Review Article

https://doi.org/10.36377/ET-0174

Artificial intelligence in dental diagnosis: evaluating CNN models for caries and periapical lesions detection

Mohammed Al-Sabri[1], Ayman Alsabry[2] ✉, Naibah Rajeh[1], Arham Alkherbash[1], Raghda Alnozaily[1], Ebtehal Sharhan[1]

[1] Sana'a University, Sana'a, Yemen

[2] International University of Technology Twintech, Sana'a, Yemen

✉ aymanalsabri@iutt.edu.ye

Abstract

INTRODUCTION. Convolutional neural networks (CNNs) show strong promise for automating dental diagnosis from radiographic images. Robust head-to-head comparisons across tasks and datasets are needed to guide model selection for clinical use.

MATERIALS AND METHODS. We compared three lightweight CNNs – EfficientNet-B0, ResNet-18, and MobileNetV3 – for two classification tasks: enamel caries on intraoral images and periapical lesions on panoramic radiographs. Data comprised the Caries-Spectra dataset (2,000 intraoral images; advanced, early-stage, and no caries) and a panoramic radiograph set with 13,071 images labeled with periapical lesion scores (PAI 3–5). For the periapical task, data augmentation was applied to the panoramic training split only, increasing its size to 17,004 training instances; validation and test splits (as well as all Caries-Spectra splits) remained at their original sizes. Models were trained via transfer learning with early stopping and evaluated using accuracy, precision, recall, F1-score, and confusion matrices.

RESULTS. EfficientNet-B0 achieved the best overall performance on both tasks, reaching 99.74% accuracy for caries detection and 69.65% accuracy for periapical lesion detection, outperforming ResNet18 and MobileNetV3 across the reported metrics.

CONCLUSIONS. Lightweight CNNs – particularly EfficientNet-B0 – are effective for dental image classification and are suitable candidates for integration into clinical diagnostic workflows. Model architecture choice and data quality materially influence performance

Keywords: artificial intelligence, dental diagnosis, convolutional neural networks, caries detection, periapical lesions

Article information: received – 03.11.2025; revised – 20.12.2025; accepted – 08.01.2026

Conflict of interest: The authors declare no conflict of interest.

Acknowledgements: There are no funding and individual acknowledgments to declare.

For citation: Al-Sabri M., Alsabry A., Rajeh N., Alkherbash A., Alnozaily R., Sharhan E. Artificial intelligence in dental diagnosis: evaluating CNN models for caries and periapical lesions detection. *Endodontics Today*. 2026;24(1):176–187. https://doi.org/10.36377/ET-0174

Искусственный интеллект в стоматологической диагностике: оценка моделей CNN для выявления кариеса и периапикальных поражений

М. Аль-Сабри[1], А. Аль-Сабри[2] ✉, Н. Раджех[1], А.А. Аль-Хирбаш[1], Р. Аль-Нозайли[1], Э.М. Шархан[1]

[1] Университет Саны, Сана, Йемен

[2] Международный технологический университет Twintech, Сана, Йемен

✉ aymanalsabri@iutt.edu.ye

Резюме

ВВЕДЕНИЕ. Сверточные нейронные сети (CNN) демонстрируют большие перспективы в автоматизации стоматологической диагностики по рентгенографическим снимкам. Для внедрения этих моделей в клиническую практику необходимы надежные сравнительные исследования на различных задачах и наборах данных.

МАТЕРИАЛЫ И МЕТОДЫ. Мы сравнили три легковесные модели CNN – EfficientNet-B0, ResNet-18 и MobileNetV3 – для двух задач классификации: выявление кариеса эмали на интраоральных снимках и периапикальные поражения на панорамных рентгенограммах. Данные включали набор Caries-Spectra (2000 интраоральных снимков: глубокий кариес, начальный кариес и отсутствие кариеса) и набор панорамных рентгенограмм из 13 071 снимка, размеченных по индексам периапикальных по-

ражений (PAI 3–5). Для задачи с периапикальными поражениями применялась аугментация данных только для обучающей выборки, что увеличило ее размер до 17 004 примеров; валидационные и тестовые выборки (как и все выборки Caries-Spectra) остались в исходном размере. Модели обучались с использованием трансферного обучения с ранней остановкой и оценивались по метрикам точности (accuracy), прецизионности (precision), полноты (recall), F1-меры и матрицам ошибок.
РЕЗУЛЬТАТЫ. EfficientNet-B0 показала наилучшую общую производительность в обеих задачах, достигнув точности 99,74% для выявления кариеса и 69,65% для выявления периапикальных поражений, превзойдя ResNet18 и MobileNetV3 по всем представленным метрикам.
ВЫВОДЫ. Легковесные CNN, в частности EfficientNet-B0, эффективны для классификации стоматологических изображений и являются подходящими кандидатами для интеграции в клинические диагностические рабочие процессы. Выбор архитектуры модели и качество данных существенно влияют на производительность.

Ключевые слова: искусственный интеллект, стоматологическая диагностика, сверточные нейронные сети, выявление кариеса, периапикальные поражения

Информация о статье: поступила – 03.11.2025; исправлена – 20.12.2025; принята – 08.01.2026

Конфликт интересов: Авторы сообщают об отсутствии конфликта интересов.

Благодарности: Финансирование и индивидуальные благодарности для декларирования отсутствуют.

Для цитирования: Аль-Сабри М., Аль-Сабри А., Раджех Н., Аль-Хирбаш А.А., Аль-Нозайли Р., Шархан Э.М. Искусственный интеллект в стоматологической диагностике: оценка моделей CNN для выявления кариеса и периапикальных поражений. *Эндодонтия Today.* 2026;24(1):176–187. https://doi.org/10.36377/ET-0174

INTRODUCTION

Enamel caries and periapical lesions are among the most prevalent oral diseases worldwide and, if left undetected or untreated, can lead to significant morbidity, functional impairment, and financial burdens for patients [1]. Conventional diagnostic methods have traditionally relied on expert interpretation of dental radiographs, which often suffer from limited reproducibility, susceptibility to human fatigue, and reduced scalability in high-demand clinical settings.

Recent advancements in artificial intelligence (AI) and deep learning – particularly convolutional neural networks (CNNs) – have demonstrated significant potential in automating image-based diagnosis across multiple medical disciplines, including dentistry. CNNs possess the capability to extract hierarchical features and detect subtle variations in radiographic images that may not be readily perceptible to the human eye. However, the performance of CNNs can vary depending on architectural design, dataset quality, and task complexity [2–5].

Several studies have explored AI-based detection of caries and periapical lesions using various CNN architectures and imaging modalities [6]. While these studies have shown encouraging results, few have conducted comprehensive, standardized comparisons across multiple CNN models using unified datasets and consistent preprocessing pipelines.

This study aims to address this gap by performing a comparative evaluation of three well-established CNN architectures – ResNet18, EfficientNet-B0, and MobileNetV3 – across two publicly available dental image datasets. The first dataset comprises intraoral images for enamel caries classification, while the second consists of panoramic radiographs for periapical lesion detection.

Through standardized training strategies and evaluation metrics – including accuracy, precision, recall, F1-score, and confusion matrices – this research seeks to identify the CNN architecture that best balances diagnostic performance and computational efficiency in the context of dental image analysis.

