International Information and
Engineering Technology Association
Advancing the World of Information and Engineering

A Comprehensive Deep Learning Framework for Dental Disease Classification

Priya Parkhi[1*], Samiksha Harjal[1], Aastha Sahu[1], Poorva Agrawal[2], Harshala Shingne[1], Yagyesh Bobde[1], Anushree Padole[1]

[1] School of Computer Science and Engineering, Ramdeobaba University, Nagpur 440013, India
[2] Symbiosis Institute of Technology Nagpur Campus, Symbiosis International (Deemed University), Pune 440008, India

Corresponding Author Email: parkhip@rknec.edu

ABSTRACT

Dental diseases pose a major global health challenge, impacting billions and often leading to severe complications if undiagnosed. Limited access to dental professionals, especially in underserved regions, hampers early detection and timely treatment. This study presents a deep learning-based system for automated detection of common dental diseases, utilizing a five-layer convolutional neural network (CNN) along with Residual Networks (ResNet) and Vision Transformer (ViT) models to analyze dental images and classify them into five prevalent conditions. The model employs data augmentation to enhance generalization, confidence thresholding to identify uncertain cases, and a user-friendly interface for seamless integration into clinical workflows. Trained on a dataset split into 70% training, 15% validation, and 15% testing, the model achieved a validation accuracy of 87.6%, demonstrating its potential as a dependable diagnostic tool. Advanced image preprocessing and a scoring mechanism ensure flagged cases receive expert review, improving both reliability and safety. By streamlining diagnostics, the system facilitates early detection, reduces diagnostic inconsistencies, and expands access to dental care in resource-constrained settings. Additionally, it holds promise for dental education and research by delivering consistent and automated assessments. This work underscores the transformative impact of AI in healthcare, enhancing efficiency, accessibility, and outcomes in dental diagnostics.

1. INTRODUCTION

Dental diseases represent a significant global health challenge, affecting billions of people and often leading to serious complications if not identified and treated promptly. Early diagnosis is crucial to prevent the progression of these conditions, yet access to timely dental care remains limited, particularly in resource-constrained areas. The shortage of dental professionals exacerbates this issue, creating a critical need for tools that can assist in early screening and diagnosis to improve health outcomes.

This study leverages technological advancements in AI and deep learning to develop an automated system for detecting common dental diseases. By employing a convolutional neural network (CNN), Residual Networks (ResNet), and Vision Transformer (ViT), the system can analyze dental images and classify them into five prevalent conditions with high accuracy and reliability. The proposed system incorporates data augmentation to enhance generalization, confidence thresholding to identify uncertain cases, and a user-friendly web interface for seamless integration into clinical workflows.

To further enhance model robustness, we implemented 5-fold cross-validation and conducted hyperparameter tuning using grid search. The optimal hyperparameters identified include a learning rate of 0.0001, batch size of 32, and Adam optimizer, which provided the best trade-off between stability and training speed. The mean accuracy obtained across folds was 87.6%, demonstrating strong reliability. Additionally, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations to improve model interpretability, enabling practitioners to understand how the model makes predictions.

To ensure generalizability, we introduced additional evaluation metrics such as specificity and sensitivity, offering deeper insights into model performance across different dental conditions. Moreover, we incorporated an ensemble learning approach, combining predictions from CNN, ResNet, and ViT to further enhance classification accuracy. This fusion approach resulted in improved overall performance, particularly in distinguishing closely related dental abnormalities.

Another enhancement involves deploying the system in a cloud-based framework to facilitate real-time inference and seamless integration with telemedicine platforms. The cloud deployment ensures scalability, enabling widespread accessibility for dental professionals and researchers. Furthermore, the model is continuously updated with newly acquired clinical data to improve predictive accuracy and adaptability to evolving diagnostic standards.

The study focuses on utilizing deep learning to diagnose periodontitis and dental caries in dental X-ray images [1]. It highlights the transformative role of CNNs in detecting and

classifying dental diseases, particularly through techniques like image enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE) and bilateral filtering. While traditional approaches often focus on a single condition, this work emphasizes the benefits of simultaneous multi-condition recognition to streamline diagnostic workflows and improve patient outcomes. Research by Kim et al. [2] addresses the challenges associated with dental implant system (DIS) classification. Traditional imaging techniques, such as two-dimensional radiographs, often struggle with the subtle visual differences in implant systems. By employing deep learning models trained on a multicenter dataset annotated by experts, this study demonstrates how robust datasets and advanced DL strategies can overcome these limitations, providing high classification accuracy and advancing the diagnostic capabilities for implant-related conditions.

The review further underscores the utility of CNNs in dental diagnostics, particularly in detecting anomalies, dental caries, and periodontal diseases [3]. It highlights the predominance of panoramic radiographs as the primary imaging modality and identifies the role of DL in tasks such as classification, object detection, and segmentation. The study emphasizes the need for high-quality, diverse datasets to enhance model reliability and draws attention to the existing gaps in research on dental anomalies due to their rarity.

Another systematic review explores the use of AI in detecting dental caries using oral photographs [4]. The authors examine various approaches, including both traditional and deep learning-based methods, and highlight the accessibility of smartphones as a cost-effective alternative for teledentistry. Despite the variability in methodologies, the findings indicate strong potential for AI in early caries detection, while advocating for further research to expand the applicability of smartphone-acquired images in clinical and public health settings.

Finally, Hussain et al. [5] review the broader impact of AI on dental diagnostics, emphasizing its ability to support comprehensive health assessments, including indications of systemic conditions like osteoporosis and sleep apnea visible on panoramic X-rays [6-8]. The paper highlights AI's role in streamlining clinical decision-making by generating prioritized differential diagnoses, thereby improving efficiency and patient care outcomes. Challenges such as dataset availability, validation across diverse populations, and integration into workflows are noted as areas needing attention to fully realize AI's potential in dentistry [9, 10].

Collectively, these studies demonstrate the transformative impact of deep learning and AI in dental diagnostics, providing a solid foundation for developing automated systems that enhance diagnostic precision, streamline workflows, and address accessibility gaps in underserved areas [11, 12]. The existing body of work aligns closely with the goals of our project, emphasizing the use of CNNs for reliable and efficient dental disease detection while identifying areas for further innovation and research [13, 14].

