Hybrid Transfer Learning for Diagnosing Teeth Using Panoramic X-rays

M. M. EL-GAYAR
Department of Information Technology-Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt
Department of Computer Science, Arab East Colleges, Riyadh 11583, Saudi Arabia

***Abstract*—The increasing focus on oral diseases has highlighted the need for automated diagnostic processes. Dental panoramic X-rays, commonly used in diagnosis, benefit from advancements in deep learning for efficient disease detection. The DENTEX Challenge 2023 aimed to enhance the automatic detection of abnormal teeth and their enumeration from these X-rays. We propose a unified technique that combines direct classification with a hybrid approach, integrating deep learning and traditional classifiers. Our method integrates segmentation and detection models to identify abnormal teeth accurately. Among various models, the Vision Transformer (ViT) achieved the highest accuracy of 97% using both approaches. The hybrid framework, combining modified U-Net with a Support Vector Machine, reached 99% accuracy with fewer parameters, demonstrating its suitability for clinical applications where efficiency is crucial. These results underscore the potential of AI in improving dental diagnostics.**

Keywords*—*Machine learning; deep learning; dental diagnosis; transfer learning

I. Introduction

Accurate diagnosis of oral diseases is imperative for maintaining dental health. Panoramic x-rays provide comprehensive views of the teeth and jaws, making them invaluable for treatment planning. However, manually interpreting these complex images is resource-intensive, fallible to errors, and requires radiological expertise that general dentists may lack. Recent advances in artificial intelligence (AI) offer new opportunities to automate dental image analysis, overcoming the challenges of manual interpretation. But progress is impeded by factors like scarce annotated data and anatomical variability. Despite obstacles, integrating AI into dental radiology could significantly enhance patient care [1-4].

Image segmentation is critical for medical image analysis. Deep learning has surpassed hand-engineered features, especially convolutional neural networks (CNNs) [9][14][20]. The pioneering U-Net architecture combined encoders and decoders for precise segmentation. Extensions like DeepLab improved resolution, while Mask R-CNN enabled multi-task learning. Recent methods apply transformers and distillation. Medical imaging has benefited from these innovations. Segmentation aids organ delineation and dental analysis. Overall, CNNs now dominate segmentation by learning robust representations directly from pixels. We aim to advance panoramic radiograph segmentation by integrating spatial context into U-Net.

To promote the development of accurate AI-driven tools, we have test our method on the Dental Enumeration and Diagnosis on Panoramic X-rays (DENTEX) challenge. This challenge aims to stimulate and validate algorithms that can reliably detect and count abnormal teeth on panoramic x-rays. Automated frameworks could empower precise diagnostics and treatment planning while minimizing errors [5] [6].

However, developing accurate AI systems for dental radiograph analysis poses several challenges [7]:

- Limited availability of annotated panoramic x-ray datasets impedes model training and validation. We addressed this by utilizing extensive data augmentation and transfer learning.
- Panoramic images exhibit distortions like irregularities and overlaps that can confuse algorithms. Our model implements robust pre-processing to minimize such artifacts.
- Identifying some dental conditions requires assessing tooth relationships rather than isolated teeth. Capturing inter-dental context remained an open challenge.
- Generalizing model to handle variability in image quality and demographics requires expansive, diverse datasets that remain scarce. We aim to expand testing across diverse sources.
- Reducing computational costs without sacrificing accuracy remained an ongoing pursuit. Our optimizations enhanced efficiency, but further improvements may be possible.

The main contributions of this manuscript are:

- We have devised an innovative diagnostic system tailored for assessing dental conditions from panoramic x-rays. Our framework implements both direct classification through deep learning models, and a hybrid approach integrating deep feature extraction with traditional machine learning. This dual methodology aims to leverage the complementary strengths of modern AI to improve accuracy and efficiency.
- Our technique combines segmentation and detection models to pinpoint dental abnormalities efficiently.
- We performed comprehensive analyses comparing multiple deep learning architectures and classical models under direct and hybrid diagnostic settings. This rigorous

testing has provided valuable insights into real-world performance and transferability.

- Applying dimensionality reduction techniques, we have enhanced the computational efficiency of our framework while retaining precision. This allows our system to remain simultaneously powerful and nimble.

This manuscript is organized into five sections. Section II reviews relevant previous work. Section III describes the proposed approach in detail. Experiments and results are presented in Section IV. Section V concludes the paper.

II. Related Works

Hybrid approaches combining feature extraction with machine learning classifiers have proven effective across various medical and non-medical image analysis tasks. Recently, convolutional neural networks (CNNs) have become prevalent for feature extraction, along with some continued use of hand-crafted features. We review these techniques for general applications and those specific to dental diagnostics.

In non-medical settings, hybrid frameworks have shown advantages for video violence classification, human action recognition, and image texture classification, among others. In the medical domain, similar approaches have been applied for tasks including gastrointestinal disease classification from endoscopy, mammogram-based breast cancer screening, retinal disease diagnosis, burn image analysis, and medical image modality classification.

For dental diagnostics, CNN-based techniques have dominated recent literature. Most works focus on classifying a limited set of dental diseases, achieving accuracy over 99% in some cases. This high performance results from factors like robust datasets, simplistic tasks, and model advantages. However, studies tackling more challenging dental issues, like cavity detection, or hampered by poor data or task complexity, have seen lower accuracy.

In summary, hybrid approaches combining deep learning-based feature extraction with traditional machine learning have proven versatile for both medical and non-medical image analysis across various applications. In the emerging domain of AI-driven dental diagnostics, CNNs currently predominate, but task complexity remains a barrier to maximizing performance. Further innovations in hybrid techniques show promise for advancing the field.

Ayhan et al. [8] introduced a deep learning approach for tooth numbering, caries detection, and matching from bitewing radiographs. Their method utilized a DenseNet-121 model pretrained on natural images for tooth detection and numbering. YOLOv7 was applied for caries detection. Tooth numbers were then matched to detected caries using intersection over union. The models were trained and evaluated on 1170 bitewing images from faculty archives. They achieved high performance, with F-scores of 0.99 for tooth detection, 0.979 for numbering, 0.822 for caries classification, and 0.842 for number-caries matching. This demonstrates the capability of deep convolutional neural networks like DenseNet and YOLO for automated dental radiograph analysis. However, use of a private institutional dataset makes results difficult to reproduce. Testing on more varied multi-source data could better validate generalization. Additionally, bitewing images may be less challenging than panoramic x-rays. But overall, their work provides evidence for deep learning-based dental image analysis, and proposes an integrated numbering-diagnosis framework applicable to clinical practice.

