IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 November 2024, accepted 25 November 2024, date of publication 28 November 2024,
date of current version 9 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3507925

RESEARCH ARTICLE

A Hybrid Transfer Learning Approach to Teeth Diagnosis Using Orthopantomogram Radiographs

AHMED ALABD-ALJABAR, (Member, IEEE), ZAIN RAISAN, MOHAMMED ADNAN, AND SALAM DHOU, (Member, IEEE)
Department of Computer Science and Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Corresponding author: Ahmed Alabd-Aljabar (akv3@outlook.com)

This work was supported in part by the Open Access Program from the American University of Sharjah under Award OPFY25-3152-OE2512.

ABSTRACT The rise in the emphasis on oral diseases has elevated the need to automate the diagnostic process of such diseases. Fortunately, the availability of modern computing devices has made the automated diagnosis of teeth readily possible using deep learning. Despite this, concerns about the accuracy and function of automated diagnosis remain among patients. To showcase the performance of such algorithms, we propose two approaches for the task of teeth diagnosis utilizing Orthopantomograms (panoramic radiographs): 1) a direct classification approach; and 2) a hybrid approach that combines a deep learning model with a traditional classifier. The results revealed that all ten chosen deep learning models experienced a similar or improved performance when used in conjunction with a machine learning classifier. In particular, Vision Transformer (ViT) performed the best with a record accuracy of 96% using both the direct and hybrid approaches. However, the hybrid framework combining AlexNet with a Support Vector Machine achieved an accuracy of 94%, and although it falls short of ViT in terms of performance, it comprises far fewer parameters. This highlights the approach's effectiveness in improving performance without the need to use a deeper model, making it well-suited for clinical adoption where efficiency is important.

INDEX TERMS Dental imaging, dental informatics, deep learning, machine learning, orthopantomography, transfer learning, vision transformer.

I. INTRODUCTION

Advances in dental technology have played a major role in treating oral diseases and improving overall oral health, which has been strongly associated with subsequent mortality [1], [2]. Naturally, this means that diagnosing these oral diseases and issues must be efficient, accurate, and fast. In traditional dentistry, there are two primary diagnostic approaches: clinical examinations and radiographs (i.e., X-rays). Clinical examinations are usually inefficient and time-consuming, requiring a dentist's manual assessment and evaluation of diseases [3].

Radiographs, on the other hand, have been used to develop Artificial Intelligence (AI)-based decision support systems that have the potential to enhance and speed the diagnostic processes. In particular, transfer learning, or the use of a pre-trained model with pre-learned weights, has been effective in its utility within dental radiographs [4].

One important type of radiograph used in dentistry is an Orthopantomogram (OPG). An OPG, sometimes called a panoramic radiograph, provides a comprehensive view of the hard tissues of the oral cavity and surrounding skeletal structures [5]. Dentists commonly use OPGs as a screening tool because they offer a complete view of the teeth and jaws, enabling oral diagnosis. Despite the potential presence of artifacts in OPGs, they usually provide a more inclusive diagnosis compared to a clinical examination, often discovering diseases that were unseen in the preceding examination [6].

After surveying 1015 patients and 250 experienced dentists in the United States in July 2023, it was found that 81% of the surveyed dentists had a positive attitude toward

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Wang.

implementing AI in applications in dentistry [7]. However, despite the numerous benefits of AI in industry, and the positive attitude from dentists, only 42% of the surveyed patients express comfort with receiving AI-based treatments while the majority remain neutral about the matter, mostly due to worries about accuracy, functional limitations, and speculative outcomes.

To address these issues, we propose a deep learning framework for teeth diagnosis using OPGs. This framework includes direct and hybrid deep learning approaches for disease classification. In the direct classification method, a full deep learning model is used and the classification is performed through the model's classification layer. In the hybrid approach, the features are extracted using a deep learning model and used as input to traditional machine learning classifiers that make the final prediction. Regardless of the approach used, the paper's main objective is to demonstrate the utility of AI in analyzing OPG radiographs. Therefore, our main contributions can be summarized as follows:

- We devised a framework for teeth diagnosis encompassing direct and hybrid deep learning approaches. The framework's performance has been demonstrated with tests on a public dataset of OPG radiographs achieving reduced computational requirements without sacrificing performance.
- We compared the performance of various models using both direct and hybrid approaches. Extensive experiments and tuning were conducted on ten deep learning models and seven traditional machine learning classifiers, which resulted in a record accuracy on the dataset used.
- We highlighted the promising applicability of the hybrid framework in dental clinics as it provides a reliable yet computationally efficient decision support system that enables clinics to utilize it with minimal computational resources for real-time inference and incremental learning.

II. LITERATURE REVIEW

The method of applying feature extraction in conjunction with a machine learning classifier has been used in many image classification and detection tasks in medical and non-medical fields. Recently, the prevalent feature extraction methods for general applications involve Convolutional Neural Network (CNN) models [8], [9], [10], [11], [12] and less traditional hand-crafted feature extraction methods [13], [14], [15]. In this section, we explore both these methods for general applications, as well as all methods applicable to dental diagnostics.

A. HYBRID APPROACHES FOR MEDICAL AND NON-MEDICAL APPLICATIONS

Hybrid approaches have shown effectiveness in various applications, outperforming other methods. In the non-medical spectrum, [8] classified violence in videos whereas [9] worked on human action recognition more generally. Reference [10] classified texture from images. On the medical side, [11] worked on the classification of gastrointestinal disease from wireless capsule endoscopy images, [12] categorized breast cancer from mammogram images, and [13] classified multiocular disease from retinal fundus images. Using only hand-crafted features, [14] diagnosed human burns from burn images, and [15] categorized the modality of medical images.

B. APPLICATIONS OF AI IN DENTAL DIAGNOSTICS

Recently, CNN-based approaches have dominated dental diagnostics. Most works in the literature classify a minimal number of diseases. The performance for this task vary, with accuracy values reaching up to 99% mainly due to good data quality, simplicity of the task, or advantages of the proposed model [3], [16], [17], [18], [19], [20], [21]. Other studies do not achieve such high accuracy values, mainly due to the detection of challenging types of diseases [22], [23], [24], [25], [26], due to lacking data quality or quantity, or a combination of both [27], [28], [29]. A detailed summary of the studies involving dental diagnosis can be found in Table 1. It is worth mentioning that we found no existing works that utilize the dataset that is used in our work. A detailed overview of the dataset is given in subsection III-A.

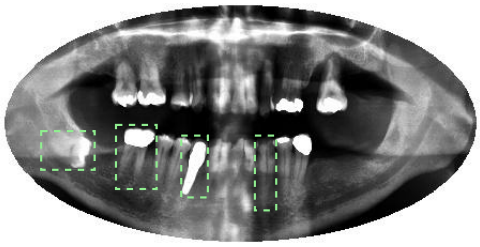

FIGURE 1. A sample OPG radiograph displaying teeth from the four classes (left to right): (1) impacted, (2) filling, (3) implant, and (4) cavity. Note that the bounding box of each tooth was included for visualization purposes.

