RESEARCH PAPER
Future Research in AI & IoT, 2025, 1(1)
Homepage: https://frai.journals.umz.ac.ir/
DOI: 10.22080/frai.2025.29272.1013

Analysis of Panoramic Dental Images for Dental Symptom Differentiation Based on Deep Learning

Hatef Hosseinpour[1], Jamal Ghasemi[2]*, Farida Abesi[3]

[1]*M.Sc. Student, Department of Engineering and Technology, University of Mazandaran, Iran, h.hossein05@umzil.umz.ac.ir.*

[2]*Associate Professor, Department of Engineering and Technology, University of Mazandaran, Iran, j.ghasemi@umz.ac.ir.*

[3]*Associate Professor, Department of Oral and Maxillofacial Radiology, Babol University of Medical Sciences, Babol, Iran, f.abesi@yahoo.com*

Article Info
Received 18 May 2025
Accepted 03 June 2025
Available online 01 June 2025

Keywords:
Deep learning;
Panoramic Images;
Caries;
Healthcare.

Abstract:
This research presents a hierarchical approach using deep learning methods to categorize panoramic dental images and identify various dental conditions. The images were initially preprocessed, and the areas of interest were cropped and labeled meticulously. During the initial stage, a pre-trained deep learning model was utilized to differentiate between images with and without caries. Afterward, the images without caries were classified into healthy and amalgam-filled categories in the second stage. Following its preparation, the dataset was preprocessed to a great level of accuracy, after which a number of models were tested and compared. Among those under test, DenseNet121 performed better in the initial stage of caries detection, with a total accuracy rate of 83.95%. Also, in the subsequent stage, EfficientNet-B4 performed best in the detection of amalgam fillings and differentiating them from healthy dentition, with an accuracy rate of 98.28%. This research shows that hierarchical methods, when integrated with deep learning models, can yield accurate and trustworthy outcomes in dental image assessment. Moreover, the application of various deep neural network structures is very helpful in enhancing classification precision. The results indicate that the suggested method shows significant potential for automatically detecting dental abnormalities, such as caries and amalgam fillings, thus serving as a complementary tool in clinical diagnostic procedures. In summary, this study presents an efficient and scalable solution for the analysis of complex dental images, which can be integrated into artificial intelligence-based healthcare systems.

***Corresponding Author:** j.ghasemi@umz.ac.ir

Supplementary information: Supplementary information for this article is available at https://frai.journals.umz.ac.ir/

Please cite this paper as: Hosseinpour, H., Ghasemi, J., & Abesi, F. (2025). Analysis of Panoramic Dental Images for Dental Symptom Differentiation based on Deep Learning. Future Research on AI and IoT, 1(1), 1-9. DOI: 10.22080/frai.2025.29272.1013.

1. Introduction

Dental diseases are now known to be widespread public health issues with significant prevalence and incidence rates in a many parts of the world [1, 2]. The most important factor is establishing an accurate and timely diagnosis to limit complications such as loss of the tooth and/or infection [3]. Diagnosing dental diseases is inherently flawed due to numerous factors, including the dentist's experience, knowledge, and skill level. The diagnosis process is often stressful for both patients and dentists due to the time involved [4]. Traditional diagnosis based on visual and manual interpretation of radiographic images is subjective and prone to interpersonal differences, often causing extensive differences in treatment planning [5]. Therefore, with all the limits and barriers to traditional diagnostics, the need for more effective ways to diagnose dental diseases is becoming increasingly important [6, 7]. Advancements in information retrieval techniques - such as deep learning - have significantly improved routine diagnostics in the medical field, and there is also an increasing number of studies on deep learning applications in oral disease diagnoses, especially dental caries [8, 9]. It is clear that Convolutional Neural Networks (CNNs) can detect caries lesions and separate healthy tissue from caries tissue accurately [10, 11]. The merger of deep learning and dentistry will likely yield higher diagnostic accuracy and efficiency, leading to personalized treatment planning and improved patient outcomes [12, 13]. In recent research, deep learning techniques have been used to classify dental diseases. In these approaches, various efforts have been made in the field of dental problem detection and

ISSN 3092-7552

classification [14]. In [15], a system for detecting dental caries in panoramic images using transfer learning has been proposed, which extracts relevant features using a capsule network. This model has shown acceptable performance in caries detection. In [16], three pre-trained models, including DenseNet-121, EfficientNet-B0, and ResNet-50, have been used on panoramic images. The ResNet-50 model has performed better than other models. In [17], a combination of CNNs and image processing technology has been used to automatically detect dental lesions. The image segmentation technique has been improved using histogram equalization and flat field correction. GoogleNet and SqueezNet models have been used to detect restored and lost teeth. In [18], a model has been presented to classify the stages of periodontitis from panoramic radiographs. This model is used to identify bone loss and the cementoenamel junction. This model is accurate in diagnosing periodontitis. In [19], a convolutional neural network-based model called PDDNet has been developed to classify three types of caries, fillings, and implants from dental radiographs. This model has shown good performance using the ADASYN oversampling method to improve classification. In [20], a CNN was used to binary classify dental caries. In this system, the quality of the input images has been improved using histogram contrast enhancement and filtering methods to overcome the problems caused by irregular illumination and low contrast, which has provided good results. In [21], a deep learning model has been developed to identify and classify dental implants from radiographic images. This model was trained using a large dataset of 156965 of 27 different types of implants. The model was evaluated on panoramic and periapical images and has shown the best performance on panoramic images with an accuracy of 88.53%. Although recent studies have demonstrated that machine learning methods can assist dentists in clinical decision-making, most of these approaches have employed single-stage classification processes. Such one-step methods may struggle to effectively distinguish between classes with similar characteristics. Implementing a multi-stage approach can help mitigate feature overlap between classes, ultimately improving system accuracy and more reliable diagnostic outcomes.