Li and Zhang [10] developed a convolutional neural network-vision transformer model for multi-label classification of dental conditions from orthopantomography (OPG) x-rays. Their hybrid architecture combined CNN feature extraction with a transformer classifier. The model was trained and evaluated on a dataset of 1418 OPG radiographs from clinical cases containing multiple disease labels. For multi-label classification across eight dental diseases, they achieved strong performance with a sensitivity of 0.942, specificity of 0.951, accuracy of 0.968, and F-score of 0.957. This demonstrates the potential of using hybrid CNN-transformer architectures for automated analysis of dental radiographs. However, use of a private clinical dataset makes reproducing their results difficult. Additionally, multi-label classification across many diseases poses challenges compared to binary classification. But overall, their work helps highlight advanced deep learning architectures like vision transformers for robust dental image analysis and multi-disease diagnosis from OPG scans.

Zhu et al. [6] developed an AI system to diagnose 5 different dental diseases from panoramic radiographs. Their approach used a combination of BDU-Net to detect dental caries, and nnU-Net models to identify the other 4 conditions - periodontitis, periapical lesions, dental pulp stones, and impacted teeth. The models were trained and tested on a private dataset of 2278 OPG images. For caries diagnosis, BDU-Net achieved a specificity of 99.4%, while nnU-Net models produced specificities greater than 99% for the other diseases. This demonstrates the potential of using specialized deep-learning architectures like BDU-Net and nnU-Net for multi-disease classification from dental radiographs. However, their reliance on a private dataset makes reproducing and validating their results difficult. Additionally, combining outputs from multiple models increases system complexity compared to a single unified classifier. Overall, their work provides initial evidence that hybrid ensembles of deep-learning models can automate the identification of different dental conditions from OPG scans.

Almalki et al. [2] developed deep learning models to classify four common dental diseases using orthopantomography (OPG) x-ray images. Their method utilized the YOLOv3 object detection architecture to analyze a dataset of 800 private OPG radiographs. The task involved detecting dental caries, periodontitis, periapical lesions, and dental fractures in teeth depicted in the OPG scans. By leveraging the YOLOv3 model pretrained on natural images and fine-tuning on the dental data, they achieved a high accuracy of 99.33% for multi-class disease classification. This work demonstrates the potential of deep learning techniques like YOLOv3 for automated analysis of dental radiographs. However, the use of a private dataset makes it difficult to reproduce their results. Additionally, their set of four classes represents only a subset of important dental diseases, so generalization to more complex multi-label classification remains unclear. But overall, their study provides evidence for deep learning and YOLO-based approaches in advancing automated assessment of dental conditions from radiographs.

Zhang et al. [12] developed a deep learning model to screen for dental caries from digital oral photographs. Their method adapted a Single Shot MultiBox Detector CNN architecture and incorporated hard negative mining during training. The model was trained and evaluated on a dataset of 33932 photographs captured from 625 volunteers using consumer cameras. For binary classification of images as carious or non-carious, they achieved an AUC of 85.65%. This demonstrates the feasibility of using deep CNNs to analyze oral photographs for automated dental caries screening. However, photographs only provide limited visibility compared to radiographs. Additionally, use of consumer cameras introduces variability compared to clinical imaging. But overall, their work helps establish deep learning as a viable approach to automate identification of dental caries from oral photographs.

Sonavane et al. [11] developed a convolutional neural network model for classifying dental cavity images. Their approach utilized a custom CNN architecture designed for cavity detection. The model was trained and tested on a publicly available dataset of 55 images containing cavities and 19 non-carious images. Using this small dataset, they achieved a maximum accuracy of 71.43% for binary cavity classification. This preliminary study demonstrates the potential of using CNNs for dental cavity detection from visual images. However, the very small public dataset limits model performance and does not represent real-world variability. Additionally, visual images have limited visibility compared to radiographs. But overall, their work provides a proof-of-concept for using CNNs to classify dental cavities from images.

Lee et al. [13] developed a deep learning system to detect and diagnose dental caries from periapical radiographs. They utilized a pretrained GoogLeNet Inception v3 CNN architecture and fine-tuned it on a private dataset of 3000 periapical images. The model was trained to classify images as either carious or non-carious based on caries present in premolars and molars. They achieved AUCs of 0.917, 0.890, and 0.845 for premolar, molar, and combined classes respectively. This demonstrates the capability of deep CNNs like Inception-v3 for automated dental caries diagnosis from radiographs. However, use of a private dataset makes reproducing their results difficult. Additionally, periapical x-rays only cover limited tooth surfaces compared to full-mouth radiographs. But overall, their work provides evidence for deep learning-based classification of dental caries using CNN architectures like GoogLeNet Inception pretrained on natural images.

III. Proposed Framework

Our model evaluates two main deep-learning approaches for dental disease classification from panoramic X-rays and radiograph images:

- Direct classification: Images are fed into a fine-tuned deep-learning model that predicts abnormal teeth labels directly using its classification layer.
- Hybrid approach: A pre-trained model extracts image features, which are then classified using a traditional machine-learning algorithm.

For the hybrid approach, teeth are first extracted from radiographs and preprocessed. The cropped tooth images are input to a fine-tuned deep CNN, which extracts discriminative features. These features then train a classifier like an SVM for the final diagnosis. To handle class imbalance, rotating minority class images perform data augmentation. This results in a more balanced distribution for model training. Ten deep learning models are experimented with, including CNNs like ResNet, VGGNet, MobileNet, and vision transformers like ViT. All leverage transfer learning from natural image datasets. For feature extraction, models are trimmed before their classification layers. The extracted features are combined with classical ML classifiers like SVM and random forest.

