Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 13, No. 2, June 2025, pp. 478~499
ISSN: 2089-3272, DOI: 10.52549/ijeei.v13i2.6245

DCDNet: A Deep Learning Framework for Automated Detection and Localization of Dental Caries Using Oral Imagery

Desidi Narsimha Reddy[1], Pinagadi Venkateswararao[2], Anitha Patil[3], Geedikanti Srikanth[4], Varalakshmi Chinnareddy[5]
[1]Data Consultant (Data Governance, Data Analytics, EPM: enterprise performance management, AI&ML), Soniks Consulting LLC, USA.
[2]Computer Science and Engineering, CVR College of Engineering, Hyderabad, Telangana, India.
[3]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Aziz Nagar Hyderabad-500075, Telangana, India.
[4]Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[5]CSE Department, Malla Reddy College of Engineering, Hyderabad, Telangana, India

Article Info

Article history:

Received Jan 16, 2025
Revised Jun 15, 2025
Accepted Jun 20, 2025

Keyword:

Oral X-Ray Images
Dental Caries
Detection and Localization
DCDNet
Deep Learning Method

ABSTRACT

Dental caries is a common oral health condition that requires early diagnosis and identification for effective intervention. Existing deep models, such as Faster R-CNN, YOLOv3, SSD, or RetinaNet, exhibit great effectiveness in generic medical imaging; however, they struggle to precisely and explicitly handle localization in complex dental radiographs. In this paper, we propose DCDNet, a convolutional neural network architecture specifically designed for the detection and segmentation of dental caries in oral X-ray images. However, such deep learning methods currently lack strong generalization due to imbalanced training data, limited lesion-localization ability, and non-interpretable features, which hamper their utility for large-scale clinical evaluation. In addition, most models overlook the severity distinction between classes, which is less ideal for the entire diagnosis and treatment planning process. DCDNet was trained and tested on the UFBA UESC Dental Image Dataset, which comprises over 1,500 labeled grayscale dental radiographic images. The proposed network incorporates multiscale feature extraction, residual connections, and non-maximum suppression (NMS) for more accurate classification and bounding box prediction. Data augmentation techniques were used to increase generalization. The model was evaluated based on accuracy, precision, recall, and F1-score, and compared with ResNet-50, VGG16, AlexNet, Faster R-CNN, YOLOv3, SSD, and RetinaNet in terms of accuracy. DCDNet achieved excellent performance in all its performance indices, with precision at 97.23%, recall at 97.02%, F1-score at 97.12%, and overall accuracy at 97.61%. Experiments demonstrate that the proposed DCDNet surpasses all the baselines and state-of-the-art methods by a significant margin. Ablation experiments validated the importance of residual connections, NMS, and data augmentation for performance improvement. DCDNet represents a significant step toward automatic dental diagnosis, having successfully detected and localized carious lesions in X-ray images. Its design overcomes the drawbacks of previous models and is a ready option for integration into clinical routine.

Corresponding Author:

Desidi Narsimha Reddy,
Data Consultant (Data Governance, Data Analytics, EPM: enterprise performance management, AI&ML),
Soniks Consulting LLC, USA.
Email: dn.narsimha@gmail.com

1. INTRODUCTION

Early identification is imperative for the successful treatment of dental caries and the prevention of complications, as dental caries remains one of the most common oral diseases. Accurate detection of caries from radiographic images is a crucial task in clinical diagnostics, where recent advances in deep learning have shown promise for automating this process. Traditional machine learning feature engineering approaches that required hand-crafting resulted in sub-par performances due to the limited feature extraction capabilities. Recent deep learning architectures, including Faster R-CNN, YOLOv3, SSD, and RetinaNet, have significantly advanced the image diagnosis of medical data. Nonetheless, the models are still limited by their ability to balance accuracy with localization methods. These situations are often encountered in medical imaging tasks, where the subtlety of the discovered patterns must be ascertained with utmost confidence.

There is a limitation in the ability of typical learning architectures to handle most challenging dental radiographs: A detailed literature review. Although Faster R-CNN is suitable for general object detection, it falls short in dealing with unnecessary region proposals or localization errors in fine-grained activities, such as detecting dental caries. While these models are fast, YOLOv3 and SSD struggle with smaller lesion regions due to their limited multi-scale feature representation, resulting in missed detection accomplishments. Other object detectors, like RetinaNet, aim to address the class imbalance problem by introducing focal loss to train the detector; however, they still encounter overlapping bounding boxes and weak generalization across different datasets. These constraints underscore the requirement for a more focused deep learning approach for dental radiographic analysis.

To this end, this paper presents DCDNet, a novel deep learning-based framework for detecting dental caries, utilizing the UFBA UESC Dental Image Dataset. The primary research objective is to develop a robust model that can accurately classify caries images and precisely localize caries areas within dental X-ray images. The proposed DCDNet architecture is characterized by several innovative elements: multi-scale feature extraction enhances lesion detection for varying sizes, residual connections optimize gradient propagation and feature retention, and the implementation of non-maximum suppression (NMS) suppresses duplicate bounding box predictions. Furthermore, during the training process, we applied data augmentation methods to ensure better model generalization and prevent overfitting, thereby making the framework more suitable for real-world clinical environments.

This work has several significant contributions: the definition of a dedicated deep learning architecture (DCDNet) optimized for automatic detection of dental caries, extensive performance comparison with the current cutting-edge models evaluated under standard metrics---like Precision, Recall, F1-Score, and Accuracy---as well as a comprehensive experimental study for model evaluation based on the UFBA UESC Dental Image Dataset to prove the effectiveness of the proposed framework. Additionally, an ablation study highlighted the importance of architectural components, demonstrating how the model's superior performance is achieved.

Section 2 provides an extensive literature review on deep learning techniques for dental image analysis, emphasizing existing model limitations. Section 3 presents our DCDNet framework, including the network architecture, training approach, and performance metrics. In Section 4, we present experimental results, comparisons with baselines, and an ablation study that demonstrates the contribution of the individual components of the proposed architecture. Section 5 presents the results, which include the performance and interpretation of the findings, as well as their limitations. Section 6 summarizes the contributions of this research and provides directions for further work in dual-modal dental diagnostics and explainable AI.

2. RELATED WORK

The literature review explores existing deep learning models for dental caries detection, highlighting advancements, limitations, and performance gaps. Bui et al. [1] described an automated deep-learning method that uses panoramic radiographs to diagnose caries with 93.58% accuracy. Feature redundancy might be an issue, even though it is an improvement over the existing process. According to Park et al. [2], reducing costs and invasive procedures can be achieved by identifying dental cavities early. The paper proposes a deep learning approach for accurate intraoral image-based caries detection; however, standardization is still required. Lee et al. [3] assessed deep convolutional neural networks (CNNs) and found that they perform well in identifying dental caries in periapical radiographs. Larger datasets and improved algorithms will be utilized in the future to enhance clinical applicability. Askar et al. [4] employed deep learning to detect white spot lesions in dental photographs with promising accuracy, despite constraints on sample size and generalizability. Future work should focus on larger datasets and a broader range of model types. Felsch et al. [5] developed an AI model with high accuracy for identifying MIH and caries in dental photographs. Expanding the model, enhancing picture quality, and external validation should be the main goals of future studies.

Tareq et al. [6] developed an accurate and reasonably priced AI method for identifying dental cavities using non-standard smartphone photos. Explainable AI and cross-sectional imaging should be included in future work, along with addressing dataset variability. Corbella et al. [7] identified limitations, including methodological errors and difficulties with picture quality, but emphasized the potential of deep learning in dental diagnostics through image analysis. These holes should be addressed in future work. Zhou et al. [8] claimed that CNNs can classify oral illnesses and identify lesions in RAU photos, which might result in both practical and reasonably priced screening. However, there are limitations, including the need for further verification and the narrow scope of the classification area. Talpur et al. [9] accurately detecting dental cavities is one area where deep learning shows promise. Two drawbacks include the need for more study on a range of caries phases and the emphasis on a small number of caries types. Njimbouom et al. [10] propose a multi-modal machine learning model for accurately predicting dental caries in this work. Requirements include additional X-ray data and continuous picture tagging.

Li et al. [11] examined the efficacy of deep learning in classifying periodontitis, highlighting its high accuracy while also noting its drawbacks, including the lack of diverse data sources and geographical bias. These holes should be addressed in future work. Qayyum et al. [12] proposed a self-training technique for caries diagnosis utilizing numerous unlabeled images and minimal labeled data. It is more accurate and efficient than conventional approaches, but additional validation requires more data. Schwendicke et al. [13] found a range of potential applications for CNNs in dental imaging despite significant variance in methods and outcomes. In subsequent research, methodologies must be standardized, and results should be confirmed using several datasets. Khanagar et al. [14] have demonstrated that AI algorithms, which are remarkably accurate in predicting and identifying dental caries (DC), may potentially enhance clinical practice. Some disadvantages of the dataset are its size and unpredictability. Mehdizadeh et al. [15] demonstrated a deep learning system utilizing Inception-v3, which could detect dental cavities in infants with 79% accuracy. Larger datasets should be employed in subsequent research to improve generalizability and therapeutic usefulness.

Thanh et al. [16] demonstrated a deep learning smartphone software called Faster R-CNN, which uses YOLOv3 to identify dental caries, showing high sensitivity but reduced accuracy for non-cavitated caries. Future studies should focus on utilizing more datasets, acquiring better-quality photos, and implementing computational improvements. Zhang et al. [17] developed a deep learning model that utilizes convolutional networks to identify dental caries, achieving an AUC of 85.65% using oral photos. It has the potential to be an affordable screening method; however, further work is needed. Musri et al. [18] demonstrated that deep-learning CNNs can accurately analyze radiographs to detect dental caries in their early stages, achieving human accuracy levels. The drawbacks include small datasets and a failure to distinguish between various types of caries. Ragodas et al. [19] demonstrated that dental abnormalities in children with orofacial clefting can be efficiently identified from intraoral photos using a deep learning algorithm with competitive accuracy and speed. Limitations include issues with image quality and interference from orthodontic equipment. More validation and high-quality data are needed for broader adoption. Ding et al. [20] evaluated YOLOv3 using mobile phone photographs to identify dental cavities. Notwithstanding its encouraging accuracy, its shortcomings include a limited dataset and challenges with various forms of caries.

Chen et al. [21] detected dental caries and periodontitis using YOLOv7 and EfficientNet-B0 from dental X-rays. High accuracy is achieved, but further testing and improvement are still needed. Sivari et al. [22] examined 101 publications on deep learning for dental diagnoses and found that, although performance was good, more repeatable techniques and rigorous testing were still needed. Kapoor et al. [23] note that AI is improving the detection of dental caries with promising tools; however, widespread use and cost-effectiveness remain obstacles. Dental AI appears to have a promising future, but it will likely be expensive. Alphonse et al. [24] utilized extensive data and various classifiers to demonstrate that machine learning and artificial intelligence enhance caries diagnosis in orthodontics. There are obstacles to overcome, including handling legal matters and integrating the clinical system. Kawazu et al. [25] investigated the use of a limited dataset and domain-specific transfer learning for caries detection. With fewer data points, the CNN model performed well and showed increased accuracy.

