INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Performance Analysis of ResNet-50, EfficientNet B4, and MobileNet V2 for Dental Caries Classification

Maulisa Oktiana [a], Maya Fitria [a,*], Hafidh Habibie [a], Khairun Saddami [a], Yasmina Elma [a], Handika Rahayu [a], Subhan Janura [a], Rizkika Putri [b], Rizki Novita [c]

[a] *Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia*
[b] *Polyclinic of Dental and Oral, Regional General Hospital dr. Zainoel Abidin (RSUDZA), Banda Aceh, Indonesia*
[c] *Faculty of Dentistry, Universitas Syiah Kuala, Banda Aceh, Indonesia*
Corresponding author: [*]mayafitria@usk.ac.id

Abstract*—Dental caries is a disease that can damage the tooth structure, leading to cavities. However, the symptoms of decay are often unnoticed until a toothache develops, which can lead to cavities. Severe cases of dental caries, which cause cavities, require direct inspection. With early diagnosis of decay, preventive measures and treatment steps can be taken sooner. This helps prevent further development of tooth damage and reduces the need for more invasive restorative treatments. Therefore, this study develops a classification model for dental caries using Convolutional Neural Network architectures, namely Residual Network (ResNet)-50, EfficientNet B4, and MobileNet V2, where images are classified into two classes: decayed and normal teeth. The dataset comprises 700 dental images. Model evaluation includes accuracy, precision, recall, and F1-score. Experimental results indicate that the EfficientNet B4 model outperforms MobileNet V2 and Residual Network (ResNet)-50 architectures in classifying dental caries. Specifically, EfficientNet B4, trained on the lower occlusal dataset with a learning rate of 10^{-5}, batch size of 16, and 200 epochs, achieves an accuracy of 77%, precision of 78%, recall of 77%, and F1-score of 77%. When the model is trained using a learning rate of 10^{-4}, it achieves a better accuracy of 81%. The superiorperformance of EfficientNet B4 comes from its depthwise separable convolutions and squeeze-and-excitation (SE) blocks, which improve how features are extracted. These results show the promise of deep learning models in identifying dental cavities, leading to better diagnostics for oral health. However, an expanding dataset diversity is needed in further research for model generalization enhancement, aiming to ensure model's robustness to apply in clinical settings.

***Keywords*—Caries classification; deep learning, CNN; oral health diagnosis; medical image processing; caries detection.**

Manuscript received 26 Jan. 2025; revised 8 Aug. 2025; accepted 24 Sep. 2025. Date of publication 31 Jan. 2026.

I. INTRODUCTION

Caries are one of the common chronic dental diseases that infect hard dental tissues [1]-[4]. This disease progressively and cumulatively causes demineralization in teeth by removing mineral ions from tooth enamel [5]. The main factors that can cause dental caries are the host (teeth and saliva), the substrate (food), bacteria or microorganisms, and time [6], [7]**.** Caries can lead to pain and infections, resulting in inflammation and, if left untreated, eventually causing swelling [8], [9]. Several clinical and epidemiological studies have proposed that dental caries can potentially be a risk factor that can lead to cognitive disorders and cardiovascular diseases [10]-[12]. In Indonesia, the population experiencing oral and dental health problems, including dental cavities and damaged teeth, accounts for 88% of the prevalence of severe cases [9], [13]-[15]. Therefore, it is essential to conduct regular dental check-ups to address dental caries promptly and to plan curative, preventive, and rehabilitative care to avoid more serious dental disorders [11], [16].

Dentists generally rely on manual inspection as a standard procedure for evaluating dental care. This process involves examining and counting decayed, missing, and restored teeth. According to the World Health Organization (WHO), this assessment can also be used to evaluate caries status within a population by calculating the Decayed, Missing, and Filled Teeth (DMF-T) index. Evaluating caries status within a population generally involves calculating the DMF-T index, which stands for Decayed, Missing, and Filled Teeth. This serves preventive, curative, and rehabilitative purposes and helps determine a community's overall dental health [5], [16]. According to interviews with dentists at Zainoel Abidin

Regional Hospital in Aceh, performing population-wide dental caries assessments using the DMF-T index is a time-intensive task. The procedure requires a thorough inspection for every tooth indicated as decayed, missing, or repaired. Subsequently, tooth conditions are recorded on an odontogram and used to calculate the DMF-T index.

Several scientific studies have focused on detecting dental caries, particularly using Convolutional Neural Networks (CNNs). Baydakar et al. [17] conducted a survey using the U-Net and VGG-16 algorithms to detect dental caries in a radiographic bitewing dataset, achieving a detection accuracy of 48%. Similarly, Kumari et al. analyzed radiographic bitewing dental images, employing image enhancement techniques such as CLAHE and FOC-KCC, followed by training with M-ResNet-RNN. However, evaluating dental cavities at the community level using radiographic images is not feasible due to the potential harm from X-ray exposure [18]. To minimize the dependency on X-rays for automated dental caries detection, Fitria et al. investigated the use of oral medical photographs for detection using a CNN architecture [19]. Their research utilized images from five perspectives of dental clinical data: anterior, right, and left buccal, and upper and lower occlusal. Using the ResNet-50 architecture, the model was trained on 1,400 augmented images. However, the performance was suboptimal due to the significant dataset variability. Additionally, missing and filled teeth needed to be accounted for, along with the classification of carious and non-carious cases**.** The presence of other anatomical features in the images, such as gums, healthy teeth, and lips, increased the complexity of the system's cavity-detection task, resulting in lower accuracy.

Novita et. al. in [20] assessed the effectiveness of deep learning models compared to clinical examinations in detecting DMF-T. The study included 50 patients, where experienced dentists conducted clinical examinations, and oral photos of the same patients were analyzed using a deep learning model based on the YOLO V5 architecture. The findings showed that the deep learning model performed comparably to clinical examinations, with no significant difference in DMF-T detection. However, the deep learning model slightly underperformed at identifying missing and filled teeth, due to challenges such as lighting conditions in the images and difficulty distinguishing between healthy teeth and those restored with tooth-colored materials. Deep learning shows promise in detecting DMF-T. Improvements to the dataset and image enhancement techniques are needed to improve accuracy further. The research suggests that with further development, deep learning models could serve as effective tools in dental health assessments, reducing the need for manual clinical examinations.

Fitria et al. [21] compare two popular object detection architectures, YOLOv5 and YOLOv8, for detecting decayed, missing, and filled teeth. The dataset used consists of 294 clinical dental images, augmented to enhance diversity. Experimental results showed that YOLOv5 and YOLOv8 models demonstrated high precision, recall, and mean average precision (mAP) in detecting dental conditions, with YOLOv5l and YOLOv8m versions performing best. YOLOv8m achieved a slightly higher mAP (90.6%) than YOLOv5l (90.4%), but it required longer training time (almost 5 hours). YOLOv5s, while slightly less accurate, was the fastest, making it suitable for real-time applications. YOLOv8 models also showed more stable training curves, indicating better training dynamics. Both YOLOv5 and YOLOv8 are promising for automated dental caries detection, but trade-offs between accuracy and computation time should be considered for practical use.