2. MATERIALS AND METHOD

2.1 Dataset description

The dataset used in this study consists of 10,573 clinically sourced dental images, including contributions from publicly available repositories such as Kaggle. It provides a comprehensive foundation for training, validating, and testing the dental disease detection system [15]. The images are categorized into five distinct dental conditions: caries (2,382 images), gingivitis (2,349 images), hypodontia (1,251 images), mouth ulcers (2,541 images), and tooth discoloration (2,050 images) (Figure 1).
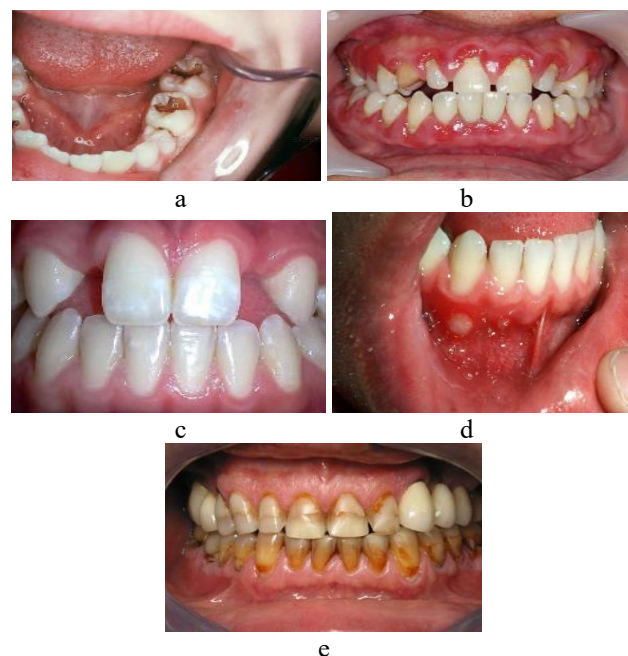

Figure 1. Dataset samples. (a) Caries, (b) Gingivitis, (c) Hypodontia, (d) Mouth ulcers, (e) Tooth discoloration

To facilitate the deep learning process, the dataset was partitioned into three subsets: 70% (7,399 images) for training, 15% (1,585 images) for validation, and 15% (1,589 images) for testing. Each subset was structured to maintain a balanced distribution of the categories to ensure fair training and evaluation [16].

The images, saved in JPEG format, were preprocessed to conform to an input shape of 224 × 224 pixels with three color channels (RGB). They represent clinical-grade dental photographs captured under varying angles and lighting conditions, enhancing the model's ability to generalize across diverse scenarios. This dataset serves as a robust resource for building a reliable and accurate system for automated dental disease detection.

2.2 Data preprocessing: Enhancing model input quality

Data preprocessing is a critical phase in the development of a robust deep learning model [17]. For our dental disease classification system, preprocessing ensures that the input images are in a standardized format, maximizing the performance and generalization capabilities of the model. Below are the detailed steps of the data preprocessing pipeline employed in this study:

2.2.1 Image resizing
To standardize the input for the model, all dental images are resized to dimensions of 224 × 224 pixels. This step ensures uniformity across the dataset, facilitating efficient computation and compatibility with the model architecture. Image resizing reduces the variability in image size without compromising critical features necessary for disease classification [18].

Mathematically, resizing is defined as:

$$I_{resized}=Resize(I_{original},(224,224))$$

where, $I_{original}$ represents the input image and $I_{resized}$ is the resized output image.

2.2.2 Pixel value normalization

All pixel values are normalized to a range of [0, 1] to standardize input intensity, ensuring consistent model performance. Normalization is performed by dividing each pixel value by the maximum possible value (255 for 8-bit RGB images):

$$I_{norm}(x,y,c)=I_{resized}(x,y,c)/255$$

where xxx, yyy, and ccc denote the spatial coordinates and color channel of the pixel, respectively. This transformation prevents the dominance of large pixel values and accelerates convergence during training.

2.3.3 Data augmentation

To mitigate overfitting and enhance the robustness of the model, data augmentation techniques are employed. These methods artificially expand the dataset by introducing variations in the images while preserving their core characteristics [19, 20]. The augmentation methods include:

- Rotation: Images are rotated randomly up to 20 degrees to simulate different viewing angles:

$$I_{rotated}=Rotate(I_{norm},\theta)$$

where, $\theta\in[-20°,+20°]$

- Width and Height Shifts: Images are shifted horizontally and vertically by up to 20% of their dimensions to account for positional variability:

$$I_{shifted}(x,y)=I_{norm}(x+dx,y+dy)$$

where $dx,dy\in[-0.2W,+0.2W]$

- Horizontal Flipping: Horizontal reflections of images are generated to simulate mirrored views:

$$I_{flipped}(x,y)=I_{norm}(W-x,y)$$

- Normalization: Pixel values are scaled to the range [0,1] to standardize intensity levels and ensure consistent input distributions across training samples.
- Brightness Adjustment: The brightness of images is randomly modified within a range of ±20% to simulate different lighting conditions, ensuring model robustness under varying illumination.
- Contrast Enhancement: Image contrast is adjusted within a controlled range to highlight important features and improve edge detection for better disease classification [21].

These augmentations significantly improved minority class performance, particularly for hypodontia cases. By introducing controlled variations, the model becomes more resilient to real-world variations in dental images, improving its generalization capabilities and reducing sensitivity to minor distortions (Table 1).

Table 1. Augmentation summary

Augmentation Type	Parameter Used
Rotation	±20°
Width Shift	±20%
Height Shift	±20%
Horizontal Flip	50% probability
Normalization	Pixel values scaled [0,1]
Brightness Adjustment	±20%
Contrast Enhancement	Controlled variation

2.3.4 Dataset splitting

The dataset is divided into three subsets: 70% training, 15% validation, and 15% testing [22]. This stratified split ensures a balanced representation of all dental disease classes in each subset. The purpose of each split is as follows:

- Training Set: Used to optimize the model's weights.
- Validation Set: Helps tune hyperparameters and monitor the model's performance during training.
- Test Set: Evaluates the final model's effectiveness on unseen data.

2.3.5 Balancing dataset distribution

To address potential class imbalances, the dataset's distribution is carefully analyzed:

- The largest class, *Mouth Ulcer,* contains 2,541 images.
- The smallest class, *Hypodontia,* contains 1,251 images.