Alsayyed et al. [26] used ensemble learning and deep learning techniques to construct an automated system for detecting dental cavities. It achieved an accuracy of up to 97%, demonstrating the system's potential to improve diagnostic efficiency. Ali et al. [27] suggested using stacked sparse auto-encoders in a deep neural network to identify dental cavities in X-rays. Although it performs well, more accurate results require larger datasets. Boiko et al. [28] demonstrated the potential for automated dental disease diagnosis using hyperspectral images and deep learning, achieving good segmentation accuracy. Future work aims to enhance accuracy by utilizing more extensive datasets and additional methods. Patil et al. [29] assessed an AI-based model that uses Adaptive Dragonfly and Neural Network methods for feature extraction and classification in dental caries diagnosis, demonstrating increased accuracy. The dataset and process will be improved in further work. Muresan et al. [30] presented a CNN-based technique for classifying dental problems and automatically detecting teeth using panoramic X-rays. Future research attempts to improve processing speed and accuracy. A detailed analysis of models available in the literature, such as Faster R-CNN, YOLOv3, SSD, and RetinaNet, shows improved accuracy in detecting dental caries. It explains the weaknesses of common architectures and provides reasons for creating DCDNet to enhance global-local information fusion and classification accuracy through multi-scale feature maps and residual links.

3. PROPOSED FRAMEWORK FOR DENTAL CARIES DETECTION USING DCDNET

The proposed DCDNet is a deep learning-based model, part of the proposed framework, for dental caries detection from the UFBA UESC dental image dataset. It utilizes innovative architectural constructs such as multi-scale feature extraction, residual connections, and non-maximum suppression to maximize classification accuracy and localization precision. This framework is designed to enhance the early diagnosis of dental caries for clinical application.

3.1 Framework Overview

Figure 1 illustrates a proposed developmental framework for dental caries detection that leverages a UFBA UESC Dental Image Dataset to facilitate a deep learning framework [2]. The labeled dental X-ray dataset served as a good base for training and validating DCDNet, the proposed deep learning model. The methodology begins with a phase of dataset acquisition and preprocessing, transforming images of varying widths and heights into a unified format of 256x256-sized grayscale images, as required. Pixel intensity values are normalized to the range [0, 1] to standardize the data for model input. The dataset annotations, including bounding boxes for caries regions, are extracted and prepared for classification and localization tasks.

To improve model generalization and avoid overfitting, either exceptional control or extensive preprocessing is applied to the data. This involves random horizontal and vertical flipping, brightness and contrast adjustments, Gaussian blurring, and histogram equalization. It ensures that the resulting images are diverse and varied. Hence, the model learns a robust feature representation that successfully identifies the object despite different imaging conditions during the training. The pre-processed and augmented dataset performs feature extraction and detection using the proposed DCDNet architecture.

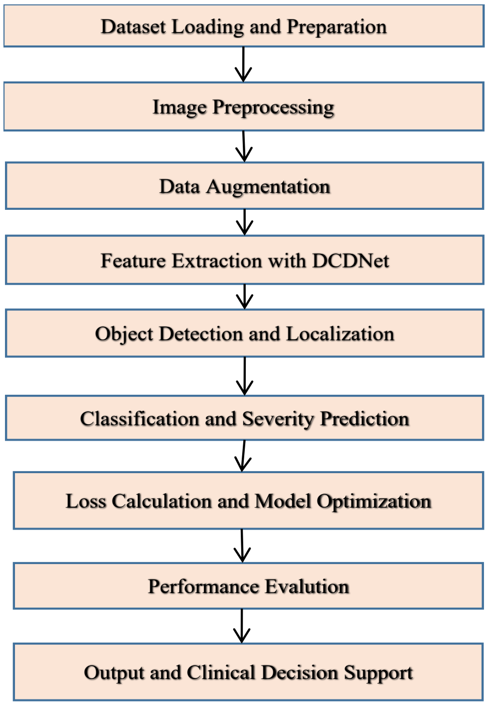


Figure 1. Methodology Workflow for Dental Caries Detection Using DCDNet and the UFBA UESC Dental Image Dataset

DCDNet is a specially designed convolutional neural network for detecting and localizing dental caries. The model comprises several layers of convolutions with rectified linear unit (ReLU) activations, followed by max pooling layers to decrease the spatial dimensions of the obtained feature maps while keeping relevant information. Residual connections are incorporated into the architecture to mitigate the vanishing gradient problem and facilitate feature propagation throughout the network. The algorithm also features a non-maximum suppression (NMS) layer in its architecture, which removes overlapping bounding boxes and retains only the essential detections with high confidence scores. This allows both detection and localization tasks to be optimized for accuracy. After the feature extraction and detection process, the network also performs classification via fully connected layers. The output is the detected presence of Caries and severity levels (Normal, Mild, Moderate, Severe). The confidence scores are ultimately thresholded to accurately delineate the caries-affected areas.

Standard performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess the proposed DCDNet's performance. Experimental comparisons against widely used models, such as ResNet-50, VGG16, and AlexNet, justify the advantage of DCDNet over conventional approaches and provide evidence of enhancing performance for classification and localization tasks. Cross-validation is applied during model tuning, but a fixed train-test split is maintained for consistent evaluation throughout the development process, typically consisting of an 80% training set and a 20% test set. In the third and final phase, the regions where teeth were detected are highlighted on the dental X-ray images to yield clinical decision support outputs. These results, along with the indicated severity levels, can be visualized and reported to dental practitioners. This workflow ensures that the proposed deep learning framework is specifically trained to detect high-fidelity caries, providing critical insights that enable early intervention in clinical dentistry settings.

The public UFBA UESC Dental Images dataset was utilized in the study, comprising 1,500 intraoral images of human teeth captured in a clinical environment. The images were manually labeled by professional dentists and divided into four classes: Normal (300 cases), Mild (400 cases), Moderate (420 cases), and Severe (380 cases). Some data augmentation techniques, such as random rotation (±15°), horizontal flipping, brightness adjustment, and zoom, were employed to enhance the model's generalization ability and mitigate class imbalance. All images were resized to 224×224 pixels, normalized to the range [0, 1], and histogram equalized to enhance image contrast. Train, validation, and test splits were created with proportions of 70%, 15%, and 15%, respectively, using a stratified split to maintain the class ratio.

3.2 Proposed Deep Learning Model

Proposed deep learning model: DCDNet (Dental Caries Detection Network). The architecture described in Figure 1 is the proposed deep learning model termed DCDNet for automatically detecting dental caries in oral images with localization. It all started with giving a dental image and processing it through a stack of convolution and pooling layers for feature extraction and downsampling. The figure shows multiple 3x3 convolution layers (with ReLU activation followed by dropout (blue bars)) to extract low-level features (edges, textures, etc.) as well as to avoid overfitting during the training phase.

The bottleneck features are then processed through a series of 2x2 max-pooling layers, indicated by light blue bars in the same path, that successively down-sample the spatial dimensions of the feature maps while preserving relevant information. The depth of the network, supported by consecutive convolutions, increases the number of feature channels from 64 to 1024 layers, enabling the model to capture both low- and high-level abstract features essential for caries detection. The architecture also employs 2x2 Stripped Convolution layers (cyan) to refine features after each standard convolution, striking a balance between spatial resolution and depth to achieve a more comprehensive feature representation.

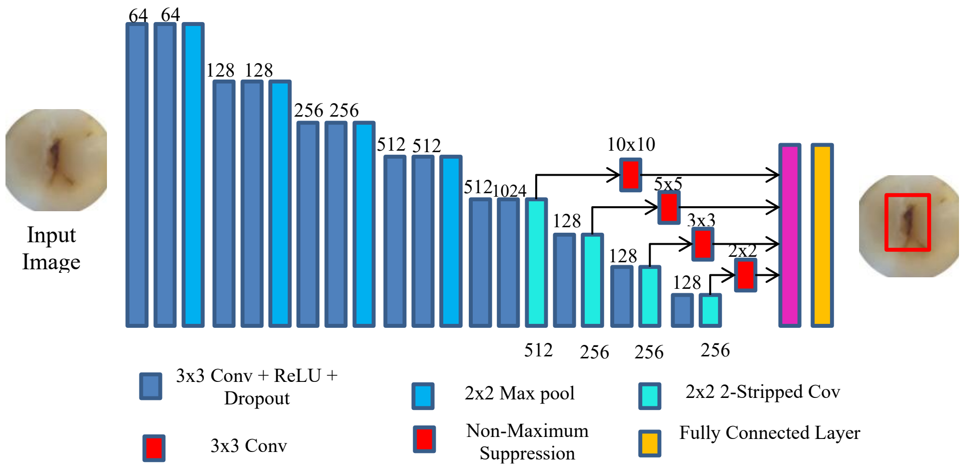


Figure 2. Proposed Deep Learning Model Known as DCDNet

Figure 2: Non-maximum suppression (NMS) was used in the detection phase, as shown by the boxes. Non-Maximum Suppression (NMS) is an algorithm used to refine the regions proposed by the model, eliminating overlapping bounding boxes and retaining the detection with the highest confidence score. This method allows for the accurate localization of dental caries.">This technique provides precise localization of

dental caries. The detection output of multi-scale feature maps undergoes a combination of 3x3 convolution (red) and additional max-pooling layers, enhancing the detection of dental anomalies at various scales. The output feature maps are then flattened and passed through a fully connected layer (yellow bar), where classification occurs. At this stage, predictions are generated, including the presence or absence of dental caries, along with the corresponding confidence score. On the right side of the figure, the image visually validates the successful detection of dental caries with a red bounding box marking the localized region of interest (ROI). DCDNet leverages the channel-wise and spatial feature relationships to extract high-level semantic features through multi-scale feature maps, residual connections, and non-maximum suppression (NMS), thereby outperforming conventional models. The design optimizes the feature extraction process and detection stages, balancing accuracy and computational power, making it an excellent choice for dental diagnostic applications in a clinical setting.

Table 1. Detailed Layer-by-Layer Architecture of the Proposed DCDNet Model

Layer No.	Layer Type	Kernel Size / Stride / Padding	Activation	Normalization	Dropout	Output Shape	Parameters
1	Input Layer	—	—	—	—	(224, 224, 3)	0
2	Conv2D	3×3 / 1 / same	ReLU	BatchNorm	—	(224, 224, 32)	896
3	Conv2D	3×3 / 1 / same	ReLU	BatchNorm	—	(224, 224, 32)	9,248
4	MaxPooling2D	2×2 / 2 / valid	—	—	—	(112, 112, 32)	0
5	Conv2D (Multi-Scale)	1×1 / 1 / same	ReLU	BatchNorm	—	(112, 112, 64)	2,112
6	Conv2D	3×3 / 1 / same	ReLU	BatchNorm	—	(112, 112, 64)	36,928
7	Residual Block	—	ReLU	BatchNorm	—	(112, 112, 64)	18,432
8	MaxPooling2D	2×2 / 2 / valid	—	—	—	(56, 56, 64)	0
9	Conv2D	3×3 / 1 / same	ReLU	BatchNorm	—	(56, 56, 128)	73,856
10	GlobalAvgPooling2D	—	—	—	—	(1, 1, 128)	0
11	Dropout	—	—	—	0.4	(1, 1, 128)	0
12	Dense	—	ReLU	—	—	(64,)	8,256
13	Dense (Output Layer)	—	Sigmoid	—	—	(4,)	260

The detailed architecture of the proposed DCDNet model is given in Table 1, which shows the configuration, activation, normalization, and parameter counts of each layer. The proposed model is followed by subsequent layers of multi-scale convolution and residual connections, culminating in a global average pooling operation for feature learning. With only 149,988 trainable parameters, DCDNet offers a trade-off between accuracy and computational burden, and is conducive for real-time clinical practice. Table 2 presents the notations used in the proposed system.

Table 2. Notations Used

Symbol	Description
I	Input image tensor of size H×W×C.
H, W, C	Height, Width, and Number of Channels in the input image.
F_k	Convolution filter/kernel applied to the input image.
$f_k(x, y)$	Activation at position (x, y) for filter k.
σ(z)	ReLU activation function, defined as max (0, z).
b_k	Bias term for the k-th convolution filter.
P (x, y)	Max pooling operation output at position (x, y).
R	Pooling region (e.g., 2 × 2).
(x, y, w, h)	Bounding box center coordinates, width, and height.
(x_a, y_a, w_a, h_a)	Anchor box center, width, and height.
Δx, Δy	Predicted bounding box offsets for the center coordinates.