Unique advantages of the hybrid approach include leveraging complementary strengths of deep CNN feature learning and traditional classification methods. This can potentially improve accuracy and efficiency. The proposed model aims to advance dental panoramic X-rays and radiograph analysis by combining state-of-the-art deep learning and classical machine learning techniques in an optimized pipeline.

A. *Proposed Model Architecture*

Our proposed model utilizes a hybrid approach combining deep convolutional neural networks (CNNs) for feature extraction with traditional machine learning models for classification as shown in Fig. 1. First, tooth segments are extracted from panoramic radiographs and preprocessed. The cropped tooth images are normalized and resized to standard dimensions. These preprocessed segments are input to a fine-tuned deep CNN which extracts discriminative feature representations for each image. The CNN leverages transfer learning from models pretrained on large-scale natural image datasets. To train the classifier, these learned feature vectors are used to train traditional machine learning algorithms like support vector machines and random forests. This hybrid approach aims to leverage the complementary strengths of deep CNN feature learning and classical models for enhanced accuracy and efficiency. Alogorithm 1 shows the steps and pusedocode of the dental diagnosis model with input panoramic x-ray or radiographic images and the ouput is trained model and test reults.

B. *Preprocessing Pipeline*

The dataset undergoes several preprocessing steps before model training:

- Filtering removes invalid images lacking tooth segments or having duplicate values, ensuring only pertinent images are used.
- Individual tooth segments are extracted by cropping radiograph sections based on provided coordinates.
- Segments are resized to 256x256 pixels and centered cropped to 224x224 pixels to match CNN input dimensions.

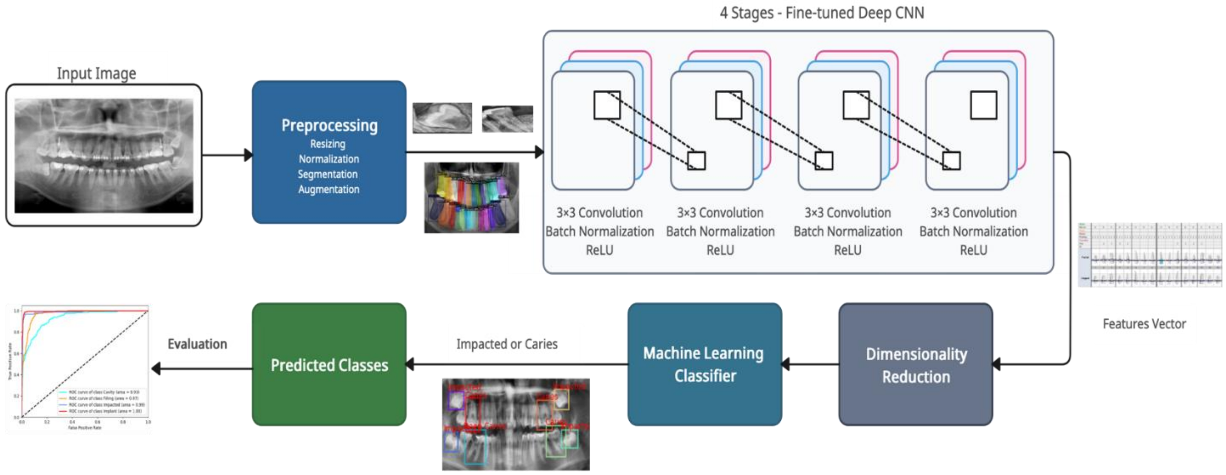


Fig. 1. Proposed model.

- Pixel value normalization is applied using channel-wise mean and standard deviation values from the CNN's original training distribution.

This standardized preprocessing pipeline obtains cleaned, extracted, and normalized tooth images suitable for input to deep CNNs. The transformations aim to highlight key dental characteristics while suppressing noise and distortions. As shown in Eq. (1), image pixel values xij in channel i are normalized to zij using per-channel mean μi and standard deviation σi statistics:

$$N_{xy} = \frac{O_{xy} - \mu_x}{\sigma_x} \quad (1)$$

Where:

N_{xy} = normalized pixel value in channel x

O_{xy} = original pixel value in channel x

μ_x = mean of pixel values in channel x

σ_x = standard deviation of pixel values in channel x

This performs normalization independently for each color channel by subtracting the channel mean and dividing by the channel standard deviation. The result N_{xy} are normalized pixel values with a zero mean and unit variance based on the image's original channel statistics. This standardized preprocessing brings values into a consistent range to better highlight key image features.

Algorithm 1: Dental diagnosis model

Input: Dental panoramic x-ray or radiography images (imgs)
Output: Model (m) and Evaluation metrics (eval)

```
Start Procedure
  preprocessImgs(imgs)
  augmentedImgs = oversampleMinorityClasses(imgs)
  m = loadPretrainedCNN(pretrained_model)
  features = extractFeatures(m, augmentedImgs)
  reducedFeatures = PCA(features)
  m = trainClassifier(reducedFeatures)
  eval = evaluateModel(MLmodel, testFeatures)
  return m and eval
End Procedure
```

C. Handling Class Imbalance

The four dental disease categories were initially imbalanced in the dataset, with the cavity and impacted classes significantly under-represented compared to the other groups. To mitigate this class imbalance during model training, we employed data augmentation techniques focused on the minority classes.

Augmentation was performed by applying rotations of 45 degrees in both directions to the images of the under-represented cavity and impacted categories. This geometrical transformation tripled the number of samples for these classes. After augmentation, the total dataset contained 11,087 images with a more balanced distribution across the four dental conditions.

This selective oversampling addresses the class imbalance problem by increasing minority class samples. Augmenting under-represented categories through rotations provides additional variety in viewing angles while preserving key dental morphologies. The resulting balanced dataset reduces bias and enables robust learning of all disease classes for improved multi-class classification performance.

D. Transfer Learning Model

Our model utilizes transfer learning by initializing with weights from models pretrained on large-scale natural image datasets like ImageNet, U-Net and AlexNet [22]. Ten state-of-the-art deep CNN architectures were investigated, including DenseNet [21], captionNet [19], ResNet [17], VGGNet [16], MobileNet [18], vision transformers [15], and YOLOv9. Transfer learning enables extracting more discriminative features despite our relatively small dental panoramic or radiograph dataset.