Although the above techniques achieve high accuracy, they are still not fully satisfactory, making sentiment analysis a continually evolving research topic. Given the wide variety of approaches to dataset enhancement, parameter tuning, and other factors, there remains a need for further research to evaluate these methods and better understand their limitations and challenges in dental caries classification. This paper contributes to the field by assessing the most popular deep learning methods to determine a configuration or a specific setup and tuning. This study evaluates the performance of several deep learning architectures for dental caries classification, including ResNet-50, EfficientNet B4, and MobileNet V2. The motivation for this study was to establish a unified framework for comparing these methods, highlighting the strengths and weaknesses of each configuration.

Previous studies on dental caries detection have primarily relied on radiographic images (e.g., bitewing X-ray images) or limited photographic datasets, often yielding suboptimal accuracy due to image variability, the presence of visual noise (e.g., lips and gums), and inconsistent lighting conditions. Moreover, existing approaches typically focus on a single model architecture, making it difficult to directly compare performance across different deep learning frameworks under a unified experimental setup. This creates a gap in understanding which architecture is most suitable for intraoral photographic caries detection in terms of accuracy, stability, and computational efficiency. To address this gap, this study evaluates and compares three widely used CNN architectures:

- ResNet-50, known for its residual connections that mitigate the vanishing gradient problem and allow training of deeper networks
- EfficientNet B4, recognized for its compound scaling method that balances depth, width, and resolution to achieve high accuracy with efficient computation
- MobileNet V2, optimized for lightweight, real-time applications through depthwise separable convolutions and inverted residuals.

By systematically comparing these architectures using the same dataset and preprocessing pipeline, this work aims to identify the most effective model for robust and accurate dental caries classification in clinical photographs.

The rest of this paper is organized as follows: Section II describes the materials and methods, including dataset preprocessing, model architecture, and evaluation metrics. Section III presents the experimental results and discussion, highlighting the performance comparison of ResNet-50, EfficientNet B4, and MobileNet V2. Section IV concludes the study and provides explicit suggestions for future work.

II. Materials and Methods

Figure 1 depicts the proposed framework, which is divided into four stages: 1) Problem analysis; 2) Dataset pre-processing; 3) Model development and verification; and 4) Results analysis. Each of these stages is explained as follows.

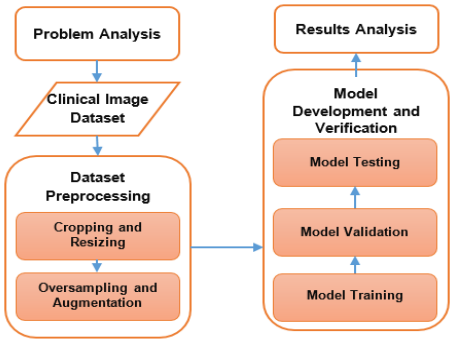


Fig. 1 Research Methodology

A. *Problem Analysis and Dataset Pre-processing*

The problem analysis stage involved reviewing related studies, assessing their drawbacks, and evaluating the latest technologies. Additionally, an interview process was conducted with experts to obtain information on dental caries, its risks, techniques, and how to determine whether someone is affected by it. In this study, the experiment was conducted using the caries dataset acquired by Fitria et al. [21]. The dataset consists of 700 dental images, divided equally into the caries and regular classes. Figure 2 showcases examples of the five tooth surfaces used as datasets in this study: labial represents the front view of upper and lower incisors (a), right buccal indicates right side of molar tooth (b), left buccal represents left side of molar teeth (c), lower occlusal represents the chewing surface of the lower teeth (d), and upper occlusal which is the chewing surface of the upper teeth (e). Each perspective provides a different view of the tooth surface condition, enabling the model to recognize caries patterns better. This variety of opinions is essential to improve the model's accuracy and generalization in distinguishing between healthy and decayed teeth. To obtain optimal images for training the caries classification model, a series of pre-processing steps, including cropping, resizing, oversampling, and augmentation, was performed to improve data quality and balance.

First, cropping was performed manually to remove irrelevant areas such as cheeks, lips, or backgrounds that do not contain important tooth information. This step helped focus on critical areas, such as tooth surfaces and caries, which are the target classification targets. Second, the large cropped images (6000x4000 pixels) were resized to 224x224 pixels (Figure 3) using bilinear interpolation to fit the standard CNN input format and reduce computational complexity during training. Third, oversampling was performed to address class imbalance, a condition in which the data distribution between the "caries" and "normal" classes is unbalanced. This technique involves randomly duplicating images from the minority class (e.g., the caries class) until the number of images in that class equals those in the other classes, thereby balancing the data distribution. Finally, data augmentation was applied using image rotation at 45°, 90°, and 180°, and brightness adjustment. This augmentation aims to increase visual diversity in the dataset, enable the model to recognize caries patterns from various viewpoints and lighting conditions, and reduce the risk of overfitting. The sample of the image rotation result is illustrated in Figure 4.

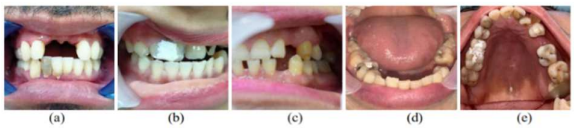


Fig. 2 Sample of the dataset shows on five sides

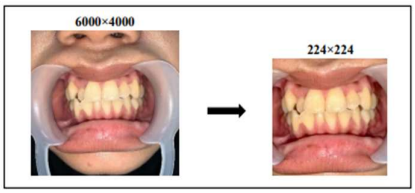


Fig. 3 Image Resizing Process

Fig. 4 Sample of Image Rotation

B. *Model Development and Verification*

The dataset used in this work is distributed across training, validation, and test sets, as shown in Table I. The model was designed to detect caries condition (normal or caries) using ResNet-50, EfficientNet B4, and MobileNet V2. The multiparameters, such as epoch, batch size, optimizer, momentum, and learning rates, were also set and shown in Table II.