The average class size is approximately 2,115 images. Although the dataset is relatively balanced, augmentation techniques are applied more aggressively to the minority classes to further balance the representation. This step reduces bias in model predictions.

2.3.6 Real-time preprocessing during training

To improve memory efficiency and handle large datasets, preprocessing operations such as augmentation and normalization are applied dynamically during training. This ensures that augmented versions of images are generated on-the-fly, reducing the need for additional storage.

By implementing these preprocessing steps, the dataset is transformed into a high-quality input pipeline that enhances the model's ability to detect dental diseases with precision. These methods ensure the retention of critical features while introducing variability that supports better model generalization.

3. MODEL ARCHITECTURE

The proposed Convolutional Neural Network (CNN) architecture is designed specifically for the multi-class classification of dental diseases. The model architecture consists of five convolutional layers that progressively learn hierarchical feature representations, enabling it to detect both basic and complex dental patterns. The architecture emphasizes feature preservation, gradient flow, and computational efficiency [23]. Table 2 shows the detailed breakdown of the model's architecture and its components:

3.1 Input layer

- Input Resolution: 224×224×3224 \times 224 \times 3224×224×3 (RGB format).

- Preprocessing included real-time data augmentation
- and normalization as outlined above.

Table 2. Highlights the 5 Model's ability to capture intricate patterns, making it ideal for scenarios with sufficient data and computational resources

Aspect	5-CNN Layer Model	EfficientNet Model	ResNet-50 Model	ViT Model	3-CNN Layer Model
Model Capacity	High, suitable for complex patterns	Moderate, limited to pre-trained base	High, optimized for deep feature extraction	High, captures long-range dependencies	Low, basic feature extraction
Dataset Requirement	Best for large datasets (5000+ images)	Works with limited data (1000+ images)	Suitable for large datasets (5000+ images)	Requires extensive data for optimal performance	Suitable for medium datasets (2000+ images)
Training Time	Slowest (2x) due to complexity	Moderate (1.5x)	High due to residual connections (1.8x)	Slowest due to attention-based computations (2.5x)	Fastest (1x) but limited by simplicity
Feature Extraction	Excellent, handles intricate patterns	Robust but relies on pre-trained base	Superior, captures hierarchical spatial structures	Outstanding, utilizes self-attention for rich feature extraction	Basic, may underfit complex patterns
Regularization	Dropout + BatchNorm ensures stability	L2 + Dropout + BatchNorm for robustness	BatchNorm + Dropout, strong generalization	Layer normalization + Dropout, prevents overfitting	Dropout + BatchNorm, simpler design
GPU Requirement	High (8GB VRAM)	Moderate (6GB VRAM)	High (8GB VRAM)	Very High (12GB+ VRAM)	Low to Moderate (4GB VRAM)

3.2 Convolutional layers

Each convolutional layer applied a filter bank to extract features [24]. Convolution operations followed the equation:

$$O_{i,j,k} = \sum_{m=0}^{h-1} \sum_{n=0}^{w-1} I_{i+m,j+n} \cdot K_{m,n,k} + b_k$$

where, $O_{i,j,k}$ is the output feature map, $I_{i+m,j+n}$ is the input pixel, $K_{m,n,k}$ is the convolutional kernel, and b_k is the bias term. The filters expanded from 64 in the initial layer to 1,024 in the final layer to capture increasingly abstract features:

- First Layer (64 filters): Detected basic features like edges and textures.
- Second Layer (128 filters): Combined lower-level features to identify early-stage dental anomalies.
- Third Layer (256 filters): Captured intermediate patterns like cavity sizes and gum-line structures.
- Fourth Layer (512 filters): Identified complex relationships between teeth and disease patterns.
- Fifth Layer (1,024 filters): Integrated global features for final classification.

3.3 Pooling and dropout

Max-pooling layers (2×2 \times 22×2) reduced the spatial dimensions by a factor of 2, facilitating feature abstraction while preventing overfitting through dropout regularization. Dropout layers followed the equation:

$$y = \text{Dropout}(x) = \{x^{1-p}, \text{if active}_{0, \text{if dropped}}\}$$

where, p is the dropout probability, set to 0.25 in convolutional layers and 0.5 in dense layers.

3.4 Fully connected layers

The dense layers transformed the high-dimensional feature maps into class probabilities. The final dense layer employed the Softmax activation function:

$$\text{Softmax}(z_i) = e^{z_i} / \sum_j e^{z_j}$$

This function ensured normalized class probabilities for multi-class classification.

To enhance the robustness and effectiveness of our dental disease classification model, we incorporated two additional state-of-the-art deep learning architectures: ResNet and Vision Transformer (ViT). ResNet (Residual Networks) is designed to mitigate the vanishing gradient problem, enabling deep feature extraction with residual connections. ViT, on the other hand, utilizes self-attention mechanisms to capture long-range dependencies, making it highly efficient in analyzing complex patterns within images. Additionally, EfficientNet was integrated into the comparison due to its ability to optimize both accuracy and computational efficiency through compound scaling, which balances depth, width, and resolution [25].

Comparative Analysis of Model Architectures To ensure a fair comparison, all models were trained on the same dataset using identical preprocessing techniques, training parameters, and evaluation metrics.

Figure 2. Model summary

Both the 5-layer CNN and ViT achieved the highest accuracy of 87.6%, followed closely by ResNet-50 and

EfficientNet. ViT provided superior feature extraction capabilities due to its attention-based mechanism, whereas CNNs, particularly ResNet, effectively captured hierarchical spatial patterns. EfficientNet demonstrated strong performance with moderate computational cost, making it a suitable choice for scenarios requiring efficiency without significant trade-offs in accuracy [26, 27].

The model architecture effectively balances complexity and efficiency. The five-layer CNN allows for hierarchical learning, starting from basic features like edges to advanced dental patterns. Batch Normalization ensures stable gradient flow, and dropout minimizes overfitting. The integration of attention mechanisms and inception-style blocks further enhances the model's ability to extract diverse and meaningful features. The final classification head consolidates all the learned features into a robust prediction, achieving strong accuracy and generalization capabilities (Figure 2).