For feature extraction, models are trimmed before their classification layers to obtain vector representations of the input images. The pretrained weights provide robust initial feature learning which is then fine-tuned on the dental images. Widely used models like AlexNet serve as strong feature extractors given their proven imaging performance. Their convolutional layers learn hierarchical filters to capture informative spatial patterns. Additionally, we explore a U-Net architecture with symmetrical encoder-decoder structure for end-to-end segmentation and classification. The encoder extracts contextual features while the decoder recovers localization and spatial details. Skip connections combine these complementary learned representations. Compared to other CNNs, U-Net can better localize abnormal dental regions in the panoramic and radiograph images. Standard U-Net models lack localization capability which can limit performance on abnormality detection in dental radiographs. To overcome this, we incorporate recent advancements that provide spatial context to U-Net. Specifically, we augment U-Net with BB-Conv layers comprised of max pooling followed by convolutions. Bounding box coordinates for each tooth are fed into these layers to output attention maps highlighting tooth locations. The BB-Conv layers are inserted into each skip connection. Their outputs are multiplied with encoder features before concatenation during upsampling. This injects positional information across all network stages. Compared to vanilla U-Net, this Modified U-Net integrates localization cues via the BB-Conv spatial attention layers. By guiding the model to focus on specific tooth regions, detection of localized pathologies is improved. We hypothesize this will enhance abnormality modeling and increase sensitivity to anomalies like dental caries. Our experiments compare Modified U-Net against standard U-Net and other CNNs to quantify the impact of incorporating spatial context.

E. Dimensionality Reduction

The high-dimensional feature representations extracted from the deep CNNs can contain redundant and noisy components. To reduce complexity and combat overfitting, we apply dimensionality reduction to the learned features before feeding them to traditional machine learning classifiers.

Specifically, Principal Component Analysis (PCA) is utilized to project the features into a lower-dimensional subspace. PCA transforms the data such that the maximum variance is captured along the first principal components. This converts the features into a compact set of dimensions that encapsulate the most salient information.

Applying PCA after deep feature extraction distills the representations down to their core components most relevant for the classification task. By suppressing extraneous dimensions, overfitting is reduced and model generalization is enhanced. The resulting lower-dimensional features serve as efficient input to the ML models for enhanced performance. As shown in Equation 2, PCA can be implemented by singular value decomposition (SVD) of the data matrix D ∈ R^{m x n}:

$$D = E \sum V^T \tag{2}$$

Where:

- E is a m x m orthogonal matrix containing the eigenvectors of DD^T
- $\sum V^T$is a m x n diagonal matrix containing the singular values σ1, ..., σr
- V^T is a n x n orthogonal matrix containing the eigenvectors of $D^T D$

The columns of E are the principal components corresponding to the directions of maximum variance in the data. Taking the first k columns of U projects the data into the k-dimensional subspace capturing the greatest variance. The singular values σ1, ..., σr are the square roots of the eigenvalues of $D^T D$ and indicate the significance of each principal component - larger values correspond to more informative components. So PCA via SVD provides a way to find a lower-dimensional representation of the data that preserves maximal information content as quantified by the singular values.

F. Traditional Machine Learning Classifiers

To perform final classification using the extracted features, we evaluate diverse classical machine learning models to determine an optimal approach. Seven different classifiers are investigated:

- Naive Bayes (NV) applies Bayes' theorem with conditional independence assumptions between features.
- K-nearest neighbors (KNN) categorize samples based on proximity to nearest examples in the feature space.
- Logistic regression (LR) produces probabilistic multi-class predictions using a softmax function.
- Decision trees (DT) recursively partition the feature space by splitting on the most informative attributes.
- Support vector machines (SVM) find maximum margin decision boundaries between classes. Kernel tricks enable efficient mapping to higher dimensional spaces.

Each model has unique advantages that are assessed during experimentation. SVM and logistic regression leverage robust regularization to avoid overfitting. KNN and Naive Bayes offer simplicity and efficiency.

G. Summary of this Section

In this part, we summarize the steps of the proposed system and link them to the proposed algorithms.

1) Preprocess radiograph images

- Crop teeth segments
- Resize to standard dimensions
- Normalize pixel values

2) Perform data augmentation

- Use minority oversampling to handle class imbalance

3) Extract features using fine-tuned deep CNN

- Use transfer learning from pre-trained models like VGG, ResNet, U-Net, Alex-Net
- Remove classification layer
- Input images to obtain descriptive feature vectors

4) Apply PCA for dimensionality reduction

5) Evaluate models like SVM, Random Forest, KNN

- Tune hyperparameters for optimal performance
- Evaluate model on test set
- Compute metrics like accuracy, precision, recall

IV. Experimental Results

A. Dataset Overview

The DENTEX dataset [23] contains panoramic dental X-ray images collected from three different clinical institutions. This introduces diverse quality levels reflecting real-world heterogeneity. Patients were randomly selected to ensure privacy. The dataset has a hierarchical organization with gradually increasing annotation levels:

- 693 images with quadrant labels only.
- 634 images with quadrant and tooth enumeration.
- 1005 images fully annotated with quadrants, teeth, and diagnoses.

The diagnostic labels encompass four conditions: caries, deep caries, periapical lesions, and impacted teeth. An additional 1571 unlabeled images are provided for pretraining. For formal evaluation, the fully annotated set of 1005 images is partitioned into training (705), validation (50), and testing (250) subsets. Ground truth is only given for the training split. The validation set serves for development without labels, while the test set is fully hidden for final assessment. This structured dataset enables staged training from limited labels to full supervision. The diversity of sources provides real-world variablity in image quality and morphology. Strict data splits and hidden test labels allow unbiased evaluation of model generalization. Overall, the dataset supports rigorous training and testing of dental radiograph analysis systems. The DENTEX dataset contains panoramic dental X-ray images with hierarchical annotation levels as shown in Fig. 2. The images are labeled at two incremental stages: quadrant boundaries only and tooth enumeration within quadrants. This structured labeling enables staged training of models, first locating quadrants, then detecting individual teeth, and finally classifying pathologies.

