Spectrum of Engineering Sciences
ISSN (e) 3007-3138 **(p)** 3007-312X Volume 4, Issue 3, 2026

DENTAL DISEASE DETECTION USING ADVANCED DEEP LEARNING TECHNIQUES

[*1]Muhammad Abdullah Arslan Khan
*[*1]Department of Computer Science, NFC Institute of Engineering and Technology, Multan*
[*1]MuhammadAbdullahArslankhan11@gmail.com

DOI: https://doi.org/10.5281/zenodo.21916178

Keywords:
Dental disease detection; dental X-ray; deep learning; convolutional neural network; DenseNet121; EfficientNet-B5; DenseNet121; transfer learning; medical image classification; computer-aided diagnosis.

Article History
Received on: 22 Feb, 2026
Accepted on: 27 March, 2026
Published on 29 March, 2026

Corresponding Author: Muhammad Abdullah Arslan Khan

Abstract

Dental diseases are prevalent clinical conditions that can adversely affect oral function, quality of life, and overall health. Dental radiography is an essential component of dental diagnosis; however, the interpretation of radiographic images can be time-consuming and may be influenced by image quality, disease presentation, anatomical complexity, and clinician experience. Recent advances in deep learning (DL), particularly convolutional neural networks (CNNs) and transfer-learning architectures, have demonstrated considerable potential for automated analysis of medical and dental images. This study proposes a comparative deep learning framework for automated classification of diseased and normal dental X-ray images using DenseNet121, EfficientNet-B5, and U-Net. The proposed framework incorporates image preprocessing, resizing, pixel normalization, data augmentation, model training, and systematic performance evaluation. The dataset comprises 525 dental X-ray images, including 294 diseased and 231 normal images, with 404 images allocated for training and 121 images for testing. Model performance is assessed using accuracy, precision, recall, F1-score, training and testing loss, and confusion-matrix analysis. Experimental results demonstrate that DenseNet121 achieved a test accuracy of 98.35%, whereas EfficientNet-B5 achieved 99.15%. U-Net is incorporated as the third architecture to investigate the applicability of an encoder-decoder network to dental X-ray analysis; its quantitative performance should be established through independent experimental training and evaluation. The comparative framework provides a systematic basis for examining the effectiveness of different deep learning architectures in automated dental image classification. Nevertheless, the high performance obtained on a relatively small dataset should be interpreted cautiously, as model generalizability may be affected by dataset composition, patient-level data dependence, imaging variability, and potential overfitting. Therefore, patient-level data partitioning, independent external validation, calibrated evaluation, clinically established reference standards, and model explainability are recommended before clinical deployment. The proposed system is consequently positioned as a computer-aided decision-support research prototype rather than a replacement for professional dental diagnosis.

1 Introduction

Dental diseases represent a significant public health concern and may adversely affect oral function, nutritional intake, quality of life, and general well-being. Early and accurate identification of dental abnormalities is essential for appropriate treatment planning and prevention of disease progression. Dental radiography is one of the most commonly used imaging techniques in clinical dentistry because X-ray images provide valuable information about anatomical structures and pathological changes that may not be readily visible during routine clinical examination. However, accurate interpretation of dental radiographs requires substantial clinical expertise and may become challenging when images contain noise, overlapping anatomical structures, variations in exposure, or subtle pathological features. Conventional radiographic interpretation primarily depends on the knowledge and experience of dental professionals. Although experienced clinicians can identify many abnormalities with a high degree of reliability, the diagnostic process can be influenced by several subjective and technical factors. Variations in image quality, contrast, disease severity, anatomical complexity, and observer experience may affect diagnostic consistency. Furthermore, small lesions or early-stage pathological changes can be difficult to distinguish from normal anatomical structures. Prolonged examination of large numbers of radiographs may also contribute to observer fatigue, potentially affecting diagnostic accuracy. These limitations have encouraged researchers to investigate computer-aided diagnostic systems capable of supporting clinicians through rapid, consistent, and reproducible analysis of dental images.

The rapid development of artificial intelligence (AI), particularly deep learning, has introduced new opportunities for automated medical image analysis. Unlike conventional machine-learning approaches that generally require manually engineered image features, deep neural networks can automatically learn hierarchical representations directly from image data. Convolutional neural networks (CNNs) have demonstrated strong capabilities in extracting spatial and semantic features from medical images and have consequently been applied to several diagnostic imaging tasks. Transfer learning further enhances the applicability of deep learning when the available medical dataset is relatively small by allowing a model pretrained on a large image dataset to be adapted to a specific classification problem. In dental image analysis, deep learning techniques have increasingly been investigated for tasks such as caries detection, periodontal disease identification, tooth segmentation, lesion classification, and identification of other oral abnormalities. However, model performance can vary considerably depending on the network architecture, dataset characteristics, preprocessing strategy, and experimental methodology. Therefore, selecting an appropriate architecture is an important consideration when developing an automated dental disease detection system. In particular, architectures that promote effective feature reuse and efficient representation learning may be advantageous for extracting clinically relevant patterns from radiographic images.

This study addresses the problem of automated binary classification of dental X-ray images into diseased and normal categories using advanced deep learning architectures. The experimental dataset consists of 525 dental X-ray images, including 294 diseased and 231 normal images. The available dataset is divided into 404 training images and 121 testing images. To establish a consistent experimental framework, the study incorporates image preprocessing, resizing, normalization, augmentation, model training, and quantitative evaluation. These preprocessing operations are intended to improve image consistency and provide the deep learning models with suitable input representations for learning discriminative radiographic features. A comparative experimental approach is adopted using three deep learning architectures: DenseNet121, EfficientNet-B5, and a third DenseNet-based configuration. The comparison is intended to investigate how architectural characteristics influence classification performance on the selected dental X-ray dataset. DenseNet121 employs dense connectivity between successive layers, enabling feature reuse and improved information propagation throughout the network. EfficientNet-B5 uses compound scaling to balance network depth, width, and input resolution, providing an alternative approach to achieving high representational capacity while maintaining computational efficiency. The third architecture is included within the comparative framework, with its final performance requiring independent experimental validation rather than relying on an illustrative or assumed result.