III. MATERIALS AND METHODS

A. DATASET

Despite the wealth of papers covering dental research, dental datasets are seldom made public [30]. A dataset that has been recently shared on Kaggle [31] is considered in this work. This dataset contains 1262 OPG radiographs each associated with multiple teeth that fall into one of four classes: cavity, filling, impacted, or implant. The initial number of individual annotated teeth is 9283, with the majority belonging to the filling class. Fig. 1 presents a sample OPG alongside its annotated teeth. Moreover, despite the dominance of the filling class, the dataset can be used to generate more images, through augmentation, which is further discussed in subsection III-D.

TABLE 1. Summary of the studies tackling dental diagnosis presented in the literature.

Paper	Purpose	Method	Dataset	Results
[3]	Classify four common dental diseases using OPG radiographs of the teeth	YOLOv3 model	800 OPG radiographs	Accuracy of 0.99
[16]	Diagnose 5 different dental disease	BDU-Net to diagnose caries and nnU-Net for the other 4 diseases	2278 OPG radiographs	Specificity greater than or equal to 0.99
[17]	Detect and number teeth, and evaluate diagnostic efficiency of detecting decayed teeth	DenseNet-121 and YOLOv7 for tooth numbering and caries detection, along with intersection over union to match numbered teeth with detected caries	1170 bitewing radiographs	F-scores of 0.99, 0.98, 0.82, and 0.84 for teeth detection, numbering, caries detection, and matching, respectively
[18]	Classify OPG radiographs	Proposed CNN-Transformer model architecture	1,418 OPG radiographs	Sensitivity of 0.94, specificity of 0.95, accuracy of 0.97, and F-score of 0.96
[19]	Detect and classify teeth using OPG radiographs for automatic filing of dental charts	DetectNet for tooth detection. ResNet for tooth type and condition classification	100 OPG radiographs	Detection sensitivity of 0.964. Type and condition classification accuracies of 0.93 and 0.98, respectively
[20]	Detect dental cysts, tumors, and abscess lesions	Grey Level Co-occurrence Matrix for feature extraction and Support Vector Machine (SVM) for classification	172 OPG radiographs	Accuracy of 0.98
[21]	Diagnose dental caries using periapical radiographs	Proposed a Multi-Input Deep Convolutional Neural Network Ensemble (MI-DCNNE) model based on score fusion	380 periapical radiographs	Accuracy of 0.99
[22]	Detect and diagnose dental caries using periapical radiographs	GoogLeNet InceptionV3 model	3000 periapical radiographs	Area Under The ROC Curve (AUC) of 0.92 on premolar, 0.89 on molar, and 0.85 on both
[23]	Classify dental caries in third molars using OPG radiographs	MobileNetV2 model	500 cropped OPG radiographs	Accuracy of 0.87 and AUC of 0.90
[24]	Diagnose children's caries using OPG radiographs	Proposed a context aware CNN that considers information among adjacent teeth for its predictions	304 OPG radiographs	Accuracy, precision, recall, F-score, and AUC of 0.83, 0.85, 0.88, 0.87, and 0.90, respectively
[25]	Diagnose true proximity between the mandibular canal and the third molar using OPG radiographs	Compared different deep learning models. EfficientDet was found to have the best performance	A total of 901 third molars from OPG radiographs	Accuracy of 0.79, sensitivity of 0.82, and specificity of 0.75
[26]	Detect the presence of supernumerary teeth in children using OPG radiographs	Compared three deep learning models. VGG16 resulted in the best performance	220 OPG radiographs	Accuracy, AUC, sensitivity, and specificity of 0.82, 0.89, 0.85, and 0.79, respectively
[27]	Detect dental caries from digital photographs	CNN adaptation from Single Shot MultiBox Detector and the hard negative mining algorithm	3932 oral photographs from 625 volunteers using consumer cameras	AUC of 0.86
[28]	Identify dental cavities from visual images of teeth	Proposed CNN model architecture	Public dataset consisting of 55 caries and 19 non-caries images	Maximum accuracy of 0.71
[29]	Identify periodontally compromised teeth	Utilized Faster R-CNN architecture	100 OPG radiographs	Precision of 0.81, recall of 0.80, and F-score of 0.81

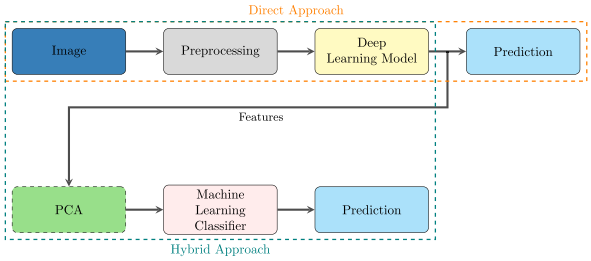

FIGURE 2. A comparison between the workflow of the direct and hybrid approaches. The hybrid approach includes an additional step of training a machine learning classifier using features extracted from a deep learning model, which, although seemingly complex, ultimately pays off as will be shown later. Note that Principal Component Analysis (PCA), shown in a dashed box, is optional.

B. ARCHITECTURAL APPROACHES

Two deep learning approaches for classification are evaluated in this paper: (1) the direct approach and (2) the hybrid approach. Fig. 2 shows a comparison between the workflows of both approaches. The approaches are detailed as follows.

1) DIRECT CLASSIFICATION APPROACH

Following this approach, the preprocessed images are fed directly into the fine-tuned deep learning model, which then predicts the disease using the model's fully-connected classification layer.

2) HYBRID APPROACH

The hybrid approach, the main emphasis of our paper, is illustrated in Fig. 3. This approach utilizes a pre-trained model for feature extraction and a traditional machine learning algorithm for the final classification. Initially, all annotated teeth are extracted from their respective radiographs using the provided coordinates (i.e., bounding box). These sub-images (i.e., the teeth) are then preprocessed and fed into a fine-tuned deep learning model, which extracts the learned features for each image. These features are subsequently used to train a traditional machine learning classifier for the final diagnosis. The general training and testing phases of the hybrid approach are shown in Fig. 4 with an optional Principal Component Analysis (PCA) step where the deep-