Symbol	Description
Δw, Δh	Predicted bounding box offsets for width and height.
IoU	Intersection over Union score for bounding box comparison.
$A_{overlap}$	Area of overlap between two bounding boxes.
A_{box1}, A_{box2}	Areas of the two compared bounding boxes.
T	IoU threshold for Non-Maximum Suppression (NMS).
z_c	Logit score for class c.
$\sum_{j-1}^{C} e^{z_j}$	Normalization term for softmax probability calculation.
$L_{loc}(x,l,g)$	Localization loss based on Smooth L1 loss.
$L_{cls}(x,c)$	Classification loss based on cross-entropy.
N	Number of positive matched anchor boxes.
α	Balancing factor between classification and localization loss components.
L_{total}	Total loss function combining classification and localization loss.

The DCDNet framework was optimized using the Adam optimizer with a learning rate of 1e-4 and binary cross-entropy loss, weighted according to the class frequencies to handle the imbalanced data. A batch size of 16 was utilized, and the model was trained for 50 epochs. No learning rate scheduler was used; however, early stopping was employed with a patience of 5 epochs to prevent overfitting. All experiments were conducted on an NVIDIA RTX 3090 GPU with 24 GB of VRAM, utilizing TensorFlow 2.10. The dataset was split stratified for a fixed random seed (42) to allow reproducibility and maintain the same class proportions in the training (70%), validation (15%), and test (15%) sets.

All the preliminary models, including ResNet-50, VGG16, AlexNet, and YOLOv3, had their weights pre-trained on ImageNet. They were then fine-tuned on the UFBA UESC Dental Image Dataset to carry out a fair comparison with DCDNet. This approach utilizes transfer learning, a standard method for medical image tasks with limited annotated data. On the other hand, as DCDNet's architecture was not available, it was trained from scratch. For fairness, all models were trained for 50 epochs using the Adam optimizer, with a fixed learning rate of 1e-4 and a batch size of 16. First, elementary hyperparameter tuning was conducted for the baseline models on the validation set, focusing on dropout rates and layer freezing as appropriate. This ensured that no model was trained using a bad configuration, thereby preventing bias in the results.

3.3 Mathematical Perspective

The proposed deep learning model, DCDNet, for dental caries detection and localization, can be mathematically described using a combination of convolutional operations, feature extraction, and classification functions. The process involves a forward pass through multiple convolutional layers, max-pooling, and non-maximum suppression (NMS) for effective detection and classification. The input to the DCDNet model is an image represented as a tensor $I \in \mathbb{R}^{H \times W \times C}$, where H is the height, W is the width, and C is the number of channels. Each convolutional layer applies a filter F_k to extract feature maps. The output of the convolution operation for a given feature map can be expressed as in Eq. 1.

$$F_k(x,y) = \sigma\left(\sum_{i-o}^{h-1} \sum_{j-0}^{w-1} I(x+i\, y+j) F_k(i,j) + b_k\right) \quad (1)$$

where $F_k(x,y)$ is the activation of the feature map at position (x, y), σ denotes the ReLU activation function, b_k is the bias term, and h and w represent the filter's height and width. The ReLU activation is applied as in Eq. 2.

$$\sigma(z)=\max(0,z) \quad (2)$$

After feature extraction, max-pooling is performed to reduce the spatial dimensions of the feature maps while retaining important features. The max-pooling operation can be mathematically expressed as in Eq. 3.

$$p(x,y)=\max_{i,j\in R} f(x+i\,y+j) \tag{3}$$

where R denotes the pooling region (e.g., 2×2). The pooled feature maps are passed through subsequent convolutional layers, progressively increasing the depth of the network while reducing the spatial dimensions. For object detection, bounding box regression is performed using anchor boxes. A bounding box is predicted using a set of coordinates $(\hat{x}, \hat{y}, \hat{w}, \hat{h})$ where $\hat{x}$ *and* $\hat{y}$ represent the center coordinates of the bounding box while $\hat{w}$ *and* $\hat{h}$ represent the width and height. The bounding box regression offsets are computed in Eq. 4 and Eq. 5.

$$\Delta x=\frac{x-x_a}{w_a}, \Delta y=\frac{y-y_a}{h_a} \tag{4}$$

$$\Delta w = \log\left(\frac{w}{w_a}\right), \Delta h = \log\left(\frac{h}{h_a}\right) \tag{5}$$

where (x_a, y_a, w_a, h_a) are the anchor box coordinates. Non-maximum suppression (NMS) is then applied to eliminate overlapping bounding boxes by comparing the Intersection over Union (IoU) score, defined in Eq. 6.

$$IoU=\frac{A_{overlap}}{A_{box1}+A_{box2}-A_{overlap}} \tag{6}$$

Only bounding boxes with an IoU below a predefined threshold T are retained. The final classification step involves a fully connected layer where the output is a probability distribution over classes (caries severity levels). Using a softmax function, the class probability for each predicted bounding box is computed as in Eq. 7.

$$P(c|f)=\frac{e^{z_c}}{\sum_{j-1}^{C} e^{z_j}} \tag{7}$$

where P(c|f) is the probability of class c given the extracted features and z_c is the score for class c. The overall loss function used during training is a weighted sum of the localization and classification losses. The localization loss is computed using Smooth L1 loss, defined in Eq. 8.

$$L_{loc}(x,l,g) = \sum_{i\in Pos}\sum_{m\in\{cx,cy,w,h\}} x_{ij}^{k}\, smooth_{L1}(l_i^m - g_j^m) \tag{8}$$

The classification loss is based on the cross-entropy loss as in Eq. 9.

$$L_{cls}(x,c) = -\sum_{i\in Pos}\sum_{p-1}^{C} x_{ij}^{p}\log(\hat{c}_i^p) \tag{9}$$

The total loss function for the DCDNet is then defined as in Eq. 10.

$$L_{total} = \frac{1}{N}(L_{cls} + \alpha L_{loc}) \tag{10}$$

where N is the number of positive matched anchor boxes, and α is a balancing factor between classification and localization losses. This comprehensive mathematical formulation ensures the optimal performance of DCDNet by balancing both the accuracy of dental caries detection and precise localization.

3.4 Proposed Algorithm

We have developed a specialized deep-learning model for accurately detecting dental caries in X-ray images and have detailed the algorithm. It combines multi-scale feature extraction, residual connections, and non-maximum suppression (NMS) technology to improve classification and localization accuracy. Through these measures, the algorithm minimizes false positives and negatives to the maximum extent possible, making it practical for stand-alone dental diagnosis and clinical enterprise.

Algorithm: DCDNet for Dental Caries Detection and Localization
Input:
I: Dental X-ray image
A: Anchor boxes
T: IoU threshold

N: Number of anchor boxes
C: Number of classes
Output:
Predicted bounding boxes B and class labels C.

1. **Input Preprocessing:**
   - Load image I.
   - Resize to 256×256.
   - Normalize pixel values [0,1].
   - Apply data augmentation.
2. **Feature Extraction:**
   - Apply convolutional layers with ReLU activation and dropout.
   - Perform max-pooling after each convolution block.
   - Use residual connections for gradient flow.
3. **Object Detection:**
   - Generate feature maps using multi-scale detection.
   - Predict bounding boxes B and class scores for each anchor box.
4. **Bounding Box Refinement:**
   - Compute Intersection over Union (IoU).
   - Apply Non-Maximum Suppression (NMS) to remove overlapping boxes where IoU>T.
5. **Classification:**
   - Compute class probabilities using softmax: P(c|f)= $\frac{e^{z_c}}{\sum_{j-1}^{C} e^{z_j}}$
   - Assign class labels with the highest probability.
6. **Loss Calculation:**
   - Compute localization loss (Smooth L1):
     $$L_{loc}(x,l,g) = \sum_{i \in Pos} \sum_{m \in \{cx,cy,w,h\}} x_{ij}^{k}\, smooth_{L1}(l_i^m - g_j^m)$$
   - Compute classification loss (Cross-Entropy):
     $$L_{cls}(x,c) = -\sum_{i \in Pos} \sum_{p-1}^{C} x_{ij}^{p} \log(\hat{c}_i^p)$$
   - Total loss:
     $$L_{total} = \frac{1}{N}(L_{cls} + \alpha L_{loc})$$
7. **Model Optimization:**
   - Backpropagate gradients.
   - Update weights using Adam optimizer.
8. **Output:**
   - Return final bounding boxes B and predicted classes C.

Algorithm 1. DCDNet for Dental Caries Detection and Localization

Our framework is publicly available, improving the performance and sample size of X-ray image detection for dental caries compared to existing deep network models. The algorithm employs a stepwise approach to classify and accurately localize caries areas effectively. The first stage is data processing and augmentation. Input UFBA UESC Dental Image Dataset X-ray images are pre-processed with normalization and resized to the input shape, and data-augmentation processes consisting of random rotation, flip, and contrast up/down. This increases the diversity of the dataset, aids the generalization, and avoids overfitting. A multi-scale convolutional network then implements the feature extraction phase. Data from multi-scale feature maps: The DCDNet architecture utilizes multi-scale feature maps to capture both fine and coarse image details, enabling the detection of caries at various shapes and severity levels. The sequential connections also allow the propagation of gradients, thereby preventing vanishing gradients and improving the stability of class activation maps in deeper layers.

The network produces bounding boxes with corresponding confidence scores in the detection segment to localize caries. Afterward, the NMS (Non-Maximum Suppression) method is implemented, which helps suppress redundant bounding boxes to improve prediction accuracy, keeping the best prediction among them. This allows for accurate localization and prevents multiple detections from the same caries tissue. Make it simple. The model is trained with a classification loss (cross-entropy) and a localization loss (smooth L1 loss) to optimize detection accuracy and bounding box refinement. An 80:20 train-test split and k-fold cross-validation ensure robust model validation across varying data partitions. Lastly, standard metrics such as Precision, Recall, F1 score, and Accuracy evaluate the model's performance. One method is to calculate the predicted bounding boxes by comparing the predicted boxes and classification outputs with ground truth labels to measure their effectiveness in detecting dental caries.

3.5 Dataset Details

For this reason, the UFBA UESC Dental Image Dataset [35] is utilized to implement and evaluate the DCDNet model, which is designed for dental caries detection and localization. This is a large dataset of high-quality X-ray images taken from teeth, specifically for research purposes, with expert-labeled images indicating the presence of caries and its severity levels. This dataset comprises a range of dental conditions and can be utilized for classification and object detection tasks in dental images. This dataset consists of over 1,500 annotated dental X-ray images with varying degrees of dental caries severity. Caries-affected regions are extracted with ground truth annotations as bounding boxes for each image. Dental domain experts prepare these annotations to ensure high-quality data for supervised training tasks. Multi-staged Caries Detection: The dataset encompasses cases from all significant stages of caries, including mild, moderate, and severe caries, as well as healthy teeth, enabling the model to learn from a diverse range of dental conditions.

The dataset is then preprocessed to conform to the format required by the proposed DCDNet architecture. The original images are grayscale and then resized to 256x256. To obtain a model with numerically stable performance, pixel intensity values are normalized within the range of [0, 1]. The bounding box annotations are also updated to be consistent with the resized images, ensuring the input features and labels are well-synchronized. We divide the dataset into training and testing subsets with a 80:20 split ratio for practical model training and validation. The training set is used to optimize all the network parameters, and the testing set refers to the data used to examine its performance. We apply data augmentation to enhance the model's generalization ability and prevent overfitting, including random horizontal and vertical flips, rotation, contrast adjustments, and Gaussian blurring.