LITERATURE REVIEW

Recent advancements in artificial intelligence have significantly contributed to the automation of dental disease detection, particularly in identifying enamel caries and periapical lesions. Numerous studies have employed deep learning models to address the challenges of dental image analysis, leveraging both two-dimensional and three-dimensional imaging modalities.

For instance, a study by [7] utilized a hybrid nnU-Net–DenseNet121 model for caries detection in panoramic radiographs, achieving a high accuracy of 98.6%, thus highlighting the effectiveness of combined segmentation-classification approaches. Similarly, ResNet-101 was employed by [8] on the DCD dataset, yielding an accuracy of 98.38%, reaffirming the robustness of deep residual networks in extracting diagnostic features from dental radiographs.

EfficientNet-B0 has also been investigated in the context of dental imaging. In [9], the model achieved 94.94% accuracy on a dataset of 1,525 dental X-ray images, demonstrating the advantage of compound scaling in CNN architectures. In contrast, a non-image-based study [10] focused on demographic and survey data to predict caries risk, using GINI and mRMR feature selection with a Gradient Boosting Decision Tree (GBDT) classifier to achieve 95% accuracy – underscoring the relevance of behavioral and demographic variables in caries risk assessment.

Three-dimensional imaging has also gained traction in recent research. In [11], a multi-input CNN was applied to cone-beam computed tomography (CBCT) images of molar teeth, achieving 95.3% accuracy. The study emphasized the diagnostic value of volumetric data in identifying complex or deep lesions. In a related effort, [12] employed ResNet-101 with a Feature Pyramid Network (FPN) and Adaptive Training Sample Selection, attaining 81.2% accuracy on panoramic radiographs – highlighting the challenges associated with sample variability and generalizability.

Regarding periapical lesion detection, [13] introduced a ResNet-based model integrated with a Spatial Attention Module (SAM) and tested it on 4,278 periapical radiographs, achieving an accuracy of 88.5%. This result illustrates the potential of attention mechanisms to enhance model focus on lesion-specific regions.

Collectively, these studies reflect the growing success of CNN-based models in dental diagnostics, with performance influenced by imaging modality (2D vs. 3D), dataset quality, and architectural enhancements such as attention mechanisms or ensemble methods. However, few works offer a unified comparative analysis across multiple CNN architectures under standardized conditions. The present study aims to fill this gap by systematically evaluating ResNet18, EfficientNet-B0, and MobileNetV3 on enamel caries and periapical lesion datasets, with a focus on balancing diagnostic performance and computational efficiency.

Table 1 provides a comparative summary of prior studies, including datasets, model architectures, and reported performance metrics for enamel caries and periapical lesion detection.

METHODOLOGY

This study aims to evaluate the performance of various Convolutional Neural Network (CNN) architectures in detecting dental pathologies, specifically caries and periapical lesions, from periapical radiographic images. As shown in Figure 1, the methodology consists of several stages, including data acquisition, preprocessing, model training, and evaluation.

Datasets

Two distinct datasets were utilized in this study:

Dataset 1 – Caries-Spectra (intraoral images): a set of 2,000 intraoral photographic images categorized into advanced caries, early-stage caries, and no caries.

Dataset 2 – Panoramic radiographs (PAI 3–5): 13,071 panoramic radiographs annotated with Periapical Index (PAI) scores 3–5 for periapical lesion detection.

Data preprocessing

All images were resized to a fixed input dimension to meet the requirements of the selected CNN models. Additionally, data augmentation techniques were selectively applied to improve model robustness and detection performance. Specifically, horizontal flipping and color jitter were applied to the Periapical Lesions dataset. This was done not merely to increase dataset diversity, but primarily to enhance the model's ability to detect small-sized objects within the radiographs, as periapical lesions tend to be subtle and localized, making them more challenging to identify compared to larger carious defects.

Dataset splitting

Each dataset was split into three subsets:

1. Training Set (80%): Used for model learning.
2. Validation Set (10%): Used for hyperparameter tuning and early stopping.
3. Test Set (10%): Reserved for final evaluation.

Model architectures

Three state-of-the-art CNN models were selected for comparative analysis:

– ResNet18;
– EfficientNet-B0;
– MobileNetV3.

Table 1. Summary of recent studies on AI-based dental disease detection

Таблица 1. Краткое описание недавних исследований в области выявления стоматологических заболеваний с помощью искусственного интеллекта

Ref.	Year	Dataset	Methods	Accuracy, %
[14]	2021	Panoramic films of Caries. A training and validation dataset (1071) and a test dataset (89)	nnU-Net and DenseNet121	0.986
[9]	2023	A total of 1525 X-ray images of Caries	EfficientNet-B0	94.94
[8]	2023	Dental Caries Detection Dataset (DCD)	ResNet-101	98.38
[10]	2023	A survey was conducted with 22,288 respondents using an oral health awareness questionnaire featuring 43 items. This questionnaire collected data on demographics (age, gender, residence), habits (snack frequency, tooth brushing frequency, oral care use, smoking experience), and assessed oral health awareness and behaviors, with "act_caries" as the key label	Integrating the GINI and mRMR algorithms with the GBDT classifier	95
[11]	2024	The CBCT image dataset included 382 molar teeth with caries and 403 non-carious molar cases	Multiple-input convolutional neural network	95.3
[12]	2023	454 objects in 357 anonymized PRs were	Adaptive Training Sample Selection and ResNet-101 FPN	0.812
[13]	2024	Periapical radiographs (4278 images)	ResNet+SAM	0.885

Training strategy

We trained each model for up to 20 epochs on Caries-Spectra and up to 10 epochs on the periapical task, with early stopping based on validation accuracy (patience = N). We monitored loss/accuracy per epoch and saved the best weights.

Evaluation metrics

Model performance was evaluated using the following standard classification metrics:

– accuracy;
– precision;
– recall;
– F1-Score.

A confusion matrix was also used to visualize the classification outcomes.

Results comparison

The final performance of each model was tabulated to allow side-by-side comparison across all metrics. This comparison facilitated the identification of the most suitable architecture for automated dental pathology detection.

DATASETS DESCRIPTION

This study utilizes two publicly available dental image datasets, each focused on a distinct pathology to enable comprehensive evaluation of deep learning models for dental diagnosis.