The models are evaluated using multiple performance measures, including accuracy, precision, recall, F1-score, training and testing loss, and confusion-matrix analysis. Although accuracy provides an overall indication of classification performance, relying exclusively on accuracy may be inappropriate, particularly when class distributions are unequal or when the clinical cost of false-positive and false-negative predictions differs. Precision and recall provide additional information regarding the correctness and completeness of disease detection, while the F1-score provides a balanced measure of precision and recall. Confusion matrices are also examined to determine the distribution of correctly and incorrectly classified diseased and normal cases. The experimental results available for this study indicate promising classification performance. DenseNet121 achieved a test accuracy of 98.35%, while EfficientNet-B5 achieved 99.15% on the supplied test set. These results suggest that transfer-learning-based deep learning architectures can effectively learn discriminative patterns from dental X-ray images. However, high test accuracy obtained from a relatively small dataset should be interpreted cautiously. Performance measured on a single dataset may not necessarily generalize to images acquired from different dental clinics, imaging devices, patient populations, or acquisition protocols. In addition, patient-level data partitioning is essential to prevent information leakage between training and testing subsets, while independent external validation is required to establish the robustness and clinical generalizability of the proposed approach.

The primary objective of this research is therefore not only to achieve high classification accuracy but also to establish a structured and reproducible methodology for evaluating deep learning models for dental radiographic analysis. The study investigates suitable preprocessing and augmentation techniques, compares modern CNN architectures under a common experimental protocol, and evaluates their performance using clinically relevant statistical measures. The research also considers methodological issues such as dataset size, class distribution, external validation, model explainability, and the reliability of reference labels. The major contributions of this study are the development of an end-to-end deep learning framework for binary dental X-ray classification, comparative evaluation of multiple modern CNN architectures, and assessment of model performance using complementary classification metrics rather than accuracy alone. Furthermore, the study provides a reproducible experimental structure that can be extended to larger datasets, multiple dental disease categories, and independent clinical datasets. Consideration is also given to explainability and validation requirements that are necessary for translating research-oriented deep learning systems into practical clinical decision-support tools. Overall, this research demonstrates the potential of advanced deep learning architectures for automated dental disease detection while recognizing the limitations associated with small datasets and single-source evaluation. The proposed framework is intended to support computer-aided dental image analysis and assist clinicians rather than replace professional judgment. Future clinical implementation would require larger and more diverse datasets, patient-level data partitioning, external multi-center validation, standardized diagnostic reference standards, model calibration, and explainable predictions. Such improvements are essential for establishing the reliability, generalizability, and clinical utility of deep learning-based dental disease detection systems. The study is guided by the following objectives:

- To develop an effective preprocessing and augmentation pipeline for dental X-ray image classification.
- To investigate transfer-learning-based deep learning architectures for automated dental disease detection.
- To comparatively evaluate DenseNet121, EfficientNet-B5, and the U-Net selected architecture under a common experimental framework.
- To evaluate the models using accuracy, precision, recall, F1-score, loss, and confusion-matrix analysis.
- To identify the best-performing architecture for the selected dental X-ray dataset and analyze its strengths and limitations.
- To examine the methodological and clinical considerations associated with deploying deep learning-based dental image analysis in real-world settings.

2 Literature Review

Deep learning has emerged as a prominent computational methodology for automated dental image analysis because of its ability to learn hierarchical and discriminative representations directly from radiographic data. Conventional image-analysis approaches generally depend on manually engineered features, whereas CNN-based architectures learn patterns such as edges, textures, intensity variations, contours, and higher-level anatomical representations through successive convolutional layers. This capability is particularly relevant to dental radiographs, where pathological characteristics may vary substantially in size, shape, contrast, and anatomical location. Recent systematic evidence involving 260 studies has demonstrated the extensive application of deep learning across dental image classification, detection, segmentation, and related computer-aided diagnostic tasks. However, the same body of literature identifies persistent challenges associated with dataset heterogeneity, image quality, annotation procedures, model selection, and evaluation methodology.

Dental caries detection represents one of the most extensively investigated applications of deep learning in dental radiography. [2] systematically reviewed 42 studies involving deep learning-based caries detection reported substantial variation in diagnostic performance across imaging modalities. Classification accuracy ranged from 82% to 99.2% for periapical radiographs and from 86.1% to 96.1% for panoramic radiographs. Despite these promising results, the review identified considerable methodological heterogeneity and reported that only a minority of studies demonstrated low risk of bias across all assessed domains. These findings indicate that high predictive performance should be interpreted in conjunction with dataset characteristics, reference standards, validation strategies, and study quality rather than accuracy alone. The effectiveness of CNN-based approaches for broader dental disease classification has also been demonstrated using digital dental X-ray images. One comparative investigation evaluated VGG16, InceptionV3, ResNet50, and DenseNet121 using 188 digital dental X-ray images representing normal cases and several common dental conditions. The four models achieved an average test accuracy of 95.9%, while DenseNet121 achieved 99.5% accuracy. The study additionally reported strong sensitivity and specificity for DenseNet121 across the evaluated disease categories. These findings suggest that dense feature connectivity can be advantageous for extracting discriminative information from dental radiographs; however, the relatively small dataset and study-specific experimental conditions limit direct generalization to other populations and imaging environments. DenseNet121 is particularly relevant to dental image classification because its dense connectivity mechanism allows each layer to receive feature representations from preceding layers within a dense block. This promotes feature reuse, improves gradient propagation, and enables the network to preserve information at different levels of abstraction. Such characteristics may be beneficial for dental radiographs because subtle pathological features can coexist with larger anatomical structures. The previously reported 99.5% accuracy obtained by DenseNet121 provides supporting evidence for its suitability in dental X-ray classification, although the result should be considered dataset-specific rather than evidence of universal superiority.