2. Materials and method

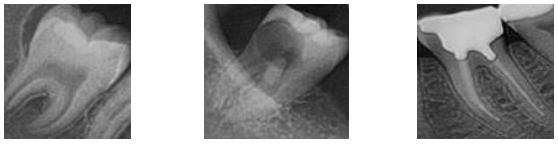
2.1. Study design

This study presents a two-step approach to improving the accuracy of multi-class classification of teeth in panoramic images. This approach can improve the system's performance by reducing interference between similar features and help dentists make more accurate diagnoses.

2.2. Dataset

The images used in this study were obtained using the Paxi Vatech panoramic radiography device. This device, manufactured by Vatech, South Korea, is designed with advanced technical specifications, including an adjustable tube voltage between 50 and 90 kV, a tube current between 4 and 10 mA, and a focal spot of 0.5 mm. The scanning time of this device was adjusted according to standard protocols; in HD mode, the scanning time was 13.5 seconds, and in normal mode, 1.10 seconds. It should be noted that this device was used to collect data for this study in 2022. To prepare the data, the panoramic images were individually cropped so that each image represented a molar or premolar tooth. These cropped images were saved in PNG format to maintain their quality and clarity. Next, dental professionals manually labeled each image. This process was performed with great precision to identify areas of suspected structural changes and assign each tooth to one of three categories: healthy, caries, or amalgam-filled. An example of these images is shown in Figure 1, which shows a sectioned and labeled tooth from the dataset.

402 images were collected and annotated in total, out of which 142 were annotated as amalgam-filled, 114 were annotated as caries, and 146 were annotated as healthy teeth. Although class distribution is reasonably balanced, the total dataset size is still small, potentially affecting the generalizability of the results to actual clinical practice. This limitation mainly owes to difficulties involved with expert-level manual annotation, data privacy issues, and availability of high-quality panoramic radiographs. To reduce the effect of small data, we used many data augmentation methods and implemented transfer learning through the usage of pre-trained models. Nevertheless, we recognize this limitation and plan to incorporate external validation by implementing larger publicly available data in future research studies.



(a) (b) (c)

Figure 1. Three different categories in the panoramic image. (a) healthy (b) Caries (c) Amalgam filling

2.3. Deep learning models

One of the main challenges in using deep learning in dentistry is the lack of high-quality and diverse data. Collecting dental radiographs with accurate labels is a time-consuming and difficult process due to clinical constraints, patient privacy, and the need for experts for labeling. In such circumstances, it is essential to select models that extract the most meaningful features from limited data and provide optimal performance.

The selection of deep learning architectures employed in this study was influenced by several considerations, including previous success in the area of medical image analysis, computational efficiency, and applicability to both low-resolution and high-resolution dental sets. Convolutional networks, specifically DenseNet121 and ResNet18, were given precedence because of their demonstrated effectiveness at identifying intricate visual patterns, whereas EfficientNet-B4 was chosen because of its compound scaling approach. Furthermore, MobileNetV3 was also employed to consider the potential for lightweight deployment, and ViT-B/32 was experimented on as a transformer-based model reference to

examine global feature extraction capabilities. Though it employed a transformer model, CNN-based methods showed better performance in this field, which may be explained by the local texture focus that is essential in dental imaging.

2.3.1. DensNet model

This model was published by Huang et al. in 2017 [22] and is considered an innovative architecture in the field of convolutional neural networks. The main feature of this architecture is the use of dense connections between layers so that the output of each layer is directly connected not only to the next layer but also to all subsequent layers. In other words, in each layer, DenseNet uses all the outputs of the previous layers as inputs. This structure allows for the reuse of features extracted at different levels and increases the richness of the internal representations of the model. These dense connections have greatly reduced problems such as vanishing gradients that usually occur in deep networks. On the other hand, due to the extensive sharing of features between layers, the need for a smaller number of filters and, consequently, the number of trainable parameters has been reduced compared to the size of conventional architectures. This model has shown excellent performance, especially in tasks that require precise, detailed, and multi-level analysis of image features. For this reason, numerous applications in sensitive and precise domains, such as image-based medical diagnostics, have been reported for this architecture. Among the different versions of DenseNet, the DenseNet-121 model is one of the most widely used and common implementations. This model consists of 121 layers and uses the dense structure introduced in the original DenseNet architecture. Figure 2. shows the general structure of this model.

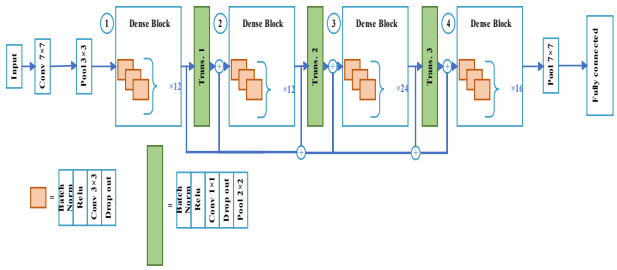


Figure 2. DenseNet-121 network architecture includes direct connections between all layers in each block.