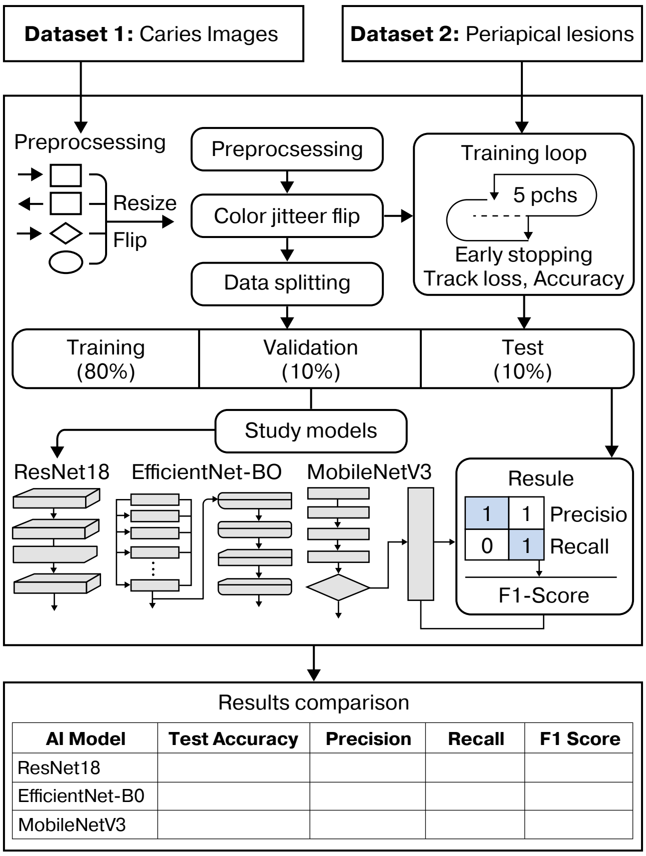

Fig. 1. Methodology

Рис. 1. Методология

Caries-spectra dataset for enamel caries detection

The Caries-Spectra dataset is a curated collection of 2,000 low-resolution intraoral images, sourced from medical clinics in Rajshahi and Dhaka, Bangladesh, between 2021 and 2022 [15].

The dataset is categorized into three distinct groups:

– **Advanced Enamel Caries** (800 images);
– **Early-Stage Enamel Caries** (800 images);
– **No Enamel Caries** (400 images).

These images represent diverse presentations of enamel caries, selected from multiple patients to ensure clinical variability. The dataset is structured into three folders corresponding to the caries categories and is intended to support fine-grained caries classification tasks.

Periapical lesions panoramic radiograph dataset

The second dataset comprises 13071 panoramic radiographs labeled with periapical lesions, obtained at the High-quality Dental Treatment Centre, Hanoi Medical University between 2016 and 2021. The radiographs were annotated by a panel of three experienced dentists based on the Periapical Index (PAI), categorizing lesions into PAI scores 3, 4, and 5 [16].

To improve lesion detection performance, particularly for small and subtle abnormalities, the dataset was extensively augmented using random scaling, mirroring, rotation, and noise addition, resulting in 17,004 images. Image annotations are provided in XML format. The dataset is structured into three main folders: original images, augmented images, and annotations.

Together, these datasets provide a robust foundation for evaluating CNN architectures on tasks related to dental caries and periapical lesion detection under varied clinical imaging conditions.

UNIFIED CNN TRAINING AND EVALUATION PIPELINE (APPLIED TO RESNET18, EFFICIENTNET-B0, AND MOBILENETV3)

To ensure consistency and fairness in evaluating different CNN architectures, a unified training and evaluation pipeline was implemented across all selected models: ResNet18, EfficientNet-B0, and MobileNetV3. This pipeline outlines the complete process from model initialization to performance evaluation on the test set. It includes data preprocessing, transfer learning, optimization, early stopping, and computation of key classification metrics.

The following pseudocode (Algorithm 1) summarizes the standardized steps followed during the development and assessment of each CNN model in this study.

Input:
– preprocessed image datasets (Train, Validation, Test);
– CNN model architecture (ResNet18 / EfficientNet-B0 / MobileNetV3);
– training hyperparameters: learning rate, batch size, number of epochs, early stopping patience.
Output:
– trained model weights;

– evaluation metrics: Accuracy, Precision, Recall, F1-Score, Confusion Matrix.

Steps:

1. Load the pretrained CNN model.
2. Replace the final classification layer to match the number of classes.
3. Apply image transforms (Resize to 224×224, normalization).
4. Load training, validation, and testing data using DataLoaders.
5. Initialize loss function (CrossEntropyLoss) and optimizer (Adam).
6. For each epoch:

a) set model to training mode;

b) for each batch in training data:

– forward pass,

– compute loss,

– backward pass and optimizer update,

– track training accuracy;

c) set model to evaluation mode;

d) validate model performance on the validation set;

e) save best model weights based on validation accuracy;

f) apply early stopping if no improvement.

7. Load best model weights.
8. Evaluate on test set and compute:

– accuracy;

– precision, Recall, F1-Score (macro average);

– confusion matrix.

RESULT AND DISCUSSION

Performance presented and discussed are in the light of application of the CNN models to the two dental image datasets considered: Caries-Spectra and Periapical Lesions. For each dataset, a comparative test with regard to performance was run for the three CNN architectures- ResNet18, EfficientNet-B0, and MobileNetV3- on standard classification metrics such accuracy, precision, recall, and F1-score.

Evaluation on the caries-spectra dataset

This section discusses the performance analysis of the CNN models for classifying caries in enamel using the Caries-Spectra dataset.

The Caries-Spectra dataset is evaluated by the efficiency of the CNN-based classification in discriminating among the three classes of enamel statuses: advanced enamel caries, initial-stage enamel caries, and healthy enamel, respectively. Among various models tested, the `ResNet18; has been the tallest performer with respect to training and validation summarized in Table 2 and illustrated in Fig. 2. This trend continued through 20 training epochs, where the ResNet18 model was observed to converge steadily. The training loss was seen to be consistently lower through the epochs: from 33.99 at epoch 1 to 3.79 by epoch 20, demonstrating good optimization. That way, the training accuracy improved considerably from under 79% to just below 98% in terms of learning the stability of the model over time. And this is where it fails as well as it goes up, reaching a peak point at 98.44% in epoch 18. This means it generalizes very well to unseen data; sporadic drops that occur at points like epochs 3 and 16 could be due to noise or variance in the dataset from which the representations are derived, since certain images would typically be diffused because they are more difficult to identify because of very slight visual cues to be identified as early-stage enamel caries representations.

The training and validation accuracy curves also run very closely thereby, confirming that the model was not seriously overfitted. The last evaluation metrics-precision, recall, and F1-score-in a way support the robustness of ResNet18 on this low-resolution caries dataset. The extreme performance is most likely due to the more distinctive separation of classes available in the dataset, as well as the residual capabilities of ResNet's connections to retain feature integrity throughout the layers.

ResNet18 proved to have a great performance classification of enamel caries on the Caries-Spectra dataset with high accuracy and low loss, thereby establishing a very good baseline for comparison with the other CNN models.

Cross the 20 training epochs of EfficientNet-B0 on the Caries-Spectra dataset, it consistently proved its impressive and stable performance. As presented in Table 3 and Fig. 3, there was a substantial decline in the training loss from 26.75 on the first epoch to 4.88 on the last, while the increases in training accuracy carried on steadily to reach the maximum of 99.27% at epoch 16.

Table 2. Epoch-wise training and validation performance of ResNet18 on the Caries-Spectra test set

Таблица 2. Результаты обучения и проверки ResNet18 на тестовом наборе Caries-Spectra по эпохам

Epoch. No	Loss	Train Acc, %	Val Acc, %
Epoch 1	33.9895	78.23	89.09
Epoch 2	20.0739	86.68	93.51
Epoch 3	15.1001	91.08	88.05
Epoch 4	14.3778	91.14	92.47
Epoch 5	12.8426	91.29	92.73
Epoch 6	12.0435	93.00	92.73
Epoch 7	10.9420	94.19	93.25
Epoch 8	7.4350	95.90	94.29
Epoch 9	9.1760	94.56	94.03
Epoch 10	8.9577	95.13	94.81
Epoch 11	6.7755	95.96	95.32
Epoch 12	7.0643	95.65	97.40
Epoch 13	5.2692	97.05	96.88
Epoch 14	7.0371	96.63	98.18
Epoch 15	5.6552	96.68	93.25
Epoch 16	7.6589	96.06	88.83
Epoch 17	11.7846	93.42	96.36
Epoch 18	5.6772	96.99	98.44
Epoch 19	3.8931	97.62	98.18
Epoch 20	3.7911	97.93	97.92

The validation accuracy was high throughout the epochs, recording a minimum of 93% at all times since epoch one and attaining 100% at epochs 13 and 16. The early and sustained high performance tells us that EfficientNet-B0 must have learned to discriminate rapidly between the three enamel caries classes, most probably due to some very modern architecture that balances depth, width, and resolution in an efficient way.

