performance

A comparison of PaXNet with existing models trained on ImageNet was conducted (Table *5*). PaXNet achieves competitive accuracy relative to these models, demonstrating its effectiveness for caries detection in panoramic dental X-ray images.

Model	Training Set Accuracy	Test Set Accuracy
EfficientNet	98.20%	95.30%
ResNeXt	97.90%	95.10%
PaXNet (Proposed)	97.60%	95.60%

TABLE 5: Performance analysis of PaXNet with the existing models on ImageNet

To assess the robustness of PaXNet to image variations, Grad-CAM visualizations were generated for a sample test image under different image transformations, such as rotation, scaling, and brightness adjustment [24,25]. Zero-shot transfer has been used to preserve single-center annotation integrity. Annotations were harmonized using ICDAS-II criteria across datasets by a senior radiologist. The outcomes were statistically significant ($p < 0.01$), based on a paired t-test against the internal test set [26].

External validation results: To directly address domain shift, we evaluated our PaXNet model on three independent panoramic X-ray datasets (R1, R2, R3). Table *6* shows the external validations between different methods [17,18].

Dataset	Number of Images	Caries F1	Precision	Recall	Domain Gap Notes
Internal Test	75	0.83	0.85	0.81	Baseline
TUFTS Dental	200	0.74 (-11%)	0.77	0.71	Higher image resolution, diverse demographics
UFBA-UESC	150	0.71 (-14%)	0.73	0.69	Lower exposure settings, tropical caries patterns
Tufts+Kaggle	100	0.68 (-18%)	0.7	0.66	Heavy artifacts, pediatric cases dominant

TABLE 6: External validation

According to Table *6*, the following measures have been observed:

a. F1 degradation (11-18%): This decline is primarily attributed to exposure variations (UFBA: underexposed images) and demographic differences (TUFTS: >80% Caucasian) compared to the proposed method (95%).

b. Test-time augmentation: Test-time augmentation like rotation, brightness ±15%, and recovered 4-6% F1 on external sets.

c. Error analysis: False negatives increased 2.3 times in posterior teeth across datasets due to consistent anatomical overlap challenges.

The analysis of the classification results highlights the effectiveness of the proposed PaXNet model in detecting dental caries in panoramic X-ray images. The high accuracy, sensitivity to different caries severity levels, and robustness to image variations demonstrate the potential of PaXNet as a valuable tool for assisting dental professionals in the diagnostic process.

Discussion

The proposed PaXNet architecture, leveraging multi-scale feature fusion from ResNet50, VGG-16, DenseNet121, and InceptionV3 backbones, was rigorously benchmarked on a 500-image panoramic X-ray dataset encompassing 7,500 teeth with expert-annotated ground truth, yielding Dice-optimized segmentation accuracies of 97.8% (upper jaw) and 98.2% (lower jaw; Table *1*) - attributable to its U-Net-like decoder's precise boundary delineation via attention-gated skip connections. Cross-methodological comparisons (Table *2*) substantiate its preeminence, with superior precision (0.986), recall (0.971), F1-score (0.974), and overall accuracy (0.980) over baselines like Jader et al. [21] (0.972) and Tuzoff et al. [22] (0.976), driven by panoramic-specific domain adaptation that mitigates modality-induced variances in periapical or CBCT or intraoral data. For caries classification, PaXNet achieved 95.6% accuracy, 94.2% sensitivity, and 96.8% specificity on a stratified test set, corroborated by confusion matrices exhibiting high diagonal dominance and low false negative rates; per-tooth heatmaps revealed nuanced posterior underperformance (e.g., molars) due to superposition artifacts and anisotropic resolution, yet robust area under the receiver operating characteristic curves affirmed gradient-aware detection across caries severities, with minor recall dips for incipient lesions stemming from low-contrast radiolucencies. Ablation studies (Table *3*) validated ensemble complementarity, maximizing at 95.63% via feature orthogonality, while ImageNet transfer comparisons (Table *4*) confirmed task-specific fine-tuning efficacy (95.6% test accuracy vs. EfficientNet's 95.3%). Grad-CAM saliency maps under affine transformations (rotation, scaling, photometric shifts) evidenced translation-invariant activations on carious hypodensities, underscoring PaXNet's generalization and clinical translational potential for automated, reproducible dentistry.

Conclusions

This study presents a novel framework for automated dental caries detection in panoramic X-ray images by combining GA-based tooth segmentation with the PaXNet deep learning architecture. The segmentation method achieved high accuracy and outperformed existing techniques, while PaXNet, integrating transfer learning and capsule networks, reached 95.6% accuracy in caries classification. Robustness to image variations and clear lesion localization through Grad-CAM further support the model’s reliability and interpretability. Comparisons with ImageNet-based models highlight PaXNet’s strong and domain-specific performance, and analysis of feature extraction modules demonstrates the complementary strengths of integrated pre-trained networks. Future work will incorporate clinical data and focus on integrating the system into dental workflows. Overall, the proposed approach offers a high-accuracy, robust, and interpretable solution for assisting dental professionals in early caries detection and improving patient outcomes.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Soma Datta

Acquisition, analysis, or interpretation of data: Soma Datta, Biswajit Modak

Drafting of the manuscript: Soma Datta

Critical review of the manuscript for important intellectual content: Soma Datta, Biswajit Modak