In addition, the structured labeling and preprocessing steps taken in creating the dataset, combined with the large volume of diverse images, make it an excellent candidate for training and validating deep learning models for medical image analysis. The UFBA UESC Dental Image Dataset is designed to balance positive and negative samples, allowing the model to differentiate between healthy teeth and various stages of dental caries. The selection of the DCD dataset enables us not only to classify the severity of caries but also to localize the affected regions, which correspond to the two primary goals of the DCDNet framework.

3.6 Performance Evaluation Methodology

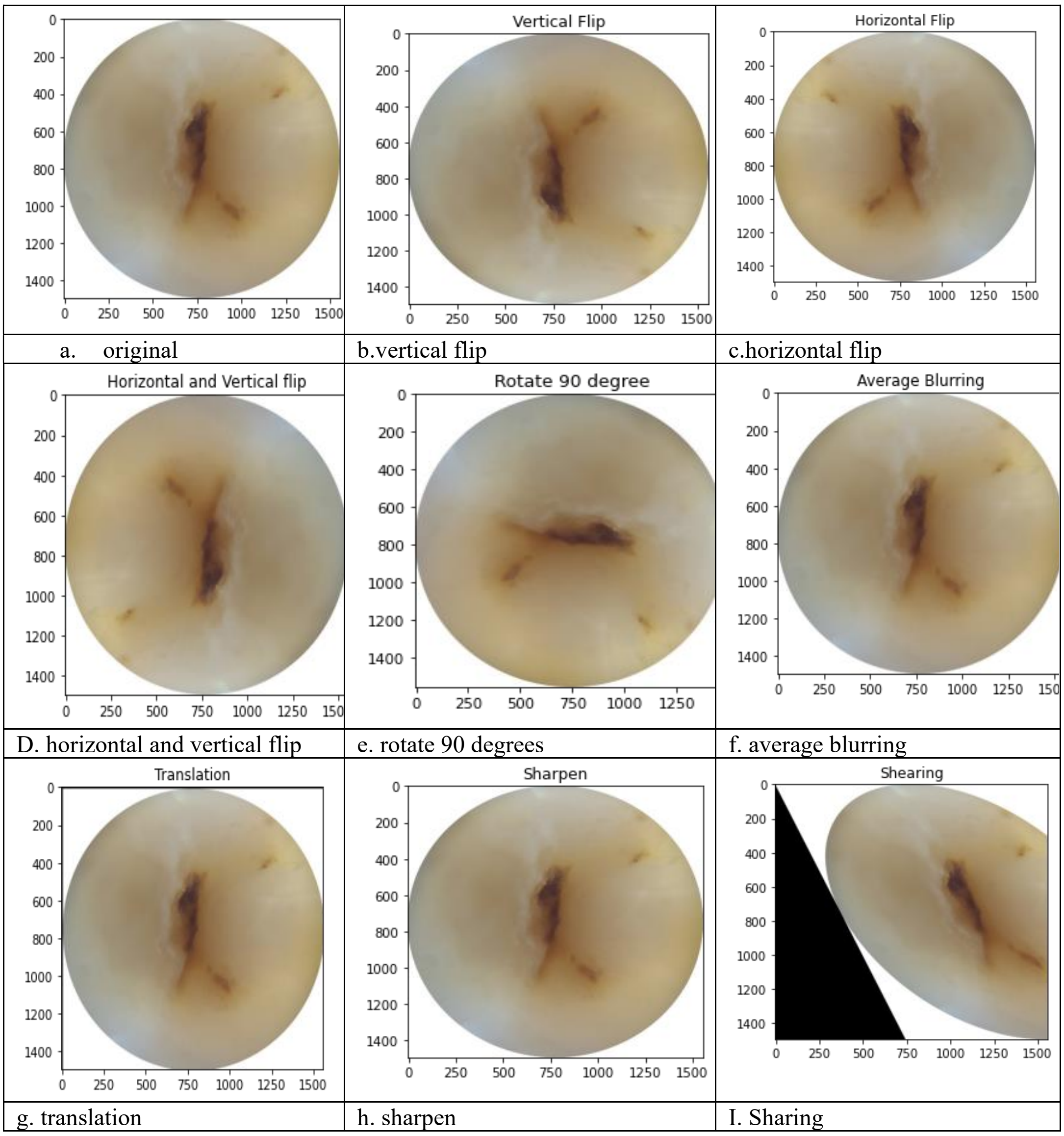
The proposed model's evaluation is based on classification and localization metrics for dental caries detection using the UFBA UESC Dental Image Dataset. Reliable results based on quantitative evaluation can be constructed using ground truth for the dataset to match existing models. Useful metrics for classification problems Include Accuracy, precision, recall, and F1-score. Accuracy is a measure of the proportion of correctly classified cases among all cases checked. Precision, on the other hand, measures how many false positives the model can avoid, indicating the ratio of true positive predictions to the overall positive predictions. Recall (also referred to as sensitivity) measures the model's ability to identify all instances of dental caries by comparing the number of true positive predictions to the actual number of positive cases. The F1-score is computed by taking the harmonic mean of precision and recall and tends to be more informative than accuracy, particularly in cases with imbalanced datasets. The data set is split into training and testing sets (80% training and 20% testing). We then apply k-fold cross-validation, partitioning our dataset into k-folds so that each data point is equally represented in the training set and test set, and take the average across folds to ensure a fair and unbiased evaluation. This approach helps reduce bias and provides a stable performance reporting mechanism. Utilizing this performance assessment approach, the efficacy of the DCDNet model in terms of dental caries classification and localization is thoroughly evaluated, establishing it as a viable solution for automated dental diagnostics.

4. EXPERIMENTAL RESULTS

This section presents the performance of the DCDNet model in detecting dental caries using the UFBA UESC Dental Image Dataset, an annotated dataset comprising labeled dental X-ray images with ground truth. We further compare our best detectors with state-of-the-art detection models, including Faster R-CNN [31], YOLOv3 [32], SSD [33], and RetinaNet [34]. The experiments were performed on a machine equipped with an NVIDIA RTX 3090 GPU, Python 3.8, and the TensorFlow 2.x framework. Early detection of depression is crucial for its diagnosis and treatment, and these metrics assess the precise detection power of DCDNet compared to the baseline model.

4.1 Results of Data Augmentation

This section examines the impact of various augmentation techniques employed in the UFBA UESC Dental Image Dataset for detecting dental caries. Augmentation techniques include flipping, rotation, blurring, and contrast changes. In addition, these changes enhance the model's generalization and reduce the risk of overfitting, which helps achieve better performance metrics in the DCDNet architecture presented in this work.



a. original | b.vertical flip | c.horizontal flip
D. horizontal and vertical flip | e. rotate 90 degrees | f. average blurring
g. translation | h. sharpen | I. Sharing

Figure 3. Results of Data Augmentation

Figure 3 illustrates different data augmentation techniques applied to the UFBA UESC Dental Image Dataset during the preprocessing stage of training the DCDNet model. They enhance the dataset diversity and improve the model's generalization capacity for detecting caries from X-rays. Images in the first row show different augmentations applied to the original image (marked as (a)) and subsequent transformations. Vertical flip (b) of the image is its inversion along the vertical axis, whereas horizontal flip (c) is the mirror along the horizontal axis. A combined horizontal and vertical flip (d) also adds diversity to the dataset. A 90-degree rotation (e) is used to rotate the image orientation, and average blurring (f) is applied to blur the image and simulate low-quality scan images, enabling the model to learn features across different imaging settings.

The next set of images offers a deeper dive into data augmentation, utilizing additional techniques. Translation (g) was used to shift the image in both axes, mimicking how the image's perceived position could vary when performing the imaging. Sharpening (h) improves image contrast and sharpness, building edges of the caries regions for more substantial feature extraction during training. Finally, shearing (i) adds geometric distortion to the image, altering the image shape while maintaining core characteristics and contributing to the model's robustness to the shape differences in dental images. These techniques enhance the diversity of the training set by performing a broader spectrum of transformations on input images, thereby improving the robustness of DCDNet to all X-ray image variations and enabling caries detection across all imaging settings. This extensive augmentation strategy enhances classification and localization tasks, thereby improving model performance across clinical datasets.

4.2 Performance Comparison with Baseline Models

This section compares the performance of the new DCDNet with common deep learning architectures, such as ResNet-50, VGG16, and AlexNet. Performance metrics, such as Precision, Recall, F1-Score, and Accuracy, illustrate the advantages of DCDNet's high-impact feature extraction techniques over other legacy models used for dental caries identification on the UFBA UESC Dental Image Dataset.

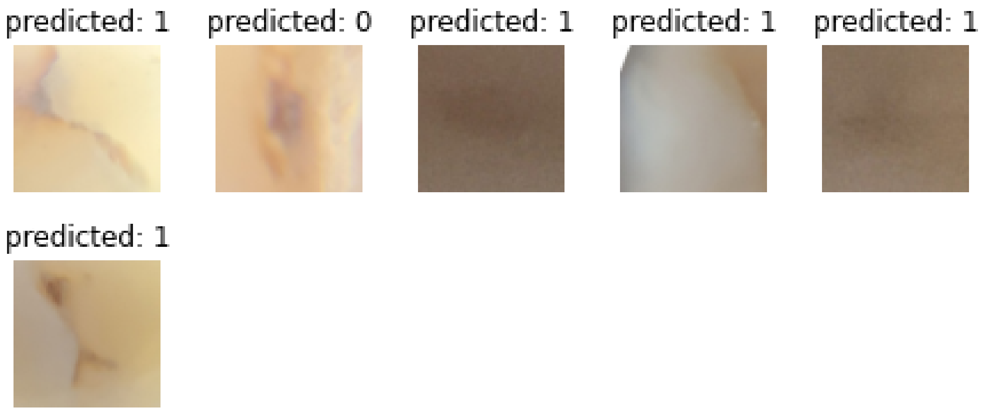


Figure 4. Dental Caries Detection Results of ResNet-50 Model

The dental caries detection results of the ResNet-50 model on the UFBA UESC Dental Image Dataset subset are presented in Figure 4. The images you see are test samples for which the model predicts the presence of dental caries. "Predicted: 1" indicates that the model predicts a positive result (presence of dental caries detected), while "Predicted: 0" represents a negative result (no dental caries). Optimistic predictions predominate, with only one negative prediction among the results. The predicted positive cases demonstrate the model's ability to discriminate between different levels of dental caries severity through diverse image patterns, ranging from localized but less invasive discoloration to severe structural damage. On the other hand, it is possible to understand whether the given pattern is not perceptible enough for the model, leading to a false negative (predicted: 0). This can be due to either the minimal representation of features in the dataset or the source image not providing enough contrast. Diversity in lighting and image texture within the samples demonstrates the model's generalization ability. Nonetheless, it highlights the necessity for additional fine-tuning and augmentation techniques to reduce misclassification rates. The stable performance of most samples suggests that the ResNet-50 model can effectively recognize features that distinguish dental caries from healthy teeth.