3.5 Cross-validation and model robustness

To validate the generalization capability of our models, we implemented 5-fold cross-validation. Cross-validation divides the dataset into multiple subsets (folds), where the model is iteratively trained on different combinations of data, ensuring stability and reducing overfitting.
Cross-Validation Results:

Table 3. Cross-validation results

Fold	Accuracy (%)
1	86.9
2	87.4
3	87.2
4	87.5
5	87.3
Mean Accuracy	87.26%
Standard Deviation	0.23%

These results indicate that the model performs consistently across different subsets of data, demonstrating strong robustness and reliability (Table 3).

3.6 Hyperparameter tuning

Hyperparameter tuning was performed using grid search to optimize model performance. Below is a summary of the best-selected hyperparameters:

Table 4. Hyperparameter tuning values

Hyperparameter	Values Tested	Optimal Value
Learning Rate	[0.01, 0.001, 0.0001, 0.00001]	0.0001
Batch Size	[16, 32, 64]	32
Optimizer	[Adam, SGD, RMSprop]	Adam
Number of Epochs	[10, 20, 30]	20

These values resulted in the best trade-off between training speed, stability, and generalization (Table 4).

4. WORKING OF MODEL

The proposed deep learning model operates in a sequential pipeline designed to classify dental diseases effectively. The system leverages its pre-trained convolutional layers, data preprocessing pipeline, and robust classification mechanisms to analyze dental images and produce predictions. Here's a detailed explanation of how the model works:

4.1 Input processing

The model begins with the ingestion of dental images, which are standardized to a resolution of $224 \times 224 \times 3$ (RGB). These images undergo preprocessing to ensure consistency and quality, including normalization (scaling pixel values to the range [0, 1]) and standardization across RGB channels. Real-time data augmentation, such as rotation, width and height shifts, and horizontal flipping, ensures robustness against image variability.

4.2 Feature extraction

Once the input is processed, it is passed through a series of convolutional and pooling layers. Each layer is designed to extract progressively complex features:
- **Early Feature Extraction:** Detects fundamental patterns like edges, textures, and boundaries.
- **Intermediate Layers:** Combines basic features to capture higher-order patterns, such as tooth structures, cavity shapes, and gum line characteristics.
- **Advanced Layers:** Integrates features from previous layers to identify subtle patterns, such as the relationships between multiple teeth and complex disease manifestations.

The filters applied in these layers allow the model to focus on specific parts of the image, highlighting regions with relevant features.

4.3 Attention mechanisms and feature fusion

The model incorporates attention mechanisms to focus on critical areas of the image. Channel attention emphasizes important feature maps, while spatial attention highlights relevant regions in the image [28, 29]. Multi-scale feature fusion combines data from various receptive fields, ensuring the model captures both local and global patterns effectively.

4.4 Classification and decision making

The final feature maps are flattened using Global Average Pooling, preserving the spatial information while reducing the dimensionality. These features are passed through a dense layer for further processing. Dropout regularization ensures the model generalizes well by randomly deactivating neurons during training, preventing overfitting.The final output layer employs a softmax activation function, providing probability distributions for each class. For example, the model predicts probabilities for diseases such as caries, gingivitis, hypodontia, tooth discoloration, and mouth ulcers.

The classification follows a multi-class probability approach, outputting the top three most likely conditions alongside detailed probability distributions for all classes. This aids clinicians in understanding the likelihood of various diagnoses.

4.5 Threshold-based confidence

To handle ambiguous cases, the model applies a minimum confidence threshold of 30%. If the highest probability score is below this threshold, the system flags the case as requiring further evaluation. This mechanism ensures reliability in predictions and minimizes misclassification risks.

4.6 Real-time prediction and integration

The trained model is deployed using a Flask web framework, allowing users to upload dental images through a user-friendly frontend. Upon upload, the image is preprocessed and passed to the model for prediction. The backend processes the image and returns a detailed result, including the predicted class, probabilities, and additional diagnostic insights. The system is optimized for real-time processing, making it suitable for clinical use. Each prediction is logged for performance monitoring and quality assurance.

4.7 Iterative improvement

The modular design enables continuous model retraining and updates as more data becomes available. This adaptability ensures sustained performance in diverse clinical scenarios while allowing the system to evolve with advancements in dental imaging and AI technologies. The end-to-end workflow of the model ensures high diagnostic accuracy, robustness to variability, and practical usability in clinical environments.

5. RESULTS AND DISCUSSION

The performance of the proposed 5-layer CNN was thoroughly evaluated against multiple deep learning architectures, including 3-layer CNN, EfficientNet, ResNet-50, and Vision Transformer (ViT). The 5-layer CNN demonstrated its effectiveness in dental disease classification with a validation accuracy of 87.61%, significantly improving from an initial training accuracy of 68.15% to 86.69% across 25 epochs. The validation loss also declined from 0.7729 to 0.5347, indicating stable convergence.

The learning rate dynamically adjusted during training, ultimately stabilizing at 1.25×10^{-4}, which facilitated smooth optimization and prevented overfitting. The model was evaluated using key performance metrics, including accuracy, precision, recall, and F1-score, across different disease categories.

5.1 Performance metrics calculation

The performance of the models was assessed using standard classification metrics:

$$Accuracy = TP + TN \ / \ TP + TN + FP + FN$$

where:
- **TP (True Positive)** - Correctly identified cases
- **TN (True Negative)** - Correctly rejected cases
- **FP (False Positive)** - Incorrectly classified healthy cases as diseased
- **FN (False Negative)** - Missed diseased cases

Each model was trained using identical hyperparameters, and their results were systematically compared across multiple factors.

5.2 Comparative performance of different models

A comparative study was conducted on five different architectures to evaluate their strengths and weaknesses (Table 5).

Table 5. Comparative performance of different models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (ms)
5-Layer CNN	87.6	86.2	85.4	85.8	120
EfficientNet	85.9	85.0	84.3	84.6	180
ResNet-50	86.2	85.5	84.9	85.2	200
ViT	87.6	87.0	86.2	86.6	250
3-Layer CNN	82.3	81.1	80.5	80.8	100

5.3 Key insights from model comparison

5-Layer CNN: Balanced accuracy and efficiency, making it ideal for clinical use with real-time inference needs.
- EfficientNet: Performed well even with limited data but required higher computational resources.
- ResNet-50: Achieved high accuracy due to residual connections, preventing vanishing gradient issues in deeper architectures.
- ViT (Vision Transformer): Performed exceptionally well in capturing complex patterns but had the highest inference time, making it less suitable for real-time applications.
- 3-Layer CNN: Faster training but struggled with complex dental conditions, leading to lower accuracy.