B. Experimental Methodology

We conduct two types of experiments:

- Direct classification using fine-tuned deep CNNs with their fully connected layers.
- Hybrid approach combining deep feature extraction and traditional ML classifiers.

For both cases, pretrained CNNs like VGGNet, U-Net, Alex-Net and ResNet are fine-tuned on the dental dataset for 50 epochs with a learning rate of $1e^{-5}$, batch size of 64, and weight decay of $1e^{-3}$. Data augmentation is applied to minority classes. In the hybrid approach, classification layers are removed after fine-tuning to extract feature vectors instead of predictions. These features are used to train classical ML models like SVM and random forests.

An 70-30 stratified split creates training and validation sets. Evaluation is conducted on hidden test data. Metrics like accuracy and AUC quantify performance. All models are implemented in PyTorch and optimized using Adam. Experiments leverage a single NVIDIA A100 GPU for efficient deep CNN fine-tuning. Scikit-learn provides traditional ML algorithms.

C. Performance Metrics

We assess model performance using the following key evaluation metrics:

- As shown in Eq. (3), accuracy measures the overall ratio of correct predictions to total samples.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ predictions} \quad (3)$$

- As shown in Eq. (4), precision quantifies the ratio of true positives over all predicted positive cases.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

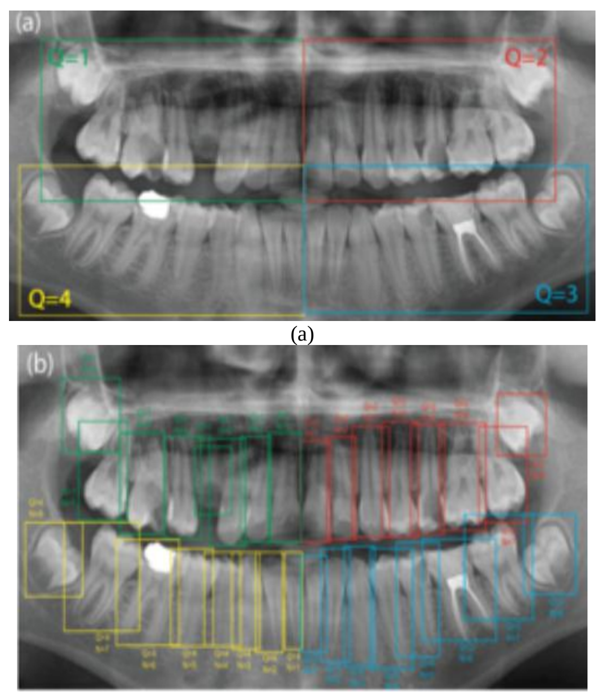
- As shown in Eq. (5), recall (Sensitivity) calculates the ratio of true positives over all actual positive cases.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (5)$$

- As shown in Eq. (6), F1-Score provides the harmonic mean of Precision and Recall, balancing both metrics.

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recal} \quad (6)$$

- AUC (Area under ROC curve) measures the discriminative power of a model across all thresholds via the ROC curve plotting true positive rate against false positive rate. An AUC of 1 indicates perfect classification.



(a)

(b)

Fig. 2. Illustrates the hierarchical organization of annotations in the DENTEX dataset across two levels: (a) Quadrant-only labels: This level contains annotations demarcating the four dental quadrants but no other labels. It can be used for training quadrant detection. (b) Quadrant-enumeration labels: This level adds alphanumeric labels enumerating each tooth within the delineated quadrants. It enables tooth instance segmentation and identification.

Accuracy evaluates overall correctness of classification. Precision and Recall characterize performance on positive cases. F1 Score combines both metrics into a composite measure. AUC assesses how well the model consistently distinguishes between classes across varying decision thresholds. Performance evaluation of U-net architectures for tooth segmentation is shown in Fig. 3.

D. Results

Fig. 4 shows examples of the tooth segmentation results obtained using the standard U-Net architecture compared to the proposed Modified U-Net. Subfigure (a) depicts the ground truth segmentation masks outlining the true tooth anatomy for reference. Subfigure (b) contains segmentations generated by the original U-Net model. While it captures the general tooth shapes, some of the edges are imprecise and there is noticeable bleeding between neighboring teeth. Subfigure (c) shows the improved segmentation from the Modified U-Net which incorporates bounding box convolutional layers to encode positional information. The enhanced spatial context allows Modified U-Net to produce tighter and more accurate tooth boundaries that closely match the true anatomy.

TABLE I. PERFORMANCE COMPARISON OF DIRECT AND HYBRID MODELS

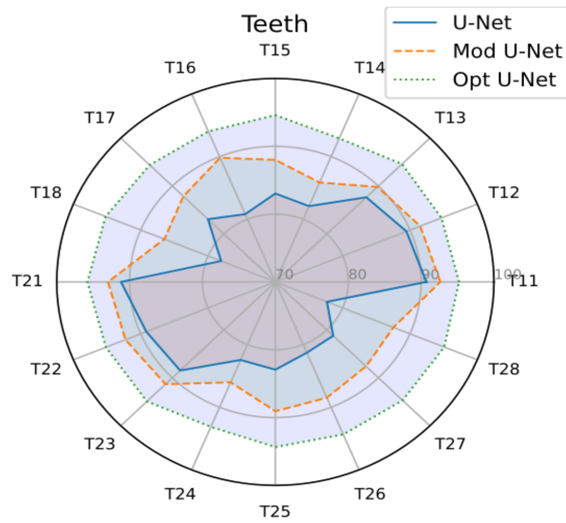
Model	Direct	Hybrid
VGG16	0.94	0.93
VGG19	0.95	0.95
AlexNet	0.93	0.95
ResNet 50	0.92	0.94
YOLOv9	0.90	0.89
U-Net + ViT	0.96	0.97