2.3.2. ResNet model

This model was proposed by Kaiming He et al. [23]. This architecture was proposed to address the problems encountered with increasing depth of networks, namely the accuracy degradation phenomenon observed with very deep networks. The employment of residual blocks is the characteristic of ResNet. In these blocks, rather than learning the target mapping explicitly, the network learns the difference or residual between the input and output of a layer. This is accomplished by a shortcut path that directly forwards a layer's output to the subsequent layers without subjecting it to nonlinear transformations. These direct links enable information and gradients to travel easily through the network during training, which mitigates the risk of vanishing gradients and enables the training of very deep networks. ResNet architecture has enabled the training of networks with over a hundred or even a thousand layers with reasonable performance without suffering a loss of accuracy. Various forms of this model have been proposed, like ResNet18, ResNet34, ResNet50, ResNet101, etc. Surprisingly, the architecture of ResNet18 has been extensively used in practical applications, particularly image classification, object detection, and medical image analysis, because of its simplicity, quick training time, and favorable efficiency. As illustrated in Figure 3, ResNet-18 architecture is made up of an input layer, a number of residual blocks, and a fully connected output layer.

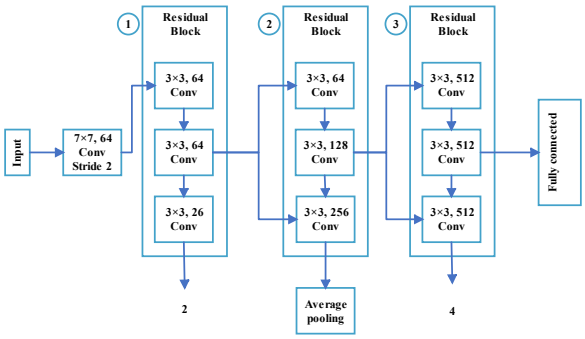


Figure 3. ResNet-18 network architecture includes residual blocks for effective deep learning

2.3.3.EfficientNet model

This architecture was originally introduced in 2019 by Quoc V. Le, Mingxing Tan et al. [24]. As a convolutional neural network, EfficientNet aims to balance efficiency and accuracy simultaneously. The characteristic of this architecture is that it applies a compound scaling method, in which the three key parameters of the network depth, width, and input image resolution are scaled in a balanced and systematic manner. Unlike traditional approaches that scale only one of these dimensions, EfficientNet harmonizes all three together and discovers a new optimal balance between model quality and computational cost. The first architecture was developed by a neural architecture search algorithm, which led to the development of the baseline model as EfficientNet-B4. Later versions of the model, named B1 to B7, were created by applying compound scaling to the base version, with each version being particularly crafted to fit different amounts of computational resources and accuracy requirements. Due to its resource-friendly architecture, EfficientNet has demonstrated improved accuracy on a variety of computer vision tasks compared to conventional architectures such as ResNet despite having fewer parameters and less memory usage. EfficientNet's features have prompted its widespread adoption across applications such as medical image classification, disease diagnosis, and other high-precision image analysis tasks. In this study, EfficientNet-B4 was used as the backbone model in panoramic dental image classification. As illustrated in Figure 4, EfficientNet-B4's structure comprises a deep path of feature extraction using MBConv modules and intermediate skip connections, terminating with a fully connected layer for classification.

This model, whose depth is higher than predecessors, can capture more complex features of image data without sacrificing computational efficiency. EfficientNet-B4 was selected due to its best trade-off between low computational cost and high accuracy, and it is found most appropriate for medical applications.

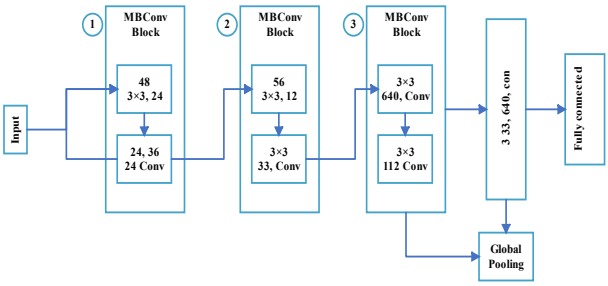


Figure 4. EfficientNet-B4 network architecture scalable model optimized in computational complexity

2.3.4. MobileNet model

The MobileNet model was first proposed by Andrew Howard et al. in 2017 [25]. It was specifically proposed to develop efficient and lightweight neural networks that can be deployed on devices with minimal computing power. The model is based on depth-wise separable convolutions inherently in its design. Instead of standard convolutions, the method separates the convolutional process into two consecutive steps: an initial step of depth-wise convolution, where one filter is applied to each individual input channel, and a subsequent step of pointwise convolution, in the shape of a 1x1 convolution that aggregates the output of the depth-wise step. This approach results in a large reduction in the number of parameters and computational complexity of the model, often with a small impact on accuracy. Subsequent versions of MobileNet, specifically MobileNetV2 and MobileNetV3, introduced substantial performance and efficiency improvements in models. MobileNetV2 employed inverted residual blocks, linear bottlenecks, and shortcut connections. MobileNetV3 architecture utilized a synergistic approach by blending automated Neural Architecture Search (NAS) with meticulous manual optimization. This version also includes new features such as Squeeze-and-Excitation (SE) modules for channel attention and employs advanced non-linear activation functions such as h-swish. These extensions enhance the model's feature extraction capacity while striving to maintain overall effectiveness. MobileNet family has some outstanding strengths, including smaller model size, low power consumption, speedy inference, and decent performance in various computer vision tasks like image classification, object detection, and semantic segmentation. These strengths especially make MobileNet architectures highly suitable for deployment on mobile phones, embedded devices, and other edge computing platforms where resources are constrained. As illustrated in Figure 5, the particular version of the MobileNet architecture under consideration, named MobileNetV3, utilizes enhanced bottleneck modules, includes SE blocks for channel attention, and h-swish as an activation function. Furthermore, its architecture has been optimized by utilizing automated architecture search methods to ensure an optimal trade-off between accuracy and computational complexity. The 'Large' version of MobileNetV3 is designed to enhance performance on challenging tasks requiring high accuracy while maintaining low resource usage principles. In the present research context, this model was selected based on its strong capabilities for rapid processing, low memory usage, and flexibility to process complex image data. Its suitability, as verified by results, demonstrated effective performance in differentiating healthy from filled teeth.