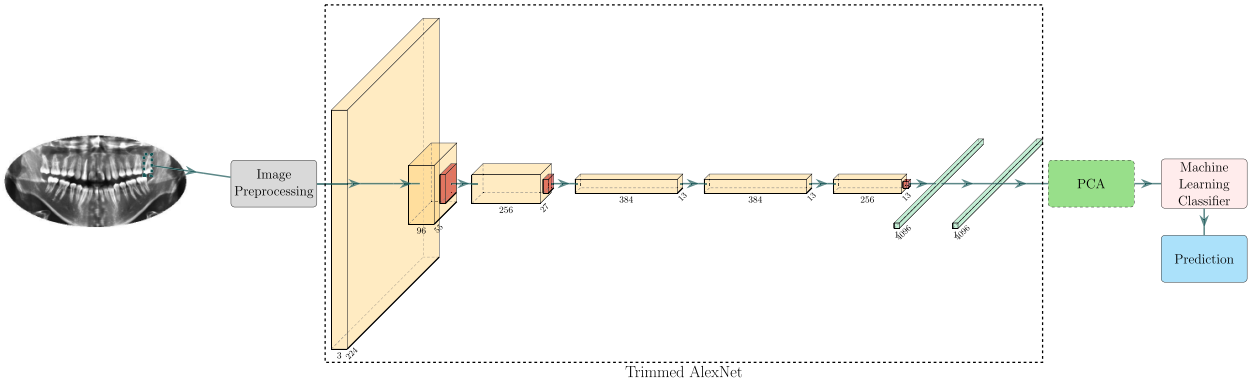

FIGURE 3. The hybrid architecture for automated teeth diagnosis using a deep learning model and a traditional machine learning classifier. Note that AlexNet is shown for illustration purposes.

learned features are reduced before training or testing the machine learning classifier.

C. DATA PREPROCESSING

Various preprocessing techniques were applied before utilizing the deep learning model on each image within the dataset. Initially, filtering of the dataset was done to ensure all the images consisted of a visible tooth with valid coordinates. This resulted in dropping ten instances due to invalid coordinates and six additional duplicate instances.

Following the filtering process, the individual teeth images were extracted and resized to 256 × 256 pixels. Each image was then cropped to a size of 224 × 224 pixels, centered around the middle of the image, as this is the input image size of the pre-trained models used. Subsequently, normalization was performed on the images using Eq. 1 with a mean and standard deviation (SD) of 0.485 (SD = 0.229), 0.456 (SD = 0.224), 0.406 (SD = 0.225) for each channel, respectively. These values are used to get the images into roughly the same data distribution as the one on which the models were trained. With these transformations, the images can be fed to the pre-trained models.

$$z_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i} \tag{1}$$

where z_{ij} is the normalized pixel value in channel i, x_{ij} is the original pixel value in channel i, μ_i is the mean of the pixel values in channel i, and σ_i is the standard deviation of the pixel values in channel i.

D. DATA AUGMENTATION

The initial distribution of images in the dataset was imbalanced, with the cavity and impacted classes being underrepresented. In particular, the cavity and impacted classes combined accounted for only 12.27% of the cleansed data. To handle the class imbalance problem, manual data augmentation using geometric transformations was performed. Since dental tasks often contain imbalanced classes, most works in the literature followed a similar augmentation approach to handle this issue [3], [16], [18], [19], [22], [23], [28], [29]. However, we remain conservative with the augmentation and do not perform it excessively, mainly to avoid the chances of overfitting. Particularly, train instances that belong to one of the two classes were rotated 45 degrees in both directions, tripling the number of instances belonging to the underrepresented classes. This resulted in a total of 10859 images after the filtering and augmentation processes, with a more balanced distribution of the four categories. Table 2 summarizes the number of instances for each class before and after augmentation. The augmentation process is crucial for the quality of the classification as it will help the model to learn about all classes equally and produce a non-biased prediction.

TABLE 2. Details of the dataset before and after the data augmentation performed on the train instances.

	Before	After
Cavity	641	1539
Filling	6083	6083
Impacted	496	1190
Implant	2047	2047
Total	9267	10859

E. TRANSFER LEARNING

Ten different models were utilized in the experiments, namely AlexNet [32], DenseNet121 [33], EfficientNet [34], MobileNetV2 [35], MobileNetV3Large [36], ResNet50 [37], VGG16 [38], VGG19 [38], Vision Transformer (ViT) [39], and You Only Look Once (YOLO) [40] with all being pre-trained on the ImageNet dataset [41] or the COCO dataset [42]. The use of the transfer learning approach allows the model to utilize knowledge gained from the other datasets, which can result in superior performance, even with our limited dataset. In regard to YOLO, the large variant of YOLOv10 was utilized in the experiments due to its size-to-accuracy trade-off that outperforms earlier versions [40].

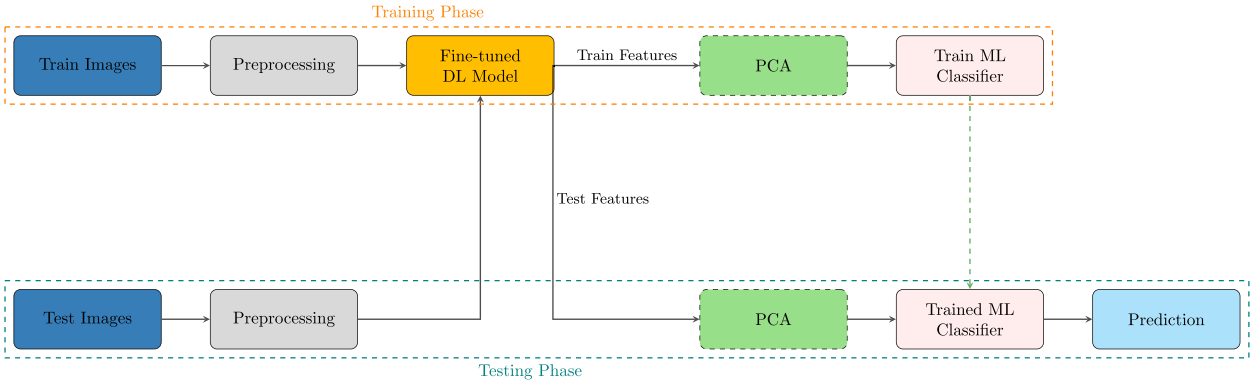

FIGURE 4. A general overview of the training and testing phases of the hybrid approach. DL: deep learning; ML: machine learning.

Apart from YOLO, all models used for direct classification are also used for feature extraction in the hybrid approach. The models are trimmed to extract the deep-learned features as opposed to a direct classification. The features extracted from the models were paired with one of seven machine learning classifiers, which is further discussed in subsection III-G.

Being one of the most established models, AlexNet has proven its performance in imaging tasks. Since all models essentially follow the same conceptual framework, the details of AlexNet exclusively are included. AlexNet comprises seven layers excluding the classification layer. The initial five layers are convolutional layers, three of which are followed by max-pooling layers, with two fully connected layers preceding the classification layer, which is trimmed in our case. The convolutional layers are the backbone of AlexNet and all other CNN models. They work by sliding a fixed window, known as the kernel, across each channel and performing the same computation at every point. The convolutional layers are needed to extract spatial information about the images by learning the different filters. In the case of AlexNet, the number and size of the kernels across the different convolutional layers vary. A visual representation of the layers with the different numbers of kernels can be found in Fig. 3.