TABLE I
DISTRIBUTION OF THE DATASET

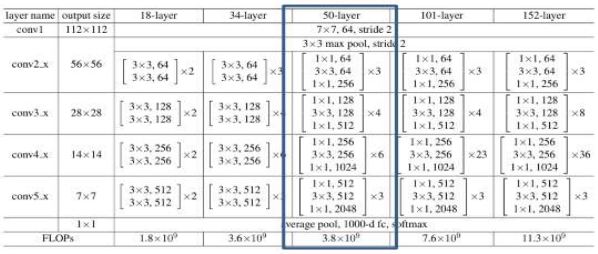
Dataset	Caries	Normal	Total
Training	210	210	420
Validation	70	70	140
Testing	70	70	140
Total	350	350	700

TABLE II
MULTIPARAMETER USED IN THE WORK

Batch Size	16, 32
Learning Rate	10^{-4}, 10^{-5}
Optimizer	Adam
Epoch	200

1) Residual Network (ResNet)-50: ResNet-50 is one of the CNN architectures that has 50 layers, containing 48 layers for convolutional, and the other two are for the MaxPool layer, Average Pool layer [22] as illustrated in Figure 5. ResNet-50 features shortcut connections, where input from the previous layer can be directly diverted to serve as input to the layer's output [23], [24]. This concept is also related to the vanishing gradient problem, which occurs when the gradient becomes smaller during backpropagation as the neural network deepens, thereby reducing the model's performance. The deeper a neural network, the more likely it is to encounter

diminishing gradients, resulting in decreased performance or accuracy [25].



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig. 2 ResNet-50 Convolutional Layer [26]

2) EfficientNet B4: EfficientNet B4 explored a scaling method for all dimensions that are the same as the combined coefficients, such as depth, width, and resolution. Therefore, EfficientNet B4 requires fewer computational resources than the Convolutional Neural Network (CNN) model and achieves higher accuracy [27], [28]. The number of layers can distinguish each EfficientNet architecture at each stage. Figure 6 shows the EfficientNet B4 architecture consisting of MBConv block layers, where the MBConv block is an inverted residual block applied using the compound scaling method through the baseline to prevent gradient loss [29]-[32].

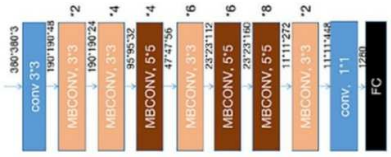


Fig. 3 EfficientNet B4 Architecture [32]

3) MobileNet V2: MobileNet V2 is specifically designed for mobile applications in a lightweight and efficient [33]. MobileNet V2 improves model performance on several tasks by combining two main components: depth-wise separable convolutions and inverted residuals. Depth-wise separable convolutions reduce the number of parameters and computations without sacrificing model accuracy, while inverted residuals improve overall efficiency and accuracy in MobileNet [34]-[37].

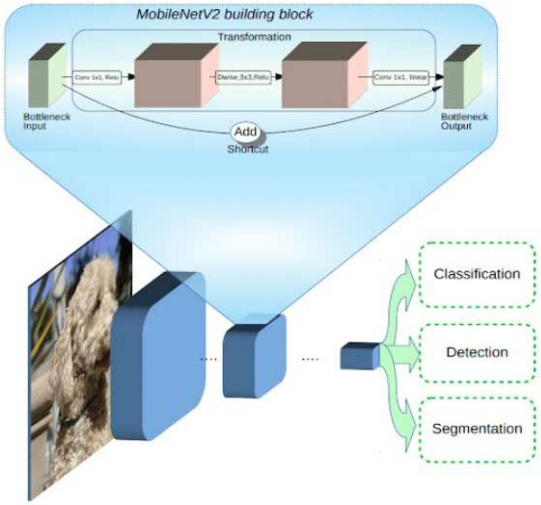


Fig. 4 MobileNet V2 Architecture [38]

In the bottleneck section of a model, input and output processes occur between models. Its function is to process the model's capacity to translate input from low-level concepts, such as pixels, to higher-level descriptors, such as image categories. *Using shortcuts between bottlenecks allows faster training and* better accuracy [38]. Figure 7 describes the MobileNet V2 architecture.

C. Evaluation Method

The evaluation of model performance includes analyzing accuracy, precision, recall, and the F1-score. Accuracy is defined as the ratio of the number of correct detections to the total number of detected images. Precision is defined as the ratio of correctly detected images, known as true positives (TP), to the total number of positive images, including false positives (FP). While precision indicates the model's accuracy in identifying the number of positive classes, recall describes how accurately the model detects actual positive instances. The recall is calculated as the ratio of correctly detected instances (TP) to all positive instances, including false negatives (FN). Higher precision and recall values result from increased TP counts and decreased FP or FN counts. F1-score combines precision and recall in a balanced way, giving an overall measure of the model's performance and indicating how well the model classifies the data. The accuracy, precision, recall, and F-1 score are obtained as equations 1-4.

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

$$precision = \frac{TP}{TP+\overrightarrow{\leftarrow}FP} \quad (2)$$

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

$$F1-score = \frac{2(precision\times recall)}{precision+recall} \quad (4)$$

III. Results and Discussion

Figures 8-12 describe the training results of ResNet-50, EfficientNet B4, and MobileNet V2 using a learning rate of 10^{-4}. Training curves for ResNet-50, EfficientNet B4, and MobileNet V2 across the five dental image perspectives, i.e., right buccal, left buccal, labial, upper occlusal, and lower occlusal, highlight several vital observations regarding overfitting, convergence behavior, and side-specific model performance. Analysis of the training results across five dental image perspectives reveals performance differences among the models, specifically in terms of side-specific differences. The lower occlusal side demonstrated the most optimal learning dynamics, with EfficientNet B4 achieving the highest validation accuracy and most stable convergence, indicating this view provides the most discriminative features for caries classification. The upper occlusal and right buccal views also produced good results, particularly with EfficientNet B4, showing steady accuracy and a low level of overfitting. On the other hand, the labial view faced many difficulties because of the interference from lips and gums, causing unstable validation lines and lower performance across all models. The left buccal side did moderately well but experienced overfitting issues with ResNet-50 and MobileNet V2. Overall, EfficientNet B4 outperformed the other architectures across all sides, maintaining a balance between fitting and generalization, whereas ResNet-50 and MobileNet

V2 were more susceptible to overfitting and training instability, particularly on image perspectives with more background noise.

While Figures 13-17 describe the training results of ResNet-50, EfficientNet B4, and MobileNet V2 using a learning rate of 10^{-5}. At a learning rate of 10^{-5}, MobileNet V2 on the right side showed decent fitting; however, towards the end of training, the model became unstable, leading to a rise in the validation loss curve and a drop in the validation accuracy curve. On the other hand, the EfficientNet B4 model performed the best, achieving a training accuracy of 0.75%, a training loss of 0.50%, a validation accuracy of 0.78%, and a validation loss of 0.48%. EfficientNet B4 demonstrated the most effective learning curve and remained quite stable in both loss and accuracy compared to the other models. This model is capable of learning from the information and patterns present in the training data. The outcomes indicate that the model performs well and can differentiate between caries and healthy images.