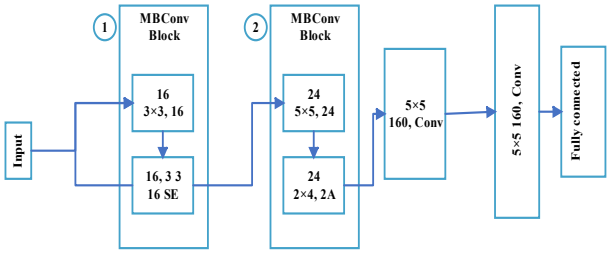


Figure 5. MobileNetV3 Large network architecture A lightweight network optimized for devices with limited processing power.

2.3.5. ViT model

Dosovitskiy et al. [26] originally introduced the Vision Transformer model. Following the significant success of the Transformer model in Natural Language Processing, ViT was designed to adapt and translate this powerful architecture to be used in computer vision.

Departing from traditional Convolutional Neural Networks, which rely to a large extent on local filter operations, ViT works differently. It accepts input images by first dividing them into a sequence of fixed-size, non-overlapping patches. They are then embedded linearly into a vector, effectively treated as a 'token' akin to words in NLP, and passed through the Transformer encoder. Specifically, in this architecture, an input image is split into a grid of small patches, usually 16x16 or 32x32 pixels. The image patches are akin to words in a sentence. Following a linear projection step, they generate a sequence of input vectors to be passed into the Transformer. The patch embedding sequence is then supplemented by a special learnable class token and positional encodings to preserve spatial context before inputting the regular Transformer encoder layers. In this architecture, the self-attention mechanism serves as the fundamental operation that allows the model to learn long-range dependencies and interdependencies between various image patches. The key strengths of ViT include the ability to overcome the inherent locality bias of CNNs, the ability to discover global and non-local relations of the entire image, and the ability to be extremely versatile when dealing with varying image data. However, one of the first issues with ViT was its enormous hunger for data, which required extremely large datasets to train from scratch effectively. This

limitation has been largely mitigated by the strategy of pre-training on large datasets followed by fine-tuning on specific downstream tasks. ViT comes in several standard variants, such as ViT-Base, ViT-Large, and ViT-Huge, which vary in model size, the number of layers, embedding dimensions, and computational complexity.

As shown in Figure 6, the ViT-B/32 model begins processing the input image by dividing it into 32x32 pixel patches. Every patch is then transformed into a vector representation by applying a linear embedding layer. Following the addition of positional encodings, this sequence of vectors is input into the multi-layer Transformer encoder. Lastly, the output vector corresponding to the token is usually taken as the aggregate image representation for the final classification head. The ViT-B/32 model's Base architecture, with patches of size 32x32, contains 12 Transformer layers and an embedding dimension of 768. It displays impressive effectiveness at picking up global image patterns, even with its relatively low patch resolution. Its application in the current study was motivated by its capacity to learn non-local relationships between various regions of an image and its potential to eliminate the necessity for complex, hand-crafted convolutional filters. As described in the results section, this process was extremely accurate.

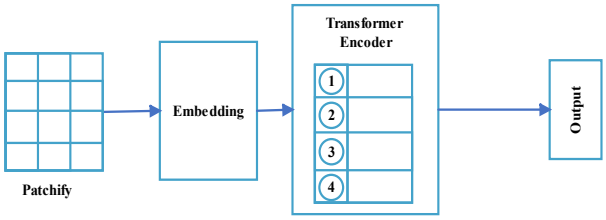


Figure 6. ViT32 network architecture dividing the image into equal patches using hierarchical transformers.

3. Methodology

This section proposes a recommended methodology based on a hierarchical framework for dental panoramic image classification. As can be seen from Figure 7, this framework is composed of two different stages. These two stages are devised specifically to achieve the twin objectives of minimizing feature interference among close classes and maximizing diagnostic accuracy.

This approach significantly minimizes the problem of overlap among closely related categories, particularly between normal teeth and those filled with amalgam. Also, the adoption of this phasic separation mechanism results in greater retrieval of meaningful features, consequently improving the system's overall performance.

3.1. Data preprocessing and preparation

Dental images undergo an initial preprocessing phase with multiple methods, including resizing to 224×224-pixel size and normalizing each color channel to a mean of 0.5 and standard deviation of 0.5. Furthermore, multiple random transformations are also carried out, including rotating up to 45 degrees and flipping horizontally with a 50% chance [27] . Not only do these procedures enable the removal of noise and quality improvement of input images, but they also assist in increasing the variety of the training dataset.

These augmentations are applied dynamically during the training phase, meaning that each image may be presented differently at every training iteration. This strategy enhances the diversity of training samples and effectively reduces the risk of overfitting. Importantly, no augmentations are applied to the validation or test sets, ensuring that evaluation is conducted on unaltered images and that performance metrics reflect the model's generalization capability.

After preprocessing, the data is split randomly into distinct subsets. 80% of the data is employed for training, 20% is reserved for validation, and the remaining 80% is for training the model. The remaining 20% of the total data is reserved as the test set for the final model assessment

3.2. Hieratical classification architecture

This study presents a two-level hierarchical framework for the classification of dental panoramic images. The proposed system operates in sequence and independently, where the first differentiates between images with and without dental caries. Images without caries in the subsequent step are further divided into two specialized categories: images with amalgam fillings and healthy teeth.