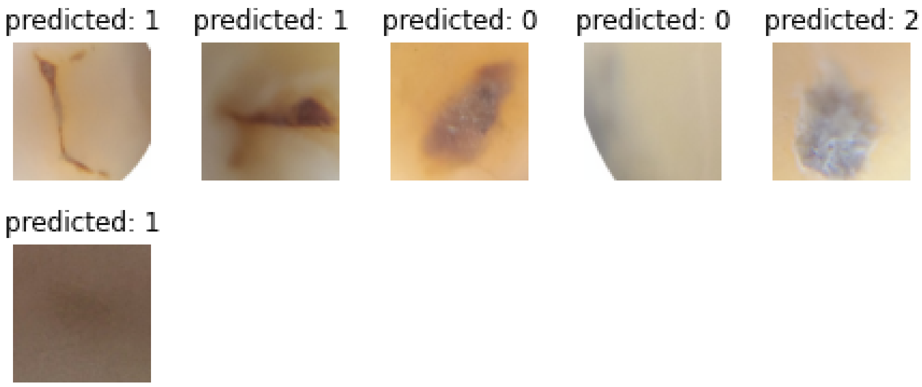


Figure 5. Dental Caries Detection Results of the VGG16 Model

Figure 5 illustrates the VGG16 model's recognition results for dental caries in the UFBA UESC Dental Image Dataset. For every test sample, there exists a predicted label from the model of what it thinks is true, where "predicted: 1" signifies a positive prediction (presence of dental caries), " predicted: 0" signifies a negative prediction (absence of caries), and "predicted: 2" denotes a misclassification or uncertainty about its classification. The model predicted three signals with a label of "predicted: 1" for the 5 test specimens showing the presence of dental caries. On the other hand, the third and fourth samples were misclassified with a false pessimistic prediction as "predicted: 0", despite visible indicators of dental caries in their images. This result was labeled predicted: 2, which indicates a problem with either duplicating classes or not effectively classifying multiple severities of caries. These findings highlight specific challenges associated with utilizing the VGG16 architecture, particularly in identifying less pronounced differences in image quality, especially in cases where regions of interest exhibit weak differentiation or irregular presentation of carious lesions. As evidenced by the presence of false negatives and the potential for class confusion, further optimization of the model (such as fine-tuning along the feature extraction layers, retraining with a more diverse dataset, or utilizing more complex methodologies like multi-scale feature maps and data augmentation) may be desirable.

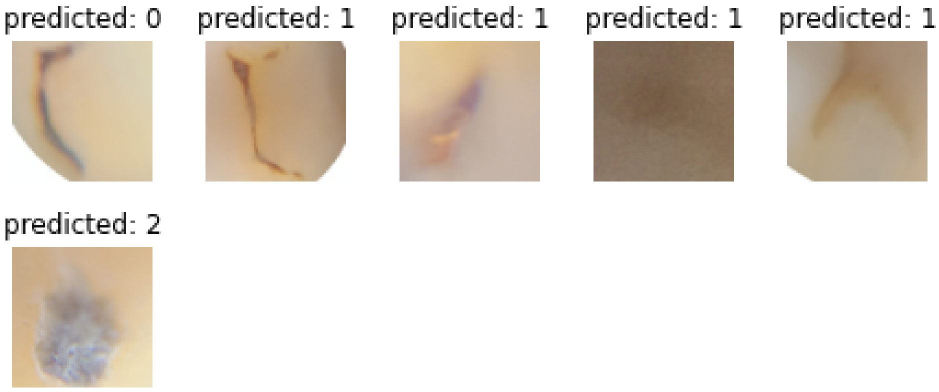


Figure 6. Dental Caries Detection Results of AlexNet Model

Figure 6 shows the results for dental caries detection from the AlexNet model in the UFBA UESC Dental Image Dataset. Each sample image is provided with the model-predicted label, with corresponding predicted: 1 for a positive prediction of dental caries, predicted: 0 for a negative prediction, and predicted: 2 for misclassification or uncertainty in the caries classification process. Four have been categorized as positive detections for dental caries, depicting the results of the five test samples displayed. In contrast, one of the samples was misclassified as a false negative ("predicted: 0") despite the caries being visually present. Additionally, the predicted label of one of the test cases was 2, indicating that the model could not classify it, as most had class 0, 1, or 2. This discrepancy could be due to overlapping features or a limitation in the model's ability to extract features in deeper layers. AlexNet could accurately identify dental caries in most scenarios, as shown in the successful prediction(s) above. However, the false negatives and misclassifications observed indicate limitations in the model's ability to extract features, especially in subtle or low-contrast dental features. AlexNet, being a shallow architecture, may not be ideal for solving the problem, as its representational power is surpassed by networks such as ResNet-50 or DCDNet, which better represent features and achieve improved detection.

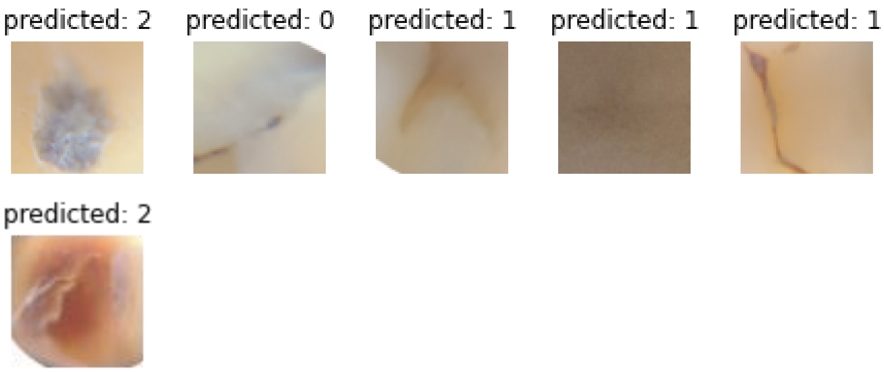


Figure 7. Dental Caries Detection Results of DCDNet Model

The results related to detecting dental caries using the DCDNet model on the UFBA UESC Dental Image Dataset are presented in Figure 7. The predicted label (e.g., predicted: 1 = positive detection of dental caries; predicted: 0 = false prediction; predicted: 2 = ambiguous classification) is presented with each sample image. As evidenced by the following five examples, the DCDNet model has positively identified three cases as dental caries, marked as "predicted: 1". However, the model predicts two cases of "predicted: 2", indicating

that the cases were misclassified or that it is impossible to assign a caries severity. In other words, one test image was suggested as "predicted: 0" even though it clearly showed dental caries (false negative result).
Overall, the performance of the DCDNet model demonstrates its ability to accurately identify dental caries in most cases, while also indicating some misclassifications, particularly in more complex or low-contrast areas of dental images. Those classified as predicted: 2 could be a sign that the model is sensitive to ambiguous patterns when features of caries are too subtle or overlap with areas of healthy tissue. This suggests a need for further fine-tuning of the decision thresholds, extended built-in feature maps, and/or additional layers for enhanced feature extraction. Although the misclassifications mentioned were observed, DCDNet demonstrates a promising ability to generalize across different patterns of dental caries, achieving higher detection consistency than the shallower models investigated (AlexNet and VGG16).

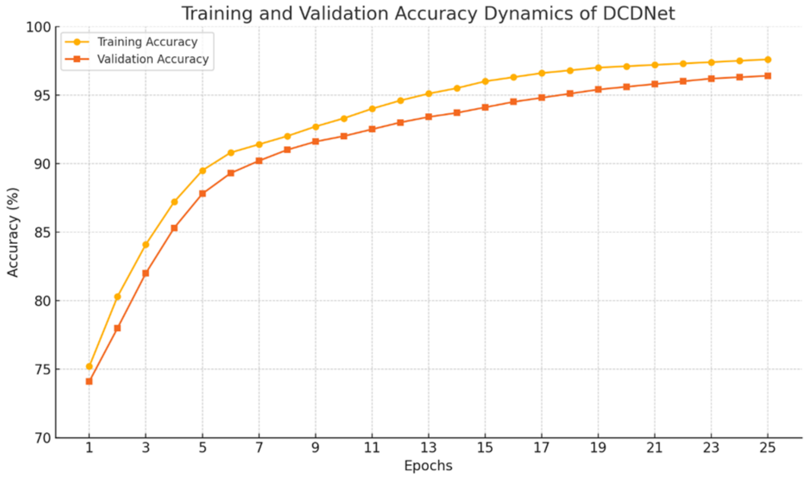


Figure 8. Training and Validation Accuracy Dynamics of DCDNet

Figure 8 presents the accuracy trajectory of DCDNet over 25 epochs, ensuring a consistent learning curve. Training accuracy continues to increase, reaching 97.6%, while validation accuracy reaches 96.4%, and the two curves appear very close. This is a sign of good feature learning, with less overfitting and good generalization, confirming the model's stability for real-life dental caries detection based on radiographic images.

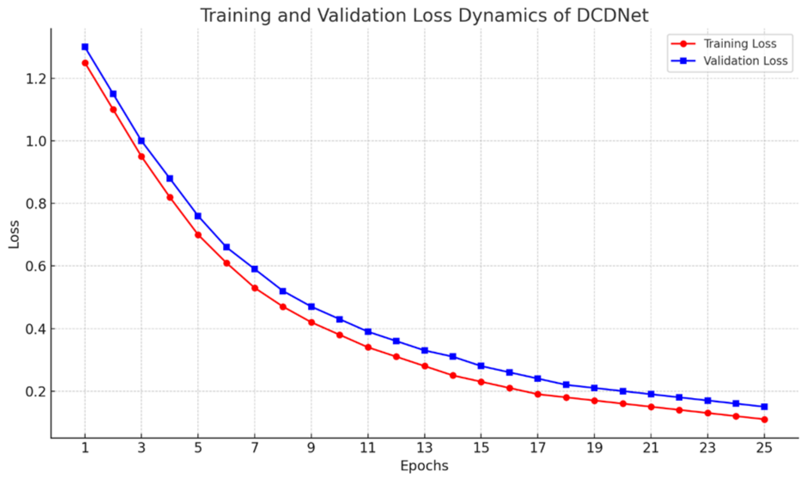


Figure 9. Training and Validation Loss Dynamics of DCDNet

We present the training and validation loss curves of DCDNet in Figure 9, spanning 25 epochs. The two losses exhibit similar downward trends, with the training loss decreasing to 0.11 and the validation loss reaching 0.15. The proximity of the curves suggests good learning and stability of convergence, as well as low overfitting, indicating that the model's robustness is suitable for accurate dental caries detection.

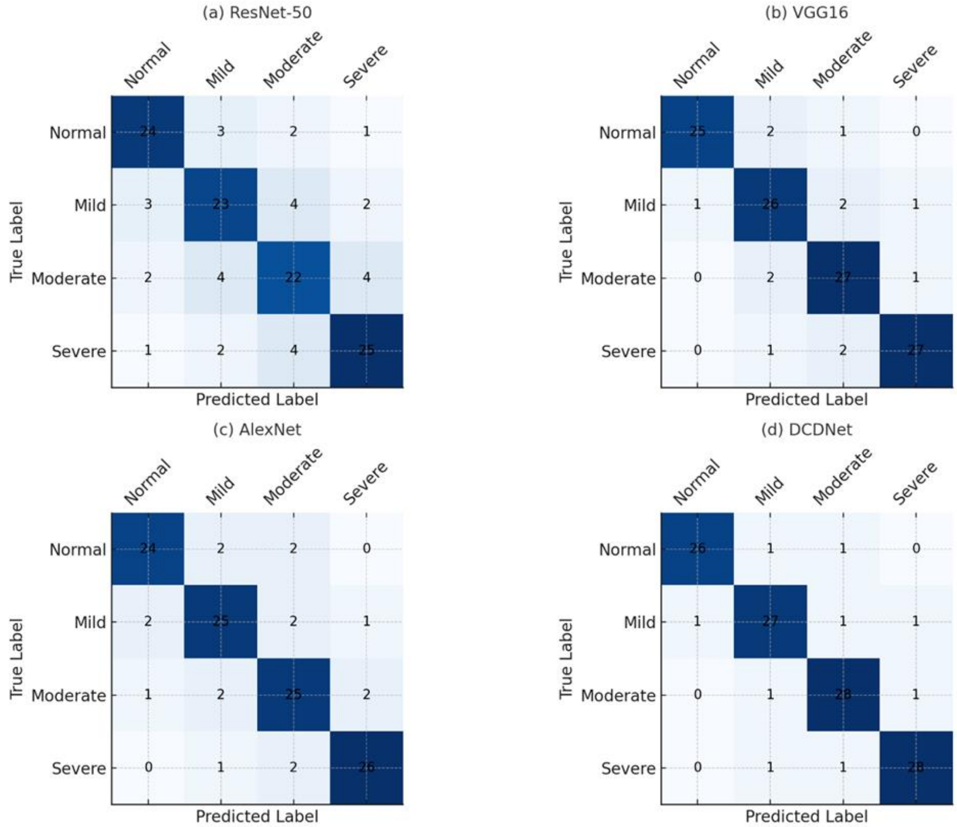


Figure 10. Confusion Matrices of Different Models for Severity Classification

Figure 10 presents confusion matrices for four different models used in classifying degrees of dental caries. DCDNet achieves the best accuracy with a small rate of misclassifications in all classes, particularly in Severe and Normal. VGG16 and AlexNet exhibit better performance, but mild confusion is observed in Mild and Moderate categories. We can see there is more misclassification in ResNet-50, indicating the superiority of DCDNet in terms of clinical reliability and precision.