One key characteristic of AlexNet lies in its Local Response Normalization (LRN) process that occurs after the ReLU activation functions in the first and second convolutional layers. LRN normalization, as seen in Eq. 2, is applied within a local neighborhood of neighboring channels across the feature dimension [32]. This technique enhances the model's robustness to overfitting, also known as the high variance problem.

$$b^i_{x,y} = \frac{a^i_{x,y}}{\left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2\right)^\beta} \tag{2}$$

where $b^i_{x,y}$ is the channel's activity after normalization, $a^i_{x,y}$ is the activity of a channel computed by applying the ReLU activation function to its input, N is the total number of kernels in the layer, n is the size of the local neighborhood, and k, α, and β are hyperparameters.

F. DIMENSIONALITY REDUCTION

PCA reduces the complexity of high-dimensional data while preserving its trends and patterns [43]. It achieves this by converting the data into a smaller number of dimensions that summarize the key features. In our framework, PCA can be used after the feature extraction process, to minimize the dimensionality of the data. PCA can be computed using Singular Value Decomposition (SVD) as follows:
Let A be a real $m \times n$ matrix with $m \leq n$. It is well known that A can be decomposed as

$$A = U \Sigma V^T \tag{3}$$

where $U^T U = V^T V = V V^T = I$ and

$$\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r) \tag{4}$$

The matrix U consists of m orthonormal eigenvectors associated with the m largest eigenvalues of AA^T, and the matrix V consists of the orthonormal eigenvectors of $A^T A$. The diagonal elements of Σ are the non-negative square roots of the eigenvalues of $A^T A$; these are called singular values.

G. MACHINE LEARNING CLASSIFIERS

Given the different perks of every classifier, seven different traditional classifiers were experimented with, namely Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) with a Linear kernel, SVM with a Radial Basis Function (RBF) kernel, K-Nearest-Neighbor (KNN), Naive Bayes (NB), and Logistic Regression (LR) [44].

- DT is a supervised machine learning algorithm used for both classification and regression tasks. DT works by recursively splitting the data into subsets based on the features that best separate the target variable.
- RF is an ensemble learning method that combines multiple DTs and takes the prominent result. In other words, it is a group of DTs, which together provide a

more accurate and stable prediction than a single tree could.

- SVMs, also known as maximal margin classifiers, solve an optimization problem by searching for the decision boundary that maximizes the margin between the data of different classes. Associated with SVMs are the "kernel tricks" that facilitate the implicit mapping of features to a dimension where they become linearly separable using a kernel function. A kernel function takes its input vectors in the original space and returns the dot product of the vectors in the feature space. This enables the problem to be solved in a higher-dimensional space, sometimes in the original input space in the case of a linear kernel, and then maps it back to the original dimension in a computationally efficient manner. In the case of a linear kernel, the decision function is the weighted sum of the features set to 0 (Eq. 5).

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0 \tag{5}$$

where $\mathbf{x}$ is the feature vector, $\mathbf{w}$ is the weight vector, and b is the bias.

- The RBF kernel, which is very popular in its utility with an SVM, can be represented using the equation:

$$K(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}} \tag{6}$$

The "$\|\mathbf{x} - \mathbf{y}\|^2$" is the squared Euclidean distance between two points in the input space, and σ is a free parameter that can be used to tune the equation. The parameter gamma, denoted as γ, can be used to replace the $\frac{1}{2\sigma^2}$ term in the equation, resulting in a simplified form:

$$K(\mathbf{x}, \mathbf{y}) = e^{-\gamma\|\mathbf{x}-\mathbf{y}\|^2} \tag{7}$$

In this case, gamma determines the influence of each training example. A high value of gamma represents close reach, in which points that are close to the boundary have a high impact on the decision boundary. On the other hand, a low value of gamma represents far reach, in which points that are far from the boundary have a high influence on the decision boundary.
C, another hyperparameter of SVM in its implementation in Scikit-learn [45], is also crucial for the performance of the model. C controls the tradeoff between smooth decision boundaries and classifying training points correctly. Similar to gamma, large values of C can make the model prone to overfitting.

- NB is a classification algorithm based on Bayes' Theorem (Eq. 8) with the "naive" assumption of independence among features. Despite its simplicity and the often unrealistic assumption of feature independence, it performs well in many scenarios, such as text classification.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}, \quad \text{where} \quad P(B) \neq 0 \tag{8}$$

- KNN is a simple classifier based on the classification of an object based on the classes of its closest neighbors. This implies finding the distance measure of the object's features with all other objects' features.
- LR, similar to Linear SVM, is also based on the equation of the line, but it utilizes the softmax function (Eq. 9) to obtain the decision boundary for multi-class classification. The softmax function returns the probability measure of a particular object belonging to each of the possible classes, all adding up to 1.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{9}$$

where $\mathbf{z}$ is the input vector, z_j is the j-th element of the input vector, K is the number of classes, and $\sigma(\mathbf{z})_j$ is the output of the softmax function for the j-th class. Each element of the output vector represents the probability that the input belongs to a particular class.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EVALUATION METRICS

In this work, we use the following evaluation metrics:

- **Accuracy**: This is the ratio of correctly predicted observations to the total observations. Mathematically, it is represented as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{10}$$

- **Precision**: This is the ratio of correctly predicted positive observations to the total predicted positive observations. It is given by:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{11}$$

- **Recall** (Sensitivity): This is the ratio of correctly predicted positive observations to all observations in the actual class. It is given by:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{12}$$

- **F-Score**: This is the harmonic mean between Precision and Recall. It is given by:

$$F - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{13}$$

- **Area Under The ROC Curve (AUC)**: The Receiver Operating Characteristic (ROC) curve is a probabilistic curve that illustrates the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) at various thresholds. The AUC measures the entire two-dimensional area underneath the ROC curve from (0,0) to (1,1). An AUC of 1 indicates perfect classification.

B. EXPERIMENTAL SETUP

Two types of experiments are performed: (1) direct classification and (2) the hybrid approach with the CNN (or transformer) feature extractor and a machine learning classifier. In both cases, we fine-tune all the pre-trained models per the settings in subsection IV-C. In (1) after fine-tuning, all models (except YOLO) are evaluated using their fully-connected classification layer. Subsequently, for (2), the classification layer is removed from the fine-tuned models to enable the extraction of features instead of a direct classification. The features of all training data are then used to train a machine learning classifier, which is evaluated using the features of the testing data.

C. IMPLEMENTATION DETAILS

All models are fine-tuned with the same setup apart from YOLO. The models are fine-tuned for 50 epochs with a learning rate of 1e-05, a batch size of 64, and a weight decay of 1e-03. YOLO was fine-tuned for 50 additional epochs with a learning rate of 0.00125 and a batch size of 32. The epochs were set such that the models would converge while avoiding overfitting. For all models, a stratified 70-10-20 split is made with the majority used for the fine-tuning process and the remaining used for validation and inference, respectively. As discussed earlier in subsection III-D, data augmentation is performed on the underrepresented classes of the training set. An Adam optimizer and a cross-entropy loss function are employed. PyTorch [46], OpenCV [47], and Scikit-learn [45] libraries are used for the experiments which are conducted on one 40 GB NVIDIA A100 GPU. All machine learning classifiers are used with the default parameters' values unless specified otherwise.