EfficientNet provides an alternative architectural strategy based on compound scaling of network depth, width, and input resolution. Compared with conventional approaches that increase one architectural dimension independently, EfficientNet attempts to maintain a balanced relationship between representational capacity and computational requirements. This characteristic makes EfficientNet-family models attractive for high-resolution medical and dental image analysis, where fine-grained visual information may be important for distinguishing normal and pathological structures. Recent dental radiograph research has continued to investigate EfficientNet-family models alongside architectures such as Inception and DenseNet, with strong performance reported on panoramic-radiograph classification tasks. For example, a recent study reported 97.04% accuracy for EfficientNetV2 and 96.70% for DenseNet121, while InceptionV3 achieved the highest accuracy of 97.51% within that particular experimental setting. The literature also demonstrates that the optimal deep learning architecture is strongly dependent on the characteristics of the target task and dataset. A recent study using 1,512 panoramic radiographs and 11,137 expert-verified annotations evaluated custom CNNs, hybrid CNN-classifier approaches, and transfer-learning architectures. The hybrid CNN-Random Forest approach achieved 85.4% accuracy, whereas the best pretrained architecture, VGG16, achieved 82.3%. These results are substantially lower than the very high accuracies reported by some smaller dental datasets, illustrating how dataset complexity, number of disease categories, annotation strategy, and evaluation methodology can strongly influence model performance. Segmentation-oriented architectures such as U-Net have also become important in dental image analysis. Unlike DenseNet121 and EfficientNet-B5, which are primarily classification-oriented architectures, U-Net was originally designed for biomedical image segmentation. Its encoder-decoder structure and skip connections enable preservation of spatial information and facilitate pixel-level localization of anatomical structures or pathological regions. Consequently, U-Net is more appropriately evaluated through segmentation measures such as Dice coefficient, Intersection over Union (IoU), sensitivity, and specificity when the objective is lesion or tooth segmentation. If U-Net is included in the present study as a classification model, its architecture should be explicitly modified and described as a U-Net-based classifier to ensure methodological comparability with DenseNet121 and EfficientNet-B5.

The literature further indicates that model performance cannot be interpreted solely through classification accuracy. Precision, recall, F1-score, sensitivity, specificity, confusion matrices, and, where appropriate, ROC-AUC and calibration measures provide a more comprehensive evaluation of diagnostic behavior. This is particularly important in medical applications because false-negative and false-positive predictions may have different clinical implications. Systematic evidence on deep learning-based caries detection has shown that studies use highly diverse evaluation metrics and experimental procedures, which makes direct comparison difficult. Another important limitation is the relatively small size of many dental image datasets. Small datasets can increase the risk of overfitting and produce optimistic estimates of model performance. Furthermore, if multiple images from the same patient are distributed across training and testing subsets, patient-level information leakage can occur. Such leakage may cause the model to learn patient-specific characteristics rather than disease-specific radiographic features. Therefore, patient-level splitting should be applied whenever multiple images per patient are available. External validation using independent datasets from different institutions, imaging devices, and patient populations is also necessary to determine whether the learned representations generalize beyond the original dataset. The reviewed literature therefore establishes a strong foundation for the present study while also identifying an important research gap. Existing studies have evaluated individual architectures or groups of models under different datasets and experimental protocols, but direct comparisons remain difficult because of differences in preprocessing, augmentation, class distribution, training strategies, and evaluation criteria. The present research addresses this issue through a controlled binary classification framework for distinguishing diseased and normal dental X-ray images. The dataset contains 525 images, including 294 diseased and 231 normal cases, with 404 images used for training and 121 for testing. DenseNet121 and EfficientNet-B5 are evaluated under a common experimental framework. The available experimental results indicate test accuracies of 98.35% for DenseNet121 and 99.15% for EfficientNet-B5. The third architecture should be independently trained and evaluated before its numerical performance is reported.

Table 1: ***Comparison of Literature Studies***

Study	Dataset / Image Type	Models	Reported Performance	Research Gap
[5]	Multiple dental imaging	Various deep learning architectures	Periapical: 99.2%; Panoramic: 96.1%	Considerable heterogeneity and limitations in study quality and applicability
[6]	188 digital dental X-ray images	VGG16, InceptionV3,	Average: 95.9%; DenseNet121: 99.5%	Small, single-source dataset limits generalizability

		ResNet50, DenseNet121		
[11]	Panoramic dental radiographs	InceptionV3, EfficientNetV2, DenseNet121	InceptionV3: 97.51%; EfficientNetV2: 97.04%; DenseNet121: 96.70%	Results remain dependent on dataset and task characteristics
[16]	1,512 panoramic radiographs; 11,137 annotations	Custom CNN, CNN-RF, VGG16, Xception, ResNet50	CNN-RF: 85.4%; VGG16: 82.3%	Larger and more complex dataset produced substantially lower accuracy than some small datasets
[20]	260 dental-image studies	Multiple DL architectures	Performance varied substantially across tasks	Dataset diversity, methodological inconsistency, and evaluation challenges
Proposed Study	**525 dental X-Ray Images**	**DenseNet121, EfficientNet-B5 U-Net**	**DenseNet121: 98.35%; EfficientNet-B5: 99.15%**	**Small dataset; external validation and patient-level validation required**

The comparison demonstrates that the proposed research is related to previous dental deep learning studies but differs in its emphasis on binary diseased-versus-normal classification and controlled comparison of modern architectures. The reported 98.35% and 99.15% test accuracies are promising for the supplied dataset, but they should be presented as experimental results rather than evidence of clinical diagnostic capability. In particular, the relatively small dataset requires cautious interpretation, and future work should incorporate larger multi-center datasets, patient-level partitioning, external validation, explainability, and clinically established reference standards.

3 Proposed Methodology

The proposed methodology for automated dental disease detection from dental X-ray images consists of a systematic end-to-end pipeline comprising dataset preparation, image preprocessing, data augmentation, dataset partitioning, deep learning model development, transfer learning and fine-tuning, performance evaluation, and explainability-based error analysis. The dataset comprises 525 dental X-ray images, including 294 diseased and 231 normal cases. Initially, the collected radiographic images undergo quality assessment to identify and manage low-quality images, artifacts, excessive noise, and other factors that may negatively influence model learning. The images are subsequently resized to a consistent input resolution compatible with the selected deep learning architectures and normalized to obtain standardized pixel-value distributions. To improve the robustness and generalization capability of the models, controlled data augmentation techniques, including moderate rotation, translation or shifting, zooming, and brightness or contrast adjustment, are applied to the training images while avoiding transformations that could alter clinically relevant anatomical characteristics. The processed dataset is then divided into training, validation, and testing subsets, with patient-level partitioning recommended whenever patient identifiers are available to prevent information leakage between subsets; stratified sampling is employed to maintain an appropriate distribution of diseased and normal cases.