Learning rate variation plays a crucial role in determining the stability and effectiveness of deep learning model training. At a learning rate of 10^{-4}, the model training process is faster, resulting in high training accuracy, but in many cases, it leads to overfitting. This is particularly evident in ResNet-50 and MobileNet V2, where the models quickly memorized the training data but failed to maintain performance on the validation data. This effect is most noticeable on the labial and left buccal surfaces, where distractions from the lips and gums make it harder for the model to understand better. However, EfficientNet B4 still demonstrates relatively better stability at this learning rate, particularly on the lower and upper occlusal surfaces. Conversely, at a learning rate of 10^{-5}, the model's learning rate is slower, allowing for a more stable and smoother convergence process. This effect was most beneficial for EfficientNet B4, which achieved the highest validation accuracy of 78% with a low validation loss, indicating the model was able to absorb data patterns without overfitting. However, for MobileNet V2, a low learning rate led to instability in the final stages of training, particularly on the right buccal side, characterized by a sudden increase in validation loss and a decrease in validation accuracy. This indicates that setting the learning rate too low can lead to the model getting trapped in local minima or struggling to adjust the weights properly as training concludes.

We can conclude that a learning rate of 10^{-4} is suitable for models with lightweight architectures like MobileNet V2 that require fast training but are at risk of overfitting. Meanwhile, a learning rate of 10^{-5} is more suitable for complex architectures like EfficientNet B4, which utilizes an incremental training process to achieve better generalization and superior training stability.

Figure 18 shows the confusion matrix for EfficientNet B4 using two learning rates: 10^{-4} (left) and 10^{-5} (right). At 10^{-4}, the model achieved better balance with 81% accuracy, while at 10^{-5}, accuracy dropped slightly to 77% due to an increase in false negatives, indicating reduced sensitivity. Figure 19 presents the ResNet-50 results. With a learning rate of 10^{-4}, accuracy was lower at 74% and showed signs of overfitting. When reduced to 10^{-5}, performance improved to 77%, with better generalization and fewer classification errors. Figure 20 displays MobileNet V2 results. At 10^{-4}, accuracy reached 77% with balanced false predictions. However, at 10^{-5}, accuracy dropped to 73%, and instability appeared in the validation phase, suggesting sensitivity to low learning rates. Overall, the figures illustrate that learning rate impacts model performance. EfficientNet B4 consistently performs best, ResNet-50 benefits from a lower learning rate, while MobileNet V2 struggles at very low rates.

The performances of the ResNet-50, EfficientNet B4, and MobileNet V2 models are summarized in Table III. EfficientNet-B4 employs depthwise separable convolutions, dividing the typical convolution operation into a pair of more basic steps. This reduces the model's parameters and computational requirements, enhancing its efficiency while still capturing key image features. Furthermore, the EfficientNet B4 design includes SE blocks that modify feature maps by emphasizing the most important channels and reducing the weight of less significant ones. This allows the model to concentrate on the most crucial aspects of images, which helps it in enhancing classification accuracy. All these aspects make EfficientNet-B4 very effective at achieving a good balance between accuracy and efficiency, leading to strong performance in tasks related to caries classification.

TABLE III
TESTING RESULTS USING LOWER OCCLUSAL DATA

Lr	Architecture	Precision	Recall	F1-Score	Accuracy
10^{-4}	EfficientNet B4	0.80	0.81	0.81	0.81
	MobileNet V2	0.76	0.79	0.77	0.77
	ResNet-50	0.71	0.67	0.72	0.74
10^{-5}	EfficientNet B4	0.73	0.86	0.79	0.77
	MobileNet V2	0.72	0.76	0.74	0.73
	ResNet-50	0.74	0.84	0.79	0.77

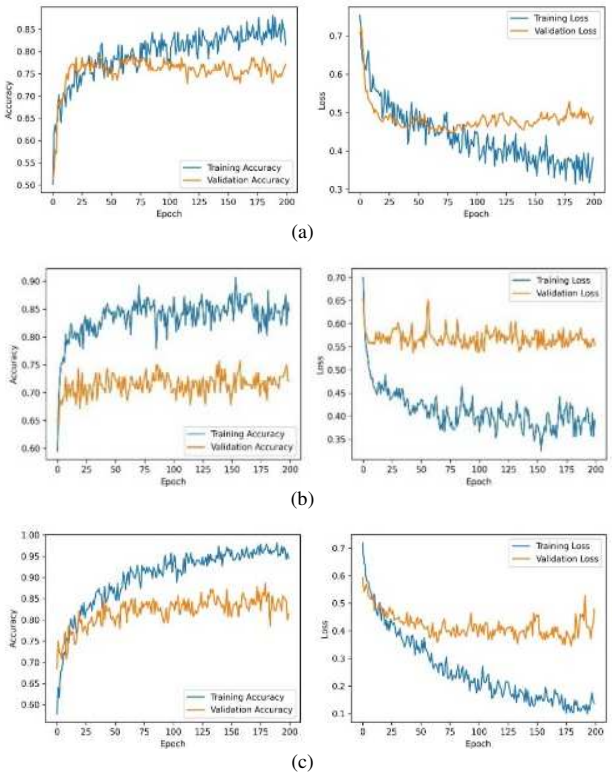
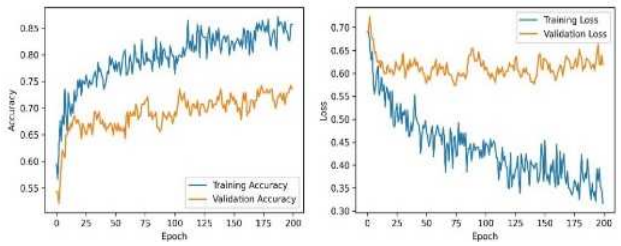


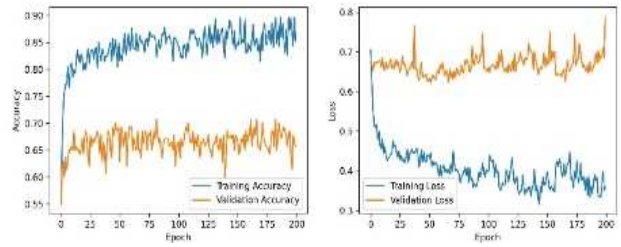
Fig. 5 Training results of Right Buccal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-4}

