SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 6, 2026

MULTI-LABEL DENTAL DISEASE DIAGNOSES USING COMPUTER VISION AND DEEP LEARNING APPROACHES

Nayab Khalid Kayani[*1], Syed M. Adnan[2], Wakeel Ahmad[3], Tahira Nazir[4], Fadia Ali Khan[5]

[*1,4]*Faculty of Computing, Riphah International University, Islamabad, Pakistan*
[*1,2,3]*Department of Computer Science, University of Engineering and Technology (UET), Taxila, Pakistan*
[5]*HITEC University, Taxila, Pakistan*

[*1]nayabkayani333@gmail.com [2]syed.adnan@uettaxila.edu.pk [3]wakeel.ahmad@uettaxila.edu.pk
[4]tahira.nazir@riphah.edu.pk [5]fadia.ali@hitecuni.edu.pk

DOI: https://doi.org/10.5281/zenodo.21126237

Keywords
Dental disease, Co-occurrence, Segmentation, multi-channel mask, feature extraction, computer vision, deep learning, Densenet-121.

Article History
Received: 25 April 2026
Accepted: 04 June 2026
Published: 21 June 2026

Corresponding Author: *
Nayab Khalid Kayani

Abstract

Dental diseases like cavities, fillings, impacted teeth, and dental implants are often found in the same image, which creates a challenging multi-label classification task because of overlapping disease patterns and irrelevant background regions. Correct diagnosis of these dental diseases is crucial for better clinical examination and diagnosis. Deep learning has demonstrated success for dental image analysis in recent years. However, there are still limitations in dealing with the co-occurrence of disease and background interference, which can lead to a decrease in classification performance. This research suggests a weakly supervised disease-oriented approach to multi-label dental disease classification of panoramic dental radiographs to solve these problems. The proposed framework leverages the weakly supervised localization technique to generate multi-channel disease masks from bounding-box annotations, enabling it to locate the exact part of the disease and co-occurrence. Besides, disease specific regions are emphasized, and the disease irrelevant background regions are suppressed by blurring the background using Gaussian blur (smoothing filter), and thus the model is focused on the important diagnostic features. In addition, images were also resized to 224 x 224 resolution, which helped to maintain consistency of features and improve classification performance. The improved disease-centered images are then classified with the DenseNet121 architecture because of its capability of feature learning and efficient feature reuse. The results of the experiments prove the effectiveness of the proposed framework; training, validation and testing accuracy of 99.86%, 93.63% and 93.16% respectively are obtained. Furthermore, the proposed model obtained F1-scores of 0.78, 0.95, 0.93, and 0.93 for Cavity, Fillings, Impacted Teeth, and Implants, respectively, showing macro-average F1-score of 0.90 and a weighted-average F1-score of 0.93. The high AUC values were confirmed by ROC analysis, demonstrating good classification ability in all the disease categories. The results obtained show that the proposed weakly supervised disease-focused framework effectively boost the accuracy of multi-label dental disease classification on dental panoramic radiographs and improves the representation of disease-specific features.

I. INTRODUCTION

Oral and dental health are very important in ensuring overall human well-being. Some of the most common oral conditions that affect millions of people worldwide include dental diseases. These are cavities, fillings, implants, and impacted teeth. When these conditions are not treated, they might result in serious issues such as loss of teeth and infection of the mouth, as well as worsening of the oral structures. It has been reported that oral diseases are among the health problems that affect approximately 3.5 billion individuals in the world, thus necessitating effective diagnostic and treatment procedures. It is thus important to detect abnormalities in the teeth at an early age to prevent the progression of the disease and improve patient outcomes [1].

Dental radiography is important in the diagnosis of oral diseases. Out of the various methods of imaging, panoramic X-ray images give a clear picture of the whole oral cavity including teeth, jawbone and surrounding tissues. Dentists usually use such radiographs to reveal the cavities, the impacted teeth, the implants, and other structural deformities. But the manual interpretation may be a difficult and time-consuming task. Dental professionals experience and expertise can occasionally dictate the diagnostic outcome that can result in an inconsistency or error in clinical decision-making. Moreover, manual evaluation is getting increasingly difficult due to the growing number of radiographic images produced in the contemporary dental practice [2].

Artificial intelligence (AI) and deep learning (DL) have become the potent means of processing medical imagery and aiding automated identification of disease in recent years. CNNs have demonstrated impressive performance in the detection of more complicated patterns in medical imaging data. Such models can automatically identify relevant features of radiographic images and help clinicians identify abnormalities at a high level of accuracy. The combination of AI and dental imaging thus presents good prospects to enhance the accuracy of the diagnosis, to minimize human error, and to increase the level of clinical performance [3].

Various investigations have been conducted to determine the use of deep learning models in analyzing dental images. Indicatively, CNN-based methods have been effectively used to identify dental caries, implants, and different tooth conditions using radiographs. Advanced networks have shown good performance in dental disease detection tasks including ResNet, VGG, EfficientNet, and DenseNet. Certain studies have indicated that the classification accuracy of automated diagnostic systems were above 90 % in the detection of dental diseases from panoramic or periapical radiographs. In this study four distinct Xray classes used for classification are illustrated in figure 1 [4].

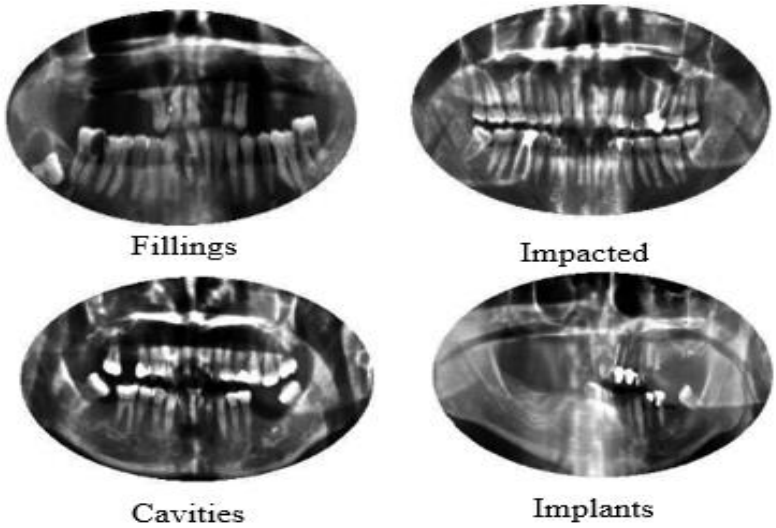


Figure 1: Dental Disease Classification Images

Even in the wake of these developments, there are still several issues in the development of credible

automated dental diagnostic systems. Most current methods give special attention to binary classification methods, thus making them less realistic about clinical conditions in real practice because more than two dental conditions can be present. Moreover, dental structures are complicated, anatomical areas may overlap, and radiographic quality can differ, which may complicate accurate feature extraction. Also, small, labeled datasets and labelling discrepancies may compromise the generalization ability of deep learning models [5].

The primary technical issues that should be tackled during the implementation of computer vision and machine learning methods to identify dental diseases automatically are as follows:

a. **Multi-label complexity**: A single radiograph may contain several dental diseases as shown in Figure 2, making the classification task more complex than traditional single-label disease classification [1, 4].
b. **Image Quality and Noise:** Dental X-ray images are prone to noise, low contrast, and overlapping anatomy, which could impact the performance of automated analysis [6].
c. **Feature Extraction:** Accurately identifying dental diseases requires capturing both local lesions
d. features and global structural relationships within panoramic radiographs [7].
e. **Data Imbalance**: Dental datasets often contain an uneven distribution of disease classes, where some conditions appear frequently while others are rare. This imbalance can negatively affect model performances and generalization [8].
f. **Data Annotation:** Deep learning models need large datasets of labeled data to be trained, but dental image annotation can need experts and may vary between different experts [9].
g. **Model Generalization:** It is not always the case that models that are trained on a certain dataset would generalize well when applied to an image that is taken by another imaging device or under a new clinical setting [10].
h. **Computational Efficiency:** To be practically useful in the clinic, the automated systems should process images in a fast fashion without affecting the diagnostic quality [11].
i. **Clinical Reliability**: It is essential to make sure that AI-based diagnostic systems can deliver reliable and understandable predictions that can be used in practice in actual clinical settings [12].
j. **Dataset limitation**: Many studies rely on relatively small datasets, which limits the generalization ability of deep learning models and may lead to overfitting [13].