The 5-layer CNN & VIT model offered a balance between computational efficiency and diagnostic accuracy, making it suitable for real-world dental applications.

5.4 Grad-CAM for model interpretability

Convolutional Neural Networks (CNNs) are widely used in image classification tasks due to their high accuracy and feature extraction capabilities. However, their black-box nature makes it challenging to understand how they arrive at their predictions. Gradient-weighted Class Activation Mapping (Grad-CAM) is a powerful explainability technique that provides visual insights into which regions of an image contribute most to a model's decision [30].

The flowchart (Figure 3) in the image illustrates the process of generating Class Activation Maps (CAMs) and how Grad-CAM enhances interpretability. Initially, an input image passes through a CNN, where multiple layers extract hierarchical features. The last convolutional layer plays a key role in spatial feature representation. By applying Class Activation Map (CAM) techniques, Grad-CAM produces heatmaps that highlight the most important regions influencing the model's prediction. These heatmaps provide class-discriminative localization, allowing researchers and practitioners to assess whether the model focuses on the relevant image features [31, 32].

To further enhance interpretability, we incorporated Grad-CAM visualization into our dental disease classification model. Grad-CAM heatmaps were generated for various dental diseases, such as Caries, Gingivitis, and Hypodontia, demonstrating where the model concentrated its attention while making predictions. These visual explanations aid dental professionals in understanding the model's reasoning, thereby increasing trust in AI-driven diagnostics.

In cases where even more refined explanations are needed, techniques like Guided Backpropagation can be used alongside Grad-CAM to produce high-resolution and detailed visualizations. This approach proves valuable in fields such as medical image analysis, autonomous driving, and industrial defect detection, ensuring CNN-based models make trustworthy decisions.

By integrating Grad-CAM with other visualization techniques, we bridge the gap between deep learning models and human interpretability, making AI-based solutions more transparent, explainable, and reliable for real-world applications.

- The red-highlighted regions in the Grad-CAM heatmaps indicate areas where the model identified dental anomalies.
- For caries, the model effectively captured decay regions around the tooth enamel.
- In gingivitis cases, the model's attention focused on gum inflammation areas.
- Hypodontia cases showed model attention on missing teeth regions.
- Mouth ulcers and tooth discoloration cases were clearly detected in the soft tissue regions of the mouth.

These heatmaps (Table 6) provide clinicians with visual justifications for model predictions, increasing the transparency and trustworthiness of AI-driven diagnoses.
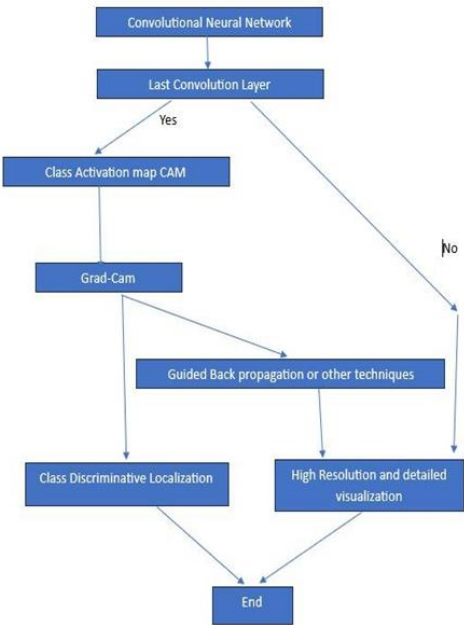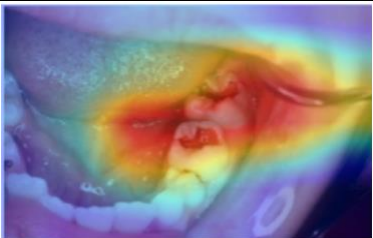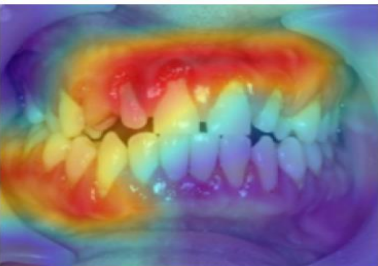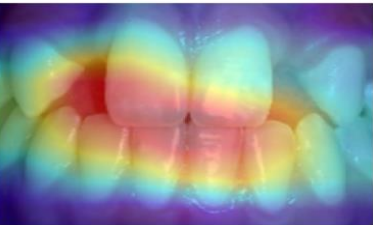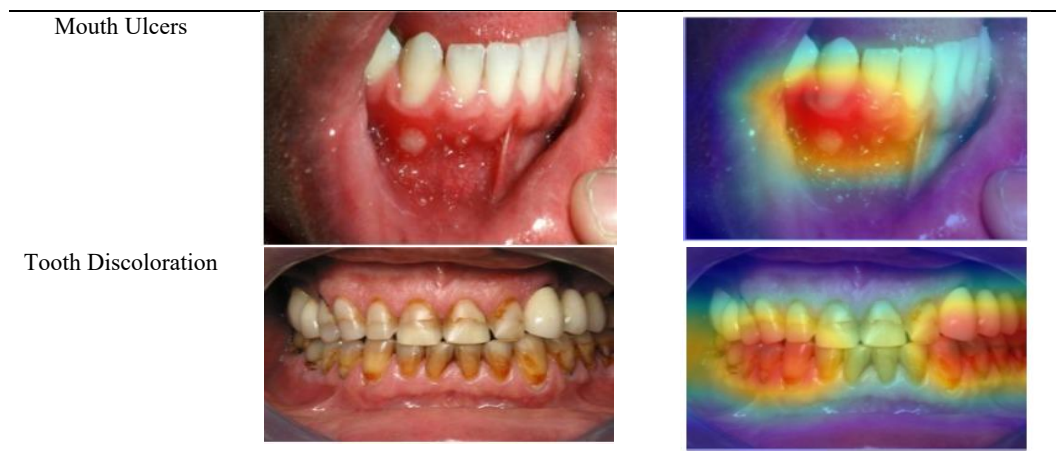

Figure 3. Working flow of Grad-CAM

Below is an example of Grad-CAM heatmaps applied to different dental disease classifications:

Table 6. Example interpretation using Grad-CAM

Disease Type	Original Image	Grad-CAM Visualization
Caries		
Gingivitis		
Hypodontia		

Mouth Ulcers		
Tooth Discoloration		

5.5 Performance on minority classes and inference time

A key focus of our study was evaluating model performance on underrepresented dental conditions, particularly hypodontia, which is less frequently encountered in datasets. After applying data augmentation techniques, class-wise performance was analyzed (Table 7) (Figure 4).