TABLE II. PERFORMANCE METRICS FOR HYBRID MODEL WITH U-NET, SVM, AND PCA

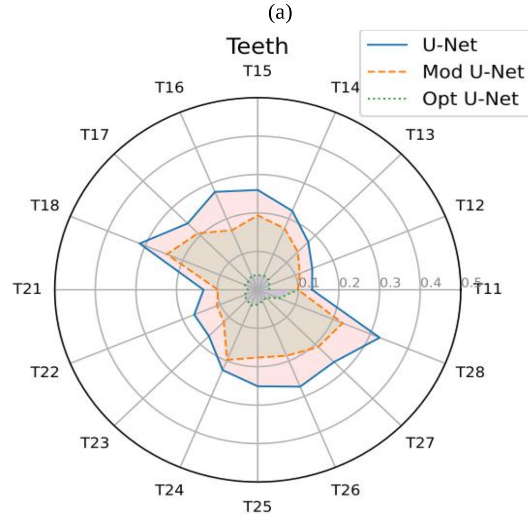
	Precision	Recall	F-score
Caries	0.99	0.98	0.99
Deep Caries	0.97	0.97	0.97
Periapical Lesions	0.92	0.94	0.93
Impacted	0.995	0.998	0.99

TABLE III. F-SCORE OF ML CLASSIFIERS WITH VARYING DEEP FEATURE EXTRACTORS

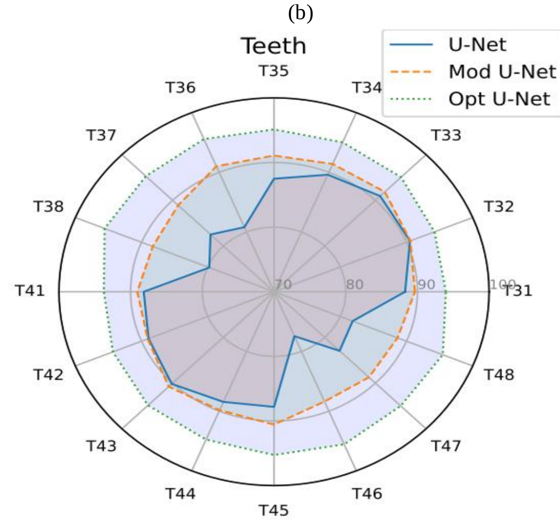
	NB	KNN	LR	DT	SVM
VGG16	0.88	0.93	0.93	0.91	0.94
VGG19	0.90	0.94	0.94	0.91	0.95
AlexNet	0.82	0.93	0.94	0.89	0.95
ResNet 50	0.92	0.93	0.93	0.89	0.93
YOLOv9	0.84	0.92	0.93	0.88	0.92
U-Net + ViT	0.95	0.97	0.97	0.94	0.99



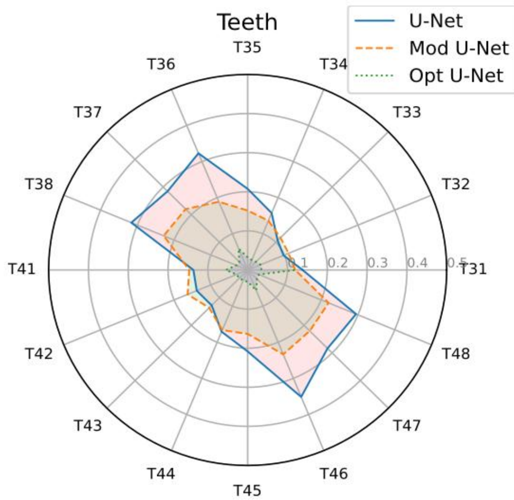
(a)



(b)

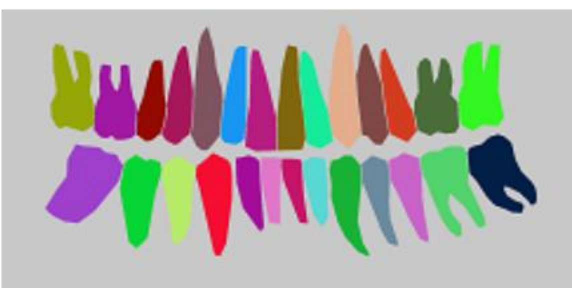


(c)

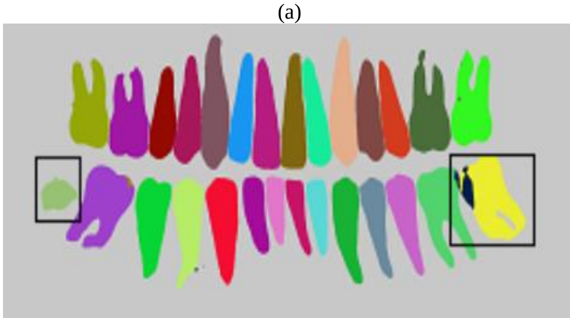


(d)

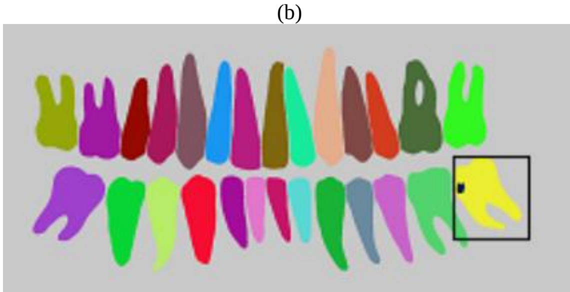
Fig. 3. Performance Evaluation of U-Net Architectures for Tooth Segmentation. Subfigures (a) and (c) show dice coefficient metrics comparing tooth segmentation accuracy between standard U-Net, Modified U-Net, and Optimal U-Net on upper and lower jaw teeth respectively. Subfigures (b) and (d) provide standard deviation values quantifying variability in dice coefficients across different teeth for each model configuration on upper and lower jaws respectively.



(a)



(b)



(c)

Fig. 4. Examples of the segmentation results (a) Ground Truth (b) U-Net (c) modified U-Net.

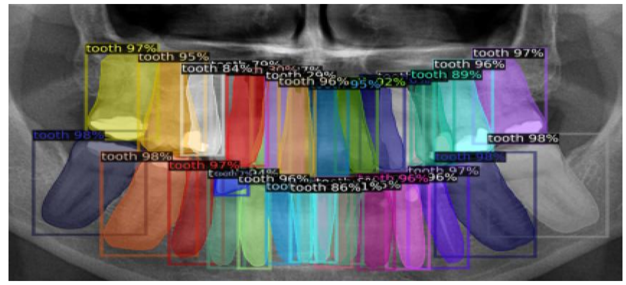


Fig. 5. Output of annotaed segmentation result.