Supervision: Biswajit Modak

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The data (and/or code) supporting this study are openly available in Panoramic Dental X-Ray Images at https://www.kaggle.com/datasets/amitabhabu/panoramic-dental-x-ray-images

References

1. World Health Organization: Global Oral Health Status Report: Towards Universal Health Coverage for Oral Health by 2030. Regional Summary of the Region of the Americas. World Health Organization, Geneva; 2023.
2. Selwitz R, Ismail A, Pitts N: Dental caries. The Lancet. 2007, 369:51-59. 10.1016/S0140-6736(07)60031-2
3. Brent Dove S: Radiographic diagnosis of dental caries. Journal of Dental Education. 2001, 65:985-990. 10.1002/j.0022-0337.2001.65.10.tb03474.x
4. Vaddi A, Parasher P, Khurana S: Beyond X-rays: Unveiling the future of dental diagnosis with dental magnetic resonance imaging. Diagnostics. 2025, 15:1153. 10.3390/diagnostics15091153
5. Izzetti R, Nisi M, Aringhieri G, Crocetti L, Graziani F, Nardi C: Basic knowledge and new advances in panoramic radiography imaging techniques: a narrative review on what dentists and radiologists should know. Applied Sciences. 2021, 11:7858. 10.3390/app11177858
6. Khanagar SB, Al-ehaideb A, Maganur PC, et al.: Developments, application, and performance of artificial intelligence in dentistry - A systematic review. Journal of Dental Sciences. 2021, 16:508-522. 10.1016/j.jds.2020.06.019
7. Schwendicke F, Golla T, Dreher M, Krois J: Convolutional neural networks for dental image diagnostics: A scoping review. Journal of Dentistry. 2019, 91:103226. 10.1016/j.jdent.2019.103226
8. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM: The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. Dentomaxillofacial Radiology. 2020, 49:20190107. 10.1259/dmfr.20190107
9. Lee J-H, Kim D-H, Jeong S-N, Choi S-H: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. Journal of Dentistry. 2018, 77:106-111. 10.1016/j.jdent.2018.07.015
10. LeCun Y, Bengio Y, Hinton GH: Deep learning. Nature. 2015, 521:436-444. 10.1038/nature14539
11. Bader J, Shugars D: A systematic review of the performance of a laser fluorescence device for detecting caries. The Journal of the American Dental Association. 2004, 135:1413-1426. 10.14219/jada.archive.2004.0051
12. Maupomé G, Pretty IA: A closer look at diagnosis in clinical dental practice: part 5. Emerging technologies for caries detection and diagnosis. Journal of the Canadian Dental Association. 2004, 70:540a-540i.
13. Zhang X, Zou J, He K, Sun J: Accelerating very deep convolutional networks for classification and detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015, 38:1943-1955. 10.1109/tpami.2015.2502579

14. Dombetzki LA: An overview over Capsule Networks. (2018). https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2018-11-1/NET-2018-11-1_12.pdf.

15. Zannah R, Bashar M, Mushfiq RB, Chakrabarty A, Hossain S, Jung YJ: Semantic segmentation on panoramic dental X-ray images using U-Net architectures. IEEE Access. 2024, 12:44598-44612. 10.1109/access.2024.3380027

16. Datta S, Chaki N, Modak B: A novel technique for dental radiographic image segmentation based on neutrosophic logic. Decision Analytics Journal. 2023, 7:100223. 10.1016/j.dajour.2023.100223

17. Barreto ML, Ichihara MY, Pescarini JM, et al.: Cohort profile: the 100 million Brazilian cohort. International Journal of Epidemiology. 2022, 51:e27-e38. 10.1093/ije/dyab213

18. Chen H, Zhang K, Lyu P, Li H, Zhang L, Wu J, Lee C-H: A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Scientific Reports. 2019, 9:1-11. 10.1038/s41598-019-40414-y

19. Zhang K, Wu J, Chen H, Lyu P: An effective teeth recognition method using label tree with cascade network structure. Computerized Medical Imaging and Graphics. 2018, 68:61-70. 10.1016/j.compmedimag.2018.07.001

20. Liu L, Xu J, Huan Y, Zou Z, Yeh S-C, Zheng L-R: A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal. IEEE Journal of Biomedical and Health Informatics. 2020, 24:898-906. 10.1109/jbhi.2019.2919916

21. Jader G, Fontineli J, Ruiz M, Abdalla K, Pithon M, Oliveira L: Deep instance segmentation of teeth in panoramic X-ray images. 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, 2018. 2018, 400-407. 10.1109/SIBGRAPI.2018.00058

22. Tuzoff DV, Tuzova LN, Bornstein MM, et al.: Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofacial Radiology. 2019, 48:20180051. 10.1259/dmfr.20180051

23. Krois J, Ekert T, Meinhold L, et al.: Deep learning for the radiographic detection of periodontal bone loss. Scientific Reports. 2019, 9:1-6. 10.1038/s41598-019-44839-3

24. Geetha V, Aprameya KS, Hinduja DM: Dental caries diagnosis in digital radiographs using back-propagation neural network. Health Information Science and Systems. 2020, 8:1-14. 10.1007/s13755-019-0096-y

25. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017, 618-626.

26. Vinayahalingam S, Xi T, Bergé S, Maal T, de Jong G: Automated detection of third molars and mandibular nerve by deep learning. Scientific Reports. 2019, 9:9007. 10.1038/s41598-019-45487-3