Training accuracy continued to rise, with smaller fluctuations in validation accuracy, further demonstrating the capability of generalization and robustness of the model. The close match in the curves for training and validation accuracy in Fig. 3 affirms the stability of the model and suggest that overfitting was avoided.

While the training loss started to show slight increases in later epochs (epoch 20 as an example), validation accuracy kept on recording its typhpeak, meaning the model remained capable of adequately classifying fresh samples even with moments of fluctuations during optimization. To put it somewhat differently, the EfficientNet-B0 was superior to the ResNet18 in terms of both speed of convergence and validation performance on the Caries-Spectra dataset. Given its efficiency and accuracy, it therefore has the potential for deployment in real life for enamel caries classification.

The MobileNetV3 classified the Caries-Spectra dataset strongly, especially in the later epochs. As seen in Table 4 and in Fig. 4, the model opened with a very low validation accuracy of 20.26% in the first epoch, suggesting it encountered great difficulty in generalizing to unseen data in the beginning. Yet, in the subsequent epochs, the validation accuracy increased drastically to above 98% by epoch 9 and 99.74% by epoch 20.

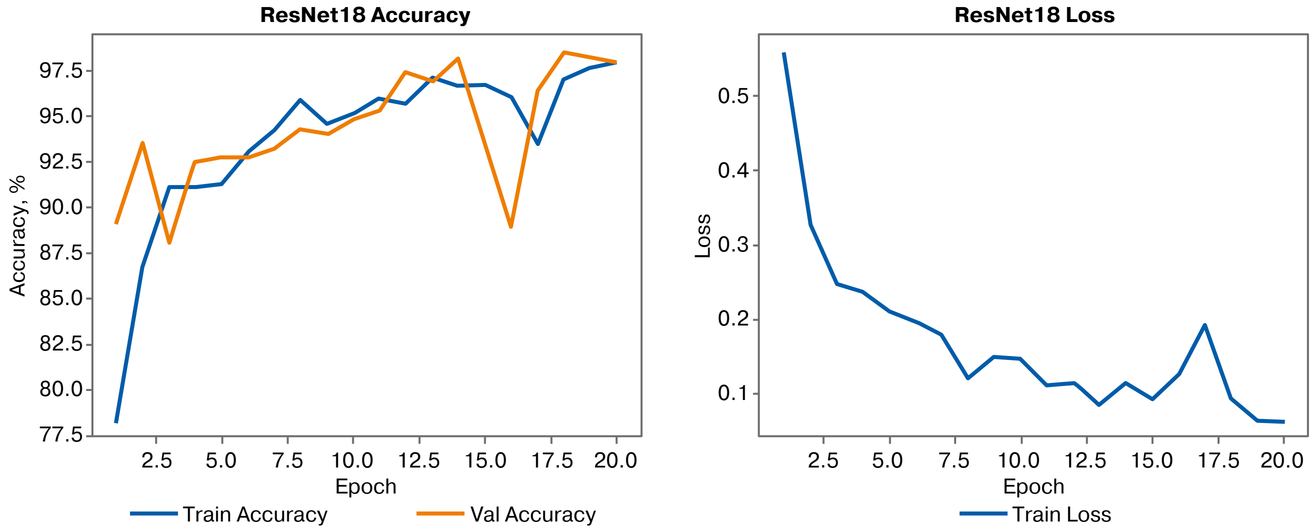

Fig. 2. Training and validation performance of ResNet18 Over 20 Epoch on the Caries-Spectra test set

Рис. 2. Результаты обучения и проверки ResNet18 за 20 эпох на тестовом наборе Caries-Spectra

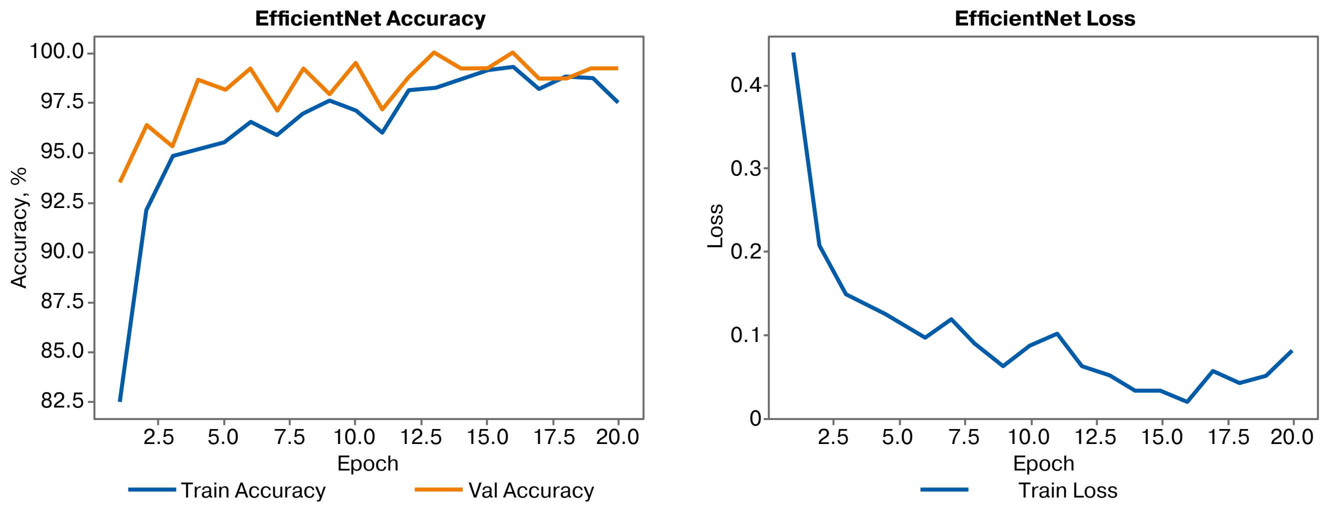

Fig. 3. Training and validation performance of EfficientNet-B0 Over 20 Epoch on the Caries-Spectra test set

Рис. 3. Эффективность обучения и проверки модели EfficientNet-B0 за 20 эпох на тестовом наборе Caries-Spectra

Table 3. Epoch-wise training and validation performance of EfficientNet-B0 on the Caries-Spectra test set

Таблица 3. Результаты обучения и проверки EfficientNet-B0 на тестовом наборе Caries-Spectra по эпохам