TABLE 3. Summary of the classification results obtained from various pre-trained deep learning models.

Model	Accuracy	F-score
AlexNet	0.92	0.92
DenseNet121	0.92	0.90
EfficientNet	0.90	0.87
MobileNetV2	0.91	0.90
MobileNetV3Large	0.90	0.88
ResNet50	0.92	0.91
VGG16	0.92	0.92
VGG19	0.93	0.93
ViT	**0.96**	**0.96**
YOLOv10	0.88	0.76

TABLE 4. Classification report of the best-performing model, ViT.

	Precision	Recall	F-score	Support
Cavity	0.92	0.70	0.80	128
Filling	0.97	0.98	0.97	1217
Impacted	0.92	0.96	0.94	99
Implant	0.97	0.98	0.98	410
Accuracy			0.96	1854
Macro Avg	0.94	0.91	0.92	1854
Weighted Avg	0.96	0.96	0.96	1854

D. EXPERIMENTAL FINDINGS AND INTERPRETATION

In this section, the experimental findings and their interpretations for each of the classification approaches considered are discussed.

1) DIRECT CLASSIFICATION APPROACH

We summarize the results of classification using the direct classification models in Table 3. The results show a clear superiority of ViT, achieving an accuracy and F-score of 0.96, which is a record accuracy on the dataset. However, as shown in Table 4, we observe a low recall for the cavity class caused by the model misclassifying the cavity class samples as filling. Fig. 5 illustrates the confusion matrix for ViT which further highlights this phenomenon. We will discuss our reasoning for these misclassifications in section IV-E.

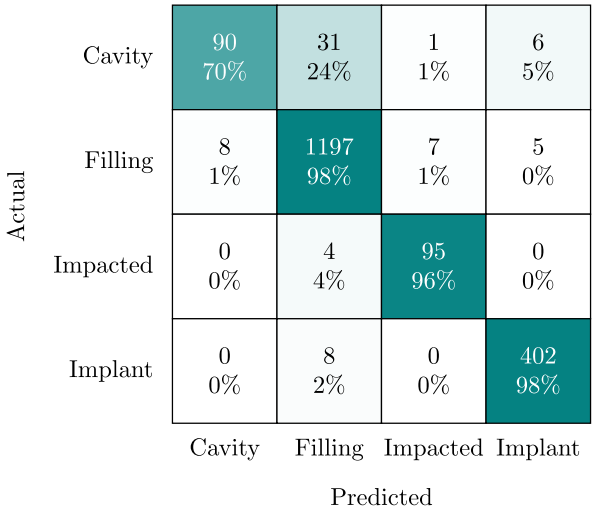

FIGURE 5. Confusion matrix of the best-performing model, ViT.

The second highest performance, achieved using VGG19, was 3% away from ViT's performance. The reason for this difference stems from ViT's attention mechanism where the model focuses on different parts of the image [39]. All other used models do not have an attention mechanism and instead focus on the entire spatial structure of the image, which can cause the model to capture irrelevant information. This is very relevant to our dataset as the tooth, especially when resized, may contain pixels that do not contribute any information (e.g., background). This exceptional performance of ViT is on par with the metrics reported by [18] using a transformer-based architecture for dental diagnosis, which in turn validates their findings on the effectiveness of transformer models within dental imaging. Additionally, YOLO has placed last in terms of F-score. This opposes other papers that chose YOLO as their model of choice [3], [17], showing that YOLO is not the best model for our particular dataset.

2) HYBRID APPROACH

When the features of the models are used in conjunction with a machine learning classifier (i.e., the hybrid approach),

TABLE 5. Comparison of the F-scores obtained from traditional machine learning classifiers using features extracted by different models.

	Decision Tree	Random Forest	Linear SVM	RBF SVM	KNN	Naive Bayes	Logistic Regression
AlexNet	0.88	0.91	**0.93**	0.92	0.92	0.80	0.92
DenseNet121	0.84	0.89	0.91	0.91	0.88	0.89	0.91
EfficientNet	0.84	0.88	0.89	0.90	0.90	0.89	0.89
MobileNetV2	0.84	0.88	0.91	0.92	0.91	0.90	0.91
MobileNetV3Large	0.81	0.87	0.88	0.91	0.91	0.87	0.90
ResNet50	0.88	0.91	0.91	0.91	0.92	0.91	0.91
VGG16	0.90	0.92	0.92	0.92	0.92	0.87	0.92
VGG19	0.90	0.92	**0.93**	**0.93**	**0.93**	0.89	**0.93**
ViT	**0.95**	**0.95**	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**

an increment in performance is observed. The results of these experiments are shown in Table 5. Furthermore, the difference with the direct classification results can be seen in Table 6, where all CNN models noticed an improved or identical F-score with at least one machine learning classifier compared to that of the direct classification approach. ViT achieved the highest F-score of 0.96, further showcasing its superiority. Additionally, both AlexNet and VGG19 achieved the second highest F-score of 0.93. While the F-scores of VGG19 are consistent within different classifiers, emphasizing the quality of its features, AlexNet achieved the same F-score of 0.93 with Linear SVM. The parameter and memory requirements of both models can be seen in Table 7 with AlexNet comprising less than half of VGG19's parameters [32], [38], [39]. Hence, since AlexNet has significantly fewer parameters than VGG19, AlexNet is deemed to be preferable over VGG19. Table 7 also showcases a comparison with ViT, the highest-performing model, with AlexNet having 26 million fewer parameters.

TABLE 6. Comparison between the F-scores of the models using the direct classification approach versus the hybrid approach with the highest-performing machine learning classifier of each model.

Model	Direct classification	Hybrid
AlexNet	0.92	0.93
DenseNet121	0.90	0.91
EfficientNet	0.87	0.90
MobileNetV2	0.90	0.92
MobileNetV3Large	0.88	0.91
ResNet50	0.91	0.92
VGG16	0.92	0.92
VGG19	0.93	0.93
ViT	**0.96**	**0.96**

TABLE 7. Comparison of the number of layers, parameters, and memory requirements of AlexNet, VGG19, ViT, and the hybrid approach.