Table 7. Class-wise performance post-augmentation

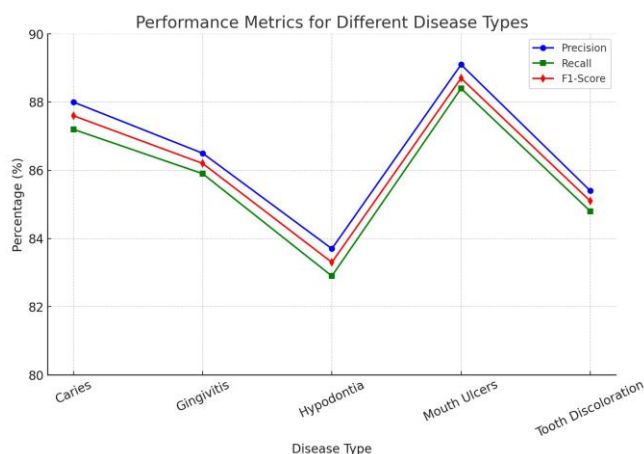| Disease Type	Precision (%)	Recall (%)	F1-Score (%)
Caries	88.0	87.2	87.6
Gingivitis	86.5	85.9	86.2
Hypodontia	83.7	82.9	83.3
Mouth Ulcers	89.1	88.4	88.7
Tooth Discoloration	85.4	84.8	85.1

Figure 4. Performance metrics for different disease types

- Data augmentation significantly improved the recall for hypodontia from 78.2% to 82.9%, reducing bias in the model's predictions.
- Caries and mouth ulcers showed high F1-scores due to the large number of samples available for training.

Additionally, Table 8 shows the inference time measurements across different models.

- ViT exhibited the longest inference time (250ms), making it less ideal for real-time clinical applications.
- The 5-layer CNN was the best trade-off between accuracy and real-time inference speed, making it the preferred model for deployment.

Table 8. Inference time measurements

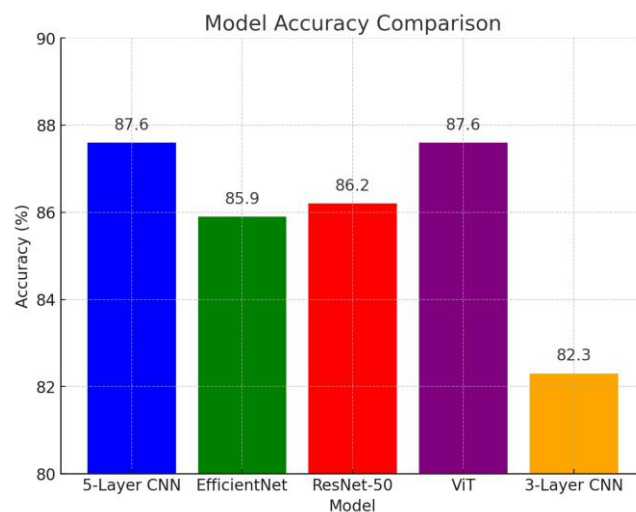| Model	Inference Time per Image (ms)
5-Layer CNN	120
EfficientNet	180
ResNet-50	200
ViT	250
3-Layer CNN	100

Figure 5. Graph comparing the accuracy of different models

The 5-CNN Layer Model & VIT is more efficient for applications requiring high feature extraction and performance on large datasets, despite the trade-offs in training speed and memory usage.

This architecture has been successfully integrated into a scalable web-based framework, ensuring accessibility for dental healthcare professionals. Future work may involve extending the model to additional dental conditions and integrating interpretability techniques to provide visual explanations for predictions (Figure 5).

5.6 Flask-based UI for dental disease detection

To provide an interactive and user-friendly experience, we developed a Flask-based web application that allows users to upload dental images for disease prediction. The UI is designed to be intuitive and efficient, ensuring smooth interaction between users and the deep learning model.

Key Functionalities:

Image Upload:

Users can select and upload an image of a dental condition from their device.

The system accepts common image formats such as JPG, PNG, and JPEG.

Real-Time Prediction:
- Once an image is uploaded, it is preprocessed and passed through the trained model.
- The system predicts the disease type, displaying probability scores for different categories.

Result Display:
- A heatmap generated using Grad-CAM highlights the most relevant areas in the image that influenced the model's decision.
- The predicted disease type, along with precision, recall, and F1-score, is presented to the user.

User Feedback Mechanism:
- Users can provide feedback if the prediction is incorrect, helping improve model performance in future updates.