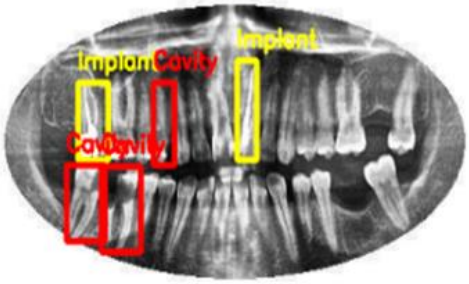


Figure 2: Co-occurrence of Dental Disease Classification Images

To counter the issue of multiple disease presence in a single tooth image shown in figure 2, this study will suggest a smart deep learning-based model to classify panoramic dental X-ray images. The system proposed is aimed at detecting four typical dental conditions, namely, cavities, fillings, implants, and impacted teeth. The proposed model will achieve high precision and reliability of dental disease classification using advanced convolutional neural networks and machine learning methods.

- Develop an AI-based framework for analyzing multiple dental conditions from panoramic radiographs.
- Improve robustness in the presence of data imbalance.
- Enhance diagnostic reliability by capturing both local and global image features.

The following are the key contributions to our proposed multi-label dental disease:

1. Using weakly supervised segmentation to generate multi-channel masks from bounding-box annotations.
2. Using Gaussian blurring to suppress background and preserve disease regions to create disease-focused images.
3. Using controlled geometric data augmentation to handle class imbalance.
4. Using DenseNet121 architecture to classify dental diseases from panoramic radiographs using multiple labels.

With such contributions, the suggested strategy would help dentists to diagnose dental diseases faster and more reliably, which would later enhance clinical decision-making and care for patients.

The remaining portions of the document are organized as follows:

Section II presents examples from the relevant literature, whereas Section III presents the details about Dataset. Section IV presents preprocessing steps. Section V concentrates on the suggested methodology. Section VI presents the results and comparison with the most recent methods. Section VII provides the conclusion.

II. RELATED LITERATURE

The history of automated dental radiography analysis has gradually switched to multi-label disease classification which resembles the clinical situation on the ground where a patient would come with several related conditions and not just with a single identical pathology. There was a tremendous transformation in the field of dental diagnostics in 2025 to multi-label classification and highly precise architectural benchmarking.

In this study, hybrid frameworks integrate entity based segmentation with Random Forest classification pipelines to achieve 85.4% accuracy on 3576 samples and still have limitations of morphological similarities between pathologies [1].According to the need to possess some form of integrated data, researchers further created the PDIAD database that comprised of 1,269 images, where the VGG19 model was accurate at 85.75 percent in identifying co-occurring dental conditions [2].

The 3,753 intra-oral photos were applied to a Vision Transformer (ViT-Base/16) to reach a high rate of accuracy of 98.94 percent through the help of explainable AI tools [3].The development of attention-enhanced CNN using a Random Forest model marked the beginning of the line of development. This multi-label model classifier classifies four different dental conditions: fillings, cavities, implants, and affected teeth using 4,764 augmented images with a precision of 90.6% [4]. A hierarchical framework with DenseNet121 and EfficientNet-B4 was suggested in which the label of healthy teeth versus amalgam-filled teeth was reduced to 98.28 percent accuracy by the two-step hierarchical method, but a limitation is the small size of the dataset (402 images) [5].This was later in the year followed by attention-driven architectures, and a custom bounding-box masking combination to further optimize multi-label pipeline, with a total peak of 90.6 percent accuracy [6].However, the constant imbalance in classes of datasets like PDIAD highlighted the reality that despite the interruption of the models, there remains low recall in regard to detecting cavities [7].

To combat this, Vision Transformers were further appended with a tri-modal explainability package Attention roll-out and LIME, to ensure that the accuracy of 98.94 percent it achieved was realistic to the practitioners and understandable [8].Despite the high metrics that ensue, comparative analyses of balanced panoramic datasets substantiated the notion that AI should be an augmentative tool because morphological similarities may tend to result in misclassification in many cases [9].

This too necessitated more specific hierarchical structures that were able to cluster visually similar classes in a strategic fashion and which, despite providing a 98.28 precision, nevertheless would require larger datasets to test in the real world [10]. Additional expansion of the technical scope of 2025 was obtained through the real-time clinical deployment tools. They have also developed a lightweight 5-layered CNN that was employed in the classification of 10,573 images in to a number

of classes with an accuracy of 87.6 percent; this offers a rapid response of 120ms to make an inference on web interfaces in the cloud [11].

Concerning the state-of-the-art detector, the initial version of the YOLO v11 architecture was utilized to find the jaw cysts with an average precision of 86 percent at the cost of the single-center performance [12].The two-stage pipeline combined with YOLOv5 and Attention U-Net increased the accuracy to 93%when trained on 500 radiographs compared to the gold-standard bitewing images [13]

Thereafter, the large-scale benchmarking on three publicly available databases demonstrated that ResNet50V2 can achieve an 82.5 percent accuracy when used to process damaged images with small features and imbalance between classes being a bottleneck [14]. Finally, authors tested a big pool of 29,815 images and a fine-tuned DenseNet201 to achieve an accuracy of 91%, yet they also concluded that the level of architectural depth is not enough to represent the bad performance of underrepresented pathologies like cavities [15].

Recent progress in 2024 focused on addressing specific dental disease conditions while improving computational efficiency. The study was an advanced investigation on the level of involvement of furcation based on 1,568 molar images, the point where the study commenced.

The researchers achieved an accuracy of 92% by predicting with a Vision Transformer (ViT) model trained using 3D CBCT standard references and could capture long-range dependencies in the anatomy, which are commonly missed by conventional networks, although it is computationally expensive [16].

The development of such complicated architectures continued with a two-stream architecture that integrated MobileNetV2 and Swin Transformers to detect different types of tooth diseases using 4,023 images with an accuracy rate of 95.6% [17].Contemporarily, other scientists have stopped attending to clinical X-rays and are exploring photographic options that are available to consumers, where ResNet50 is being trained on 3,392 natural photographs and still, the uneven illumination remains a major limitation of the approach [18].

In 2024, technological innovation was also devoted to the preprocessing of the data and the extraction of certain features. The customized 5-block CNN proposed in one of the studies using SMOTE TOMEK oversampling called PDCNET reached the highest accuracy of 98.39% by effectively cleaning and balancing panoramic data [19].

The other team focused on intraoral X-rays and, rather than traditional pooling, was a new 86.4% correct "CVApool" layer constructed on the principal Component Analysis to a greater number of discriminating features among 20 clinical classes [20].Automation was also furthered to another level in which the three-stage pipeline of VGG16, YOLOv7, and ResNet50, which is capable of recognizing and classifying molar impaction at 92.17 percent without involving the human element [21].

The year was perhaps the most distinctive approach a cross-modal diagnostic model using ChatGPT on 29 pediatric radiographs to chop the pictures into text to be categorized by NLP, which yielded 84 percent outcomes, although the sample size is so small it is highly likely to experience overfitting [22].Multi-label classification of the already co-occurring conditions became one of the main topics of the course towards the end of the year. Researchers used certain bounding-box masking and attention models on 4,764 images and attained a 90.6%accuracy, and this limitation is because healthy images are missing to make comparisons [23].

Similarly, the fusion of different datasets was used to overcome the fact of clinical reality, and VGG19 was able to achieve 85.75 percent accuracy in a multi-label format, but it remained with bad cavity recall [24].

Intra-oral RGB images ensured that the model used a pure Vision Transformer with a tri-modal explainability with 98.948% accuracy to ensure that the decision taken by the model was comprehensible to the clinicians [25].Elsewhere, entity based segmentation was combined with hybrid networks to provide the spatial localization of pathologies with 85.4 percent accuracy when morphological similarities of conditions like cavities and implants also resulted in misclassifications being reported [26].