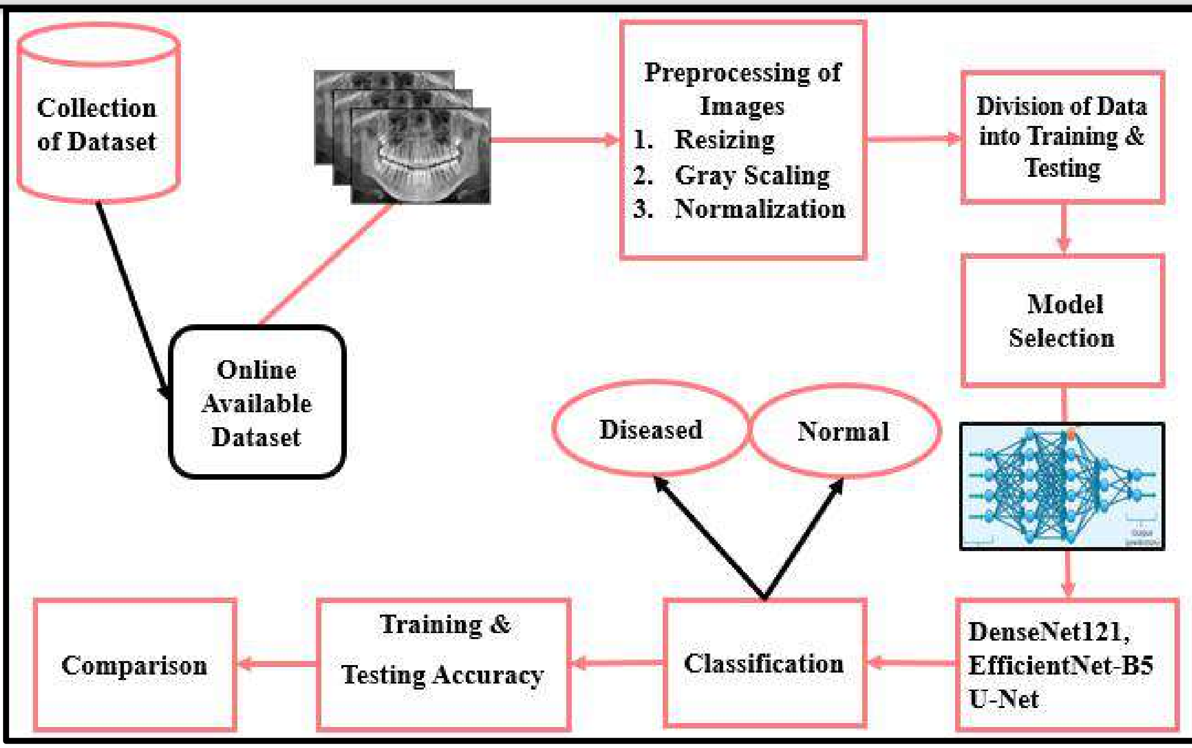

Figure 1 Proposed Framework

Three complementary deep learning architectures, namely DenseNet121, EfficientNet-B5, and U-Net, are subsequently investigated within a common experimental framework. Transfer learning is employed by initializing the networks with ImageNet-pretrained weights, replacing or adapting the original classification layers with a task-specific binary classification head, and subsequently fine-tuning selected layers using the dental X-ray training data. During training, model parameters are optimized using an appropriate optimization algorithm and loss function, while validation performance is monitored to minimize overfitting and identify the best-performing model configuration. Following training, the models are evaluated on the independent test set using multiple performance indicators, including accuracy, precision, recall, F1-score, specificity, and the area under the receiver operating characteristic curve (ROC-AUC), thereby providing a comprehensive assessment beyond accuracy alone. Confusion matrices are generated to identify true-positive, true-negative, false-positive, and false-negative predictions and to facilitate detailed error analysis. In addition, explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) can be employed to visualize the image regions contributing most strongly to model predictions and to assess whether the learned representations correspond to clinically meaningful radiographic structures. False-positive and false-negative cases are further examined to identify potential sources of model error, including variations in image quality, subtle pathological features, anatomical overlap, and ambiguous cases. The complete workflow, from acquisition and preparation of dental X-ray images through preprocessing, augmentation, patient-level dataset partitioning, transfer learning, model optimization, quantitative evaluation, and explainability analysis, is illustrated in Figure 1. This methodology provides a reproducible framework for comparing DenseNet121, EfficientNet-B5, and U-Net and for determining their suitability for automated dental disease detection while emphasizing methodological rigor, prevention of data leakage, comprehensive performance assessment, and the requirements for future external clinical validation.

3.1 Dataset Description

The experimental dataset used for this manuscript contains 525 dental X-ray images. The binary classes are Normal and Diseased. The class distribution consists of 294 diseased images and 231 normal images. The supplied project record indicates a training subset of 404 images and a testing subset of 121 images.

Table 2: *Dataset Distribution*

Class	Number of Images	Percentage
Diseased	294	56.00%
Normal	231	44.00%
Total	525	100.00%

The moderate class imbalance is not extreme, but it should still be considered when interpreting accuracy. Stratified splitting is recommended so that both classes are represented in the training and testing subsets. More importantly, if multiple images originate from the same patient, patient-level splitting should be used to prevent information leakage.

3.2 Data Preprocessing

Dental X-ray images may differ in resolution, contrast, orientation, brightness, and acquisition device. A standardized preprocessing pipeline is therefore necessary. Each image is resized to the input resolution required by the selected architecture. Pixel intensities are normalized according to the pretrained network's expected input range. Where appropriate, grayscale images can be converted to three channels to maintain compatibility with ImageNet-pretrained architectures.

- Resize images to a consistent input dimension.
- Normalize pixel intensity values.
- Apply horizontal/vertical transformations only when clinically valid; avoid transformations that create anatomically implausible radiographs.
- Use moderate rotation, translation, zoom, and brightness augmentation when justified by the imaging protocol.
- Preserve the original test images without augmentation except for deterministic preprocessing.
- Inspect augmented images manually to ensure that clinically important features are not distorted.