IV. CONCLUSIONS

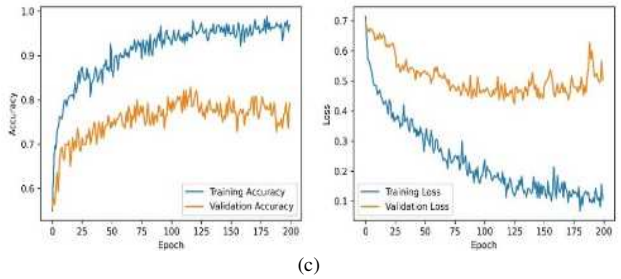
This study evaluates the performance of ResNet-50, EfficientNet B4, and MobileNet V2 for dental caries classification using intraoral clinical images. The results indicate that EfficientNet B4 consistently achieved superior performance, yielding the highest accuracy of 81%, and strong precision, recall, and F1-score, attributing to its efficient architecture that effectively balances model depth, width, and resolution. In contrast, ResNet-50 and MobileNet V2 showed limitations in generalizing to new data, as evidenced by overfitting and performance instability across different image perspectives, particularly under certain learning rate configurations. These models, while capable of high training accuracy, struggled with validation accuracy, indicating reduced robustness. In the future, it is recommended to enhance the dataset quality by applying background removal and more accurate annotation techniques. These improvements are expected to minimize noise, improve feature focus, and ultimately support better generalization and accuracy in caries detection tasks across different clinical conditions.



(a)

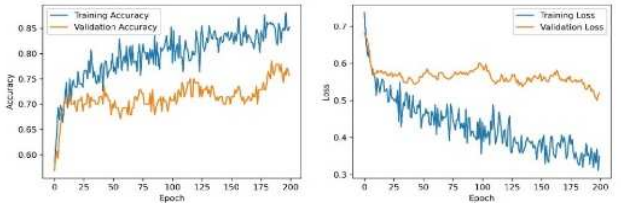


(b)

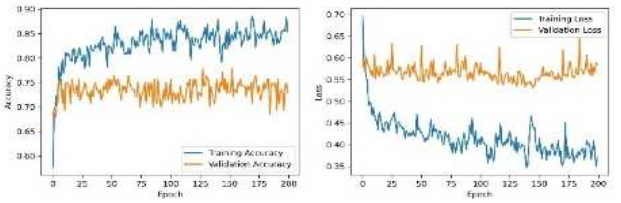


(c)

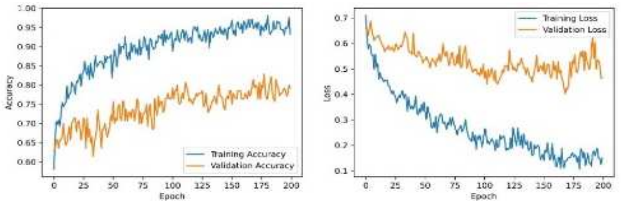
Fig. 6 Training results of Left Buccal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-4}



(a)

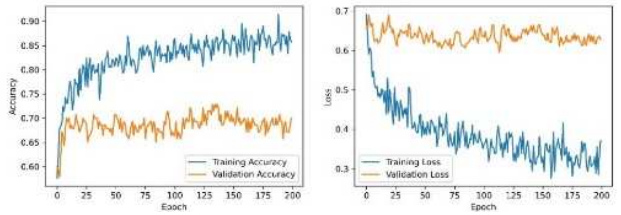


(b)

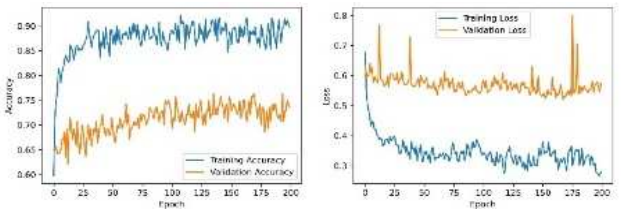


(c)

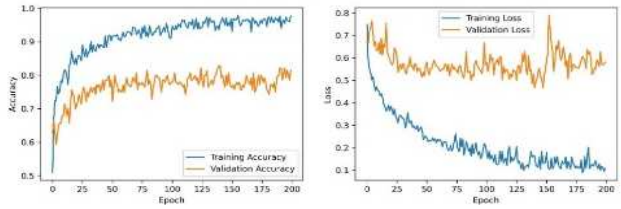
Fig. 7 Training results of Labial Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-4}



(a)



(b)



(c)

Fig. 8 Training results of Upper Occlusal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-4}

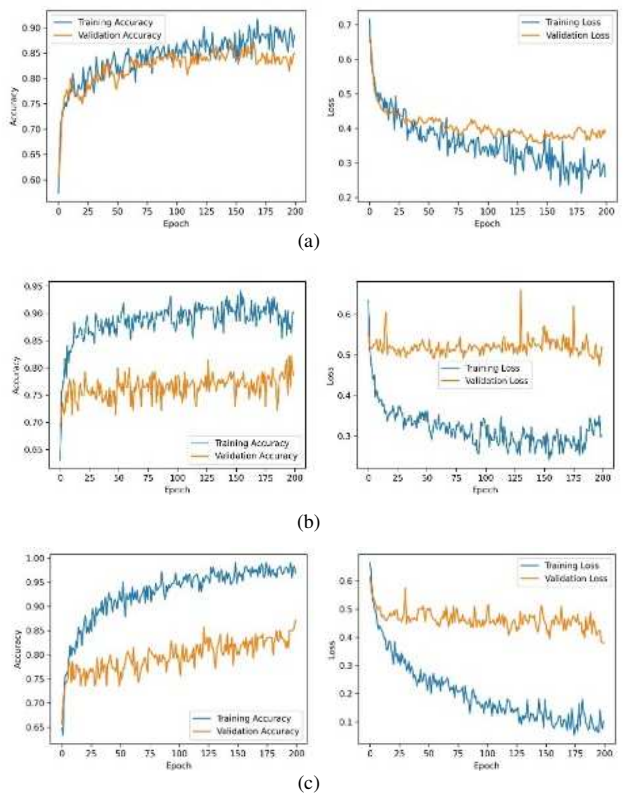


Fig. 9 Training results of Lower Occlusal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-4}

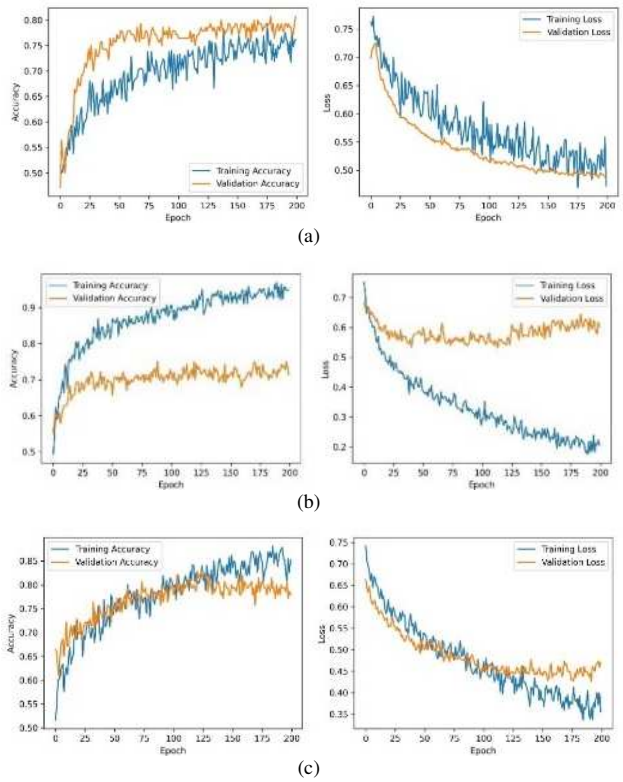


Fig. 10 Training results of Right Buccal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-5}