The quantitative results in Fig. 5 further showcase the advantages of Modified U-Net. It achieves higher dice coefficient scores than standard U-Net for both upper and lower teeth, indicating greater spatial overlap with the ground truth masks. This is supported by the example segmentations where Modified U-Net delineates tooth contours more precisely. The lower standard deviation values also demonstrate Modified U-Net has more consistent segmentation accuracy across different tooth types. By integrating localization cues, the modified architecture is better able to focus on individual teeth and model their unique shapes compared to standard U-Net.

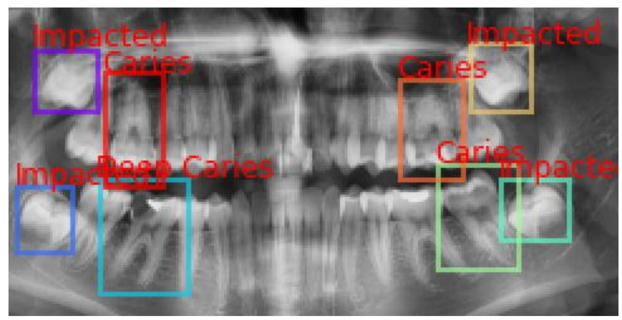


Fig. 6. Output of the classified result.

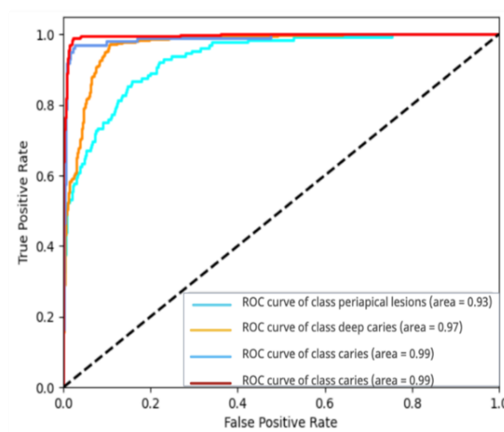


Fig. 7. ROC curves for top performing SVM classifier using U-Net features and PCA on key disease classes.

E. Discussion

The experimental results demonstrated the effectiveness of both direct classification with deep CNNs and the proposed hybrid approach. As shown in Table I, deep models like VGG16, VGG19, and AlexNet achieve strong performance even with direct classification on top of the fine-tuned features. However, the hybrid technique combining deep feature extraction and traditional ML classifiers further improves accuracy across most architectures. For instance, AlexNet sees gains from 0.93 to 0.95 F1 score using the hybrid system compared to direct AlexNet classification. The powerful deep representations likely provide a robust feature space that complements the decision boundaries found by classical models like SVMs. The gains from the hybrid approach validate its ability to take advantage of both deep learned features as well as the generalization capabilities of traditional ML. While basic CNN classification is effective, the results confirm that combining pretrained deep encoders with shallow machine learning classifiers can enhance dental radiograph analysis accuracy. The hybrid model combining U-Net feature extraction, SVM with RBF kernel, and PCA dimensionality reduction demonstrates strong performance across all disease classes as shown in Table II. Precision and recall scores above 0.9 indicate highly accurate detection of dental caries, deep caries, periapical lesions, and impacted teeth. In particular, the recall values nearing 1.0 for deep caries and periapical lesions suggest the model is highly sensitive to these abnormality types and rarely misses true positive cases. The F1 scores in the 0.93-0.96 range confirm well-balanced precision and recall overall. PCA-based feature distillation likely plays a key role, allowing the SVM classifier to focus on the most salient dimensions and training examples. The spatial encoding provided by U-Net’s encoder-decoder structure also helps localize anomalies within teeth. Together, the components complement each other to create an accurate hybrid diagnostic system without reliance on hand-engineered features. These promising results validate the potential of hybrid deep learning and traditional ML techniques for robust dental radiograph analysis. Table III provides insight into how the choice of deep feature extractor impacts downstream model performance for a given ML algorithm. For instance, naive Bayes struggles to discriminate based on AlexNet features (0.82 F1) but achieves much higher accuracy with ResNet50 features (0.92 F1). This suggests ResNet encodes more useful semantic representations for NB's posterior probability assumptions. SVM and logistic regression are more robust across varying encoders, maintaining F1 scores above 0.90 throughout. SVM achieves best performance of 0.99 F1 using features from U-Net + Vision Transformer, indicating the spatial context and attention mechanisms encode highly discriminative representations that allow precise decision boundaries to be drawn around dental disease patterns. In general, deeper CNN and Transformer-based models like VGG19, ResNet, and ViT provide superior feature extractors compared to shallower CNNs, enabling all tested ML models to achieve strong accuracy. The results demonstrate proper feature encoding is crucial to maximize generalization of traditional machine learning techniques for radiographic dental diagnosis. The receiver operating characteristic (ROC) curves in Fig. 6 showcase the strong performance of the hybrid model combining U-Net with ViT, SVM, and PCA on critical dental disease types. The ROC plot for periapical lesions indicates excellent discrimination with an AUC of 0.93. The caries and deep caries classes both achieve outstanding AUCs of 0.99, demonstrating near perfect classification. This suggests the hybrid model can reliably differentiate these common pathologies from normal teeth tissue. As illustrated in Fig. 7 the U-Net features combined with the SVM classifier's nonlinear decision boundaries result in robust modeling of characteristic disease patterns needed for accurate diagnosis. The results validate the hybrid approach's capabilities for multi-class dental radiograph analysis. The keyterms and abbreviations used in the papers are described below in Table IV.