Model	Layers	Parameters	Parameter Memory
AlexNet	8	∼60 million	∼233 MB
VGG19	19	∼144 million	∼548 MB
ViT	12	∼86 million	∼336 MB
Hybrid (AlexNet with RBF SVM)	8	∼60.2 million	∼234 MB

More tests were conducted with AlexNet, exclusively, given that it has resulted in the highest performance within the scope of our work encompassing a hybrid approach with memory constraints. Since multiple machine learning classifiers achieved comparable results on features obtained from AlexNet, we performed hyperparameter tuning on them for a potential increase in performance. We considered tuning the hyperparameters for Linear SVM, RBF SVM, and KNN since they are AlexNet's top-performing classifiers. When varying the number of neighbors (i.e., K) for the KNN classifier, no significant improvement in the metrics is seen with some signs of overfitting at small values of K. Moreover, the results of SVM showed that the model is overfitting at certain values of C with the linear kernel. However, the RBF kernel, with specific C and gamma, showed promising results. In particular, when C was set to 7 using the RBF kernel, the model performance was at its highest while also avoiding overfitting. Therefore, we picked that hyperparameter value for our final experiment.

Using PCA, we tested using different numbers of principal components. It is found that RBF SVM (with $C = 7$ and default gamma) performed the best at around 90 components, carrying 89.5% of the explained variance. This means that out of the total variation in the original 4096 features obtained from AlexNet, 89.5% was explained by only 90 components which is a great reduction of dimensionality. For illustration, Fig. 6 shows the 12 principal components of the highest explained variance holding 68.3% of the total explained variance while the lower 80 components hold 21.2% of the total explained variance. The results of the best-performing RBF SVM classifier, configured with the chosen parameters utilizing AlexNet's features, are summarized in the confusion matrix depicted in Fig. 7 with a final accuracy of 94%. The precision, recall, and F-score for each class are shown in Table 8. Similar to ViT's direct classification, this model also has low recall for the cavity class, which is also discussed in section IV-E.

TABLE 8. Classification report of the hybrid architecture encompassing AlexNet and SVM with an RBF kernel and PCA for dimensionality reduction after hyperparameter tuning.

	Precision	Recall	F-score	Support
Cavity	0.81	0.52	0.63	128
Filling	0.95	0.97	0.96	1217
Impacted	0.86	0.92	0.89	99
Implant	0.95	0.97	0.96	410
Accuracy			0.94	1854
Macro Avg	0.89	0.84	0.86	1854
Weighted Avg	0.93	0.94	0.93	1854

Additionally, the ROC curve shown in Fig. 8 highlights the dominant performance of the model in predicting all

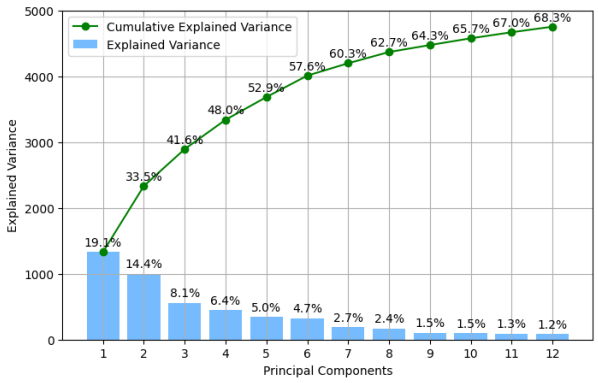

FIGURE 6. Explained variance of the top 12 principal components.

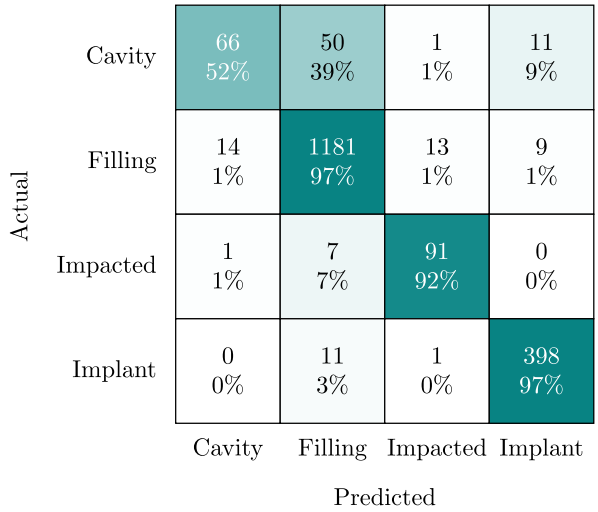

FIGURE 7. Confusion matrix of the best-performing classifier using AlexNet's features with RBF SVM and PCA.

four classes, with areas close to 1. However, the cavity class shows the poorest performance with an area of 0.93, which demonstrates the model's slight difficulty in predicting instances from that class as opposed to the other classes.

E. DISCUSSION AND LIMITATIONS

As shown by the confusion matrices depicted in Fig. 5 and Fig. 7 for the direct and hybrid approaches, respectively, both models experience some difficulty in distinguishing cavities from fillings. While ViT does a better job at this task, a considerable number of cavities are still being confused for fillings. We suspect multiple underlying causes for this behavior. Despite the filling class being the dominant class, the class imbalance is unlikely to be the reason since instances from the impacted class are generally not being confused as fillings. One reason, however, could be attributed to the low-resolution images that make it hard to distinguish textures. As a consequence of these low-resolution images, we suspect the models are simply not capturing the subtle differences that distinguish a cavity from a filling. This raises a question about the overall reliability of AI-based decision support systems and the practicality of their deployment in a clinical setting. However, despite decision support systems being prone to mistakes, they ultimately minimize human-induced errors and are meant to complement radiologists rather than replace them [48].

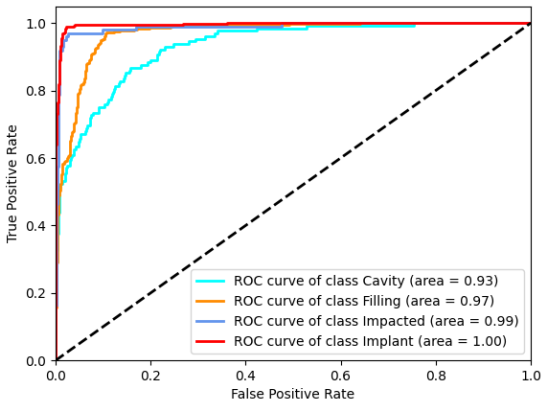

FIGURE 8. ROC curve of the best-performing classifier using AlexNet's features with RBF SVM and PCA.

TABLE 9. Comparison of the training and inference time between the direct approach (ViT) and the hybrid approach (AlexNet with PCA and RBF SVM). Inference was performed using a CPU on one random batch of size 64.