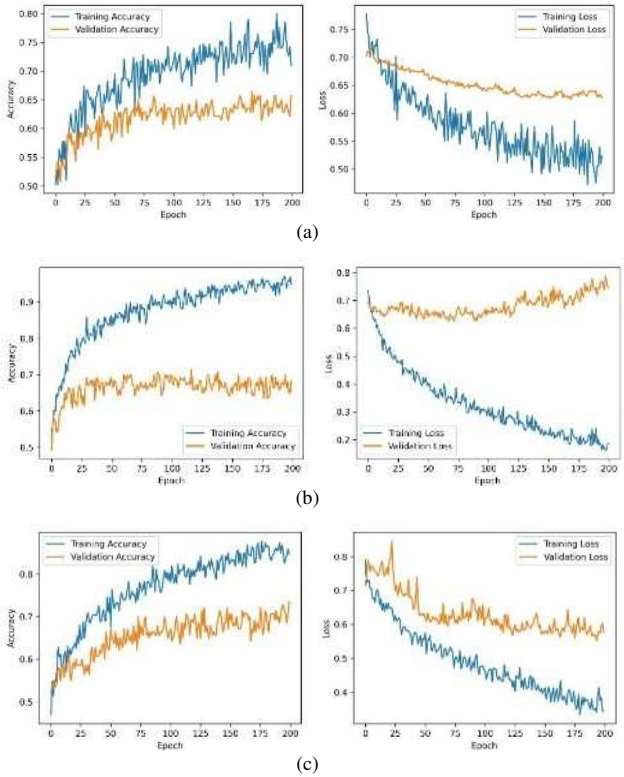


Fig. 11 Training results of Left Buccal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-5}

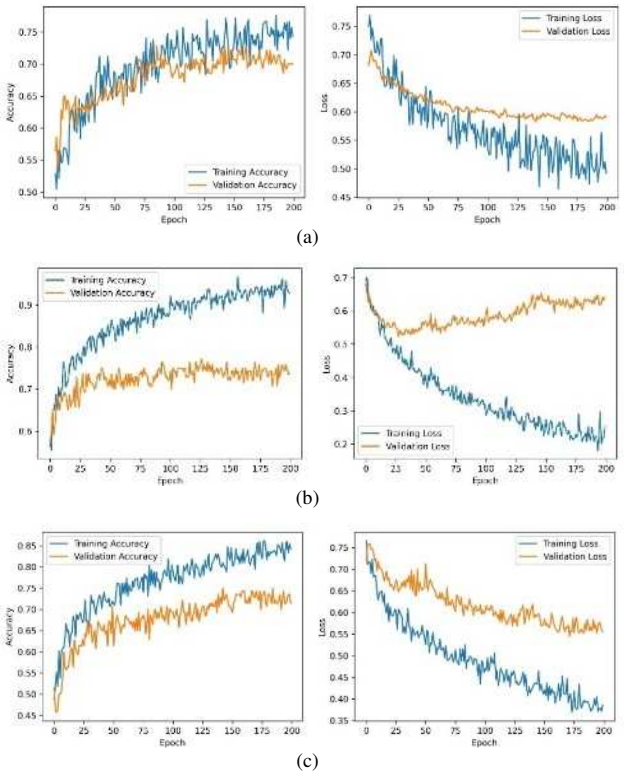
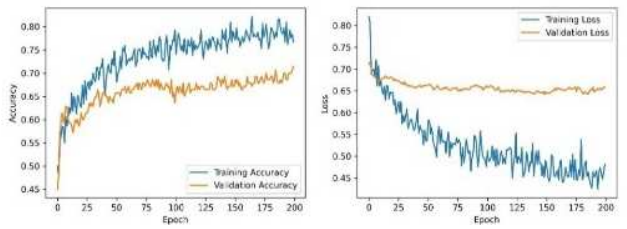
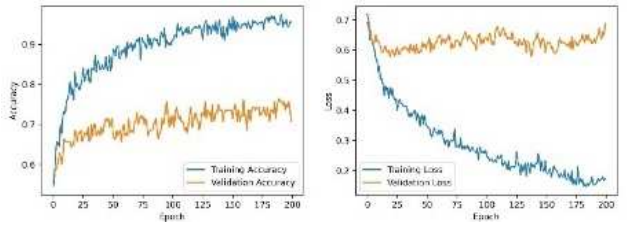


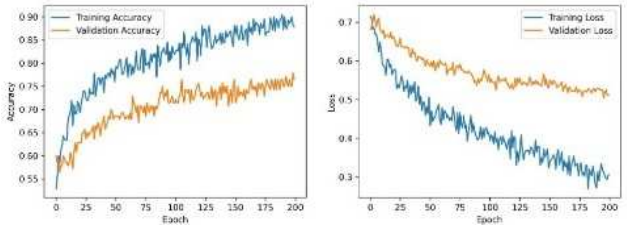
Fig. 12 Training results of Labial Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-5}



(a)

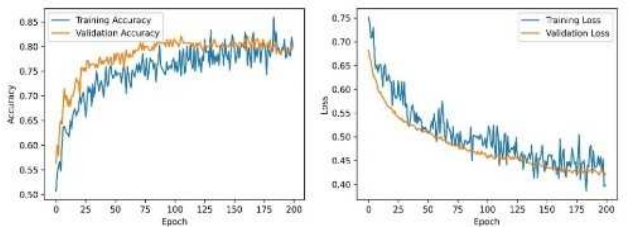


(b)

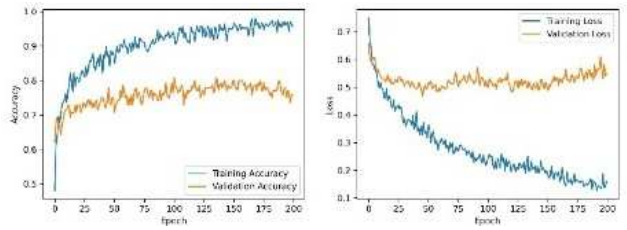


(c)