Final advances achieved in 2024 solved specialized

combinations of diseases and resource efficiency. The hybrid pipeline, which was premised on the integration of eight trained CNNs and XGBoost classifiers, was trained to detect up to 46 combinations of diseases with a certain level of accuracy of up to 93.2% [27].

With heavy networks, a 5-layer CNN was reduced to the level of automation needed to evaluate the quality of constrained datasets, and achieved a 97.87 percent accuracy with only 126 images [28].In conclusion of the marvels of the year, the complete benchmarking of the EfficientNet family along with Borderline SMOTE was successfully completed to 98.32 percent accuracy with 1,317 images, and EfficientNet-B5 is the optimal architectural tradeoff in the field of dental radiography [29].

It concentrated on the field of architecture and the enhancement of diagnostic transparency in order to manage the complex pathologies of the teeth by 2023.

In this study, the researchers had a distinctive two-step hierarchical design that reduced the feature interference in case of visually similar conditions. Researchers also identified the caries with 83.95% accuracy and amalgam-filled teeth with 98.28% accuracy by processing 402 panoramic images with DenseNet121 and EfficientNet-B4, respectively [30]. Although this technique was effective in reducing the overlap of classes, it showed a narrow generalizability due to the small sample size [30].

At the same time, another 2023 study was more focused on interpretability and based on fine-tuned MobileNetV2 and 7506 images and six clinical classes. This model achieved 89.02% accuracy and included visual justifications with Explainable AI (XAI) algorithms, such as LIME and Grad-CAM, but could not differentiate between cases of gingivitis and calculus that are visually similar [31].

In 2022, it was determined by a concern of model resilience on realistic imaging constraints. In this research, it was observed that scientists used about 3,383 pictures to carry out binary classification, knowingly exposing models to stressful conditions. The highest accuracy of 82.5% was obtained by the study with the help of testing the architectures such as the ResNet50V2 with noise-reduced and blurred images. It is a comparative stress-testing that offered a critical novelty, as it simulated low-resolution clinical conditions; and finally, the main bottlenecks to a trustworthy dental AI were found to be inadequate dataset sizes and inconsistent image resolutions [32].

Although there have been significant advances in the analysis of dental images using hybrid frameworks, transformer models, ensemble approaches, and augmentation methods, the efficient way to classify panoramic radiographs into multiple classes is still difficult. Despite the high accuracy provided by previous research, most of them are limited by:

1. Limited or proprietary datasets with limited external validity.
2. The absence of features to address the issue of class imbalance.
3. The absence of instructed features learning to remove interference between visually similar dental pathologies.

Consequently, classification models can tend not to be robust and reliable in a real clinical environment.

Our proposed work overcomes these limitations by presenting an automated multiclass dental disease diagnosis system based on DenseNet121 and panoramic X-rays.

- The foundation of the extraction of the strong and complex attributes based on the dental X-ray images was a DenseNet121 network.
- The multi-channel masks is produced using the weakly supervised bounding-box label segmentation because it does not need pixel-based assignments.
- To help the model to concentrate on the key areas, the disease specific images were formed by blurring the background, however maintaining the disease areas.
- The data augmentation is conducted on the training set with rotation, flipping, scaling, shearing, and translation to equalize classes and make classes less generalized.
- The suggested method maintains the simplicity and efficiency of annotations of the system and enhances the classification

A summary of techniques for the classification of multi-label dental disease images is shown in Table 1.

TABLE 1. Classification accuracies of existing techniques for the classification of multi-label dental conditions.

#	Year	Dataset	Technique	Classifier	Acc%
1	2026	VZRAD2(8429 images) **(multi-label)**	EffiecientNet-B4+CBAM	Fully Connected layer (11 outputs) + Sigmoid	85.67
2	2025	1512 panoramic dental Xray **(multi-label)**	CNN(VGG16)	RF	85.4
3	2025	PDIAD **(multi-label)**	CNN(VGG19)	CNN(transfer learning)	85.75
4	2025	1512 Panoramic Radiographs **(multi-label)**	CNN + Attention	RF	90.6
5	2025	9573 panoramix Xray 19 classes **(multi label)**	CNN/Transformer Features (RT-DETR)	YOLO,RT-DETR	65.4 63.0
6	2023	1067 dental images **(multi class)**	CNN with attention+SE+ Resedual+ASPP	Fully connected + Softmax	92.76
7	2022	500 panoramic dental Xray **(multi label)**	Transfer learning ensemble(CNN's)	XGBoost	92-93

As shown in the literature presented in Table 1, the recent literature has widely examined the techniques of deep learning in panoramic dental X-ray analysis. Other work like [1]and [2]have used CNN based feature extraction based on VGG networks, with moderate results using Random Forest and transfer learning methods. Random forest and CNN[4] achieve better results through the consideration of attention mechanisms, whereas hybrid and transformer-based models [14] In dealing with more complex multi-class cases with significantly reduced accuracy. It is worth noting that the ensemble and advanced architectures of [31], [32] show better performance, with an accuracy of over 92, which proves that the results of using deep learning and ensemble strategies can be considerably improved.

III. DATASET

In this study, a panoramic dental X-ray dataset was obtained from Roboflow [1, 4] is utilized. The dataset consists of 1,512 panoramic dental radiographs, each with a spatial resolution of 512 × 256 pixels. Every image was carefully annotated by experts for four dental conditions: Fillings, Cavities, Dental Implants, and Impacted Teeth. In total, the dataset contains 11,137 bounding-box annotations.

TABLE 2. Dataset splitting into test, train and valid

Dataset	Folders	No of images	Annotations
Panoramic Xray Radiographs (disease.v5i.coco)	Train Valid Test	1318 121 73	9884 780 473
Total		**1512**	**11137**

The dataset is divided into three subsets: training, validation, and testing. A smaller portion of images was reserved for validation and testing, while the majority were used for training. The distribution of images and annotations is summarized in Table 2.

TABLE 3. Class-wise Imbalance Statistics of training dataset

Dataset	Classes	Imbalance images
Panoramic Xray Radiographs (disease.v5i.coco)	Fillings Cavities Implants Impacted	1037 473 495 727
Total		**2732**

After reviewing table 3, it is detected that there is a huge class imbalance in the Training dataset.

IV. PREPROCESSING

A. GRAYSCALE CONVERSION:

Dental X-ray images naturally contain grayscale information; however, the images in the dataset were stored in JPG format with three channels (H × W × 3). Since these channels carried the same intensity information, they introduced unnecessary redundancy. To address this, all images were converted to a single-channel grayscale format (H × W × 1). This reduced the data size by approximately 66.6% while preserving all clinically relevant features, resulting in lower GPU memory usage and more efficient model training.

$$I(x,y) = 0.299R(x,y) + 0.587G(x,y) + 0.114B(x,y) \quad (1)$$

Where:

- $I(x,y)$: resulting grayscale intensity
- $R(x,y), G(x,y), B(x,y)$: Red, Green, blue channel intensities at pixel

This is used to reduce image complexity and focus on intensity based features instead of color variations.

B. ADJUSTMENT OF BRIGHTNESS

A straightforward linear transformation was used to modify brightness and contrast to enhance the original dental X-ray images visual quality. The following formula was used to adjust each pixel's intensity:

$$I'(x, y) = \alpha \cdot I'(x,y) = \alpha \cdot I(x,y) + \beta \quad (2)$$

where $\alpha\beta$ modifies the brightness and contrast α regulates the contrast (gain), β controls brightness and $I'(x, y)$ gives enhanced image. The values used in this investigation were α = 15 β = 15 and α = 1.5 α = 1.5.

C. Contrast Enhancement using CLAHE:

To enhance local contrast without over-amplification of noise, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used. Histogram equalization is carried out locally using a contrast-limiting threshold τ once the image is broken into small tiles:

$$H_{clip}(i) = (H(i), \tau) \quad (3)$$

Where:

- H(i): Histogram value at intensity level i
- τ: Threshold value

Extra histogram counts are equitably dispersed. This preserves radiographic realism while enhancing minor disease patterns like cavities and fillings, as shown in figure 3.