3.2.1. First stage: Caries detection

In this phase, a binary classifier model is trained to predict the presence or absence of dental caries in the images. For this purpose, the classes comprising amalgam fillings and sound teeth are merged into a single class described as caries-free. Thus, the goal of the model's objective in this phase is to categorize images into two classes: caries and non-caries.

In implementing the model, a number of pre-trained architectures are used, including EfficientNet-B4, ResNet18, ViT-B/32, MobileNetV3-Large, and DenseNet121. The architectures are fine-tuned with the training data for this particular phase, where their output layers are modified to depict the two defined classes. After training, the model with the best performance is selected based on the best accuracy obtained from the test set.

3.2.2.Second stage: Amalgam and healthy detection

In the second stage, images that were previously marked as caries-free in the first stage are processed again to differentiate between amalgam-filled and healthy teeth. Another binary classification model is trained exclusively to categorize these images into two groups: those with amalgam restorations and healthy teeth.

Using an approach consistent with that followed in phase one, the same list of pre-trained models, EfficientNet-B4, ResNet18, ViT-B/32, MobileNetV3-Large, and DenseNet121, is utilized. Training is performed on the subset of the dataset pertaining to the two classes of sound and filled teeth, and the optimal model is determined by its performance on the validation set.

3.2.3.Training process

The models are trained with the Adam optimization algorithm, which is set to have an initial learning rate of

0.001. This particular algorithm is chosen because of its adaptive learning rate adjustment capability and quick

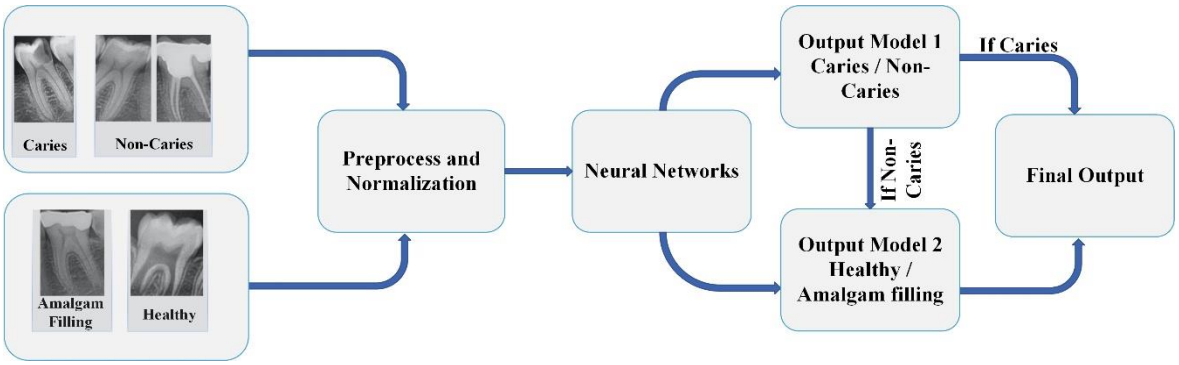


Figure 7. Schematic diagram of the proposed method for hierarchical method of dental images.

convergence in deep learning-related tasks. The binary cross-entropy loss function is used, given its appropriateness for binary classification issues and its capability to reduce the divergence between the estimated probabilities and the actual labels.

Every model is trained for 30 epochs. During each epoch, the model is fine-tuned on the training dataset, and its performance is evaluated on a validation set at the end of the epoch. The model with the best performance, as evaluated by the validation results, is determined and saved. Figure 7 presents a schematic overview of the training pipeline.

3.2.4. Base method

A multiclass classification method is addressed to provide a contrast to the hierarchical process. In a multiclass approach, a single model learns to classify the sample directly into one of three groups: healthy, caries, or amalgam-filled. The same pre-trained models used in the hierarchical approach are used in this case, and the output layers are adjusted to suit the three target classes.

Data preprocessing and training are implemented in the same way as the hierarchical approach, with the only exception being that no stepwise separation is performed in the multiclass approach.

4. Results

Classical classification metrics, such as overall accuracy, class precision, recall, and the F1 score, were used to compare the performance of the two methods. A confusion matrix has also been included to give an additional perspective. True positive (TP) within these evaluations refers to the number of positive examples the model has correctly picked up. True negative (TN) represents the number of negative examples identified correctly. False positive (FP) represents negative samples wrongly predicted as positive, while false negative (FN) represents positive samples incorrectly predicted as negative. Precision measures the proportion of correctly predicted positive instances to the total instances predicted as positive. Accuracy is the proportion of samples correctly classified out of all predictions made by the model. Recall is the proportion of correctly predicted positive samples in relation to all actual positive instances. The F1 score has been utilized as an integrative measure that balances precision and recall. To examine model performance more closely, a confusion matrix has also been used. This matrix breaks down correctly and incorrectly classified samples and gives a closer look at the model's behavior.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

In the multiclass strategy, one model was developed to classify dental images into one of three distinct classes: healthy, caries, or amalgam filling. The performance of different models on the test dataset is reported in Table 1. The findings presented in the table indicate that the DenseNet121 model performed best overall, with an accuracy of 91.35%, surpassing all the other models considered in the study. Through its densely connected architecture and feature reuse across the various layers, this model had a greater capacity to detect and distinguish the intricate patterns within the images and, hence, better classification performance.