Epoch. No	Loss	Train Acc, %	Val Acc, %
Epoch 1	26.7512	82.53	93.51
Epoch 2	12.6605	92.07	96.36
Epoch 3	9.0410	94.82	95.32
Epoch 4	8.1270	95.13	98.70
Epoch 5	6.9896	95.54	98.18
Epoch 6	5.8580	96.53	99.22
Epoch 7	7.2084	95.85	97.14
Epoch 8	5.3080	96.94	99.22
Epoch 9	3.7269	97.62	97.92
Epoch 10	5.2328	97.10	99.48
Epoch 11	6.1185	96.01	97.14
Epoch 12	3.7166	98.13	98.70
Epoch 13	3.1427	98.24	100.00
Epoch 14	2.0137	98.70	99.22
Epoch 15	1.8328	99.12	99.22
Epoch 16	1.1737	99.27	100.00
Epoch 17	3.3485	98.19	98.70
Epoch 18	2.5278	98.81	98.70
Epoch 19	3.0224	98.70	99.22
Epoch 20	4.8824	97.56	99.22

Table 4. Epoch-wise training and validation performance of MobileNetV3 on the Caries-Spectra test set

Таблица 4. Результаты обучения и проверки MobileNetV3 на тестовом наборе Caries-Spectra по эпохам

Epoch. No	Loss	Train Acc, %	Val Acc, %
Epoch 1	26.2387	82.17	20.26
Epoch 2	11.0225	93.78	57.40
Epoch 3	8.3453	95.80	64.16
Epoch 4	7.5284	95.23	56.10
Epoch 5	7.2374	96.06	87.53
Epoch 6	5.1619	96.79	81.82
Epoch 7	2.3749	98.76	88.05
Epoch 8	4.5275	97.87	94.55
Epoch 9	2.5562	98.50	98.96
Epoch 10	3.7563	98.34	97.92
Epoch 11	1.9478	98.91	96.36
Epoch 12	1.9185	99.12	97.14
Epoch 13	3.3504	98.60	97.40
Epoch 14	2.7350	98.91	100.00
Epoch 15	5.8561	96.73	97.14
Epoch 16	1.5765	99.02	96.10
Epoch 17	6.4405	96.94	98.96
Epoch 18	3.4517	98.13	97.14
Epoch 19	0.6250	99.69	99.48
Epoch 20	0.8073	99.74	99.74

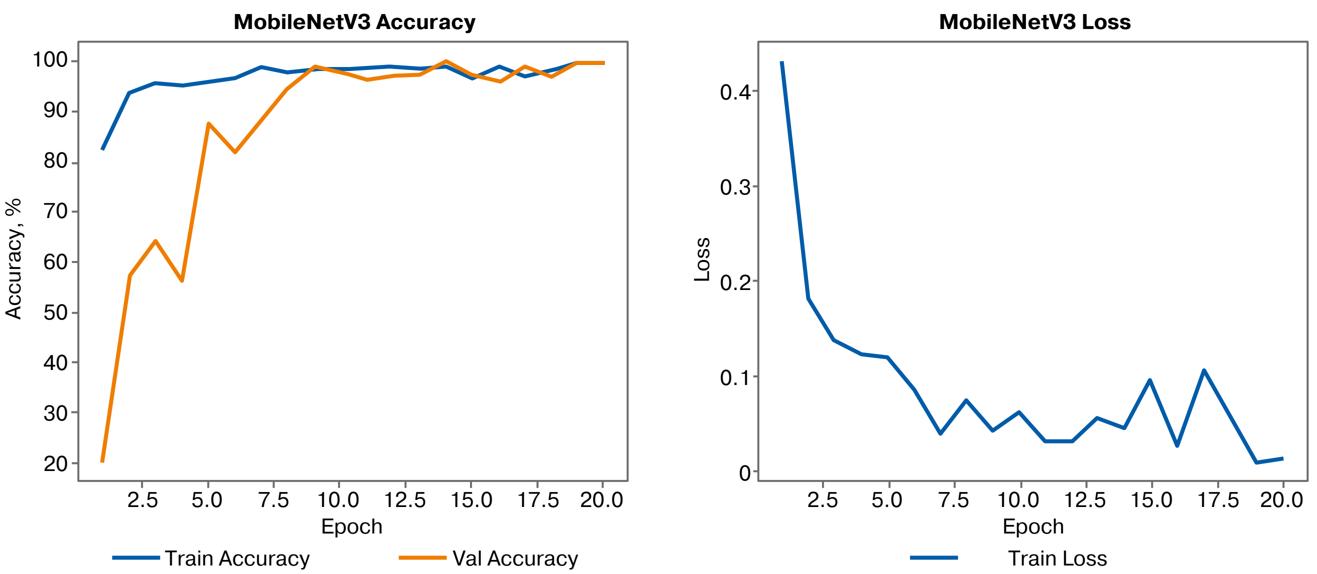

Fig. 4. Training and validation performance of MobileNetV3 Over 20 Epoch on the Caries-Spectra test set

Рис. 4. Эффективность обучения и проверки MobileNetV3 в течение 20 эпох на тестовом наборе Caries-Spectra

The training loss had a steep decline from 26.23 to 0.81, with the training accuracy reaching 99.74%, which confirmed that learning was indeed performed efficiently. The early fluctuations in the validation accuracy (for instance, from epochs 2 to 4) might have indicated early overfitting or the model's inability to accommodate the features of early-stage enamel caries that were more subtle in appearance. However, this early oscillation settled with a lot of stability from epoch 5 onward.

As illustrated in Fig. 4, both curves converged toward one another around epoch 10. The convergence indicates an improvement in generalization performance and reduced overfitting with extended training. The mild fluctuations in the training loss during the later epochs (epochs 15 to 18) did not adversely affect the validation score, indicating a robust optimization process for the model.

After training and validating the models, three CNNs – ResNet18, EfficientNet-B0, and MobileNetV3 – were tested on the held-out test set. As shown in Table 5, all models performed exceedingly well in all metrics since they generalized well to unseen data.

EfficientNet-B0 performed the best: it displayed an outstanding test result of 99.74%, being a perfect 0.9974 in macro-averaged precision, recall, and F1-score. This supports previous judgments made during training about the network's generalization ability and powerful learning. The compound scaling method of EfficientNet-B0 was likely responsible for obtaining a trade-off on the feature extraction throughout resolution and depth.

ResNet18 also performed quite well, achieving 98.94% accuracy and perfectly balanced precision and recall at 0.9896 and 0.9894, respectively. In conclusion, the residual connections that supported stable learning in their ResNet18 architecture were less efficient for parameter utilization in their case, giving slightly higher performances to EfficientNet-B0.

But for the most lightweight model, which is still classed as a CNN, which will be best adapted for real-time or embedded dental diagnostic systems, test accuracy of 98.68%, with consistent measures of precision, recall, and F1-score (0.9868) were obtained from MobileNetV3. The model proved to be quite effective despite fewer parameters and optimizations for mobile applications, especially for deployment with limited resources.

In conclusion, all three CNNs demonstrated quite good classification abilities over the Caries-Spectra test set, with EfficientNet-B0 being declared first among equals by any measure. At the same time, MobileNetV3 serves as the best trade-off between accuracy and computational efficiency, an attractive choice for real-time or embedded dental diagnostic systems.

Evaluation on the periapical lesions dataset

The findings of detecting periapical lesions on panoramic radiographs are discussed in this section of the second dataset.