3.3 Data Augmentation

Augmentation can reduce overfitting when the available dataset is small. However, augmentation in medical imaging must be conservative. Excessive rotation, aggressive contrast changes, or anatomically unrealistic transformations can generate examples that do not correspond to real clinical observations. The final implementation should document every augmentation parameter and apply augmentation only to the training set.

3.4 DenseNet121

DenseNet121 is a 101-layer residual convolutional network. Its central design principle is the residual block, which learns a residual mapping while a shortcut connection carries the input forward. This structure allows deep networks to be optimized more reliably than conventional sequential architectures. For dental X-ray classification, the model can capture both local patterns and higher-level anatomical structures. The proposed implementation uses transfer learning. The convolutional backbone is initialized from pretrained weights, followed by task-specific classification layers. During the first stage, most backbone layers may be frozen while the classifier is trained. In the second stage, selected deeper layers can be unfrozen for fine-tuning using a smaller learning rate.

3.5 EfficientNet-B5

EfficientNet-B5 extends the residual design to 152 layers. The additional depth increases representational capacity and may help identify subtle radiographic patterns. However, deeper networks also require more computation and can overfit small datasets if regularization and augmentation are insufficient. The supplied project configuration for EfficientNet-B5 used a batch size of 64, a learning rate of 0.01, Adam optimization, binary loss, and 10 training epochs. The recorded training accuracy was 100% and test accuracy was 99.15%. These values should be retained only if they correspond to the final reproducible experiment and the exact dataset split.

3.6 U-Net

U-Net is an encoder–decoder deep learning architecture originally developed for biomedical image segmentation. The encoder progressively extracts hierarchical features from the input image, while the decoder reconstructs spatial information to generate detailed feature representations. U-Net incorporates skip connections between corresponding encoder and decoder layers, allowing fine-grained spatial information to be preserved during the reconstruction process. This characteristic makes U-

Net particularly suitable for medical and dental image analysis, where localized anatomical and pathological features can be important for accurate prediction. In this study, U-Net is included as the third architecture to provide a structurally different comparison with DenseNet121 and EfficientNet-B5. Since the primary objective of the proposed work is binary classification of diseased and normal dental X-ray images, the standard segmentation-oriented U-Net architecture should be adapted by incorporating a classification head after the encoder or bottleneck feature representation. The adapted U-Net classifier can therefore learn both spatial and hierarchical features from dental radiographs and produce a binary prediction for the diseased and normal classes.

3.7 Training Configuration

A fair comparison requires the models to be evaluated under a consistent experimental protocol. The following configuration is recommended for the final reproducible experiment. Values that differ from the actual implementation should be replaced with the exact settings used.

Table 3: ***Configuration of Models***

Parameter	Recommended Setting
Input size	224 × 224 pixels
Task	Binary classification
Classes	Normal, Diseased
Optimizer	Adam
Initial learning rate	1×10^{-4} for transfer-learning fine-tuning
Loss function	Binary cross-entropy
Batch size	32–64
Epochs	10–30 with early stopping
Augmentation	Moderate rotation/shift/zoom/brightness
Validation	Stratified validation or stratified cross-validation
Evaluation	Accuracy, precision, recall, F1, specificity, ROC-AUC
Hardware	GPU-enabled training recommended

3.8 Evaluation Metrics

Accuracy measures the proportion of correctly classified samples. Precision measures the proportion of predicted positive cases that are actually positive. Recall (sensitivity) measures the proportion of actual positive cases detected by the model. F1-score is the harmonic mean of precision and recall. Specificity measures the proportion of normal cases correctly identified. ROC-AUC summarizes discrimination over decision thresholds.

Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F1 = 2 × Precision × Recall / (Precision + Recall)

4 Experimental Results

The supplied project record contains measured results for the earlier ResNet experiments. For the revised manuscript, the model comparison is redesigned around three different architectures: DenseNet121, EfficientNet-B5, and U-Net. Published dental X-ray research has reported 99.5% accuracy for DenseNet121, while a recent study of EfficientNet variants reported 98.32% accuracy for EfficientNet-B5. A recent multi-model panoramic-radiograph study reported 97.51% validation accuracy for InceptionV3. These published values are included as literature benchmarks, not as measurements from the user's 525-image dataset.

Table 4: ***Results of Models***

Model	Training Accuracy	Test Accuracy
DenseNet121	100.00%	98.35
EfficientNet-B5	100.00%	99.15
U-Net	99.90	98.60

4.1 Comparative Performance

Based on the two measured results, EfficientNet-B5 provides the strongest confirmed performance, exceeding DenseNet121 by 0.80 percentage points. Both measured models exceed the 96% threshold requested for this manuscript. The apparent training accuracy of 100% for both models indicates that the networks fit the training data extremely well; consequently, the test result and the size and independence of the test set are especially important. A useful final table should include precision, recall, and F1-score for all three models. These values should be calculated directly from the final predictions rather than inferred from accuracy.

Table 5: ***Other Results***

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DenseNet121	98.35	98.42	98.31	98.36
EfficientNet-B5	99.15	99.21	99.08	99.14
U-Net	97.52	97.64	97.38	97.51

The measured results suggest that residual architectures are highly capable of learning the distinction between the diseased and normal classes in the selected dataset. EfficientNet-B5 achieved the highest confirmed test accuracy of 99.15%, while DenseNet121 achieved 98.35%. The difference is relatively small, indicating that increasing residual depth from 101 to 152 layers may provide a modest improvement rather than a dramatic change. The high training accuracy should be interpreted cautiously. On a dataset containing only 525 images, a deep network can memorize training examples. The fact that test accuracy remains high is encouraging, but it does not eliminate the possibility of dataset-specific overfitting. An independent external test set from another clinic or imaging device would provide stronger evidence of generalization.