Table 1. Comparison of performance metrics for different deep learning models on the multi label approach

Model	Accuracy	Precision	Recall	F1 Score
DENSENET121	**0.914**	**0.915**	**0.913**	**0.9137**
RESNET18	0.901	0.904	0.901	0.899
EFFICIENTNET-B4	0.864	0.870	0.864	0.863
MOBILENETV3-LARGE	0.815	0.884	0.814	0.775
VIT_B_32	0.753	0.766	0.753	0.699

The ResNet18 model showed performance highly similar to that of DenseNet121, with a general accuracy of 90.12%. By employing residual blocks and shortcut connections in

the feature propagation, the model maintained key structural information in the images and provided acceptable outputs. Then, EfficientNet-B4 placed third with an accuracy of 86.41%. The improved architecture achieved a good trade-off between computational efficiency and accuracy. The MobileNetV3-Large model was less effective, achieving an accuracy of 81.48%. This is due to its optimized design and emphasis on minimizing computational expense. Nevertheless, it could still be a viable choice for time-critical applications or edge computing environments. Lastly, the ViT-B/32 model performed the worst in this approach, with an accuracy of 75.30%. This is due to the Vision Transformers' uncommon architecture and their reliance on big datasets to optimally learn. Moreover, the utilization of bigger image patches may have caused the omission of significant fine-grained details in small dental structures essential for correct classification. In general, the outcomes show that DenseNet and ResNet, which are architectures derived from densely connected convolutional networks, are superior to the other models in handling the complicated and similar features of panoramic dental images. This shows the necessity of having deep and recursive structures to adequately extract necessary features in multiclass classification problems in dental imaging. Within the hierarchical framework, panoramic dental radiograph classification occurred in several successive steps. At the initial step, emphasis was placed on differentiating between non-caries and caries teeth. For this purpose, a binary classifier model was created. Both healthy teeth and teeth with amalgam restorations were categorized into one class to enable the model to specialize in detecting regions that may possess signs of caries. The performance of different models for this phase is shown in Table 2. It was at this point that the DenseNet121 model achieved the highest overall accuracy at 83.95%. The model's densely connected structure, along with its capacity to recycle features from earlier layers that had been extracted, enabled the model to recognize caries-associated patterns within the images more effectively. Taking the runners-up position was the EfficientNet-B4 model with an accuracy of 80.24%, indicating its capacity to blend visual recognition capabilities with computational efficiency. The ViT-B/32 and ResNet18 models also had very acceptable performance with accuracies of 67.90% and 76.54%, respectively, but were ranked below the top models. Surprisingly, the MobileNetV3-Large model performed the worst, with an accuracy of 65.43%. This lower performance is most likely due to its lightweight architecture, which restricts its ability for extensive deep feature extraction. In the next stage of the hierarchical structure, the images that were labeled as "non-caries" in the first step were further processed to separate the healthy teeth from the teeth with amalgam restorations. The results of this binary classification are shown in Table 3. At this point, the EfficientNet-B4 model excelled the most, with an accuracy of 98.28%. By using a uniform compound scaling technique on different network sizes, the model was able to successfully extract and classify amalgam filling-related features. The ResNet18 and ViT-B/32 models also did well, with 96.55% and 94.83% accuracy, respectively. DenseNet121 achieved a moderate accuracy of 93.10%, but the performance was marginally less compared to the first stage. On the contrary, MobileNetV3-Large achieved the lowest accuracy of 86.21%, which may likely be because it cannot identify the fine-grained detail characteristic of amalgam restorations. The outcomes achieved in both phases of the hierarchical method indicate that classification task specification, along with the utilization of suitably chosen models in every phase, improves the system's overall accuracy. EfficientNet-B4's phenomenal performance in the second phase, coupled with the astounding outcome achieved by DenseNet121 in the first phase, simply underscores the significance of model selection in streamlining the efficiency of hierarchical classification.

Table 2. Results of the first stage of caries diagnosis.

model	Accuracy	Precision	Recall	F1 Score
DENSENET121	**0.840**	**0.851**	**0.901**	**0.876**
EFFICIENTNET-B4	0.802	0.786	0.901	0.875
RESNET18	0.765	0.863	0.745	0.800
VIT_B_32	0.679	0.727	0.784	0.754
MOBILENETV3-LARGE	0.654	0.645	1.0	0.784

Table 3. Results of the second stage of distinguishing healthy from amalgam fillings.

model	Accuracy	Precision	Recall	F1 Score
EFFICIENTNET-B4	**0.983**	**0.971**	**1.0**	**0.985**
RESNET18	0.966	1.0	0.941	0.967
VIT_B_32	0.948	0.969	0.941	0.955
DENSENET121	0.931	0.894	1.0	0.944
MOBILENETV3-LARGE	0.862	0.988	0.764	0.866

4.1. Confusion matrix

The confusion matrix has been identified as the basic tool for assessing the performance of classification models in that it enables close examination of the predictions on a per-class basis. In the current research, since there were three classes, caries, amalgam-filled, and healthy teeth, and two methods (multiclass and hierarchical) implemented independently, the examination of the confusion matrix is of paramount importance for discussing performance differences between these methods. This type of analysis is particularly helpful in ascertaining the extent of misclassification between visually close classes.

Subsequently, the confusion matrices of the top models obtained from each method are examined, and the respective results are shown.

4.1.1. Multi classes approach

Figure 8 illustrates the confusion matrix related to this approach. It illustrates the strong ability of the model to correctly identify the amalgam-filled class; however, challenges are evident in distinguishing between the healthy and caries classes. Specifically, many caries samples were misclassified as healthy, a scenario that may be attributed to the visual similarities evident in the dataset's images. Regardless of these limitations, the model's overall performance is still considered acceptable.