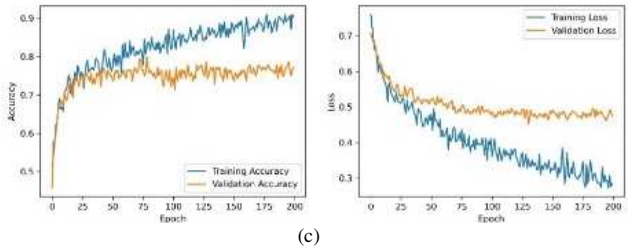
Fig. 13 Training results of Upper Occlusal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-5}



(a)



(b)



(c)

Fig. 14 Training results of Lower Occlusal Side Employing ResNet-50 (a), EfficientNet B4 (b), and MobileNet V2 (c) using learning rate 10^{-5}

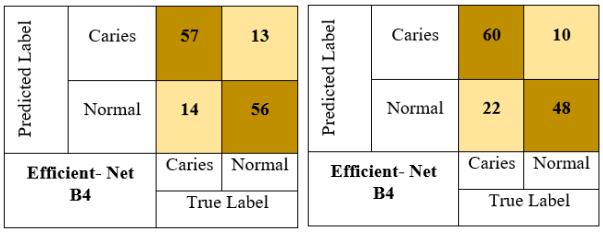


Fig. 15 Confusion Matrix for Lower Occlusal Dataset Using EfficientNet B4 with learning rate of 10^{-4} (left) and 10^{-5} (right)

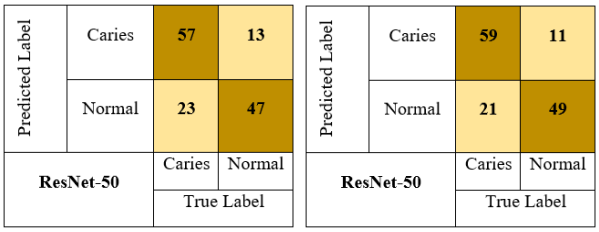


Fig. 16 Confusion Matrix for Lower Occlusal Dataset Using ResNet-50 with learning rate of 10^{-4} (left) and 10^{-5} (right)

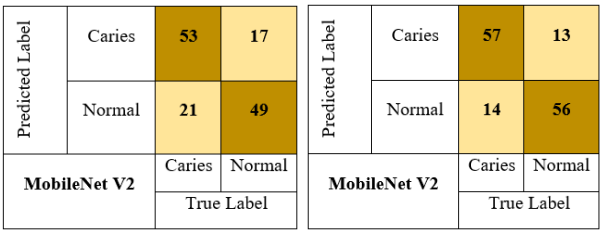


Fig. 17 Confusion Matrix for Lower Occlusal Dataset Using MobileNet V2 with learning rate of 10^{-4} (left) and 10^{-5} (right)

ACKNOWLEDGMENT

We express our gratitude to all dental and medical experts of Dental and Oral, Regional General Hospital dr. Zaineol Abidin (RSUDZA), who assisted throughout the research process.

REFERENCES

[1] A. Warreth, "Dental caries and its management," *Int. J. Dent.*, vol. 2023, p. 9365845, 2023, doi: 10.1155/2023/9365845.

[2] L. Cheng *et al.*, "Expert consensus on dental caries management," *Int. J. Oral Sci.*, vol. 14, no. 1, p. 17, 2022, doi: 10.1038/s41368-022-00167-3.

[3] D. Cogulu and C. Saglam, "Genetic aspects of dental caries," *Front. Dent. Med.*, vol. 3, p. 1060177, 2022, doi:10.3389/fdmed.2022.1060177.

[4] M. Widbiller *et al.*, "Biology of selective caries removal: A systematic scoping review protocol," *BMJ Open*, vol. 12, no. 2, p. e061119, 2022, doi: 10.1136/bmjopen-2022-061119.

[5] J. Flemming, C. Hannig, and M. Hannig, "Caries management—The role of surface interactions in de- and remineralization processes," *J. Clin. Med.*, vol. 11, no. 23, p. 7044, 2022, doi: 10.3390/jcm11237044.

[6] A. S. Soro *et al.*, "Dental caries," in *Molecular Medical Microbiology*, 3rd ed., vol. 1, Y.-W. Tang, M. Sussman, D. Liu, I. Poxton, and J. Schwartzman, Eds. San Diego, CA, USA: Academic Press, 2024, pp. 915–930, doi: 10.1016/B978-0-12-818619-2.00051-7.

[7] E. Fibryanto and W. Widyastuti, "Effect of brushing the teeth before and after meals on salivary pH: A quasi-experimental study," *J. Int. Oral Health*, vol. 14, no. 2, pp. 163–167, 2022, doi:10.4103/jioh.jioh_286_21.

[8] X. Huang *et al.*, "Microenvironment influences odontogenic mesenchymal stem cells mediated dental pulp regeneration," *Front. Physiol.*, vol. 12, p. 656588, 2021, doi: 10.3389/fphys.2021.656588.

[9] T. Kikuiri *et al.*, "Occurrence of subcutaneous emphysema during a caries filling procedure: A case report," *Pediatric Dent. J.*, vol. 32, no. 3, pp. 211–215, 2022, doi: 10.1016/j.pdj.2022.06.002.

[10] R. A. Giacaman *et al.*, "Understanding dental caries as a non-communicable and behavioral disease: Management implications," *Front. Oral Health*, vol. 3, p. 764479, 2022, doi:10.3389/froh.2022.764479.

[11] M. Zhou, J. Dong, L. Zha, and Y. Liao, "Causal association between periodontal diseases and cardiovascular diseases," *Genes*, vol. 13, no. 1, p. 13, 2021, doi: 10.3390/genes13010013.

[12] I. Abramovitz *et al.*, "Cognitive performance and its associations with dental caries: Results from the dental, oral, medical epidemiological (DOME) records-based nationwide study," *Biology*, vol. 10, no. 3, p. 178, 2021, doi: 10.3390/biology10030178.

[13] M. J. Y. Yon *et al.*, "Medical model in caries management," *Dentistry J.*, vol. 7, no. 2, p. 37, 2019, doi: 10.3390/dj7020037.

[14] F. Chairunisa *et al.*, "Oral health status and oral health system in Indonesia: A narrative review," *J. Int. Soc. Prevent. Commun. Dent.*, vol. 14, no. 5, pp. 352–361, 2024, doi: 10.4103/jispcd.jispcd_73_24.

[15] L. Andayasari *et al.*, "Association between tobacco smoking and dental caries in the Indonesian population: Results of a national study in 2018," *J. Prev. Med. Public Health*, vol. 46, no. 4, p. 357, 2023, doi:10.3961/jpmph.22.417.