ResNet18 exhibited a relatively moderate performance in classifying periapical lesions into PAI categories 3, 4, and 5 on panoramic radiographs. As seen in Table 6, the training accuracy shows a steep increase from 57.13% in epoch one to 96.93% by epoch six, indicating that effective learning has happened on the training set. However, with respect to validation accuracy, it kept fluctuating in an unstable manner between 58.71% and 64.32%, which thereafter led to early stopping after six epochs with no other improvement.

The model's limitations are further elaborated in Fig. 5, where the confusion matrix talks about the limitations in generalizing to validation data never seen before. PAI category 3 was identified more or less accurately with 651 correct predictions, while there was tremendous confusion distinguishing PAI 4 and PAI 5, where 180 instances of PAI 4 were incorrectly predicted as PAI 3, and 89 instances of PAI 5 were incorrectly assigned to PAI 3. Such imbalance indicates that ResNet18, despite the signature residual learning architecture for small, low-contrast lesions, is probably not suitable for differential diagnosis in the apical periodontitis examination. The overlapping radiographic characteristics across PAI categories might have contributed to such impairment of the model in extracting relevant features, particularly due to fewer training epochs and lack of sophisticated regularization.

ResNet18 converged well on the training, but generalization on the validation and test was problematic, probably owing to lack of discriminative capability for the complex patterns of radio-opaque lesions.

Table 5. Comparative evaluation of CNN models on the Caries-Spectra test set

Таблица 5. Сравнительная оценка моделей CNN на тестовом наборе Caries-Spectra

AI Model	Test Accuracy, %	Precision, %	Recall, %	F1 Score, %
ResNet18	98.94	98.96	98.94	98.94
EfficientNet-B0	99.74	99.74	99.74	99.74
MobileNetV3	98.68	98.68	98.68	98.68

Table 6. Epoch-wise training and validation performance of ResNet18 on the Periapical Lesions test set

Таблица 6. Результаты обучения и проверки ResNet18 на тестовом наборе данных с периапикальными поражениями

Epoch. No	Loss	Train Acc, %	Val Acc, %
Epoch 1	313.1151	57.13	58.71
Epoch 2	259.0123	64.93	60.41
Epoch 3	156.4190	80.84	64.32
Epoch 4	65.5993	93.21	63.30
Epoch 5	37.1525	96.33	59.64
Epoch 6	29.4193	96.93	63.98
🔴 Early stopping at epoch 6 (no improvement for 3 epochs)			

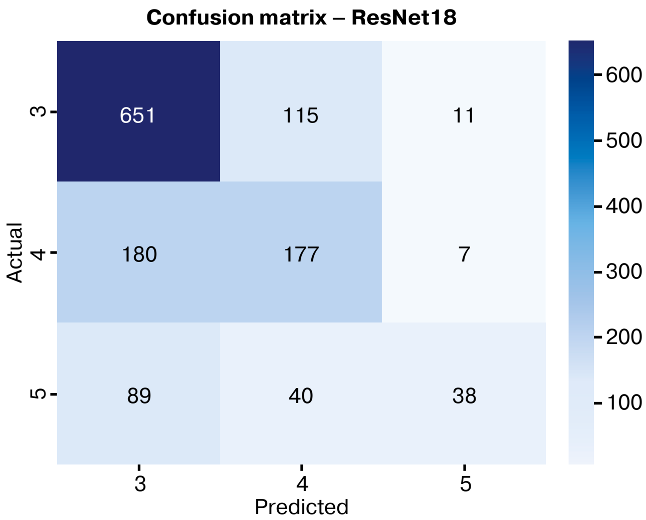

Fig. 5. Confusion matrix of ResNet18 on the periapical lesions test Set

Рис. 5. Матрица неточностей ResNet18 на тестовом наборе данных с перифокальными поражениями

Table 7. Epoch-wise training and validation performance of EfficientNet-B0 on the Periapical lesions test set

Таблица 7. Результаты обучения и проверки EfficientNet-B0 на тестовом наборе данных с периапикальными поражениями по эпохам

Epoch. No	Loss	Train Acc, %	Val Acc, %
Epoch 1	309.1451	58.89	59.47
Epoch 2	276.6413	61.88	58.88
Epoch 3	231.8796	69.73	61.17
Epoch 4	168.7719	79.76	59.90
Epoch 5	113.4400	87.42	66.61
Epoch 6	76.2439	91.82	71.11
Epoch 7	61.9546	93.24	70.60
Epoch 8	46.5596	95.09	70.01
Epoch 9	43.2456	95.42	71.62
Epoch 10	37.1156	96.48	70.18

The architecture of EfficientNet-b0 gives generalization superior to ResNet18 and stability on training data compared to other architectures, for the dataset of Periapical Lesions. As training data would affix between the first and tenth epochs of training, the aforementioned average accuracy of training, for instance, would rise regularly from 58.89% up to 96.48%, with an apparently upsurging validation accuracy from 59.47% toward 71.62%. Thus, it implicates better convergence and less overfitting in the training stage especially within later epochs (Table 7).

This enhancement is further backed by the confusion matrix in Figure 6 which demonstrates clearer classification patterns across all three PAI categories as indicated:

1. For PAI 3, the model has correctly classified 602 samples, though it still misclassifies 159 as PAI 4, thereby revealing some confusion with intermediate cases.

2. For PAI 4, there are performance numbers: 241 were classified accurate while very few misclassified.

3. For PAI 5, 68 predictions out of total are correct while confusion with PAI 3 and 4 trips along as differentiation between the bad discoveries proves to be much harder with overlapping clinical features.

In comparison to ResNet18, EfficientNet-B0 serves a better flexibility in synergetic training and validation performance, the compound scaling has probably increased its ability to learn multi-scale lesion features, an essence being important for detecting radiographically subtle differences within PAI 3 and 4 categories.

EfficientNet-B0 performed best and returned the highest test-time reliability, especially in moderate grades of periapical lesion identification. Interclass level of confusion still exists between classes, more obviously between categories of grade 5 and lower, which calls for improved feature discrimination, or perhaps balancing the dataset.

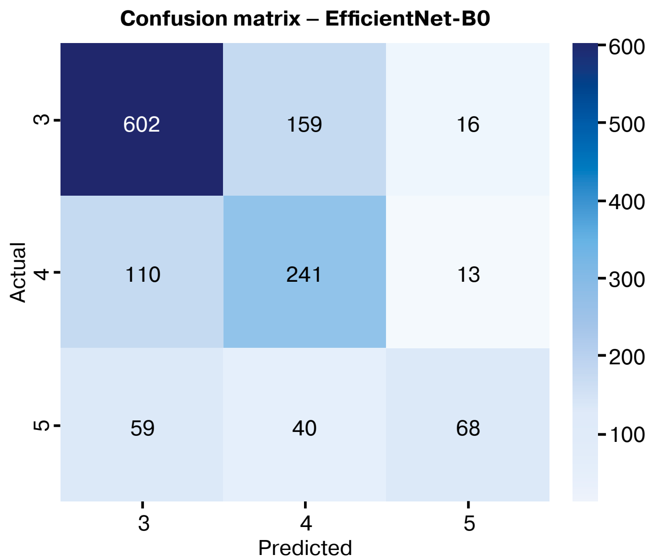

Fig. 6. Confusion Matrix of EfficientNet-B0 on the Periapical Lesions Test Set

Рис. 6. Матрица неточностей EfficientNet-B0 на тестовом наборе данных с периапикальными поражениями

MobileNetV3 was applied for periapical lesions dataset, it yielded a fair yet slightly inconsistent performance. The improved result is shown in Table 8, which reflects training accuracy that steadily increases reach to the maximum of 97% by epoch 10 from 59.06%. It shows good optimization though the validation accuracy varies in epochs, between 54.72% and 62.87%, a neutral direction is not expressed. This indicates that there is difficulty in generalizing the model even though it has converged to the training data evidence.