D. Weakly Supervised Segmentation

a) Multi-Channel Mask Generation

Bounding-box labels were used for weakly supervised segmentation because pixel-level annotations were not available.

$$M(x{:}\, x + w,\ y{:}\, y + h,\ c) = 1 \quad (4)$$

A multi-channel mask $M \in R^{H \times W \times C}$ was created for every image of size H×W where:

- W, H: width and height of the region.
- M is the mask region in the image
- C represents the number of disease classes.

For bound boxes all remaining pixels were set to zero. This approach provides approximate spatial localization of disease regions with minimal annotation overhead. The generated masks were stored in. npy format for efficient loading during training.

(b) Disease-Focused Image Generation

In order to highlight areas specific to a class, disease-focused images were produced. The focused image I_f was calculated using a grayscale image I and a disease mask D_c.

$$I_f = I \cdot D_c + G(I) \cdot (1 - D_c) \quad (5)$$

where G(I) denotes an image that has been blurred using Gaussian. The model's focus is directed toward clinically significant regions by this operation, which maintains disease regions while suppressing extraneous background information. Images pertaining to diseases were arranged into folders according to class as shown in Table 4.

E. Data Augmentation

Only the training set was subjected to data augmentation using gentle geometric changes in order to reduce class imbalance and enhance model generalization: Rotation (±10°), Flipping horizontally, Translation on both axes (±7%),Scaling (±10%), Shearing(±5°). Let a stochastic augmentation operator be represented by

T(·):

$$I_{aug} = T(I) \quad (6)$$

Where:

- I_{aug} : $Augmented\ Image$
- $T(I)$: $Transformation\ function$

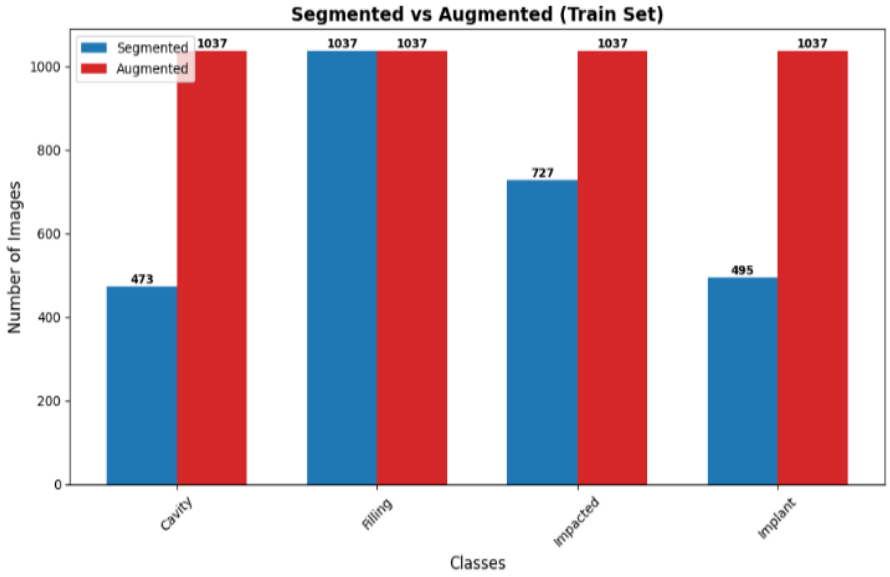
To maintain anatomical plausibility and ensure balanced class representation, each class was increased until a predetermined goal sample count was reached.

TABLE 4. Class-wise Balance Statistics of the training dataset

Images	Fillings	Cavities	Implants	Impacted	Total
Segmented	1037	473	495	727	**2732**
Augmented	0	564	542	310	**1416**
Total	1037	1037	1037	1037	**4148**

After applying augmentation data is now balanced in train dataset class wise. Each folder contains 1037 images shown in Table 4.

Figure 3. Segmented and augmented train set dental disease classification visual representation.



The difference between segmented and augmented images are shown in figure 3. Now train dataset is fully balanced and ready for training.

F. Image Resizing:

Area-based interpolation was used to resize each image to a fixed resolution of 224 × 224 pixels:

$$I_{resize} = Resize(I_{aug}, 224{,}224) \quad (7)$$

Lastly, the range [1] was used to normalize the pixel intensities:

$$I_{final} = \frac{I_{resize}}{255} \quad (8)$$

This normalization converts pixel values from [0,255] to [0,1] enhances deep learning model convergence and stabilizes training.

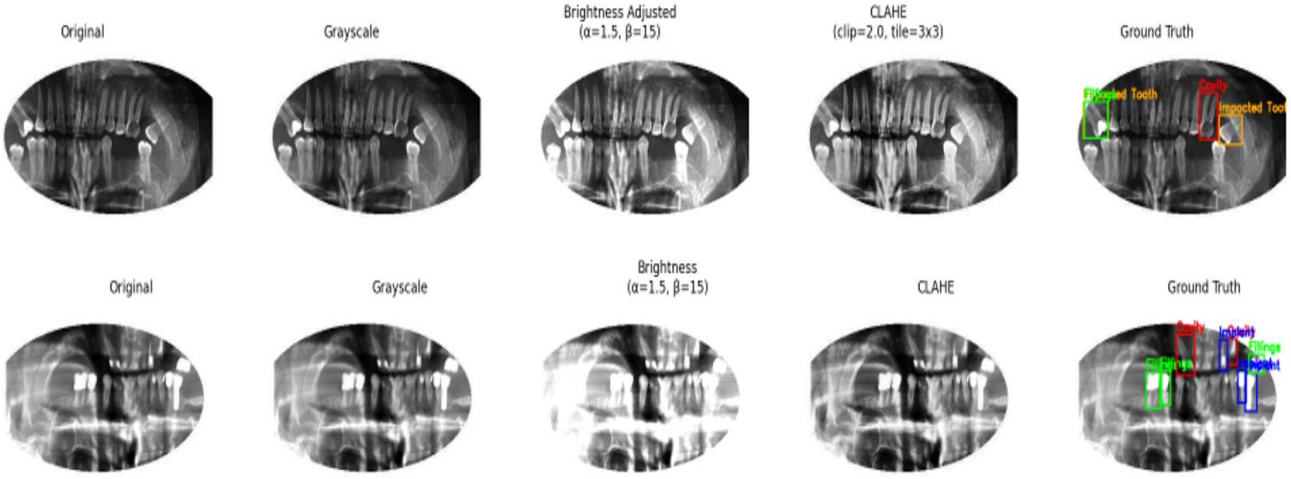


Figure 4. Preprocessing steps and ground-truth bounding box visualization of dental disease classes.

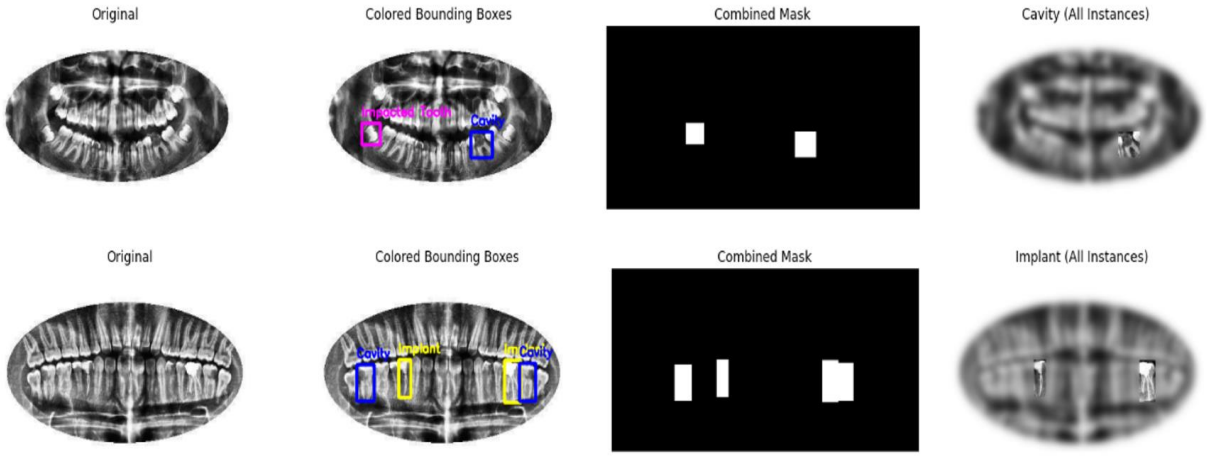


Figure 5. Multi-Channel Mask Generation and disease focused on all instances of dental disease classes.