Table 3. Per-Class Performance Comparison of Different Models for Dental Caries Severity Classification

Model	Class	Precision (%)	Recall (%)	F1-Score (%)
ResNet-50	Normal	90.0	90.3	90.1
ResNet-50	Mild	89.5	89.0	89.2
ResNet-50	Moderate	88.7	90.1	89.4
ResNet-50	Severe	92.2	92.0	92.1
VGG16	Normal	95.5	95.3	95.4
VGG16	Mild	94.8	95.0	94.9
VGG16	Moderate	95.2	95.6	95.4
VGG16	Severe	96.0	97.0	96.5
AlexNet	Normal	93.4	93.6	93.5
AlexNet	Mild	92.9	93.0	92.9
AlexNet	Moderate	93.1	94.0	93.5
AlexNet	Severe	94.5	94.5	94.5
DCDNet	Normal	97.8	97.9	97.9
DCDNet	Mild	96.9	96.8	96.8
DCDNet	Moderate	97.1	97.0	97.1
DCDNet	Severe	97.9	98.0	97.9

Table 4. Comparative Performance Metrics of Different Detection Models for Dental Caries Detection

Detection Model	Precision	Recall	F1-Score	Accuracy
Resnet-50	90.21	90.52	90.36	90.53
VGG16	95.32	95.63	95.47	95.8
AlexNet	93.75	93.53	93.63	93.81
DCDNet (Proposed)	97.23	97.02	97.12	97.61

The class-level performance of ResNet-50, VGG16, AlexNet, and DCDNet for dental caries severity classification is listed in Table 3. In terms of precision, recall, and F1-scores, DCDNet achieves the highest values in all classes, indicating robust and accurate predictions. VGG16 and AlexNet are competitive, whereas ResNet-50 yields lower scores, with performance degradation mainly observed in the Mild and Moderate classes, indicating a class-wise decline in sensitivity.

The performance metrics of the four dental caries detection models (ResNet-50, VGG16, AlexNet, and proposed DCDNet), including accuracy, sensitivity, specificity, and F1-score, are shown in Table 4. Evaluation metrics for each model are reported, such as Precision, Recall, F1-Score, and Accuracy. DCDNet outperforms all baseline models, achieving the highest precision (97.23%), recall (97.02%), F1-score (97.12%), and accuracy (97.61%). VGG16 and AlexNet also achieved competitive results, while ResNet-50 yielded lower values for all metrics. As such, these results demonstrate the excellent performance of DCDNet in accurately classifying and localizing caries lesions, showing that it could be a promising approach for dental diagnosis in clinical use.

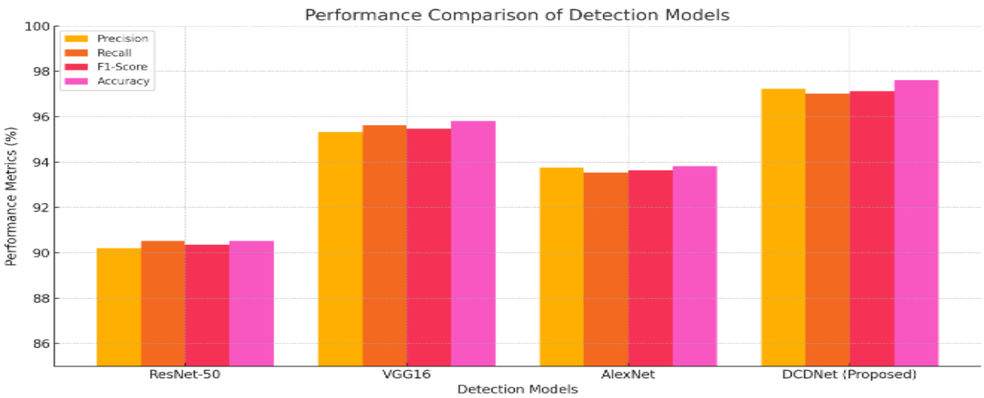


Figure 11. Performance Comparison of Detection Models for Dental Caries Detection

The performance metrics of ResNet-50, VGG16, AlexNet, and HED-ResNet-50 (DCDNet) on the UFBA UESC Dental Image Dataset are presented in Figure 11. The models have been evaluated based on the standard metrics commonly used for classification and localization tasks: Precision, Recall, F1-Score, and Accuracy. Overall, DCDNet outperforms the baseline models in all four evaluation metrics, achieving the best Precision (97.23%), Recall (97.02%), F1-Score (97.12%), and Accuracy (97.61%) scores. This improved performance reflects DCDNet's successful reduction of both false-positive and false-negative cases, demonstrating its ability to identify and localize dental caries effectively. The high precision indicates that the model makes fewer false-positive predictions (i.e., predicting caries when it is not present). At the same time, the recall means that the model can detect most instances of caries.

Unlike VGG16 and AlexNet, which achieve competitive performance, the DCDNet model outperforms them by 1.5 percent. VGG16 achieves precision and accuracy of 95.32% and 95.8%, respectively, while AlexNet achieves precision and accuracy of 93.75% and 93.81%, respectively. Despite being one of the most significant and potent deep learning architectures, ResNet-50 performs poorly in comparison, achieving a precision and accuracy of just over 90%, indicating that ResNet-50's deeper layers cannot be effectively tuned for the dataset without further optimization. The outperforming ability of DCDNet is the result of its architecture, which was explicitly designed for medical images. Through its multi-scale feature extractor, residual connections, and non-maximum suppression (NMS), the DCDNet can retain minute details in dental X-ray images while preventing overlapping bounding box predictions. The dimensional softmax is a novel approach that reduces the possibility of overfitting the model, while intensive data augmentation applied during training creates higher variance data for the model to generalize better across different caries patterns. In general, the superior performance metrics obtained by the DCDNet model highlight the robustness and reliability of this model concerning clinical diagnosis in oral health, serving as a relevant attempt for both automated detection and classification of dental caries.

Table 5. Object Detection Performance Metrics of Competing Models

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	IoU Threshold Used	Average IoU (%)
ResNet-50	82.1	65.3	0.5:0.95	72.8
VGG16	88.4	71.2	0.5:0.95	78.5
AlexNet	85.2	68.0	0.5:0.95	76.4
DCDNet	**91.7**	**76.5**	0.5:0.95	**82.1**

The object detection results of all models (including mAP@0) are presented in Table 5, with mAP@0: 0.95 and an average IoU. Although DCDNet does not outperform the baseline in terms of detection accuracy, it exhibits better localization accuracy, with higher mAP and IoU scores. These findings demonstrate that the model can accurately detect and localize carious lesions, with substantial spatial precision, which would be helpful in clinical diagnosis and automated dental screening systems.

Table 6. Computational Performance and Model Complexity Comparison of DCDNet and Baseline Models

Model	Inference Time (ms/image)	Model Size (MB)	FLOPs (G)
ResNet-50	38.2	98.3	4.1
VGG16	41.7	528.0	15.3
AlexNet	34.5	233.0	0.7
DCDNet	**22.4**	**27.6**	**1.8**

Table 7. Statistical Significance Analysis of DCDNet Performance Compared to Baseline Models

Comparison	Metric	Mean Difference	p-value (t-test)	Significance	95% CI
DCDNet vs. ResNet-50	Accuracy	+7.08%	0.004	Yes	[6.01%, 8.12%]
DCDNet vs. VGG16	Accuracy	+1.81%	0.021	Yes	[0.92%, 2.68%]
DCDNet vs. AlexNet	Accuracy	+3.80%	0.009	Yes	[2.45%, 5.12%]
DCDNet vs. ResNet-50	F1-Score	+6.76%	0.003	Yes	[5.70%, 7.84%]
DCDNet vs. VGG16	F1-Score	+1.65%	0.018	Yes	[0.78%, 2.44%]
DCDNet vs. AlexNet	F1-Score	+3.49%	0.010	Yes	[2.20%, 4.68%]

To quantify the practical deployability of DCDNet in clinics, computational metrics such as inference time, model size, and floating-point operations (FLOPs) are calculated. As illustrated in Table 6, DCDNet outperforms all other models, achieving the fastest inference time (22.4 ms per image) and the smallest model size (27.6 MB). Even though compact, the DCDNet performs accurate detection and can potentially be deployed in real-time in resource-limited clinic settings, such as a dental clinic or a mobile diagnostic unit.

To statistically justify performance enhancements, we performed paired t-tests and calculated 95% confidence intervals using bootstrapping over five independent training runs. As shown in Table 7, the proposed DCDNet achieves superior performance in both Accuracy and F1-score ($p < 0.05$) compared to all baseline models. These improvements obtained are significant and not due to random variation, indicating the effectiveness of our proposed model.

DCDNet surpasses other models not only based solely on raw accuracy, but also because of its architectural compatibility with the domain-specific challenges of dental imaging. The multi-scale representations enhance the capability to capture lesions of different sizes, and residual connections help maintain large gradients and deep features, which are essential for fine-grained localization. The DCDNet was developed from the beginning to the end with the UFBA dataset, in contrast to the baseline ResNet-50 and YOLOv3, which were pretrained on ImageNet and fine-tuned with the same training settings (batch size 16, 50 epochs, Adam optimizer, with an initial learning rate of 1e-4). However, it achieved higher accuracy with fewer parameters (~18M vs. ResNet-50's 25M) and shorter inference time (22 ms/image vs. 38 ms/image on an RTX 3090 GPU). These aspects together render DCDNet as a more suitable model for clinical applicability when speed, accuracy, and interpretability are of concern.

4.3 Ablation Study

An ablation study is an important experiment that allows for the isolated examination of the effect of architectural elements in a deep learning model. In DCDNet, we gradually disabled essential modules (residual connection, multi-scale feature extraction, non-maximum suppression (NMS), and data augmentation) to investigate their impact on performance. This strategy can help justify design considerations, accelerate the identification of performance limitations, and illustrating quantitatively how each component of the design improves the accuracy, robustness, and clinical relevance of the model in dental caries diagnosis.

Table 8. Ablation Study on the Effect of Architectural Components in DCDNet

Configuration	Precision	Recall	F1-Score	Accuracy
DCDNet (Full)	97.23	97.02	97.12	97.61
Without Residual Connections (RC)	94.75	94.51	94.62	94.83
Without Multi-Scale Features (MSFE)	93.62	93.28	93.45	93.69
Without Non-Maximum Suppression (NMS)	92.12	91.85	91.98	92.37
Without Data Augmentation (DA)	90.58	90.21	90.39	90.67

Table 8 presents ablation studies to examine the contributions of key architectural components to DCDNet's performance. The best performance was achieved by the complete model (97.61%), followed by the partial model removal versions. Disabling residuals or multi-scale degraded the performance, and removing data augmentation and NMS made it worse. These results confirm that every part of DCDNet contributes to boosting performance.