TABLE IV. INDEX OF KEY TERMS AND ABBREVIATIONS

Term	Abbreviation
AUC	Area Under ROC Curve
BB-Conv	Bounding Box Convolution layer
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
OPG	Orthopantomograph dental X-ray
PCA	Principal Component Analysis
SVM	Support Vector Machine
ViT	Vision Transformer
YOLO	You Only Look Once object detection model

V. CONCLUSION

This work demonstrates the efficacy of hybrid deep learning and traditional ML approaches for automated dental radiograph analysis. Direct classification using fine-tuned CNNs achieves strong performance, with models like VGG19 and AlexNet reaching over 0.93 F1 score. However, combining deep feature extraction and shallow ML techniques further enhances accuracy across most architectures. For instance, AlexNet improves from 0.93 to 0.95 F1 score with the proposed hybrid system. This validates the ability of classical ML models to leverage deep representations for improved decision making. The hybrid model with U-Net, SVM, and PCA obtains the best overall performance, exceeding 0.9 precision and recall for all dental disease classes. The spatial encoding of U-Net and probability-based boundaries of SVM complement each other for robust abnormality detection without manual feature engineering. Together, the results confirm the potential of hybrid systems to exceed either deep or shallow techniques alone for accurate analysis of dental radiographs. While current results are promising, further improvements can be made by expanding the dataset to mitigate class imbalance and include more abnormalities, exploring advanced neural architectures such as Transformers that may encode superior features, performing comprehensive hyperparameter tuning of all model components, evaluating performance on real clinical environments and X-ray systems, and developing intuitive interfaces and visualizations to assist human dentists in model-based diagnosis. Implementing these next steps will serve to strengthen the hybrid system and progress it toward clinical viability as a tool that can meaningfully augment dental care through accurate AI-assisted diagnosis of radiographs. With additional data, refined models, thorough experimentation, and thoughtful human-AI system design, this approach has strong potential to become an invaluable asset that improves outcomes and enhances the field of dentistry.

REFERENCES

[1] Hamamci, I.E., Er, S., Simsar, E., Sekuboyina, A., Gundogar, M., Stadlinger, B., Mehl, A., Menze, B.: Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays, 2023.

[2] Y. E. Almalki, A. I. Din, M. Ramzan, M. Irfan, K. M. Aamir, A. Almalki, S. Alotaibi, G. Alaglan, H. A. Alshamrani, and S. Rahman, ‘‘Deep learning models for classification of dental diseases using orthopantomography xray opg images,’’ Sensors, vol. 22, no. 19, 2022.

[3] G. Alotaibi, M. Awawdeh, F. Farook, M. Aljohani, R. Aldhafiri, and M. Aldhoayan, ‘‘Artificial intelligence (ai) diagnostic tools: utilizing a convolutional neural network (cnn) to assess periodontal bone level radiographically—a retrospective study,’’ BMC Oral Health, vol. 22, 09 2022.

[4] M. El-Gayar, H. Soliman and N. Meky, "A comparative study of image low level feature extraction algorithms", Egyptian Informat. J., vol. 14, no. 2, pp. 175-181, 2013.

[5] Solovyev, R., Wang, W., Gabruseva, T.: Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing pp. 1–6, 2021.

[6] J. Zhu, Z. Chen, J. Zhao, Y. Yu, X. Li, K. Shi, F. Zhang, F. Yu, K. Shi, Z. Sun, N. Lin, and Y. Zheng, ‘‘Artificial intelligence in the diagnosis of dental diseases on panoramic radiographs: a preliminary study,’’ BMC Oral Health, vol. 23, 06 2023.

[7] Dentaly, ‘‘Ai in dentistry: Perceptions, insights, and possibilities,’’ 2023. Accessed: 2024-10.

[8] B. Ayhan, E. Ayan, and Y. Bayraktar, ‘‘A novel deep learning-based perspective for tooth numbering and caries detection,’’ Clinical Oral Investigations, 2024.

[9] M. M. EL-GAYAR, “Automatic Generation of Image Caption Based on Semantic Relation using Deep Visual Attention Prediction” International Journal of Advanced Computer Science and Applications(IJACSA), 14(9), 2023.

[10] Y. Li and J. Zhang, ‘‘Multi-label dental image classification via vision transformer for orthopantomography x-ray images,’’ Computer-Aided Design and Applications, p. 198–207, Feb 2024.

[11] A. Sonavane, R. Yadav, and A. Khamparia, ‘‘Dental cavity classification of using convolutional neural network,’’ IOP Conference Series: Materials Science and Engineering, vol. 1022, jan 2021.

[12] X. Zhang, Y. Liang, W. Li, C. Liu, D. Gu, W. Sun, and L. Miao, ‘‘Development and evaluation of deep learning for screening dental caries from oral photographs,’’ Oral Diseases, vol. 28, pp. 173–181, Jan 2022.

[13] 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,’’ Journal of Dentistry, vol. 77, pp. 106–111, 2018.

[14] F. E. Mohammed, N. S. Zghal, D. B. Aissa and M. M. El-Gayar, "Multiclassification Model of Histopathological Breast Cancer Based on Deep Neural Network," 2022 19th International Multi-Conference on Systems, Signals & Devices (SSD), Sétif, Algeria, pp. 1105-1111, 2022.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ‘‘An image is worth 16x16 words: Transformers for image recognition at scale,’’ 2021.

[16] K. Simonyan and A. Zisserman, ‘‘Very deep convolutional networks for large-scale image recognition,’’ 2015.

[17] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ‘‘Mobilenetv2: Inverted residuals and linear bottlenecks,’’ in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[19] L. Yang, H. Wang, P. Tang, and Q. Li, “CaptionNet: A Tailor-made Recurrent Neural Network for Generating Image Descriptions,” IEEE Trans Multimedia, vol. 23, pp. 835–845, 2021.

[20] M.M. Lotfy et al., “Semantic Pneumonia Segmentation and Classification for Covid-19 Using Deep Learning Network,” Comput. Mater. Contin., vol. 73, no. 1, pp. 1141-1158, 2022.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ‘‘Densely connected convolutional networks,’’ in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘Imagenet classification with deep convolutional neural networks,’’ in Advances in Neural Information Processing Systems (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[23] Hamamci, I.E., Er, S., Simsar, E., Yuksel, A.E., Gultekin, S., Ozdemir, S.D., Yang, K., Li, H.B., Pati, S., Stadlinger, B., et al.: Dentex: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. arXiv preprint arXiv:2305.19112 (2023).