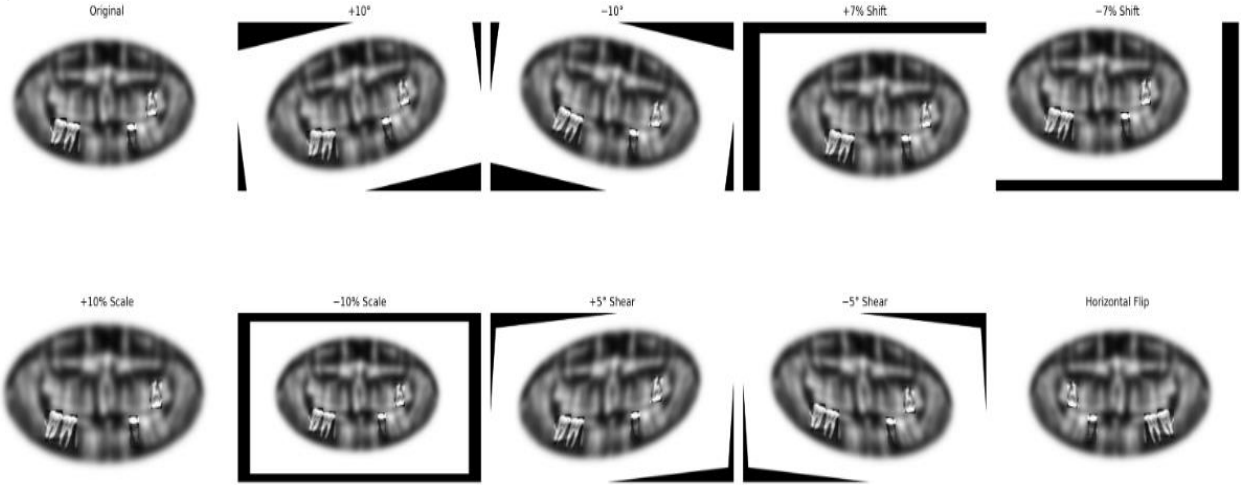


Figure 6. Augmentation version of multilabel dental disease classification.

Figure 7. Resized and normalized dental disease visualization of four classes

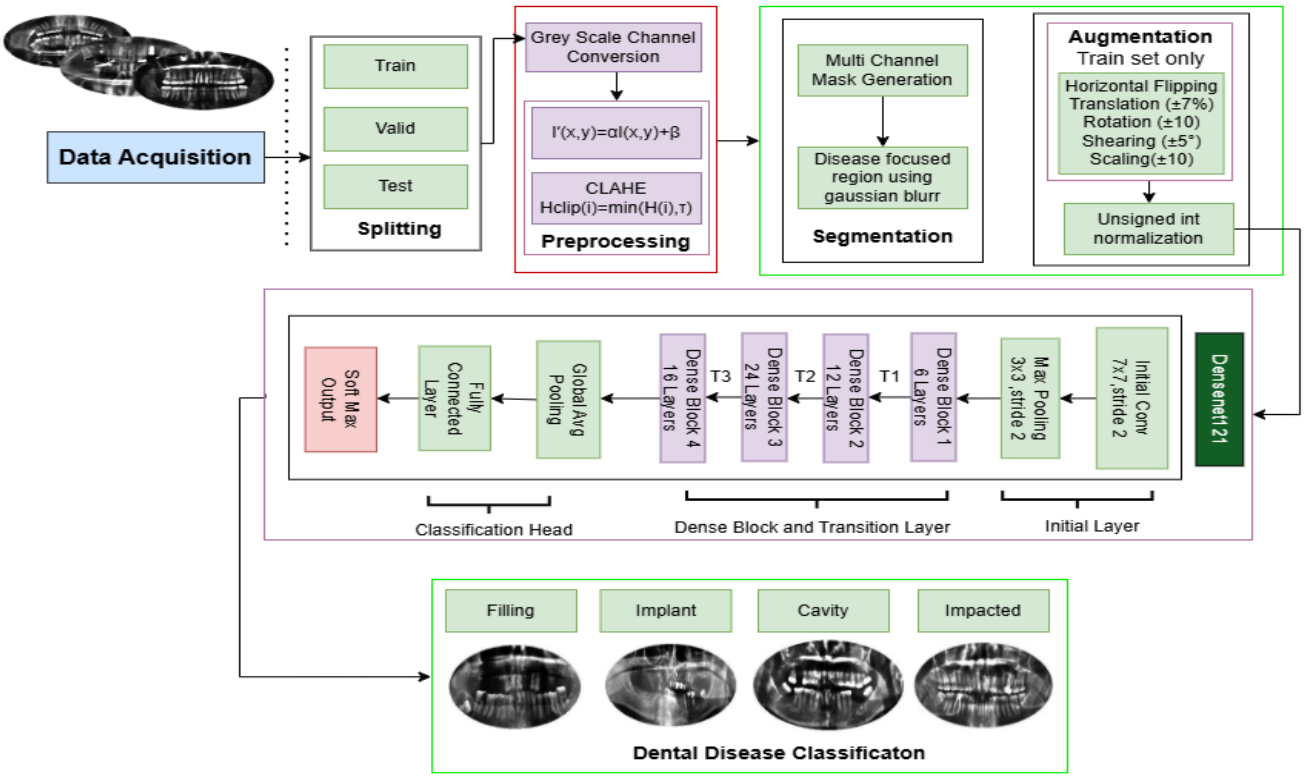


Figure 8. Methodology Diagram of dental disease classification

V. PROPOSED METHODOLOGY

The suggested deep learning framework for the multi-class categorization of dental abnormalities using panoramic radiographs is presented in this section. Data preparation, network architecture adaptation, training strategy, optimization, and evaluation are all included in the methodology. The DenseNet121 design serves as the foundation for the suggested model, which has been modified for the four-class classification of dental diseases, as shown in figure 8.

A. Data preparation and preprocessing

Four diseases of panoramic dental radiographs cavity, filling, impacted, and implant, were separated into subsets for testing, validation, and training, and saved in the npy format. To meet DenseNet121's input requirements, the grayscale radiographs were transformed into a three-channel format. To speed up convergence

and stabilize training, input images were normalized as follows:

$$\hat{x} = \frac{x-\mu}{\sigma} \quad (9)$$

x stands for the original pixel intensity, μ=0.5 shows mean of the dataset, and σ=0.5 is the standard deviation of the dataset.

$\hat{x}$ is the normalized value. Controlled variance and a zero-centered input distribution are guaranteed by this normalization.

B. Network Architecture

DenseNet121 serves as the backbone feature extractor in the suggested framework. For each layer to get feature maps from every layer that came before it within the same dense block, DenseNet provides dense connection. For a dense block with layers, the output of the l^{th} layer is defined as:

$$x_l = H_l([x_0, x_1, x_2, \ldots, x_{l-1}]) \quad (10)$$

- where:
- $x_l$denotes the output feature map of the l^{th}layer,
- $H_l(\cdot)$represents a composite function consisting of Batch Normalization (BN), ReLU activation, and convolution
- [.]indicates concatenation of feature maps
- This extensive connectedness lessens redundancy, promotes feature reuse, and enhances gradient flow.

1)Dense Blocks

A bottleneck structure made up of the following is applied by each dense layer:

- Channel reduction, 1x1 convolution
- extraction of features,3x3 convolution

2) Layers of Transition

Transition layers are used between thick blocks to minimize spatial dimensions:

$$x' = AvgPool(Conv_{1\times1}(x)) \quad (11)$$

This operation prevents excessive computational growth and controls model complexity.