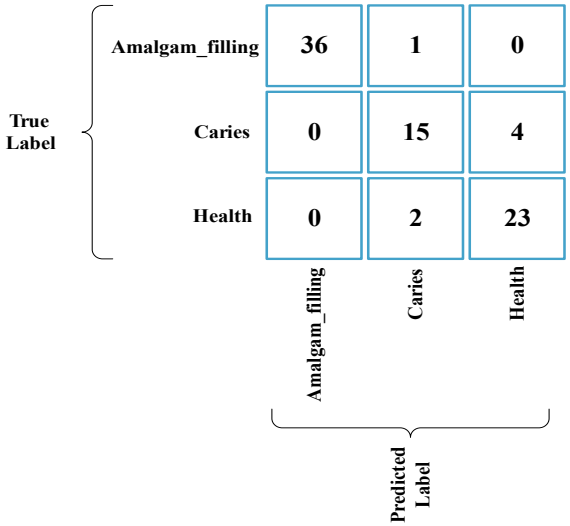


Figure 8. Confusion Matrix for based method multi label classification.

4.1.2. First stage hierarchical approach

Figure 9 illustrates the confusion matrix for this method. The model correctly labeled the majority of the non-caries samples, that is, the healthy and the amalgam-filled classes. There was a large overlap between the two caries classes, with mislabeling of the same into different classes. This might have resulted from the visually similar appearance of some of the caries regions with filled teeth.

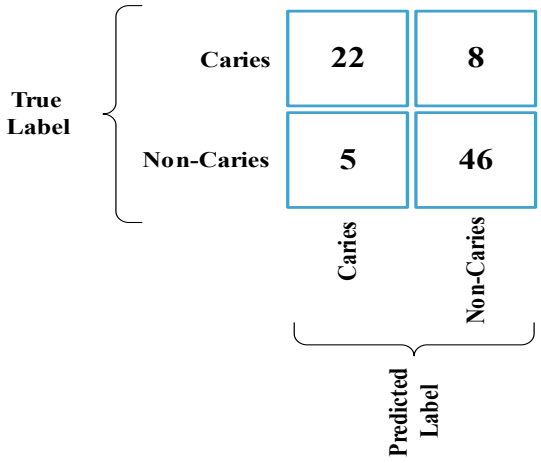


Figure 9. Confusion matrix for hierarchical approach to distinguish Caries from non-Caries.

4.1.3. Second stage hierarchical approach

The confusion matrix of this method is presented in Figure 10. The findings from the confusion matrix reveal that the model has a very positive performance in discriminating between the amalgam-filled and healthy classes. The very high proportion of correct classification for both classes, along with a negligible level of overlap, implies a stable and credible performance of this model in this particular classification problem.

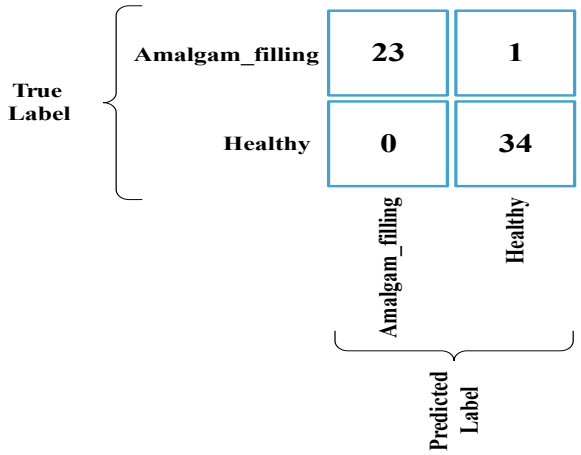


Figure 10. Confusion matrix for hierarchical approach to healthy diagnosis of amalgam fillings.

5. Conclusion

In this study, a hierarchical classification model was proposed for panoramic dental images to reduce misclassification among classes with similar visual characteristics. By utilizing deep learning models and pre-trained architectures, the framework consistently identified complex structures and achieved successful separation among dental classes. The two-step methodology applied in this framework enabled an initial determination of overall trends pertaining to dental caries, succeeded by a more nuanced analysis aimed at spotting the existence of amalgam restorations. The methodological difference not only improved the feature extraction process but also offered an improved training trajectory for the models. Since there is a great visual similarity between numerous dental characteristics, this hierarchical approach helped to minimize classification faults and enhance the reliability of the findings. Moreover, a varied set of deep learning models used across varying stages enabled the customized optimization of each model independently, thereby permitting greater flexibility for processing varied and complex datasets. The modular structure also offers room for reuse in other image-based medical diagnostic tasks and can potentially act as a template for further related research efforts. Not just in the structural model but even in the strategic design of the system, the notion of hierarchy was implemented, resulting in a consolidated and scalable system for dental image classification. The system can be expanded further in the future by adding more stages or more specialized classification systems. The conclusions of this research have the potential to contribute significantly to the design of assistive systems for clinical decision-making. The deployment of these systems in clinical environments has the potential to enhance diagnostic precision and speed up evaluation procedures, thereby optimizing treatment processes. With the proven performance of deep learning models, it is expected that their use in every field of medicine, including dentistry, will continue to grow, enabling the development of sophisticated diagnostic systems. Finally, this study is a feasible, scalable, and comprehensive method, which is a major breakthrough in using artificial intelligence to facilitate accurate assessment of panoramic dental images.

This work can potentially induce a paradigm shift in oral disease diagnosis and improve the quality of dental healthcare services.

6. References

[1] Luo, S.-C., Wei, S.-M., Luo, X.-T., Yang, Q.-Q., Wong, K.-H., Cheung, P. C., & Zhang, B.-B. (2024). How probiotics, prebiotics, synbiotics, and postbiotics prevent dental caries: an oral microbiota perspective. *npj Biofilms and Microbiomes*, *10*(1), 14.

[2] Radha, R., Raghavendra, B., Subhash, B., Rajan, J., & Narasimhadhan, A. (2023). Machine learning techniques for periodontitis and dental caries detection: A narrative review. *International journal of medical informatics*, *178*, 105170.