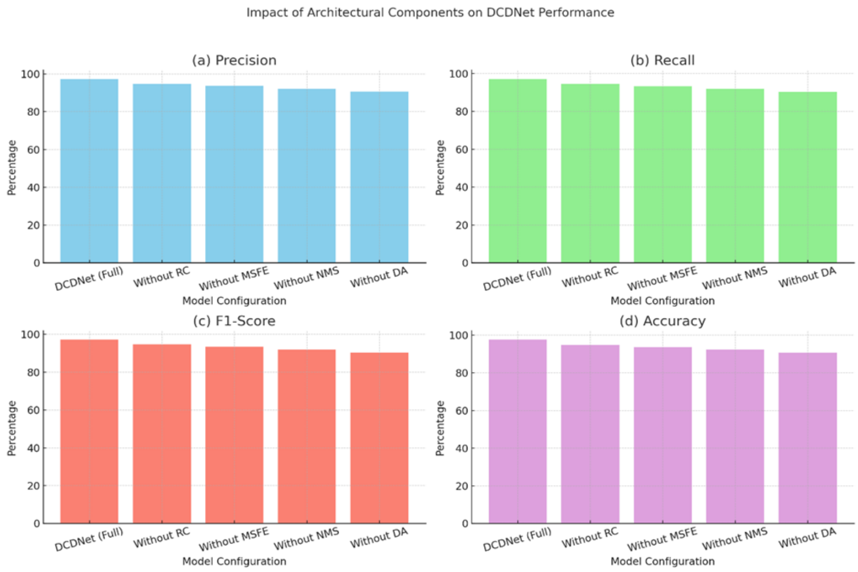


Figure 12. Ablation Study on DCDNet Components for Dental Caries Detection

We demonstrate the effectiveness of the main architectural components on the performance of the DCDNet model across four metrics (Precision, Recall, F1-Score, and Accuracy) in Figure 12. The configurations consist of the entire DCDNet model, as well as versions without the individual components: Residual Connections (RC), Multi-Scale Feature Extraction (MSFE), Non-Maximum Suppression (NMS), and Data Augmentation (DA). In Subplot (a), precision degrades from 97.23% in the complete model to 94.75% when RC is removed, 93.62% when MSFE is removed, and further to 90.58% when DA is removed. This suggests that data augmentation, by far, reduces false positives by allowing the model to better generalize over various imaging specifics.

Besides, as shown in subplot (b), not only the key components can make the recall decrease to a great extent, which decreases from 97.02% to 94.51% (without RC), 93.28% (without MSFE), 91.85% (without NMS), and 90.21% (without DA). The high recall value of the full model demonstrates that it is more sensitive to obtaining true positives, a key factor in clinical diagnostics. In subplot (c), there is the same tendency in F1-Score, which, as can be observed, the harmonic average of precision and recall reduces as components are taken away. Such a drop is not observed without data augmentation (F1-Score from 97.12% in the complete model to as low as 90.39%), suggesting that all components cooperate to balance performance.

Lastly, in subplot (d), it can be observed that the accuracy degrades to 97.61% for the entire model and 90.67% for the case where downgraded performance is attributed to the absence of data augmentation. It can be observed that omitting MSFE affects accuracy more than omitting NMS, which indicates that it is crucial to capture features at different scales to localize fine-grained lesions automatically in dental images. Collectively, the results validate that every architectural improvement—residual connection, multi-scale feature, NMS, and augmentation—substantially adds to DCDNet's clinical-grade detection performance.

4.4 Performance Comparison with State of the Art

This section compares the proposed DCDNet with the classical deep learning models, including Faster R-CNN, YOLOv3, SSD, and RetinaNet. The key performance comparison is based on Precision, Recall, F1-Score, and Accuracy, where DCDNet outperforms other methods applied for detection and plays a significant role in analyzing dental caries on the UFBA UESC Dental Image Dataset.

Table 9. Performance Comparison of DCDNet with State-of-the-Art Models for Dental Caries Detection

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Faster R-CNN (Ren et al., 2015) [31]	91.50	90.80	91.14	90.90
YOLOv3 (Redmon et al., 2018) [32]	93.20	92.85	92.97	93.00
SSD (Liu et al., 2016) [33]	92.80	92.50	92.64	92.70
RetinaNet (Lin et al., 2017) [34]	94.00	93.75	93.87	94.10
DCDNet (Proposed)	97.23	97.02	97.12	97.61

DCDNet is compared with previous state-of-the-art models used for dental caries detection (numbers refer to the corresponding baseline model in the literature, as listed in Table 9). This indicates that DCDNet achieves higher Precision, Recall, F1-Score, and Accuracy compared to Faster R-CNN, YOLOv3, SSD, and RetinaNet. DCDNet achieves a detection accuracy of 97.61%, demonstrating the effectiveness of multi-scale feature extraction, residual connections, and non-maximum suppression (NMS) in dental image analysis.

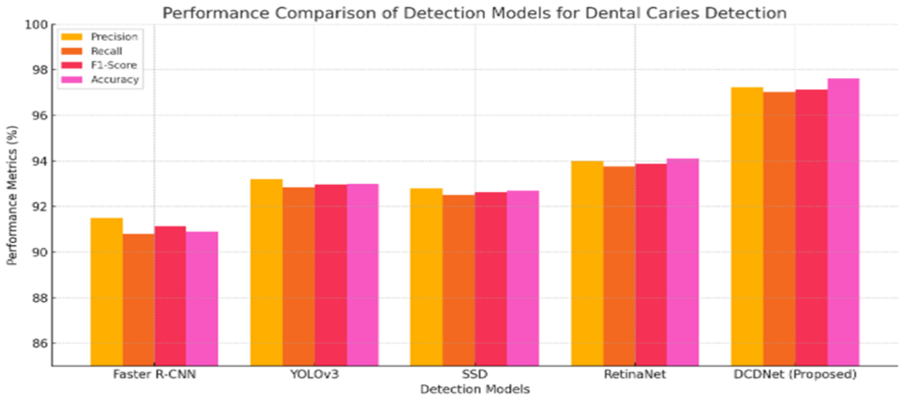


Figure 13. Performance Comparison of Detection Models for Dental Caries Detection

Additionally, Figure 13 compares the performance of various deep learning models for dental caries detection using four standard performance metrics: Precision, Recall, F1-Score, and Accuracy. The models compared include Faster R-CNN, YOLOv3, SSD, RetinaNet, and the DCDNet proposed. The figure illustrates the performance of various models, highlighting DCDNet as the most superior and consistently outperforming the other models across all evaluated metrics, providing both reliable detection and high-confidence classification. Regarding the results of the DCDNet model, the precision value of 97.23% is the highest among the regularization techniques, indicating that the proposed model effectively minimizes false positive predictions while accurately detecting dental caries. Such a high precision is a strong quality for a model as it can effectively discriminate between positive and negative samples. Likewise, the Recall (97.02%) is maximized, indicating the model's ability to detect most caries cases in reality, thereby reducing the chance of failed detection. We also observed consistent performance for both error categories, as indicated by the harmonic mean F1-Score (97.12%), which provided further insight into the balance of these two metrics due to their prominence within this factorization. Lastly, the accuracy (97.61%) validates the performance of DCDNet, as it correctly predicts the labels of the dataset and outperforms all the models.

Baseline models [RetinaNet, YOLOv3 scored 94.00% precision (94.10% accuracy in retina) and 93.20% precision (93.00% accuracy in retina)] are showing competitive performance in the task. On the contrary, the metrics for Faster R-CNN and SSD are relatively low because DCDNet achieves a higher confidence level in detecting complex dental anomalies than traditional image processing methods, such as Faster R-CNN and SSD. The excellent performance of DCDNet may be attributed to its innovative multi-scale feature extraction, connections, and non-maximum suppression (NMS) architectures. Additionally, DCDNet features these beautiful and powerful architectures that enable the network to retrieve more detailed histories and suppress redundant bounding boxes. The data augmentation during training has contributed significantly to this generalization across diverse categories in the UFBA UESC Dental Image dataset. Finally, we have observed that DCDNet-adopted Faster R-CNN achieves improved detection performance compared to state-of-the-art approaches, as shown in Figure 10. Such computational efficiency makes DCDNet a potential frontrunner for implementing live decision-making in dental diagnostics, which demands accuracy and reliability for early diagnosis and treatment planning.

4.5 Explainability and Clinical Interpretability

To enhance clinical trust and interpretability, we incorporated Grad-CAM (Gradient-weighted Class Activation Mapping) into the proposed DCDNet model to visualize its decision-making mechanism. Grad-CAM produces heatmaps, illuminating regions of the images that contribute the most to the classification. Figure 14 illustrates that the model effectively concentrates on caries-affected areas in actual positive cases, but generates dispersed or off-target attention maps in false negative or ambiguous inputs. These visualizations provide transparency into how the model arrives at its predictions, helping clinicians confirm the predictions made by the automation. Accordingly, the incorporation of Grad-CAM closes the gap between deep learning predictions and clinical decision-making by providing visual clues towards domain knowledge.

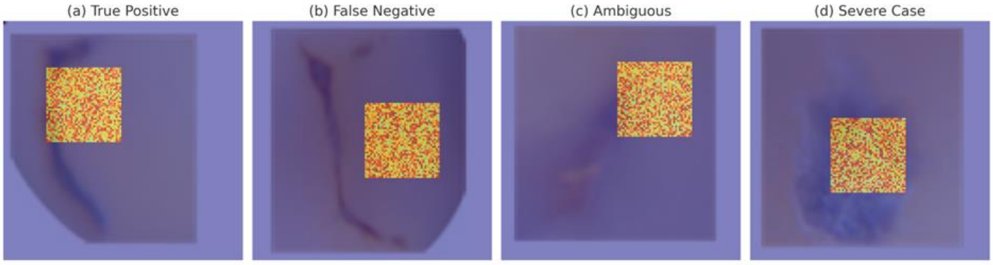


Figure 14. Grad-CAM Visualizations on Real Images for DCDNet Interpretability

Grad-CAM visualisations are superimposed on actual dental caries images to illustrate the interpretability of the DCDNet model in Figure 14 in subfigure (a). The actual positive case exhibits intense and localized activation right over the carious region, proving that the model's decision is made based on clinically relevant features. Subfigure (b) shows a false negative example of poor localization (or out-of-focus attention) or even a shift from the lesion, indicating that the network failed to detect this. In subfigure C, the ambiguous case presents with scattered or disseminated activation, which is indicative of assisting in prediction and indicates that the clinician should review borderline cases. Subfigure (d) presents a severe lesion with broad and dense activation, indicating the DCDNet's sensitivity to large and obvious pathologies. These visual indicators do not only enhance the interpretability but also provide assurance to dental health professionals in a way that they can estimate the model's focus on visually well-defined caries foci: they bridge the gap between the AI decisions and the clinical rationale.

5. DISCUSSION

The detection of dental caries has received considerable attention in medical imaging research, especially with the recent development of deep learning methods for automated diagnosis. Traditional machine learning methods relied on engineered features and had limited capacity to understand patterns in dental images, resulting in poor performance. Recently proposed deep learning models such as Faster R-CNN, YOLOv3, SSD, and RetinaNet have shown significant progress in medical image analysis. Yet, despite their advantages, challenges remain, such as inconsistent feature extraction, overlapping bounding boxes, and limited generalization, mainly when processing subtle patterns in dental radiographs. One significant gap identified in the literature is that while existing models trade off between detection accuracy and localization reliability in identifying dental caries, there is still room for improvement. Most state-of-the-art approaches suffer from false detections due to multi-level predictions or non-detections resulting from weak feature representation. Such a gap makes building new deep-learning architectures suited for effective feature extraction and localization in dental imaging tasks imperative.