3) Modified Classification Head

A fully connected layer with four output neurons was used in place of the original DenseNet121 classifier, which had 1000 neurons:

$$z = Wx + b \quad (12)$$

where $W \in R^{4\times d}$, $b \in R^4$, and $d$is the number of extracted features.

The SoftMax function is used to calculate the anticipated class probabilities:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{4} e^{z_j}} \quad (13)$$

C. Loss Function

The Cross-Entropy Loss was used to maximize the performance of multi-class classification:

$$L_{CE} = -\sum_{i=1}^{4} y_i log(\hat{y}_i) \quad (14)$$

where:

- $y_i$is the ground-truth label (one-hot encoded),
- $\hat{y}_i$is the predicted probability for class i
- The training's goal is to reduce L_{CE}*across the dataset.*

D. Optimization Strategy

For parameter updates, the Adam optimizer was used. The definition of the parameter update rule is:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (15)$$

where:

- $\theta_t$represents model parameters at iteration t,
- $\eta$is the learning rate
- $\hat{m}_t$and $\hat{v}_t$are bias corrected first and second moment estimates,
- $\epsilon$is a small constant for numerical stability.

E. Hyperparameter Configuration

The hyperparameters indicated in Table 5 were used to train the model. The final optimized model was the one that achieved the maximum validation accuracy.

Table 5: Hyperparameter Configuration

Parameter	Value
Backbone Model	DenseNet121
Input Resolution	(224×224)
Batch Size	16
Learning Rate	0.001
Epochs	30
Optimizer	Adam
Loss Function	Cross-Entropy

Algorithm 1 Algorithm of Multi-label dental disease diagnoses and Classification

Input: X-ray image of the panoramic dental image.
Output: Class of dental disease (e.g., Caries, Fillings, Implant, Impacted).

1. Read the image.
2. Preprocess the image.
3. Generate weakly supervised bounding-box masks.
4. Turn masks into multi-channel.
5. Use weakly supervised segmentation to point out disease areas.
6. Disease-related images were created by applying gaussian blur to the background.
7. Use structured data augmentation (rotation, flipping, scaling, shearing, translation).
8. Get features with DenseNet121.
9. Learn discriminative features using dense connections and localization.
10. Use classification to predict dental disease in a multi-class setting.

VI. EXPERIMENTATION METHODS AND RESULTS

An unseen test dataset was used to assess the trained model. The total accuracy was calculated as follows:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Samples} \quad (16)$$

Furthermore, each class's F1-score, Precision, and Recall were determined:

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

$$F1-Score = 2 \cdot \frac{Precision \cdot Recall}{Precsion+Recall} \quad (19)$$

where True Positives (TP), False Positives (FP), and False Negatives (FN) are denoted, correspondingly.
Additionally, inter-class misclassification behavior was examined using a confusion matrix.
We have performed all the experiments in Python by using Google Colab, while the images are uploaded to Google Drive, divided into four classes: Cavity, fillings, impacted and implant, as mentioned earlier. Precision, Recall, Accuracy, and F1 Score are calculated for VitBase_16, CNN, VIT, Resnet18, DenseNet121, EfficientNet-B4,

MobileNetV2, UNET as shown in Table 6.

Table 6: Performance metrics summary of dental disease classification using different model

#	Method	Classes	Precision	Recall	F1-score
1	**Vit-Base 16**	Cavity	0.72	0.78	0.75
		Fillings	0.88	0.87	0.87
		Impacted	0.96	0.92	0.92
		Implant	0.81	0.81	0.81
		Macro Avg	0.84	0.85	0.84
		Weighted AVG	0.85	0.85	0.85
2	**EfficientNet-B4**	Cavity	0.72	0.66	0.69
		Fillings	0.93	0.94	0.93
		Impacted	0.93	0.96	0.95
		Implant	0.87	0.82	0.84
		Macro Avg	0.86	0.85	0.85
		Weighted AVG	0.90	0.90	0.90
3	**UNET + Finetune**	Cavity	0.70	0.58	0.64
		Fillings	0.93	0.89	0.91
		Impacted	0.84	0.94	0.89
		Implant	0.94	1.00	0.97
		Macro Avg	0.85	0.85	0.85
		Weighted AVG	0.89	0.90	0.89
4	**Densenet121**	**Cavity**	0.81	0.75	0.78
		Fillings	0.94	0.96	0.95
		Impacted	0.90	0.96	0.93
		Implant	1.00	0.88	0.93
		Macro Avg	0.91	0.89	0.90
		Weighted AVG	0.93	0.93	0.93

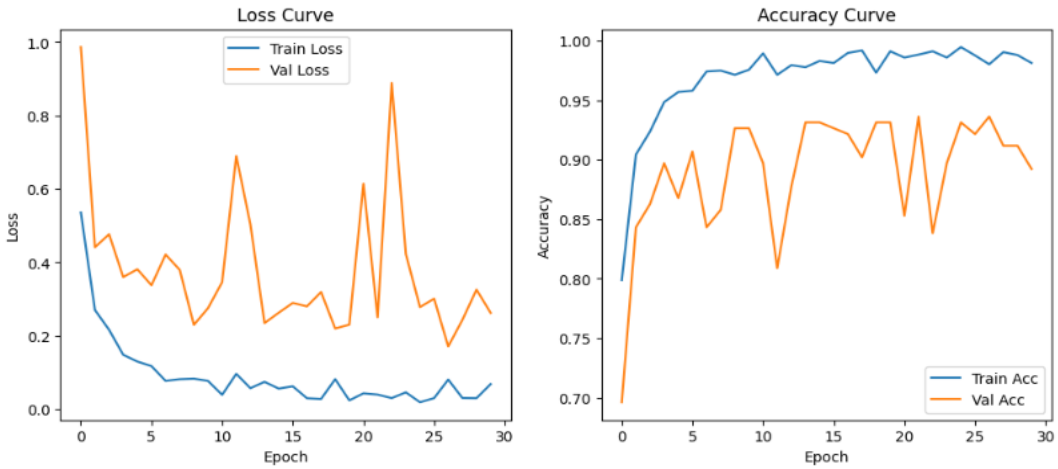


Figure 9: Training and validation loss curves using densenet121

DenseNet121 demonstrated balanced precision and recall across all classes, achieving the best overall class-wise performance with a weighted F1-score of 0.93 and a macro F1-score of 0.90. It performed very well on Impacted. (F0.93) and Fillings (F1 = 0.95), with perfect precision (1.00)

on Implant. While U-Net with fine-tuning demonstrated relatively poorer performance, particularly in the Cavity class, EfficientNet-B4 and ViT-Base_16 obtained moderate performance.

Table 7: Model results for dental disease classification

#	Method	Train	Val	Test
1	Vit-Base_16	96.31%	84.80%	84.62%
2	CNN+VIT	87.62%	73.04%	82.05%
3	Resnet18+vit	87.17	86.27%	8.18%
4	EfficientNet-B4	99.61%	95.10%	90.60%
5	UNET + Finetune	93.54%	87.75%	89.74%
6	**Proposed**	**99.86%**	**93.63%**	**93.16%**

EfficientNet-B4 demonstrated less generalization than DenseNet121, with slightly greater validation accuracy but lower test accuracy. Strong test results were obtained by the Although U-Net and transformer-based models performed competitively, none of them consistently outperformed DenseNet121 in table 7.

DenseNet121 was the best option for the suggested framework since it showed the most consistent and balanced performance across all datasets. Effective learning is indicated by the loss curve, which demonstrates that training loss drops gradually and reaches a very low value.

Validation loss changes in subsequent epochs, indicating slight instability, even though it initially diminishes. According to the accuracy curve, training accuracy hovers around 99%, whereas validation accuracy fluctuates slightly between 90 and 93%. With just slight overfitting, the model performs well overall and converges nicely shown in figure 9.

The DenseNet121 model achieved good classification performance for all classes of dental diseases in the ROC analysis. The AUC values of the model were 0.9532, 0.9696, 0.9985 and 0.9965 for Cavity, Filling, Impacted, and Implant classes respectively, which showed that the model had good discriminative performance shown in figure 10.