The confusion matrix (Fig. 7) sheds more insight into the classification behavior. For instance, PAI 3 was yet again the most accurately classified category (547 correct predictions). However, a lot of confusion is found with PAI 4 and PAI 5. For instance, there were 175 cases of PAI 3 wrongly predicted as PAI 4, whereas 55 of these classes were misidentified as PAI 5. PAI 4 shows moderate accuracy (196 predictions correct), but with misclassification among both adjacent classes. PAI 5 is problematic as well, with a distribution that almost evenly split at the average of correct predictions (57) to misclassifications (60 as PAI 3, 50 as PAI 4).

Table 8. Epoch-wise training and validation performance of MobileNetV3 on the periapical lesions test set

Таблица 8. Результаты обучения и проверки MobileNetV3 на тестовом наборе данных с периапикальными поражениями

Epoch. No	Loss	Train Acc, %	Val Acc, %
Epoch 1	307.6318	59.06	59.22
Epoch 2	288.7250	60.78	59.30
Epoch 3	262.6076	64.98	54.72
Epoch 4	220.9292	71.48	57.18
Epoch 5	165.3464	79.58	60.49
Epoch 6	116.3207	86.60	56.67
Epoch 7	77.4871	91.37	62.87
Epoch 8	55.1178	94.01	60.49
Epoch 9	36.3399	96.36	62.62
Epoch 10	29.2710	97.00	62.11

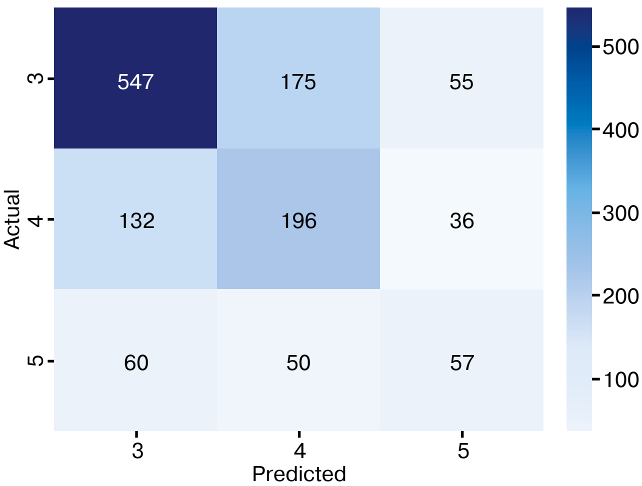

Fig. 7. Confusion Matrix of MobileNetV3 on the periapical lesions test set

Рис. 7. Матрица ошибок MobileNetV3 на тестовом наборе данных с периапикальными поражениями

Though MobileNetV3 is a lightweight architecture designed for high speed and mobile applications, it might not possess sufficient representational depth to capture the subtle differences between complex radiographic features, especially for more severe lesion classes. In conclusion, however high MobileNetV3 might have gotten in terms of training accuracy, the validation would not hold true as well as the configured disparity from the confusion matrix would indicate that it is weaker than EfficientNet-B0 and similar to ResNet18, in such cases. Nonetheless, the computational efficiency exhibited makes it a valuable candidate for deploying systems where real-time inference takes precedence over very marginal accuracy losses.

Table 9 presents the comparative evaluation among the three CNNs – ResNet18, EfficientNet-B0, and MobileNetV3 – on the Periapical Lesions test set. It shows that there is a large difference in the classification effectiveness between models when it comes to precision, recall, and F1-score.

The performance of EfficientNet-B0 is particularly high among all these models, achieving the maximum test accuracy of 69.65%, precision of 0.6765, recall of 0.6147, and F1-score of 0.6308. All these results confirm that the current model outperformed others in generalization and identification between PAI classes, especially in the presence of overlapping radiographic traits and highly ambiguous inter-class differences.

ResNet18 has had a comparable accuracy of 66.21% but has significantly low recall (0.5172) and F1-score (0.5389), indicating that it appears to misclassify a higher proportion of true positive samples. It indicated a moderate precision (0.6398) that demonstrates the correctness of many predictions, thus failing to detect several other cases with lesions, especially in the severe categories.

MobileNetV3, which is appropriate in terms of computation efficiency, gave the weakest results having the test accuracy of 61.16%. MobileNetV3 scored the lowest under all performance measures. Its precision (0.5303) and F1-score (0.5276) seem to reflect its limited discrimination capacity, which could be due to its lightweight features and shallow depth of feature extraction with less use for fine grained medical classification tasks.

Table 9. Comparative evaluation of CNN models on the periapical lesions test set

Таблица 9. Сравнительная оценка моделей сверточных нейронных сетей на тестовом наборе данных с периапикальными поражениями

AI Model	Test Accuracy, %	Precision, %	Recall, %	F1 Score, %
ResNet18	66.21	63.98	51.72	53.89
EfficientNet-B0	69.65	67.65	61.47	63.08
MobileNetV3	61.16	53.03	52.79	52.76

In summary, the best performance in terms of balance and reliability was achieved by EfficientNet-B0, making it the most competitive model for periapical lesion detection. In terms of speed and simplicity in architecture, ResNet18 is a good contender as well. MobileNetV3 best fits those whose priority is speed of inference and use of fewer resources over the precision of classifications.

CONCLUSION

In this study, the development of deep learning models for the automated classification of two dental conditions, namely enamel caries and periapical lesions, was undertaken. A total of two publicly available radiographic datasets were used. A consistent methodology was chosen to train and validate the three different CNN architectures: ResNet18, EfficientNet-B0, and MobileNetV3, such that a distinguishable fair comparison could generate the most valid results.

From these experimental results, it is clear that all considered models could learn meaningful features from dental images. However, in terms of the metrics used to consider generalization ability applied to both datasets, uses EfficientNet-B0 stood in prestige over its counterparts. Test accuracy of 99.74% on the Caries-Spectra dataset and 69.65% on the periapical lesions dataset display its capabilities of generalization. On the other hand, ResNet18 was not as advanced but gave due respect to the performance and ease of creating training. Also, MobileNetV3 was found to be computationally efficient, but its classification performance was poor-worse in sophisticated periapical cases.

The conclusion emphasizes the need to architect models according to the complexity of the diagnostic task in hand and the characteristics of their training dataset. The usage of lightweight models, such as MobileNetV3, would constitute a fair advantage in real-time tasking applications or where resources are scarce. On the contrary, where the main emphasis and concern are on accuracy would even call for the use of models like the EfficientNet-B0 in the diagnosis in health care settings.