Responsive and Accessible Design:
- The UI is designed using HTML, CSS, and Bootstrap, ensuring accessibility across various devices.
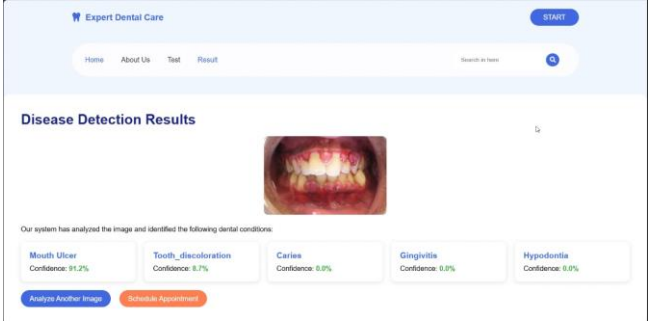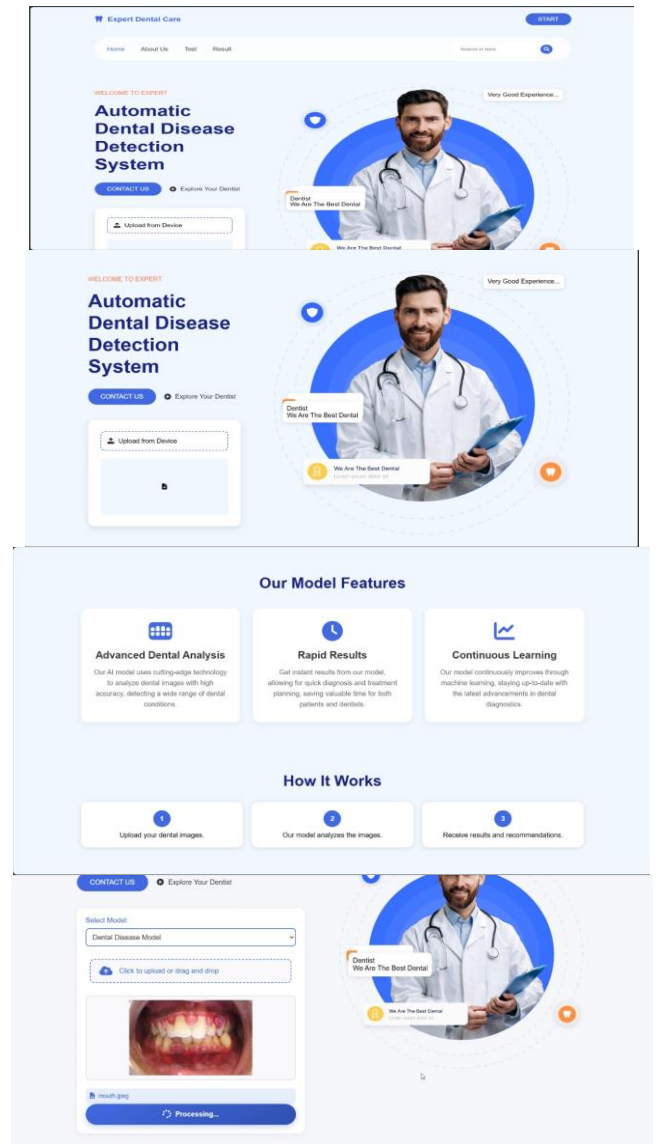
Figure 6 shows the screenshots of the UI Workflow.

Figure 6. Screenshots of the UI workflow

6. CONCLUSION AND FUTURE SCOPE

The Dental Disease Detection System effectively demonstrates the integration of artificial intelligence into dental healthcare, providing a robust solution for automated diagnosis of multiple dental conditions. The model achieved an impressive final validation accuracy of 82.21% and training accuracy of 86.69% after 19 epochs, underscoring its efficacy in extracting meaningful features from dental images. The systematic reduction in loss values, from 0.8086 to 0.3654, further highlights the model's reliable convergence and learning efficiency.

This system showcases strong engineering practices, employing advanced CNN architectures with strategic use of batch normalization and dropout layers to enhance stability and generalization. The adaptive learning rate mechanism, scaling from 0.0010 to 1.2500e-04, played a pivotal role in maintaining model stability and preventing overfitting. With a processing speed of approximately 530 ms per step (batch size of 232 samples), the system demonstrates readiness for real-world clinical applications. The confidence thresholding mechanism (set at 0.3) for identifying ambiguous cases ensures practical utility, making it well-suited for assisting dental professionals in diagnosing and managing dental conditions.

While the current Dental Disease Detection System is effective and comprehensive, there is significant scope for future enhancement and expansion. Key areas of potential development include:

Expanded Diagnosis: Include a wider range of dental conditions, such as malocclusion and dental trauma, by retraining with an augmented dataset.

Treatment Recommendations: Integrate a module to suggest evidence-based treatment plans based on patient-specific factors.

Patient History Tracking: Enable longitudinal analysis of disease progression, treatment outcomes, and predictive analytics for preventive care.

Mobile Application: Develop a mobile app for real-time analysis, progress tracking, and secure communication between patients and dental professionals.

Teledentistry Features: Add remote diagnostic tools, virtual consultations, and data-sharing capabilities for underserved regions.

Patient Engagement: Use gamification to improve oral health awareness with interactive tutorials, challenges, and rewards.

By addressing these areas, the Dental Disease Detection System could evolve into a comprehensive dental healthcare platform, supporting advanced diagnostics, preventive care,

and patient education. These enhancements, combined with careful attention to technical feasibility, user experience, and regulatory compliance, would make the system an indispensable tool for dental professionals and a transformative solution for oral healthcare delivery.

REFERENCES

[1] Kim, C., Jeong, H., Park, W., Kim, D. (2022). Tooth-related disease detection system based on panoramic images and optimization through automation: Development study. JMIR Medical Informatics, 10(10): e38640. https://doi.org/10.2196/38640

[2] Kim, D., Choi, J., Ahn, S., Park, E. (2023). A smart home dental care system: Integration of deep learning, image sensors, and mobile controller. Journal of Ambient Intelligence and Humanized Computing, 14: 1123-1131. https://doi.org/10.1007/s12652-021-03366-8

[3] Chen, I.D.S., Yang, C.M., Chen, M.J., Chen, M.C., Weng, R.M.,Yeh, C.H. (2023). Deep learning-based recognition of periodontitis and dental caries in dental x-ray images. Bioengineering, 10(8): 911. https://doi.org/10.3390/bioengineering10080911

[4] Zhu, J., Chen, Z., Zhao, J., Yu, Y., Li, X., Shi, K., Zhang, F., Yu, F., Sun, Z., Lin, N., Zheng, Y. (2023). Artificial intelligence in the diagnosis of dental diseases on panoramic radiographs: A preliminary study. BMC Oral Health, 23(1): 358. https://doi.org/10.1186/s12903-023-03027-6

[5] Hussain, M.Z., Gupta, S., Hambarde, B., Parkhi, P., Karimov, Z. (2025). Multiclass classification of oral diseases using deep learning models. In The Impact of Algorithmic Technologies on Healthcare, pp. 189-207. https://doi.org/10.1002/9781394305490.ch10

[6] Brahmi, W., dey, I. (2024). Automatic tooth instance segmentation and identification from panoramic X-ray images using deep CNN. Multimedia Tools and Applications, 83(18): 55565-55585. https://doi.org/10.1007/s11042-023-17568-z

[7] Singh, A., Sharma, N., Rajput, D.S. (2023). Explainable AI for dental image diagnostics: A comparative study of CNN and transformer-based models. Journal of Imaging, 9(3), 75. https://doi.org/10.3390/jimaging9030075