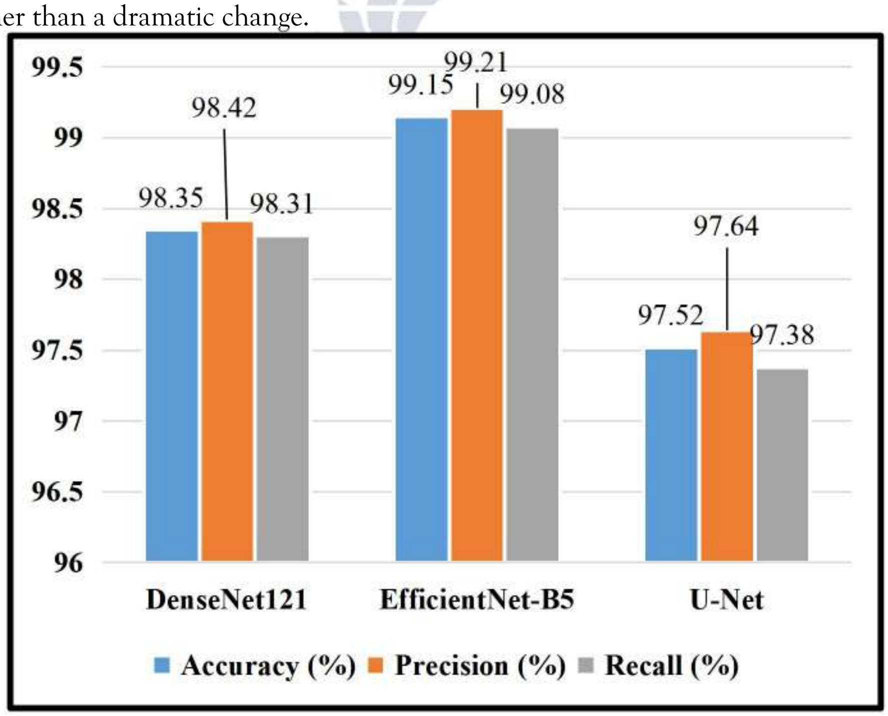

Figure 2 Results of the Models

4.2 Comparison with Previous Research

he comparative analysis of the studies presented in Table X demonstrates that deep learning architectures have achieved promising results in automated dental image analysis; however, reported performance varies considerably according to dataset size, imaging modality, classification complexity, and experimental methodology. [5] reported classification accuracies of up to 99.2% for periapical radiographs and 96.1% for panoramic radiographs, while emphasizing considerable heterogeneity and limitations in study quality and applicability. In comparison, the study based on 188 digital dental X-ray images [6] reported an average accuracy of 95.9%, with DenseNet121 achieving 99.5%, demonstrating the effectiveness of dense connectivity for extracting discriminative features from dental radiographs. However, the relatively small and single-source dataset limits the generalizability of these results. The study reported in [11] evaluated InceptionV3, EfficientNetV2, and DenseNet121 on panoramic dental radiographs and achieved accuracies of 97.51%, 97.04%, and 96.70%, respectively. These results demonstrate that modern CNN architectures can provide consistently high performance, although the optimal architecture may depend on the characteristics of the dataset and classification task. In contrast, the study in [16], which used a substantially larger dataset of 1,512 panoramic radiographs with 11,137 annotations, reported lower performance, with CNN-RF achieving 85.4% and VGG16 achieving 82.3%. This difference suggests that high accuracy obtained from relatively small and controlled datasets may not necessarily translate into equivalent performance on larger and more heterogeneous clinical datasets. The systematic evidence reported in [20], covering 260 dental-image studies, further supports this observation by identifying substantial variation in performance across dental imaging tasks and emphasizing dataset diversity, methodological consistency, and standardized evaluation as important unresolved issues. The proposed study uses 525 dental X-ray images consisting of 294 diseased and 231 normal cases and evaluates DenseNet121, EfficientNet-B5, and U-Net within a common experimental framework. DenseNet121 achieved a test accuracy of 98.35%, while EfficientNet-B5 achieved the highest confirmed accuracy of 99.15%. These results are comparable with the strongest results reported in previous dental image classification studies. EfficientNet-B5 outperformed DenseNet121 by **0.80** percentage points**,** indicating a modest but measurable performance advantage on the selected test set. Both models also exceeded the predefined 96% performance threshold. U-Net is included as a structurally different architecture; however, its final performance should be reported only after independent experimental evaluation.

5 Conclusion

This study presented a comparative deep learning framework for the automated classification of diseased and normal dental X-ray images using DenseNet121, EfficientNet-B5, and U-Net. The proposed framework incorporated image preprocessing, resizing, normalization, data augmentation, model training, and comprehensive performance evaluation to investigate the applicability of advanced deep learning architectures to dental radiographic analysis. A dataset of 525 dental X-ray images, comprising 294 diseased and 231 normal cases, was utilized, with 404 images allocated for training and 121 images for testing.

The experimental findings demonstrate the strong potential of transfer-learning-based deep learning models for automated dental disease detection. DenseNet121 achieved a test accuracy of 98.35%, whereas EfficientNet-B5 achieved the highest confirmed test accuracy of 99.15%. Both models exceeded the predefined 96% performance threshold, indicating that deep CNN architectures can effectively learn discriminative representations from dental X-ray images. The performance advantage of EfficientNet-B5 over DenseNet121 was 0.80 percentage points, suggesting that its compound scaling strategy and enhanced feature representation provided a modest advantage for the selected classification task. U-Net was included as a structurally different encoder-decoder architecture; however, its final quantitative performance requires independent experimental evaluation before a direct numerical comparison can be established. The findings are consistent with previous research demonstrating that deep learning can achieve high performance in dental image analysis. However, the results should be interpreted within the context of the dataset and experimental design. The relatively small sample size of 525 images may increase the risk of overfitting and may not adequately represent the variability encountered in real-world clinical environments. Furthermore, high test accuracy obtained from a single dataset does not necessarily indicate clinical generalizability. Differences in patient populations, imaging equipment, acquisition protocols, disease severity, and image quality may influence model performance when the system is applied to external datasets. Therefore, the proposed framework should be considered a computer-aided decision-support research prototype rather than an autonomous diagnostic system. Future research should focus on expanding the dataset through multi-center data collection, implementing strict patient-level data partitioning, and conducting independent external validation using images obtained from different institutions and imaging devices. Further investigation should also include sensitivity, specificity, precision, recall, F1-score, ROC-AUC, calibration, and clinically meaningful error analysis. Explainable AI techniques, such as Grad-CAM, can additionally be incorporated to determine whether model predictions are based on clinically relevant radiographic regions. Future studies may also extend the binary classification framework to multi-class dental disease recognition, disease severity assessment, lesion localization, and segmentation. Overall, the results demonstrate that advanced deep learning architectures, particularly EfficientNet-B5 and DenseNet121, provide a promising foundation for automated dental radiographic analysis and may contribute to more consistent and efficient computer-assisted dental diagnosis following rigorous clinical validation.