[16] A. Alami *et al.*, "Investigation of dental caries prevalence, decayed, missing, and filled teeth (DMFT and dmft indexes) and the associated factors among 9-11 years old children," *Res. Sq.*, 2020, doi:10.21203/rs.2.21545/v1.

[17] I. S. Bayrakdar *et al.*, "Deep-learning approach for caries detection and segmentation on dental bitewing radiographs," *Oral Radiol.*, vol. 38, no. 4, pp. 468–479, 2022, doi: 10.1007/s11282-021-00577-9.

[18] Y. Al-Hadeethi, M. I. Sayyed, H. Mohammed, and L. Rimondini, "X-ray photons attenuation characteristics for two tellurite based glass systems at dental diagnostic energies," *Ceram. Int.*, vol. 46, no. 1, pp. 251–257, 2020, doi: 10.1016/j.ceramint.2019.09.258.

[19] M. Fitria *et al.*, "Development of intraoral clinical image dataset for deep learning caries detection," in *Proc. 2nd Int. Conf. Comput. Syst., Inf. Technol., Electr. Eng. (COSITE)*, Banda Aceh, Indonesia, 2023, pp. 194–198, doi: 10.1109/COSITE60233.2023.10249428.

[20] R. Novita *et al.*, "Performance analysis of DMF teeth detection using deep learning: A comparative study with clinical examination as quasi experimental study," *Padjadjaran J. Dent.*, vol. 36, no. 1, pp. 17–24, 2024, doi: 10.24198/pjd.vol36no1.52357.

[21] M. Fitria *et al.*, "The deep learning model for decayed-missing-filled teeth detection: A comparison between YOLOv5 and YOLOv8," *Jordanian J. Comput. Inf. Technol.*, vol. 10, no. 3, pp. 335–349, 2024, doi: 10.5455/jjcit.71-1710834785.

[22] M. Suganthy, F. R. Tabassum, S. Kaviya, and E. Karpagambal, "Classification of COVID-19 using convolutional neural network-ResNet-50," *Int. J. Health Sci.*, vol. 6, no. S4, pp. 4645–4657, 2022, doi: 10.53730/ijhs.v6nS1.5950.

[23] S. Pramanik and H. A. B. Dahlan, "Age estimation using shortcut identity connection of ResNet50 based on convolutional neural network," in *Proc. Int. Conf. Electr. Eng. Informat. (ICEEI)*, Banda Aceh, Indonesia, 2021, pp. 1–7, doi:10.1109/ICEEI52609.2021.9611146.

[24] A. Shabbir *et al.*, "Satellite and scene image classification based on transfer learning and fine tuning of ResNet50," *Math. Problems Eng.*, vol. 2021, p. 5843816, 2021, doi: 10.1155/2021/5843816.

[25] F. Nashrullah, S. A. Wibowo, and G. Budiman, "Epoch parameter investigation on ResNet-50 architecture for pornography classification," *J. Comput., Electron., Telecommun.*, vol. 1, no. 1, pp. 1–8, 2020, doi: 10.52435/complete.v1i1.51.

[26] A. V. Ikechukwu *et al.*, "ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images," *Glob. Transit. Proc.*, vol. 2, no. 2, pp. 375–381, 2021, doi: 10.1016/j.gltp.2021.08.027.

[27] R. Preetha, M. J. P. Priyadarsini, and J. S. Nisha, "Automated brain tumor detection from magnetic resonance images using fine-tuned EfficientNet-B4 convolutional neural network," *IEEE Access*, vol. 12, pp. 100889–100901, 2024, doi: 10.1109/ACCESS.2024.3442979.

[28] L. Li, Z. Tan, and X. Han, "An improved EfficientNet model and its applications in pneumonia image classification," *J. Eng. Sci. Technol. Rev.*, vol. 15, no. 6, pp. 49–54, 2022, doi: 10.25103/jestr.156.07.

[29] A. Deokar and M. El-Sharkawy, "Efficientnext: Efficientnet for embedded systems," in *Proc. 7th Int. Congr. Inf. Commun. Technol. (ICICT)*, London, U.K., 2023, pp. 449–459, doi: 10.1007/978-981-19-2394-4_41.

[30] Q. Lin, Q. Leng, Z. Ding, C. Yan, and X. Xu, "Optimizing contrail detection: A deep learning approach with EfficientNet-b4 encoding," in *Proc. 4th Int. Conf. Electron. Technol., Commun. Inf. (ICETCI)*, Changchun, China, 2024, pp. 504–508, doi:10.1109/ICETCI61221.2024.10594699.

[31] D. Alici-Karaca and B. Akay, "An efficient deep learning model for prostate cancer diagnosis," *IEEE Access*, vol. 12, pp. 150776–150792, 2024, doi: 10.1109/ACCESS.2024.3472209.

[32] C. Y. Zhu *et al.*, "A deep learning based framework for diagnosing multiple skin diseases in a clinical environment," *Front. Med.*, vol. 8, p. 626369, 2021, doi: 10.3389/fmed.2021.626369.

[33] Z. Xiaolong, J. Tian, and D. Hao, "A lightweight network model for human activity classification based on pre-trained MobileNetV2," in *Proc. IET Int. Conf. Cyber Secur. Resilience (CSR)*, Nanjing, China, 2021, pp. 1483–1487, doi: 10.1049/icp.2021.0595.

[34] S. Zhang, D. Jiang, and C. Yu, "A mixed depthwise separation residual network for image feature extraction," *Wireless Netw.*, vol. 28, no. 6, pp. 2631–2642, 2022, doi: 10.1007/s11276-021-02665-4.

[35] L. Ai, Z. Luo, C. Wang, and Y. Wu, "MobileNet investigation: Its application and reproducing edge detectors using depth-wise separable convolution," in *Proc. 2nd Int. Conf. Mach. Learn. Comput. Appl. (ICMLCA)*, Shenyang, China, 2021, pp. 1–6, doi:10.1109/ICMLCA54677.2021.00010.

[36] P. Shi and Z. Zhang, "A lightweight traffic signal video stream detection model based on depth-wise separable convolution," *Electronics*, vol. 14, no. 22, p. 4396, 2025, doi:10.3390/electronics14224396.

[37] H. Chen *et al.*, "Classification and identification of agricultural products based on improved MobileNetV2," *Sci. Rep.*, vol. 14, no. 1, p. 3454, 2024, doi: 10.1038/s41598-024-53454-6.

[38] S. Pooja and S. Preeti, "Face mask detection using AI," in *Predictive and Preventive Measures for COVID-19 Pandemic*, S. K. Saxena, Ed. Singapore: Springer, 2021, pp. 293–305, doi: 10.1007/978-981-33-4236-1_16.