	Direct (ViT)	Hybrid (AlexNet with RBF SVM)
Training	3300 s	285 s
CPU Inference	5.06 s	0.34 s

Furthermore, the results obtained with the hybrid framework exceed expectations, demonstrating superiority in efficiently detecting dental diseases with a reduced parameter count as opposed to other state-of-the-art methods as seen in Table 7. While it might look more complicated, the hybrid approach may be looked at as a simplification of the direct classification approach, as it utilizes a machine learning classifier which is generally simpler and faster than the classification layers of a deep learning model. The speed of a model, particularly for inference, is significant for its deployment in a practical setting. As seen in Table 9, the hybrid framework (AlexNet with RBF SVM) is significantly faster than ViT taking 285 and 0.34 seconds for training and CPU inference, respectively as opposed to 3300 and 5.06 seconds with ViT. Additionally, while all training experiments were conducted on a 40 GB GPU, the framework only utilized 5% of the GPU's memory, meaning it could be trained on a laptop with a minimum of 2 GB of GPU memory. This implies the ability to adopt incremental learning, where a clinic's laptop can be used on-demand to further train the model on new data. The importance of incremental learning has been shown in the past [49] and the hybrid approach not only facilitates real-time inference but also facilitates incremental learning where clinics would not have to worry about needing a large-scale GPU to actively train the framework on new instances. This reduction in computational requirements while retaining a similar performance to that of a more complex model suggests that the hybrid approach may

be better suited for clinical adoption as opposed to other state-of-the-art deep learning models.

TABLE 10. Comparison of results from various OPG classification-based research studies with our results. Note that different datasets were used in each study, making the comparison between the accuracies inconclusive.

Paper	Method	Accuracy
[3]	YOLOv3	0.99
[16]	BDU-Net and nnU-Net	0.93 (calculated)
[18]	CNN-Transformer	0.97
[19]	ResNet	0.98
[23]	MobileNetV2	0.87
[24]	Context aware CNN	0.83
Ours (Hybrid)	AlexNet with PCA and RBF SVM	0.94
Ours (Direct)	ViT	0.96

Table 10 compares our work and other studies that target a condition classification task using OPGs. However, the studies included each utilize a different dataset that has not been made public. This explains the inflated accuracy of some studies, as they only include relevant radiographs and exclude problematic ones (e.g., those with artifacts or rare diseases) in their private datasets. In our case, however, we only excluded duplicated and invalid instances. As mentioned earlier, we could not find other studies that utilize the dataset we used in our work.

The limitations of our work are directly correlated with the available data, as dental annotations are both expensive and time-consuming. This includes the quantity of annotations, the diversity of the annotated categories, and the tasks to which the annotations target. This is particularly relevant due to the natural imbalance of the data when it comes to rare dental diseases, which can cause the model to perform poorly if not handled appropriately [50].

Future work could involve utilizing a larger and more diverse dataset which would alleviate the mentioned issues while also enhancing the framework's generalizability and applicability across different populations and conditions. Finally, integrating different types of dental imaging modalities in addition to OPGs, such as dental Cone Beam Computed Tomography (CBCT) scans, would be the next approach for a more comprehensive and well-rounded framework to be incorporated in a real-world setting.

V. CONCLUSION

In this work, we followed an experimental approach to the diagnosis of dental diseases using OPG radiographs, leveraging transfer learning. The hybrid approach involves the use of a pre-trained deep learning model for feature extraction followed by a traditional machine learning algorithm for disease prediction. The best accuracy of 96%, which is also the highest recorded accuracy on the dataset used, was achieved by utilizing ViT with both the direct and hybrid approaches. Furthermore, the hybrid framework utilizing AlexNet's features followed by SVM with an RBF kernel achieved an accuracy of 94% with a large cut in the number of parameters. These promising results emphasize the effectiveness of a deep learning-based diagnostic decision support system. Most notably, the balance between complexity and performance makes the hybrid approach favorable for clinical adoption with it enabling on-demand inference and incremental learning.

ACKNOWLEDGMENT

This article represents the opinions of the authors and does not mean to represent the position or opinions of the American University of Sharjah.

REFERENCES

[1] I. Appollonio, C. Carabellese, A. Frattola, and M. Trabucchi, "Dental status, quality of life, and mortality in an older community population: A multivariate approach," *J. Amer. Geriatrics Soc.*, vol. 45, no. 11, pp. 1315–1323, Nov. 1997.

[2] S. Ajwani, K. J. Mattila, T. O. Narhi, R. S. Tilvis, and A. Ainamo, "Oral health status, C-reactive protein and mortality—A 10 year follow-up study," *Gerodontology*, vol. 20, no. 1, pp. 32–40, Jul. 2003.

[3] 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 X-ray OPG images," *Sensors*, vol. 22, no. 19, p. 7370, Sep. 2022.

[4] G. Alotaibi, M. Awawdeh, F. F. Farook, M. Aljohani, R. M. 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, no. 1, p. 399, Sep. 2022.

[5] J. Cosson, "Interpreting an orthopantomogram," *Austral. J. Gen. Pract.*, vol. 49, no. 9, pp. 550–555, Sep. 2020.

[6] D. McComb and L. Tam, "Diagnosis of occlusal caries: Part I. Conventional methods," *J. Can. Dental Assoc.*, vol. 67, no. 8, pp. 1–4, 2001.

[7] Dentaly. (2023). *Ai in Dentistry: Perceptions, Insights, and Possibilities*. Accessed: May 2024. [Online]. Available: https://www.dentaly.org/us/research/ai-in-dentistry/#:~:text=of%20surveyed%20dentists%20have%20implemented%20AI%20in%20their,in%20dental%20clinics%20could%20be%20performed%20by%20AI

[8] Z. Meng, J. Yuan, and Z. Li, "Trajectory-pooled deep convolutional networks for violence detection in videos," in *Proc. Int. Conf. Virtual Storytelling*, 2017, pp. 437–447.

[9] K. Paramasivam, M. M. R. Sindha, and S. B. Balakrishnan, "KNN-based machine learning classifier used on deep learned spatial motion features for human action recognition," *Entropy*, vol. 25, no. 6, p. 844, May 2023.

[10] S. R. Mathu Sudhanan, K. Priya, and P. Uma Maheswari, "Deep learnt features and machine learning classifier for texture classification," *J. Phys., Conf. Ser.*, vol. 2070, no. 1, Nov. 2021, Art. no. 012108.

[11] D. Varam, R. Mitra, M. Mkadmi, R. A. Riyas, D. A. Abuhani, S. Dhou, and A. Alzaatreh, "Wireless capsule endoscopy image classification: An explainable AI approach," *IEEE Access*, vol. 11, pp. 105262–105280, 2023.

[12] T. S. Priya and T. Ramaprabha, "Resnet based feature extraction with decision tree classifier for classificaton of mammogram images," *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 2, pp. 1147–1153, Apr. 2021.

[13] M. Raveenthini and R. Lavanya, "Multiocular disease detection using a generic framework based on handcrafted and deep learned feature analysis," *Intell. Syst. with Appl.*, vol. 17, Feb. 2023, Art. no. 200184.

[14] D. P. Yadav, A. Sharma, M. Singh, and A. Goyal, "Feature extraction based machine learning for human burn diagnosis from burn images," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–7, 2019.

[15] S. Khan and S.-P. Yong, "A comparison of deep learning and hand crafted features in medical image modality classification," in *Proc. 3rd Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Aug. 2016, pp. 633–638.