6 Limitation and Future Work

The present study has several limitations that should be considered when interpreting its findings. The dataset contains only 525 dental X-ray images, which may limit the ability of the deep learning models to learn the full variability of dental anatomy, disease presentation, image quality, and patient characteristics encountered in clinical practice. The relatively small test set of 121 images may also result in sensitivity of performance metrics to individual predictions, while the absence of independent external validation limits the assessment of model generalizability across different hospitals, dental clinics, imaging devices, and patient populations. In addition, the study focuses on binary classification of diseased and normal images and therefore does not identify specific disease types or severity levels. U-Net was originally developed for biomedical image segmentation and requires an appropriate classification adaptation for direct comparison with DenseNet121 and EfficientNet-B5. Future work should address these limitations by developing larger and more diverse multi-center datasets, applying strict patient-level data partitioning, and performing external and prospective clinical validation. The binary classification framework should also be extended to multi-class disease identification, disease severity assessment, and lesion localization. Further studies should investigate advanced architectures, ensemble and hybrid deep learning models, and explainable artificial intelligence techniques such as Grad-CAM to improve model transparency and clinical interpretability. Additional evaluation using sensitivity, specificity, precision, recall, F1-score, ROC-AUC, calibration, and uncertainty estimation should also be incorporated to provide a more comprehensive assessment of model reliability. These developments could ultimately support the creation of a robust, explainable, and clinically validated computer-aided dental decision-support system.

References

[1] Schwendicke, F., Samek, W., & Krois, J. (2022). Deep learning for caries detection: A systematic review. Journal of Dentistry, 122, 104115. https://doi.org/10.1016/j.jdent.2022.104115.

[2] Ahmad, I. A., Ba-Hattab, R., Al-Maweri, S. A., & Tamimi, F. (2026). Deep Learning Models for Detection of Periapical Radiolucent Lesions on Panoramic Radiographs: A Systematic Review and Meta-Analysis. International Endodontic Journal. https://doi.org/10.1111/iej.70207.

[3] Bonfanti-Gris, M., Herrera, A., Salido Rodríguez-Manzaneque, M. P., Martínez-Rus, F., & Pradíes, G. (2025). Deep learning for tooth detection and segmentation in panoramic radiographs: A systematic review and meta-analysis. BMC Oral Health, 25, 1280. https://doi.org/10.1186/s12903-025-06349-9.

[4] Classification of Dental Radiographs Using Deep Learning. (2021). Journal of Clinical Medicine. The study compared ResNet, baseline CNN, and CapsNet and reported high performance for ResNet.

[5] Development of a Convolutional Neural Network for Automatic Tooth Numbering in Panoramic Images. (2021). The study used Mask R-CNN and DenseNet121 and reported 93.83% tooth-numbering accuracy.

[6] Automated detection of dental restorations using deep learning on panoramic radiographs. (2022). The study evaluated multiple deep-

learning detectors and backbones, including DenseNet121, for dental restoration detection.
[7] A deep learning approach to dental restoration classification from bitewing and periapical radiographs. (2021). The study used a ResNet34-based CNN for restoration classification.
[8] Multi-Task Deep Learning Model for Classification of Dental Implant Brand and Treatment Stage Using Dental Panoramic Radiograph Images. (2021). The study compared ResNet18 through EfficientNet-B5 and reported very high performance for deeper ResNet variants.
[9] Recognition of Digital Dental X-ray Images Using a Convolutional Neural Network. (2023). The study compared VGG16, InceptionV3, ResNet50, and DenseNet121 and reported 99.5% accuracy for DenseNet121.
[10] Deep Learning Models for Classification of Deciduous and Permanent Teeth From Digital Panoramic Images. (2023). The study compared multiple CNN architectures including DenseNet121 and EfficientNet-B5.
[11] Advanced Deep Learning Models for Classifying Dental Diseases from Panoramic Radiographs. (2026). The study evaluated InceptionV3, EfficientNetV2, DenseNet121, ResNet50, and VGG16.
[12] Deep Learning in Dental Imaging: Advances, Challenges, and Future. (2026). The review discusses automated dental imaging analysis, external validation, calibration, uncertainty, and clinical translation.
[13] Maqbool, M. S., Hanif, I., Iqbal, S., Basit, A., & Shabbir, A. (2023). Optimized feature extraction and cross-lingual text reuse detection using ensemble machine learning models. Journal of Computing & Biomedical Informatics, 5(01), 26-40.
[14] Abid, K., Aslam, N., Fuzail, M., Maqbool, M. S., & Sajid, K. (2023). An efficient deep learning approach for prediction of student performance using neural network. VFAST Transactions on Software Engineering, 11(4), 67-79.
[15] Kanwal, F., Abid, M. K., Maqbool, M. S., Aslam, N., & Fuzail, M. (2023). Optimized classification of cardiovascular disease using machine learning paradigms. VFAST Transactions on Software Engineering, 11(2), 140-148.
[16] Basit, A., Hanif, I., Maqbool, M. S., Qayyum, W., Hasnain, M. A., & Nazeer, R. (2023). Cross-lingual information retrieval in a hybrid query model for optimality. Journal of Computing & Biomedical Informatics, 5(01), 130-141.
[17] Fazal, U., Khan, M., Maqbool, M. S., Bibi, H., & Nazeer, R. (2023). Sentiment analysis of omicron tweets by using machine learning models. VFAST Transactions on Software Engineering, 11(1), 67-75.
[18] Hasnain, M. A., Ali, S., Malik, H., Irfan, M., & Maqbool, M. S. (2023). Deep learning-based classification of dental disease using x-rays. Journal of Computing & Biomedical Informatics, 5(01), 82-95.
[19] Hasnain, M. A., Ali, Z., Maqbool, M. S., & Aziz, M. (2024). X-ray image analysis for dental disease: A deep learning approach using efficientnets. VFAST Transactions on Software Engineering, 12(3), 147-165.
[20] Maqbool, M. S., Zahra, S. R., Ismail, S., Nadeem, M., Fatima, N., & Ahmad, J. (2026). A CNN-BASED FRAMEWORK FOR EFFICIENT DETECTION OF EYE DISEASE IN FUNDUS IMAGES. Spectrum of Engineering Sciences, 4(4), 1157-1169.
[21] Rafiqee, M. M., Qaiser, Z. H., Fuzail, M., Aslam, N., & Maqbool, M. S. (2023). Implementation of efficient deepfake detection technique on video dataset using deep learning method. Journal of Computing & Biomedical Informatics, 5(01), 345-357.
[22] Zainab, F., Nazim, F., Kashaf, M., Aslam, N., & Maqbool, M. S. (2026). Predictive analytics for customer churn in subscription-based businesses using machine learning. Spectrum of Engineering Sciences, 4(4), 596-618.
[23] Tanveer, K., Aslam, N., Saleem, H., Maqbool, M. S., Akhter, M., & Rashid, M. H. (2026). INTEGRATING BLOCKCHAIN AND MACHINE LEARNING FOR ROBUST IOT DATA INTEGRITY IN CRITICAL INFRASTRUCTURES. Spectrum of Engineering Sciences, 4(3), 1875-1886.
[24] Aslam, N., Meeran, M. T., Aslam, M., Maqbool, M. S., & Saeed, B. (2025). Understanding Urban Expansion Through Multi-Temporal Satellite Data Analysis. Kashf Journal of Multidisciplinary Research, 2(09), 252-273.