[3] Akhter, M. N., Hussain, S. S., Riaz, N., & Zulfiqar, R. (2023). Using Technological Diagnostic Tools to Find Early Caries: A Systematic Review. *Dinkum Journal of Medical Innovations*, *2*(07), 271-283.

[4] Chauhan, R. B., Shah, T. V., Shah, D. H., Gohil, T. J., Oza, A. D., Jajal, B., & Saxena, K. K. (2023). An overview of image processing for dental diagnosis. *Innovation and Emerging Technologies*, *10*, 2330001.

[5] Walsh, T., Macey, R., Ricketts, D., Carrasco Labra, A., Worthington, H., Sutton, A. J., Freeman, S., Glenny, A. M., Riley, P., Clarkson, J., & Cerullo, E. (2022). Enamel Caries Detection and Diagnosis: An Analysis of Systematic Reviews. *Journal of Dental Research*, *101*(3), 261-269. https://doi.org/10.1177/00220345211042795

[6] Zhu, J., Chen, Z., Zhao, J., Yu, Y., Li, X., Shi, K., Zhang, F., Yu, F., Shi, K., & Sun, Z. (2023). Artificial intelligence in the diagnosis of dental diseases on panoramic radiographs: a preliminary study. *BMC Oral Health*, *23*(1), 358.

[7] Kühnisch, J., Meyer, O., Hesenius, M., Hickel, R., & Gruhn, V. (2021). Caries Detection on Intraoral Images Using Artificial Intelligence. *Journal of Dental Research*, *101*, 158 - 165.

[8] Fawaz, P., El Sayegh, P., & Vannet, B. V. (2023). What is the current state of artificial intelligence applications in dentistry and orthodontics? *Journal of Stomatology, Oral and Maxillofacial Surgery*, *124*(5), 101524.

[9] Yoon, K., Jeong, H.-M., Kim, J.-W., Park, J.-H., & Choi, J. (2024). AI-based dental caries and tooth number detection in intraoral photos: Model development and performance evaluation. *Journal of Dentistry*, *141*, 104821.

[10] Qayyum, A., Tahir, A., Butt, M. A., Luke, A., Abbas, H. T., Qadir, J., Arshad, K., Assaleh, K., Imran, M. A., & Abbasi, Q. H. (2023). Dental caries detection using a semi-supervised learning approach. *Scientific Reports*, *13*(1), 749.

[11] Chen, I. D. S., Yang, C.-M., Chen, M.-J., Chen, M.-C., Weng, R.-M., & Yeh, C.-H. (2023). Deep learning-based recognition of periodontitis and dental caries in dental x-ray images. *Bioengineering*, *10*(8), 911.

[12] Batra, A. M., & Reche, A. (2023). A new era of dental care: harnessing artificial intelligence for better diagnosis and treatment. *Cureus*, *15*(11).

[13] Anil, S., Porwal, P., & Porwal, A. (2023). Transforming dental caries diagnosis through artificial intelligence-based techniques. *Cureus*, *15*(7).

[14] Huang, C., Wang, J., Wang, S., & Zhang, Y. (2023). A review of deep learning in dentistry. *Neurocomputing*, *554*, 126629.

[15] Haghanifar, A., Majdabadi, M. M., Haghanifar, S., Choi, Y., & Ko, S.-B. (2023). PaXNet: Tooth segmentation and dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier. *Multimedia Tools and Applications*, *82*(18), 27659-27679.

[16] Oztekin, F., Katar, O., Sadak, F., Yildirim, M., Cakar, H., Aydogan, M., Ozpolat, Z., Talo Yildirim, T., Yildirim, O., & Faust, O. (2023). An explainable deep learning model to prediction dental caries using panoramic radiograph images. *Diagnostics*, *13*(2), 226.

[17] Chen, S.-L., Chen, T.-Y., Huang, Y.-C., Chen, C.-A., Chou, H.-S., Huang, Y.-Y., Lin, W.-C., Li, T.-C., Yuan, J.-J., & Abu, P. A. R. (2022). Missing teeth and restoration detection using dental panoramic radiography based on transfer learning with CNNs. *IEEE Access*, *10*, 118654-118664.

[18] Shon, H. S., Kong, V., Park, J. S., Jang, W., Cha, E. J., Kim, S.-Y., Lee, E.-Y., Kang, T.-G., & Kim, K. A. (2022). Deep Learning Model for Classifying Periodontitis Stages on Dental Panoramic Radiography. *Applied Sciences*, *12*(17), 8500. https://www.mdpi.com/2076-3417/12/17/8500

[19] Hasnain, M. A., Ali, Z., Saeed, A., Aijaz, S., & Khurram, M. S. (2024). PDDNet: Deep Learning Based Dental Disease Classification through Panoramic Radiograph Images. *VFAST Transactions on Software Engineering*, *12*(4), 180-198.

[20] Son, J.-Y., Park, Y., Park, J.-Y., Kim, M.-J., & Han, D.-H. (2024). Overdiagnosis of dental caries in South Korea: a pseudo-patient study. *BMC Oral Health*, *24*(1), 1-10.

[21] Park, W., Huh, J.-K., & Lee, J.-H. (2023). Automated deep learning for classification of dental implant radiographs using a large multi-center dataset. *Scientific Reports*, *13*(1), 4862.

[22] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

[23] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

[24] Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning,

[25] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[26] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., & Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[27] Fariza, A., Asmara, R., Rojaby, M. O. F., Astuti, E. R., & Putra, R. H. (2022). Evaluation of Convolutional Neural Network for Automatic Caries Detection in

Digital Radiograph Panoramic on Small Dataset. 2022 International Conference on Data and Software Engineering (ICoDSE),