[16] 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, no. 1, p. 358, Jun. 2023.

[17] B. Ayhan, E. Ayan, and Y. Bayraktar, "A novel deep learning-based perspective for tooth numbering and caries detection," *Clin. Oral Investigations*, vol. 28, no. 3, p. 178, Feb. 2024.

[18] Y. Li and J. Zhang, "Multi-label dental image classification via vision transformer for orthopantomography X-ray images," *Comput.-Aided Des. Appl.*, vol. 21, pp. 198–207, Feb. 2024.
[19] C. Muramatsu, T. Morishita, R. Takahashi, T. Hayashi, W. Nishiyama, Y. Ariji, X. Zhou, T. Hara, A. Katsumata, E. Ariji, and H. Fujita, "Tooth detection and classification on panoramic radiographs for automatic dental chart filing: Improved classification by multi-sized input data," *Oral Radiol.*, vol. 37, no. 1, pp. 13–19, Jan. 2021.
[20] V. S. Kumar, P. R. Kumar, P. K. Yadalam, R. V. Anegundi, D. Shrivastava, A. A. Alfurhud, I. T. Almaktoom, S. A. A. Alftaikhah, A. H. L. Alsharari, and K. C. Srivastava, "Machine learning in the detection of dental cyst, tumor, and abscess lesions," *BMC Oral Health*, vol. 23, no. 1, pp. 1–11, Nov. 2023.
[21] A. Imak, A. Celebi, K. Siddique, M. Turkoglu, A. Sengur, and I. Salam, "Dental caries detection using score-based multi-input deep convolutional neural network," *IEEE Access*, vol. 10, pp. 18320–18329, 2022.
[22] 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. Dentistry*, vol. 77, pp. 106–111, Oct. 2018.
[23] S. Vinayahalingam, S. Kempers, L. Limon, D. Deibel, T. Maal, M. Hanisch, S. Bergé, and T. Xi, "Classification of caries in third molars on panoramic radiographs using deep learning," *Sci. Rep.*, vol. 11, no. 1, p. 12609, Jun. 2021.
[24] X. Zhou, G. Yu, Q. Yin, Y. Liu, Z. Zhang, and J. Sun, "Context aware convolutional neural network for children caries diagnosis on dental panoramic radiographs," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–8, Sep. 2022.
[25] K. J. Jeon, H. Choi, C. Lee, and S.-S. Han, "Automatic diagnosis of true proximity between the mandibular canal and the third molar on panoramic radiographs using deep learning," *Sci. Rep.*, vol. 13, no. 1, p. 22022, Dec. 2023.
[26] Y. Mine, Y. Iwamoto, S. Okazaki, K. Nakamura, S. Takeda, T. Peng, C. Mitsuhata, N. Kakimoto, K. Kozai, and T. Murayama, "Detecting the presence of supernumerary teeth during the early mixed dentition stage using deep learning algorithms: A pilot study," *Int. J. Paediatric Dentistry*, vol. 32, no. 5, pp. 678–685, Sep. 2022.
[27] 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, no. 1, pp. 173–181, Jan. 2022.
[28] A. Sonavane, R. Yadav, and A. Khamparia, "Dental cavity classification of using convolutional neural network," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1022, no. 1, Jan. 2021, Art. no. 012116.
[29] B. Thanathornwong and S. Suebnukarn, "Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks," *Imag. Sci. Dentistry*, vol. 50, no. 2, p. 169, 2020.
[30] S. E. Uribe, A. Sofi-Mahmudi, E. Raittio, I. Maldupa, and B. Vilne, "Dental research data availability and quality according to the FAIR principles," *J. Dental Res.*, vol. 101, no. 11, pp. 1307–1313, Oct. 2022.
[31] I. K. Team, "Dental radiography," Kaggle, San Francisco, CA, USA, Tech. Rep., 2023.
[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., 2012, pp. 1097–1105.
[33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[34] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[36] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, *arXiv:1409.1556*.
[39] 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," 2020, *arXiv:2010.11929*.
[40] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "YOLOv10: Real-time End-to-End object detection," 2024, *arXiv:2405.14458*.
[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[42] T. Lin, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[43] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Math.*, vol. 14, no. 5, pp. 403–420, Apr. 1970.
[44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), 2nd ed., New York, NY, USA: Springer, 2009.
[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikitlearn: Machine Learning in Python*, 2011.
[46] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. NIPS*, Dec. 2019, pp. 8024–8035.
[47] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, vol. 120, pp. 122–125, 2000. [Online]. Available: https://github.com/opencv/opencv/wiki/CiteOpenCV
[48] Z. M. Semerci and S. Yardımcı, "Empowering modern dentistry: The impact of artificial intelligence on patient care and clinical decision making," *Diagnostics*, vol. 14, no. 12, p. 1260, Jun. 2024.
[49] Y. Luo, L. Yin, W. Bai, and K. Mao, "An appraisal of incremental learning methods," *Entropy*, vol. 22, no. 11, p. 1190, Oct. 2020.
[50] F. Pethani, "Promises and perils of artificial intelligence in dentistry," *Austral. Dental J.*, vol. 66, no. 2, pp. 124–135, Jun. 2021.

AHMED ALABD-ALJABAR (Member, IEEE) is currently pursuing the B.S. degree in computer science, with a minor in data science with the American University of Sharjah (AUS), United Arab Emirates. He is a Computer Vision Research Assistant with the Department of Computer Science and Engineering. He has been placed on the Dean's List for all his semesters at AUS. He also received the Undergraduate Research Grant for further research. His current research interests include computer vision and biomedicine.

ZAIN RAISAN received the dual B.S. degree (Hons.) in computer engineering and computer science from the American University of Sharjah (AUS), United Arab Emirates. He was a Research Assistant with the Department of Electrical Engineering. His current research interests include artificial intelligence, computer architecture, embedded systems, and cyber security. He is a member of the Alpha Lambda Delta Honor Society and the IEEE-Eta Kappa Nu Honor Society.

MOHAMMED ADNAN was born in Baghdad, Iraq, in 2003. He is currently pursuing the dual B.S. degree in computer engineering and mathematics with the American University of Sharjah (AUS), United Arab Emirates. His current research interests include computer vision, machine learning, computer hardware, and embedded systems. He is a member of the Tau Beta Pi Engineering Honor Society and the Alpha Lambda Delta Honor Society.

SALAM DHOU (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from Jordan University of Science and Technology, Jordan, in 2004 and 2007, respectively, and the Ph.D. degree in electrical and computer engineering from Virginia Commonwealth University, USA, in 2013. She was a Postdoctoral Research Fellow with the Department of Radiation Oncology, Harvard Medical School, USA, from 2013 to 2016. She is currently an Associate Professor with the Department of Computer Science and Engineering and the Biomedical Engineering Graduate Program, American University of Sharjah (AUS), United Arab Emirates. Her research interests include medical imaging and informatics, machine learning, and data mining.