Future research directions may open ensembles for models, means of attention, and frameworks for explainability to improve diagnostic framework reliability and clinician trust in AI systems. Likewise, the increase of new data types and magnitudes for training datasets, especially for the less represented lesion categories, would give a lot in strengthening the more robust and generalizable performance for real-world dental practice.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. Gliga A., Imre M., Grandini S., Marruganti C., Gaeta C., Bodnar D. et al. The limitations of periapical X-ray assessment in endodontic diagnosis – A systematic review. *J Clin Med*. 2023;12(14):4647. https://doi.org/10.3390/jcm12144647
2. Ba-Hattab R., Barhom N., Osman S.A.A., Naceur I., Odeh A., Asad A. et al. Detection of periapical lesions on panoramic radiographs using deep learning. *Appl Sci*. 2023;13(3):1516. https://doi.org/10.3390/app13031516
3. Song I.-S., Shin H.-K., Kang J.-H., Kim J.-E., Huh K.-H., Yi W.-J. et al. Deep learning-based apical lesion segmentation from panoramic radiographs. *Imaging Sci Dent*. 2022;52(4):351–357. https://doi.org/10.5624/isd.20220078
4. Panyarak W., Wantanajittikul K., Suttapak W., Charuakkra A., Prapayasatok S. Feasibility of deep learning for dental caries classification in bitewing radiographs based on the ICCMS™ radiographic scoring system. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2023;135(2):272–281. https://doi.org/10.1016/j.oooo.2022.06.012
5. Reyes L.T., Knorst J.K., Ortiz F.R., Ardenghi T.M. Machine learning in the diagnosis and prognostic prediction of dental caries: A systematic review. *Caries Res*. 2022;56(3):161–170. https://doi.org/10.1159/000524167
6. Imak A., Celebi A., Siddique K., Turkoglu M., Sengur A., Salam I. Dental caries detection using score-based multi-input deep convolutional neural network. *IEEE Access*. 2022;10:18320–18329. https://doi.org/10.1109/ACCESS.2022.3150358
7. Lian L., Zhu T., Zhu F., Zhu H. Deep learning for caries detection and classification. *Diagnostics*. 2021;11(9):1672. https://doi.org/10.3390/diagnostics11091672
8. Qayyum A., Tahir A., Butt M.A., Luke A., Abbas H.T., Qadir J. et al. Dental caries detection using a semi-supervised learning approach. *Sci Rep*. 2023;13(1):749. https://doi.org/10.1038/s41598-023-27808-9
9. Chen I.D.S., Yang C.M., Chen M.J., Chen M.C., Weng R.M., Yeh C.H. Deep learning-based recognition of periodontitis and dental caries in dental X-ray images. *Bioengineering*. 2023;10(8):911. https://doi.org/10.3390/bioengineering10080911
10. Kang I.A., Njimbouom S.N., Kim J.D. Optimal feature selection-based dental caries prediction model using machine learning for decision support system. *Bioengineering*. 2023;10(2):245. https://doi.org/10.3390/bioengineering10020245
11. Esmaeilyfard R., Bonyadifard H., Paknahad M. Dental caries detection and classification in cbct images using deep learning. *Int Dent J*. 2024;74(2):328–334. https://doi.org/10.1016/j.identj.2023.10.003
12. Çelik B., Savaştaer E.F., Kaya H.I., Çelik M.E. The role of deep learning for periapical lesion detection on panoramic radiographs. *Dentomaxillofac Radiol*. 2023;52(8):20230118. https://doi.org/10.1259/dmfr.20230118
13. Liu Y., Xia K., Cen Y., Ying S., Zhao Z. Artificial intelligence for caries detection: a novel diagnostic tool using deep learning algorithms. *Oral Radiol*. 2024;40(3):375–384. https://doi.org/10.1007/s11282-024-00741-x
14. Tibúrcio-Machado C.S., Michelon C., Zanatta F.B., Gomes M.S., Marin J.A., Bier C.A. The global prevalence of apical periodontitis: a systematic review and meta-analysis. *Int Endod J*. 2021;54(5):712–735. https://doi.org/10.1111/iej.13467
15. Himel G.M.S., Islam M.M., Hannan U.H. *Caries-Spectra: A dataset of enamel caries*. 2023. https://doi.org/10.17632/9jnf2jvghy.2
16. Do H.V., Vo T.N.N., Nguyen P.T., Luong T.H.L., Cu N.G, Le H.S. A Dataset of apical periodontitis lesions in panoramic radiographs for deep-learning-based classification and detection. *Data Brief*. 2024;54:110486. https://doi.org/10.1016/j.dib.2024.110486

INFORMATION ABOUT THE AUTHORS

Mohammed Al-Sabri – Department of Restorative and Esthetic Dentistry, Faculty of Dentistry, Sana'a University, Sana'a, Yemen; https://orcid.org/0009-0005-6062-3458

Ayman Abduh Qaaed Mohammed Alsabry – Department of Computer Science, International University of Technology Twintech, Sana'a, Yemen; https://orcid.org/0000-0003-4835-4054

Naibah Ali Yahya Rajeh – Department of Restorative and Esthetic Dentistry, Faculty of Dentistry, Sana'a University, Sana'a, Yemen

Arham Mohammed Abdulqawi Alkherbash – Department of Restorative and Esthetic Dentistry, Faculty of Dentistry, Sana'a University, Sana'a, Yemen; https://orcid.org/0009-0007-8922-2476

Raghda Ameen Alnozaily – Department of Restorative and Esthetic Dentistry, Faculty of Dentistry, Sana'a University, Sana'a, Yemen; https://orcid.org/0009-0001-4960-5137

Ebtehal Mogahed Sharhan – Department of Restorative and Esthetic Dentistry, Faculty of Dentistry, Sana'a University, Sana'a, Yemen

ИНФОРМАЦИЯ ОБ АВТОРАХ

Мохаммед Аль-Сабри – кафедра реставрационной и эстетической стоматологии, стоматологический факультет, Университет Саны, Сана, Йемен; https://orcid.org/0009-0005-6062-3458

Айман Аль-Сабри – кафедра компьютерных наук, Международный технологический университет Twintech, Сана, Йемен; https://orcid.org/0000-0003-4835-4054

Наиба Али Яхья Раджех – кафедра реставрационной и эстетической стоматологии, стоматологический факультет, Университет Саны, Сана, Йемен

Архам Мохаммед Абдулкави Аль-Хирбаш – кафедра реставрационной и эстетической стоматологии, стоматологический факультет, Университет Саны, Сана, Йемен; https://orcid.org/0009-0007-8922-2476

Рагда Амин Аль-Нозайли – кафедра реставрационной и эстетической стоматологии, стоматологический факультет, Университет Саны, Сана, Йемен; https://orcid.org/0009-0001-4960-5137

Эбтехаль Могахед Шархан – кафедра реставрационной и эстетической стоматологии, стоматологический факультет, Университет Саны, Сана, Йемен

AUTHOR'S CONTRIBUTION

All the authors made equal contributions to the publication preparation in terms of the idea and design of the article; data collection; critical revision of the article in terms of significant intellectual content and final approval of the version of the article for publication.

ВКЛАД АВТОРОВ

Все авторы внесли равноценный вклад в подготовку публикации в части замысла и дизайна исследования; сбора данных; критического пересмотра статьи в части значимого интеллектуального содержания и окончательного одобрения варианта статьи для опубликования.