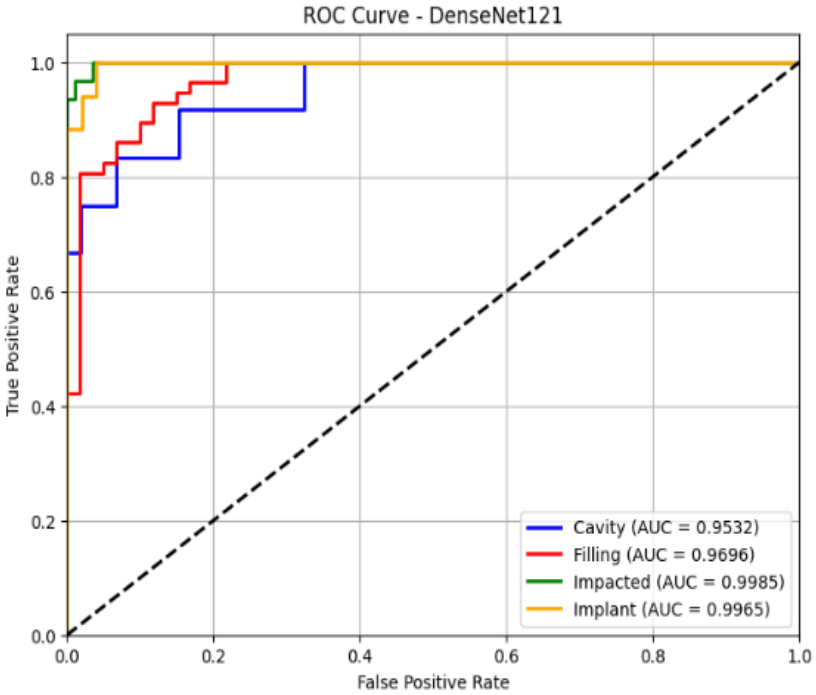


Figure 10: ROC Curve of DenseNet121 for Multi-Class Dental Disease Classification

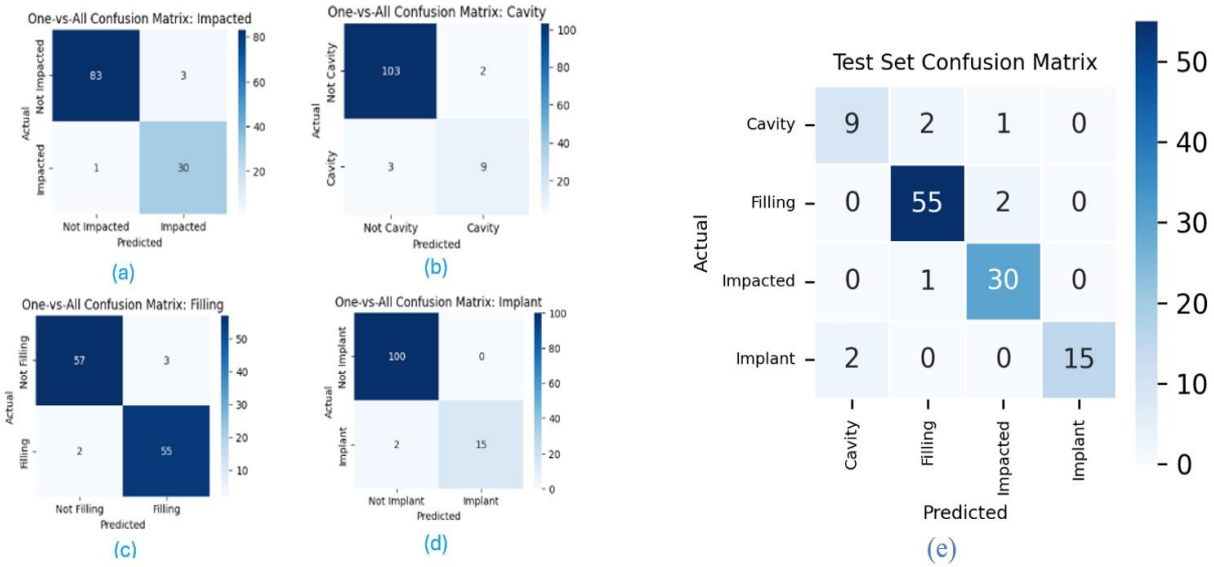


Figure 11: Class wise confusion matrix (a)Impacted (b)cavity (c)Filling (d)Implant (e)Test set overall confusion matrix

The confusion matrices indicate the effectiveness of the model in predicting various dental diseases in terms of correct and incorrect predictions. In the one-vs-all, the predictions are mostly correct, and this indicates that the model is effectively detecting instances of a disease (true positives) and also detecting when there is none (true negatives). Very few cases are misclassified, with either the model labeling a disease incorrectly (false positives) or other cases (false negatives) being a genuine disease. In the case of Impacted and Implant, the model has very high performance with high number of correct predictions and low number of errors, which means that it has a high number of true positives and true negatives. The results in the Filling class are also balanced, and there is not much misclassification as shown in figure 11.

Table 8: Comparison with different existing studies

Ref	Dataset	Tech	Acc%
[1]	Panoramic dental Xray	CNN(VGG16)	85.4
[2]	PDIAD	CNN(VGG19)	85.75
[4]	Panoramic Radiographs	CNN + Attention	90.6
[14]	Panoramic Xray	CNN/Transformer Features (RT- DETR)	65.4
[31]	Dental images	CNN with attention+SE+ Resedual+ASPP	92.76
Proposed	**Panoramic dental Xray**	**Weakly supervised Segmentation,Gaussian BlurrDensenet 121**	**93.16**

In the case of Cavity, the number of wrong predictions, i.e. false positives and false negatives, is a bit higher than other classes, which makes it a bit harder.

These findings are supported by the overall multi-class confusion matrix, where most of the values are on the diagonal which indicates that each class is correctly predicted. The small number of off-diagonal values signifies misclassifications of similar classes. In general, the model is highly accurate with a good capacity to accurately identify diseases and low misdiagnoses, as shown in figure 11.

Prior works on multi-label panoramic dental X-ray datasets had moderate to high accuracy, such as VGG16 [1] with 85.4%, and VGG19 [2] with 85.75% as shown in table 8. The performance was further enhanced with the addition of the attention mechanism [4] to 90.6%, and unexpectedly reduced by the hybrid CNN-Transformer (RT-DETR) [14] to 65.4%, probably because of global features being overemphasized. To the best of our knowledge, the state-of-the-art results from previous works is a multi-layered CNN incorporating attention, SE blocks, residual connections and ASPP [31] with a result of 92.76%. Our proposed approach, which is based on weakly-supervised segmentation, Gaussian blur pre-processing and a DenseNet-121 backbone, is able to reach 93.16% accuracy, outperforming all previous approaches with its simpler design. This shows that the use of the efficient feature reuse along with mild regularization can outperform more complex architectures in the case of multi-label dental X-ray analysis.

VII. CONCLUSION

We proposed a novel multi-label dental disease diagnosis framework designed for the timely detection of co-occurrence in single teeth using dental X-ray images rather than requiring pixel-level annotated data that needs specialized effort. The suggested approach uses the features obtained based on a DenseNet121 network, as well as weakly supervised segmentation and disease-oriented learning, and attains the training accuracy of 99.86% and validation accuracy of 93.63% and test accuracy of 93.16%, respectively. In future practice, we can use our framework in other forms of dental imaging modalities, too. Since our proposed method is based on strong feature extraction and incorporates region-focused learning, this innovative combination can potentially be generalized to address periodontal diseases.

STATEMENTS AND DECLARATIONS

- **COMPETING INTERESTS**

There are no conflicts of interest to declare.

- **ETHICAL AND INFORMED CONSENT FOR DATA USED**

Ethical Considerations: A 1512 panoramic dental X-ray dataset was obtained from Roboflow from publicly available datasets. There is no personally identifiable information collected or used. Throughout the research process, we followed the relevant ethical guidelines to protect the privacy and confidentiality of the people whose images are already included in the datasets.

- **DATA AVAILABILITY AND ACCESS**

The particular dataset utilized in this research is the Multi-label dental diseases, 1512 panoramic Xray from Roboflow, which is openly accessible at: https://universe.roboflow.com/yolocthel/disease-xaijn.