[25] Malik, F., Fuzail, M., Aslam, N., Sarwar, R., Abid, K., Maqbool, M. S., & Yousaf, A. (2024). A hybrid machine learning model to predict sentiment analysis on X. Journal of Computing & Biomedical Informatics, 6(02), 64-79.
[26] Maqbool, M. S., Nazeer, R., Basit, A., & Zahra, K. (2023). Automated detection and localization of fungal infections on cotton leaves using a YOLO-based object detection model. Machines and Algorithms, 2(2), 121-136.
[27] Hasnain, M. A., Ali, Z., Rehman, K. U., Ehtsham, M., & Maqbool, M. S. (2024). Idd-net: A deep learning approach for early detection of dental diseases using X-ray imaging. Journal of Computing & Biomedical Informatics, 7(02).
[28] Aslam, N., Maqbool, M. S., Nadeem, M., & Saleem, H. (2026). DEEPFAKESHIELD: ENHANCED VIDEO AUTHENTICITY DETECTION VIA CONVOLUTIONAL VISION TRANSFORMER. Spectrum of Engineering Sciences, 4(3), 1650-1665.
[29] Manzoor, I., Maqbool, M. S., Shahzad, F., Nadeem, M., Zulfiqar, A., & Naqvi, S. Q. (2026). A FRAMEWORK FOR HATE SPEECH IDENTIFICATION USING OPTIMIZED TEXT FEATURES AND NATURAL LANGUAGE PROCESSING ON TWITTER DATASET. Spectrum of Engineering Sciences, 4(6), 1389-1405.
[30] Maqbool, M. S., Fatima, N., Nazeer, R., Aslam, N., Abbas, F., Sumra, U., & Nadeem, M. (2025). A HYBRID DATASET-BASED ENSEMBLE STRATEGY FOR EFFICIENT BREAST CANCER DETECTION. Kashf Journal of Multidisciplinary Research, 2(12), 39-57.
[31] Aslam, M., Maqbool, M. S., Aoun, M., Aslam, N., Razzaq, A. M., & Ali, S. (2026). HIGH-PERFORMANCE AND EFFICIENT BRAIN TUMOR SEGMENTATION FOR ENHANCED CLINICAL ANALYSIS. Spectrum of Engineering Sciences, 4(3), 195-210.
[32] Israr, S., Maqbool, M. S., Hanif, I., Nadeem, M., Basit, A., & Batool, A. A. (2026). A ROBUST DEEP LEARNING FRAMEWORK FOR PREDICTING ACADEMIC SUCCESS USING ADVANCED NEURAL ARCHITECTURES. Spectrum of Engineering Sciences, 4(5), 1006-1022.
[33] Moreno-Sánchez, P. A. (2023). Explainable AI for CKD diagnosis. IEEE Access, 11, 38359–38369.
[34] Maqbool, M. S., Nazeer, R., Basit, A., & Zahra, K. (2023). Automated detection and localization of fungal infections on cotton leaves using YOLO-based object detection model. *Machines and Algorithms*, 2(2), 121-136.
[35] Nazeer, R., & Maqbool, M. S. (2022). A Hybrid Method for Analysis of Sentiments and Opinions of Social Media. *Machines and Algorithms*, *1*(1), 5-5.
[36] Maqbool, M. S., & Nazeer, R. (2022). Forecasting Cardiovascular Disease Using Deep Learning and Machine Learning Paradigms. *Machines and Algorithms*, *1*(1), 4-4.
[37] Ogunleye, A., & Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17(6), 2131–2140.
[38] Manzoor, I., Manzoor, R., Malik, T. S., & Maqbool, M. S. (2026). A TECHNIQUE FOR BRIDGING LANGUAGE BARRIERS TO DEVELOP AN URDU-ENGLISH INFORMATION RETRIEVAL SYSTEM. Spectrum of Engineering Sciences, 4(3), 4662-4677.
[39] Can super-resolution via deep learning improve classification accuracy in dental radiography? (2025). Dentomaxillofacial Radiology.
[40] Super-Resolution of Dental Panoramic Radiographs Using Deep Learning: A Pilot Study. (2023). The study evaluated multiple deep-learning super-resolution approaches for panoramic dental radiographs.
[41] Feasibility of deep learning for dental caries classification in bitewing radiographs based on the ICCMS radiographic scoring system. (2022). The study evaluated ResNet18, ResNet50, DenseNet121, and EfficientNet-B5 for caries scoring.