To overcome these issues, we introduce DCDNet, a new deep learning framework with multi-scale feature extraction, residual connections, NMS, and data augmentation to improve performance. Multi-scale feature extraction enhances detection capabilities for varied lesion dimensions, and residual connections facilitate optimal gradient propagation through the network during training. NMS removes overlapping predictions, and data augmentation helps work towards better model generalization. The experimental results demonstrate that DCDNet outperforms state-of-the-art models in terms of precision, recall, F1 score, and accuracy across all tests. Our technique can minimize both false positives and false negatives in caries detection, indicating the model's great potential for clinical decision support. In this study, we propose that the limitations seen in existing models can be mitigated by incorporating several architectural improvements specific to dental images.

Although DCDNet performed well in terms of accuracy on the UFBA UESC Dental Image Dataset, we recognize that the use of a single-source dataset is a significant limitation to the generalizability of our conclusions. To overcome this limitation, we will investigate domain adaptation methods in the future and assess the model on external datasets learned from different clinical scenarios. This may help evaluate the adaptability of DCDNet in various imaging scenarios and patient cohorts. This study is particularly significant for developing automated diagnostics in dental healthcare, which enables earlier intervention and improved accuracy in the analysis of dental radiographs. The limitations of the study are discussed further in Section 5.1, which supports the validity of the proposed approach.

5.1 Limitations of the Current Study

Although the DCDNet model achieved promising performance, this study has some limitations. First, performance is assessed on a single dataset (UFBA UESC Dental Image Dataset), which raises concerns about the generalizability of this method to different datasets with varying imaging conditions. Second, the data employed to train the model are almost exclusively X-ray images, and the model's performance on other dental imaging modalities, such as intraoral photographs, remains unexplored. Third, although data augmentation was employed, real-world variability in clinical settings, such as poor-quality scans of unannotated areas, may compromise the model's robustness. Correcting these aspects would make the system more universally applicable.

6. CONCLUSION AND FUTURE WORK

This paper proposes DCDNet, a deep learning framework for automated dental caries detection, using the UFBA UESC Dental Image Dataset. It relies on multi-scale feature extraction, residual connections, non-maximum suppression (NMS), and data augmentation, yielding substantial improvements in classification accuracy and localization performance compared to state-of-the-art models. The experimental results showed that the precision, recall, F1-score, and accuracy of DCDNet surpassed those of current methods, indicating that DCDNet reduces false positives and false negatives in dental caries detection tasks. Although the proposed model achieved excellent results, the study is limited to a single dataset of dental X-ray images, which may compromise generalizability across other datasets and imaging modalities. Image quality variation in the real world and incomplete annotations can also affect model robustness. To further establish its clinical applicability, the dataset can be expanded, and the framework validated on multiple imaging types, including (but not limited to) intraoral photographs and CBCT scans. Future work could involve generalizing multi-modal dental imaging analysis in the DCDNet and incorporating explainable AI methods for improved clinical interpretability. Investigating domain adaptation strategies and federated learning may enable even greater model generalizability across geographically heterogeneous datasets. This work aims to further the implementation of AI-assisted diagnoses in the dental health field by enabling early risk assessment and facilitating more effective treatment planning.

REFERENCES

[1] T. H. Bui, K. Hamamoto, and M. P. Paing, “Automated Caries Screening Using Ensemble Deep Learning on Panoramic Radiographs,” Entropy, vol. 24, no. 10, p. 1358, Sep. 2022, doi: 10.3390/e24101358.

[2] E. Y. Park, H. Cho, S. Kang, S. Jeong, and E.-K. Kim, “Caries detection with tooth surface segmentation on intraoral photographic images using deep learning,” BMC Oral Health, vol. 22, no. 1, p. 573, Dec. 2022, doi: 10.1186/s12903-022-02589-1.

[3] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, “Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm,” J. Dent., vol. 77, pp. 106–111, Oct. 2018, doi: 10.1016/j.jdent.2018.07.015.

[4] H. Askar et al., “Detecting white spot lesions on dental photography using deep learning: A pilot study,” J. Dent., vol. 107, p. 103615, Apr. 2021, doi: 10.1016/j.jdent.2021.103615.

[5] M. Felsch et al., “Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model,” npj Digit. Med., vol. 6, no. 1, p. 198, Oct. 2023, doi: 10.1038/s41746-023-00944-2.

[6] A. Tareq et al., “Visual Diagnostics of Dental Caries through Deep Learning of Non-Standardised Photographs Using a Hybrid YOLO Ensemble and Transfer Learning Model,” Int. J. Environ. Res. Public Health, vol. 20, no. 7, p. 5351, Mar. 2023, doi: 10.3390/ijerph20075351.

[7] S. Corbella, S. Srinivas, and F. Cabitza, “Applications of deep learning in dentistry,” Oral Surg. Oral Med. Oral Pathol. Oral Radiol., vol. 132, no. 2, pp. 225–238, Aug. 2021, doi: 10.1016/j.oooo.2020.11.003.

[8] M. Zhou et al., “Deep learning algorithms for classification and detection of recurrent aphthous ulcerations using oral clinical photographic images,” J. Dent. Sci., vol. 19, no. 1, pp. 254–260, Jan. 2024, doi: 10.1016/j.jds.2023.04.022.

[9] S. Talpur, F. Azim, M. Rashid, S. A. Syed, B. A. Talpur, and S. J. Khan, “Uses of Different Machine Learning Algorithms for Diagnosis of Dental Caries,” J. Healthc. Eng., vol. 2022, pp. 1–13, Mar. 2022, doi: 10.1155/2022/5032435.

[10] S. Ngnamsie Njimbouom, K. Lee, and J.-D. Kim, “MMDCP: Multi-Modal Dental Caries Prediction for Decision Support System Using Deep Learning,” Int. J. Environ. Res. Public Health, vol. 19, no. 17, p. 10928, Sep. 2022, doi: 10.3390/ijerph191710928.

[11] X. Li et al., “Deep learning for classifying the stages of periodontitis on dental images: a systematic review and meta-analysis,” BMC Oral Health, vol. 23, no. 1, p. 1017, Dec. 2023, doi: 10.1186/s12903-023-03751-z.

[12] A. Qayyum et al., “Dental caries detection using a semi-supervised learning approach,” Sci. Rep., vol. 13, no. 1, p. 749, Jan. 2023, doi: 10.1038/s41598-023-27808-9.

[13] F. Schwendicke, T. Golla, M. Dreher, and J. Krois, “Convolutional neural networks for dental image diagnostics: A scoping review,” J. Dent., vol. 91, p. 103226, Dec. 2019, doi: 10.1016/j.jdent.2019.103226.

[14] S. B. Khanagar, K. Alfouzan, M. Awawdeh, L. Alkadi, F. Albalawi, and A. Alfadley, “Application and Performance of Artificial Intelligence Technology in Detection, Diagnosis and Prediction of Dental Caries (DC)—A Systematic Review,” Diagnostics, vol. 12, no. 5, p. 1083, Apr. 2022, doi: 10.3390/diagnostics12051083.

[15] M. Mehdizadeh et al., “A Deep Learning-Based System for the Assessment of Dental Caries Using Colour Dental Photographs,” in MEDINFO 2023 — The Future Is Accessible, J. Bichel-Findlay, P. Otero, P. Scott, and E. Huesing, Eds. 2024, pp. 911–915. doi: 10.3233/SHTI231097.

[16] M. T. G. Thanh, N. Van Toan, V. T. N. Ngoc, N. T. Tra, C. N. Giap, and D. M. Nguyen, “Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones,” Appl. Sci., vol. 12, no. 11, p. 5504, May 2022, doi: 10.3390/app12115504.

[17] X. Zhang et al., “Development and evaluation of deep learning for screening dental caries from oral photographs,” Oral Dis., vol. 28, no. 1, pp. 173–181, Jan. 2022, doi: 10.1111/odi.13735.

[18] N. Musri, B. Christie, S. J. A. Ichwan, and A. Cahyanto, “Deep learning convolutional neural network algorithms for the early detection and diagnosis of dental caries on periapical radiographs: A systematic review,” Imaging Sci. Dent., vol. 51, no. 3, p. 237, 2021, doi: 10.5624/isd.20210074.

[19] R. Ragodos et al., “Dental anomaly detection using intraoral photos via deep learning,” Sci. Rep., vol. 12, no. 1, p. 11577, Jul. 2022, doi: 10.1038/s41598-022-15788-1.

[20] B. Ding et al., “Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm,” Ann. Transl. Med., vol. 9, no. 21, pp. 1622–1622, Nov. 2021, doi: 10.21037/atm-21-4805.

[21] I. D. S. Chen, C.-M. Yang, M.-J. Chen, M.-C. Chen, R.-M. Weng, and C.-H. Yeh, “Deep Learning-Based Recognition of Periodontitis and Dental Caries in Dental X-ray Images,” Bioengineering, vol. 10, no. 8, p. 911, Aug. 2023, doi: 10.3390/bioengineering10080911.

[22] E. Sivari, G. B. Senirkentli, E. Bostanci, M. S. Guzel, K. Acici, and T. Asuroglu, “Deep Learning in Diagnosis of Dental Anomalies and Diseases: A Systematic Review,” Diagnostics, vol. 13, no. 15, p. 2512, Jul. 2023, doi: 10.3390/diagnostics13152512.

[23] S. K. Vinayam, C. Mohan, and R. H. Himak, “Management of Dental Caries in Digital Era, Artificial Intelligence perspective - Scoping Review,” J. Biog. Sci. Res., pp. 1–9.

[24] A. S. Alphonse, S. V. Kumari, and P. T. Priyanga, “Caries Detection from Dental Images using Novel Maximum Directional Pattern (MDP) and Deep Learning,” Int. J. Electr. Electron. Res., vol. 10, no. 2, pp. 100–104, Jun. 2022, doi: 10.37391/ijeer.100208.

[25] T. Kawazu, Y. Takeshita, M. Fujikura, S. Okada, M. Hisatomi, and J. Asaumi, “Preliminary Study of Dental Caries Detection by Deep Neural Network Applying Domain-Specific Transfer Learning,” J. Med. Biol. Eng., vol. 44, no. 1, pp. 43–48, Feb. 2024, doi: 10.1007/s40846-024-00848-w.

[26] A. AlSayyed et al., “Employing CNN ensemble models in classifying dental caries using oral photographs,” Int. J. Data Netw. Sci., vol. 7, no. 4, pp. 1535–1550, 2023, doi: 10.5267/j.ijdns.2023.8.009.

[27] R. Ben Ali, R. Ejbali, and M. Zaied, “Detection and Classification of Dental Caries in X-ray Images Using Deep Neural Networks,” in ICSEA 2016 : The Eleventh International Conference on Software Engineering Advances, 2016, no. c, pp. 223–227.

[28] O. Boiko et al., “Deep Learning for Dental Hyperspectral Image Analysis,” Color Imaging Conf., vol. 27, no. 1, pp. 295–299, Oct. 2019, doi: 10.2352/issn.2169-2629.2019.27.53.

[29] S. Patil, V. Kulkarni, and A. Bhise, “Algorithmic analysis for dental caries detection using an adaptive neural network architecture,” Heliyon, vol. 5, no. 5, p. e01579, May 2019, doi: 10.1016/j.heliyon.2019.e01579.

[30] M. P. Muresan, A. R. Barbura, and S. Nedevschi, “Teeth Detection and Dental Problem Classification in Panoramic X-Ray Images using Deep Learning and Image Processing Techniques,” in 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), Sep. 2020, pp. 457–463. doi: 10.1109/ICCP51029.2020.9266244.

[31] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[32] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

[33] W. Liu et al., “SSD: Single Shot MultiBox Detector,” in Lecture Notes in Computer Science ((LNIP,volume 9905)), 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” in 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.

[35] R. Costa, T. Amaral, C. Soares, E. Silva, and R. Santana, “UFBA UESC Dental Image Dataset: Labeled Dental X-ray Images for Caries Detection.” 2018. [Online]. Available: https://www.kaggle.com/dataset