REFERENCES

"<1ST...2508.21088v1 .. 2025.pdf>."

R. Tarek, A. Elshenawy, M. I. Assadwy, and M. A. Madkour, "Automated Diagnosis of Dental Diseases Using Deep Learning on Radiographic Images," *SN Computer Science*, vol. 6, no. 6, 2025, doi: 10.1007/s42979-025-04264-y.

S. Masri and A. Hasasneh, "Explainable Transformer-Based Approach for Dental Disease Prediction," *Computer Systems Science and Engineering*, vol. 49, no. 1, pp. 481-497, 2025, doi: 10.32604/csse.2025.068616.

"<4th...2508.21088v1 .. 2025.pdf>."

"<5th... FRAI_Volume 1_Issue 1_Pages 1-10.pdf>," doi: 10.22080/frai.2025.29272.1013 10.22080/frai.2025.29272.1013 10.22080/frai.2025.29272.1013.1.

S. Pornprasertsuk-Damrongsri *et al.*, "Clinical application of deep learning for enhanced multistage caries detection in panoramic radiographs," *Sci Rep*, vol. 15, no. 1, p. 33491, Sep 29 2025, doi: 10.1038/s41598-025-16591-4.

X. Zhang *et al.*, "Enhancing furcation involvement classification on panoramic radiographs with vision transformers," *BMC Oral Health*, vol. 25, no. 1, p. 153, Jan 29 2025, doi: 10.1186/s12903-025-05431-6.

S. B. Singh, A. Laishram, K. Thongam, and K. Manglem Singh, "Automatic Detection and Classification of Dental Anomalies and Tooth Types Using Transformer-Based Yolo With GA Optimization," *IEEE Access*, vol. 13, pp. 59326-59338, 2025, doi: 10.1109/access.2025.3556523.

C. Vipin Kumar and S. Dr. Nagendra Pratap, "Deep Learning and Machine Learning Techniques in Dental Disease Detection and Classification," *International Journal of Scientific Research in Science, Engineering and Technology*, vol. 12, no. 2, pp. 716-721, 2025, doi: 10.32628/ijsrset25122199.

O. F. Kaygisiz, O. Uranbey, B. Gursoytrak, Z. B. Gur, A. Cicek, and M. A. Canbal, "A deep learning approach based on YOLO v11 for automatic detection of jaw cysts," *BMC Oral Health*, vol. 25, no. 1, p. 1518, Oct 2 2025, doi: 10.1186/s12903-025-06767-9.

P. Parkhi *et al.*, "A Comprehensive Deep Learning Framework for Dental Disease Classification," *Journal Européen des Systèmes Automatisés*, vol. 58, no. 3, pp. 511-521, 2025, doi: 10.18280/jesa.580309.

S. K. Veerabhadrappa, S. Vengusamy, S. Padarha, K. Iyer, and S. Yadav, "Fully automated deep learning framework for detection and classification of impacted mandibular third molars in panoramic radiographs," *Journal of Oral Medicine and Oral Surgery*, vol. 31, no. 1, 2025, doi: 10.1051/mbcb/2025008.

T. D. Pham, 2025, doi: 10.1101/2025.01.30.25321418.

A. Hussein, H. G. Güneç, K. C. Aydin, and H. F. Ates, "Deep Learning based Tooth Multi-Disease Detection in Dental Diagnostics," presented at the 2025 33rd Signal Processing and Communications Applications Conference (SIU), 2025.

P. Kaushik and P. Sharma, "DenseNet201-Powered Dental Radiograph Classification: A Deep Learning Approach for Multi-class Dental Pathology Detection," presented at the 2025 Fourth International Conference on Power, Control and Computing Technologies (ICPC2T), 2025.

J. Priya, S. K. S. Raja, and S. Sudha, "An intellectual caries segmentation and classification using modified optimization-assisted transformer denseUnet++ and ViT-based multiscale residual denseNet with GRU," *Signal, Image and Video Processing*, vol. 18, no. 6-7, pp. 5213-5227, 2024, doi: 10.1007/s11760-024-03227-9.

Y. M. Alsakar, N. Elazab, N. Nader, W. Mohamed, M. Ezzat, and M. Elmogy, "Multi-label dental disorder diagnosis based on MobileNetV2 and swin transformer using bagging ensemble classifier," *Sci Rep*, vol. 14, no. 1, p. 25193, Oct 24 2024, doi: 10.1038/s41598-73297-9.

"<18th(2024)...Automated+Teeth+Disease+Classification+using+Deep+Learning+Models.pdf>."

A. Bilal, A. Haider Khan, K. Almohammadi, S. A. Al Ghamdi, H. Long, and H. Malik, "PDCNET: Deep Convolutional Neural Network for Classification of Periodontal Disease Using Dental Radiographs," *IEEE Access*, vol. 12, pp. 150147-150168, 2024, doi: 10.1109/access.2024.3472012.

Z. Can, S. Isik, and Y. Anagun, "CVApool: using null-space of CNN weights for the tooth disease classification," *Neural Computing and Applications*, vol. 36, no. 26, pp. 16567-16579, 2024, doi: 10.1007/s00521-024-09995-2.

M. A. Hasnain, Z. Ali, A. Saeed, S. Aijaz, and M. S. Khurram, "PDDNet: Deep Learning Based Dental Disease Classification through Panoramic Radiograph Images," *VFAST Transactions on Software Engineering*, vol. 12, no. 4, pp. 180-198, 2024, doi: 10.21015/vtse.v12i4.2028.

"<22nd...624 (2024).pdf>," doi: 10.56979/702/2024.

M. A. Hasnain, Z. Ali, M. S. Maqbool, and M. Aziz, "X-ray Image Analysis for Dental Disease: A Deep Learning Approach Using EfficientNets," *VFAST Transactions on Software Engineering*, vol. 12, no. 3, pp. 147-165, 2024, doi: 10.21015/vtse.v12i3.1912.

N. Thalji, E. Aljarrah, M. H. Almomani, A. Raza, H. Migdady, and L. Abualigah, "Segmented X-ray image data for diagnosing dental periapical diseases using deep learning," *Data Brief*, vol. 54, p. 110539, Jun 2024, doi: 10.1016/j.dib.2024.110539.

X. Li *et al.*, "A Multi-center Dental Panoramic Radiography Image Dataset for Impacted Teeth, Periodontitis, and Dental Caries: Benchmarking Segmentation and Classification Tasks," *J Imaging Inform Med*, vol. 37, no. 2, pp. 831-841, Apr 2024, doi: 10.1007/s10278-024-00972-8.

S. Okazaki *et al.*, "RadImageNet and ImageNet as Datasets for Transfer Learning in the Assessment of Dental Radiographs: A Comparative Study," *J Imaging Inform Med*, vol.38, no. 1, pp. 534-544, Feb 2025, doi: 10.1007/s10278-024-01204-9.

Y. Ikhwani, E. Noersasongko, Purwanto, and M. A. Soeleman, "Comparative Performances of the Convolutional Neural Network based Transfer Learning Models for Classification of Dental Disease," presented at the 2024 International Seminar on Application for Technology of Information and Communication (iSemantic), 2024.

S. Taskin and I. Ferdib Al, "Transfer Learning-based Fine Tuned MobileNetV2 Model with Explainable Artificial Intelligence for Identifying Dental Diseases," presented at the 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), 2024.

N. R. P, K. B, M. B, and S. R, "Deep Transfer Learning Based Multi-Classification of Dental Diseases X-Ray Images," presented at the 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), 2024.

Journal of Computing & Biomedical Informatics, vol. 5, no. 01, 2023, doi: 10.56979/501/2023.

H. Shafiq, G. Gilanie, M. Sajid, and M. Ahsan, "Dental radiology: a convolutional neural network-based approach to detect dental disorders from dental images in a real-time environment," *Multimedia Systems*, vol. 29, no. 6, pp. 3179-3191, 2023, doi: 10.1007/s00530-023-01169-9.

"<32nd...2022-Multi_Oral_Disease_Classification_from_Panoramic_Radiograph1.pdf>."