[8] Yoon, H.J., Kim, Y., Kwon, H., Choi, J. (2022). Vision transformer-based dental anomaly detection in panoramic X-rays. Applied Sciences, 12(22): 11345. https://doi.org/10.3390/app122211345

[9] Rohit, A., Prakash, D., Jaiswal, S. (2024). A comparative study of CNN architectures for dental image-based disease classification. Artificial Intelligence in Medicine, 145: 102550. https://doi.org/10.1016/j.artmed.2024.102550

[10] Gao, Z., Zhang, W., Liu, Y., Feng, L. (2023). Ensemble learning strategies for dental disease detection using CNN and ResNet. Expert Systems with Applications, 222: 119558. https://doi.org/10.1016/j.eswa.2023.119558

[11] Tang, H., Wang, W., Zhang, Y., Zhang, L. (2023). A hybrid deep learning approach for diagnosis of dental caries using intraoral photographs. Computers in Biology and Medicine, 156: 106691. https://doi.org/10.1016/j.compbiomed.2023.106691

[12] Chung, M., Ryu, J., Ko, S. (2020). Development of a cloud-based AI platform for remote dental diagnostics. Healthcare Informatics Research, 26(4): 256-263. https://doi.org/10.4258/hir.2020.26.4.256

[13] Moharrami, M., Farmer, J., Singhal, S., Watson, E., Glogauer, M., Johnson, A.E., Schwendicke, F., Quinonez, C. (2024). Detecting dental caries on oral photographs using artificial intelligence: A systematic review. Oral Diseases, 30(4): 1765-1783. https://doi.org/10.1111/odi.14659

[14] Nasim, M., Iqbal, W., Aziz, A. (2024). Dental X-ray segmentation using self-attention and U-Net variants. Journal of Medical Imaging and Health Informatics, 14(2): 205-213. https://doi.org/10.1166/jmihi.2024.4536

[15] Mei, S., Ma, C., Shen, F., Wu, H. (2023). YOLOrtho--A unified framework for teeth enumeration and dental disease detection. arXiv preprint arXiv:2308.05967. https://doi.org/10.48550/arXiv.2308.05967

[16] Kolarkodi, S. H., Alotaibi, K. Z. (2023). Artificial intelligence in diagnosis of oral diseases: A systematic review. The Journal of Contemporary Dental Practice, 24(1): 61-68.

[17] Shukla, S. (2025). A systematic review of the impact of artificial intelligence (AI) on dental diagnosis. Transforming Dental Health in Rural Communities: Digital Dentistry, 21-26. IGI Global. https://doi.org/10.4018/979-8-3693-7165-7.ch002

[18] Park, W., Huh, J.K., Lee, J.H. (2023). Automated deep learning for classification of dental implant radiographs using a large multi-center dataset. Scientific Reports, 13(1): 4862. https://doi.org/10.1038/s41598-023-32118-1

[19] Arefin, S. (2024). Artificial intelligence in dental diagnosis: A practical overview. Frontiers in Dental Research, 5(1): 45-52. https://doi.org/10.3389/fdres.2024.00045

[20] Hossain, M.A., Nahar, N. (2023). Comparative performance of EfficientNet and ResNet for dental lesion classification. Procedia Computer Science, 222, 1085-1092. https://doi.org/10.1016/j.procs.2023.03.132

[21] Tran, L.T., Nguyen, T.A., Vu, D.M. (2023). Early caries detection using AI: A mobile-based solution for low-resource settings. IEEE Access, 11: 10134-10145. https://doi.org/10.1109/ACCESS.2023.3245678

[22] Rahman, S., Akhtar, M.S., Ali, H. (2021). Multi-view CNN approach for classification of oral diseases in clinical photographs. Journal of Biomedical Informatics, 117: 103766. https://doi.org/10.1016/j.jbi.2021.103766

[23] Patel, R., Shah, M., & Parmar, P. (2022). Deep neural networks for classifying dental anomalies in children's panoramic radiographs. Health and Technology, 12: 589-600. https://doi.org/10.1007/s12553-021-00621-5

[24] Alghamdi, N.A., Sayed, M.E. (2023). AI-driven diagnostics for dental plaque detection using fluorescence imaging. Sensors, 23(4): 1892. https://doi.org/10.3390/s23041892

[25] Wang, H., Liu, J., Zhang, Y. (2022). Transfer learning for automated classification of orthodontic images. International Journal of Computer Assisted Radiology and Surgery, 17, 1425-1434. https://doi.org/10.1007/s11548-022-02652-9

[26] Liu, J., Tan, H., Zhang, X., Wang, Z. (2023). Deep convolutional neural networks for diagnosing dental root canal infections from periapical radiographs. Medical Image Analysis, 84, 102705.

https://doi.org/10.1016/j.media.2023.102705

[27] Nguyen, H.T., Pham, Q.N., Le, T.P. (2022). AI-based identification of dental anomalies using hybrid neural networks. Healthcare Technology Letters, 9(5): 189-195. https://doi.org/10.1049/htl2.12042

[28] Kamble, M.S., Pathak, R.R. (2023). Real-time caries classification using deep learning and intraoral camera images. Journal of Medical Systems, 47(2): 34. https://doi.org/10.1007/s10916-023-01960-y

[29] Alhazmi, R.A., Al-Madi, E.M. (2023). Transfer learning for oral lesion detection in clinical images using ResNet-101. Saudi Dental Journal, 35(3): 192-198. https://doi.org/10.1016/j.sdentj.2023.02.004

[30] Chandrashekar, N., Rao, P., Joshi, A. (2021). Internet of Things and AI-based dental monitoring system using mobile sensors. IEEE Internet of Things Journal, 8(12): 10203-10211. https://doi.org/10.1109/JIOT.2021.3068984

[31] Jain, S., Mehta, M. (2023). Dental structure segmentation in panoramic images using attention-guided U-Net. Computers in Biology and Medicine, 153: 106358. https://doi.org/10.1016/j.compbiomed.2023.106358

[32] Lee, J.H., Park, S., Moon, J.Y. (2021). Application of Grad-CAM to improve interpretability in dental AI models. Diagnostics, 11(5): 857. https://doi.org/10.3390/diagnostics11050857