Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 572-590 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Ensemble Voting Method to Enhance the Performance of a Dental Caries Detection System using Convolutional Neural Network.

Putri Rizkiah, Maulisa Oktiana, Khairun Saddami, Maya Fitria, Fitri Arnia, and Yunida Yunida

Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Indonesia

Corresponding author: Maulisa Oktiana. (e-mail: maulisaoktiana@usk.ac.id), **Author(s) Email**: Putri Rizkiah (e-mail: rizkiah@mhs.usk.ac.id), Khairun Saddami (e-mail: khairun.saddami@usk.ac.id), Maya Fitria (e-mail: mayafitria@usk.ac.id), Fitri Arnia (e-mail: f.arnia@usk.ac.id), Yunida Yunida (e-mail:yunida@usk.ac.id)

Abstract **Individual classification models for caries detection still face significant challenges, including limited accuracy and unstable predictions, which can hinder diagnosis, delay clinical decisions, and increase the risks associated with patient care. To overcome these limitations, this study proposes an ensemble voting method that combines five deep learning models, such as ResNet-152, MobileNetV2, InceptionV3, NASNetMobile, and EfficientNet-B5. This approach aims to enhance the accuracy and stability of caries detection by leveraging the complementary strengths of the individual models while mitigating their weaknesses. Each model was trained and tested on the same dataset of dental images, categorized into caries and regular classes. Their predictions were aggregated using hard and soft voting techniques. The ensemble's performance was evaluated using accuracy, precision, recall, and F1-score. The ensemble voting demonstrates a notable improvement in classification performance over individual models. Hard and soft voting have excellent classification performance and consistently outperform the best individual models. The accuracy increased from EfficientNetB5 0.8485 to 0.8864 and 0.8712, representing increases of 4.46% and 2.68%, respectively. The precision increased from MobileNetV2 0.8182 to 0.8493 and 0.8551, representing increases of 3.81% and 4.52%. For recall, EfficientNetB5 ranked highest among individual models with a score of 0.9242. Hard voting increased 1.64% to 0.9394, and soft voting decreased slightly by 3.28% to 0.8939. The F1 score of EfficientNetB5 is 0.8592. Hard and soft voting increased 3.83% and 1.73% to 0.8921 and 0.8741. The proposed ensemble improves the F1-score by 3.83 percentage points compared to the best individual model. The ensemble voting method effectively leverages the complementary strengths of each deep learning model to improve the stability and accuracy of fast, reliable dental caries early detection prediction.**

Keywords **Caries detection; Deep learning; Ensemble voting; Hard voting; Soft voting.**

I. Introduction

Teeth are complex organs that play an essential role in eating and speaking. The structure of teeth consists of several layers. Each layer has a specific function [1]. Dental caries, also known as tooth decay or cavities, is a common dental issue that impacts the complex structures of the teeth. Tooth decay is a condition where the tooth structure is damaged from the outer surface to the pulp, caused by the accumulation of bacteria from food residues. Failure to maintain oral hygiene can lead to tooth decay, causing the teeth to become brittle, develop cavities, or break [2]. This condition is important to pay attention to because untreated dental caries can lead to serious problems, including severe pain, tooth abscesses, and even gum disease [3].

Dental caries is diagnosed by reviewing the patient's history and using dental instruments for direct inspection. Dentists desire automation to simplify caries diagnosis, but existing techniques are considered inefficient and time-consuming. Deep learning in dental health is expected to streamline the diagnosis process and reduce the time required. The Convolutional Neural Network (CNN) is a deep learning technique used for image data. CNN is the most widely used algorithm for detecting, grouping, and classifying diseases based on related organs in medical images [4]. In previous studies, intraoral photographs have helped diagnose dental caries. At the same time, deep

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

learning has been applied to distinguish caries from normal teeth via CNN-based tooth surface segmentation, achieving 83% accuracy [5].

Meanwhile, D.L. Doung [6] researched early detection of occlusal caries using smartphone images. The approach involved contour-finding techniques to locate caries lesions using machine learning models and CNN. The machine learning model, specifically a Support Vector Machine (SVM), achieved 88.76% accuracy, 92.31% sensitivity, and 85.21% specificity. Therefore, the SVM model was considered more optimal for this task. Besides, L. Zheng et al. [7] evaluated and compared three CNN architectures. An analysis of the depth of radiographic penetration of carious lesions was performed to assist clinical diagnosis based on clinical parameters, and the dataset was obtained using radiographic images of carious lesions. The results showed that the Residual Network (ResNet)18 model performed best among the Visual Geometry Group (VGG)19 and InceptionV3 models.

M. Moral et al. [8] also applied a CNN model to radiographic images to detect approximal dental caries in bitewing radiographs and categorize them by lesion severity. The severity of the lesions was classified based on normal, new, and advanced stages. Two CNN architectures, namely ResNet and Inception, were used in this research. The Inception model obtained the highest performance, attaining an accuracy of 73.3% on the test dataset. The other related study looked at early detection of dental caries using individually captured images to compare the accuracy of several CNN architectures, namely VGG16, VGG19, InceptionV3, and ResNet50, using dental images. Five hundred dental images were divided into two categories: normal and caries. The test results showed that the InceptionV3 architecture achieved the highest accuracy, at 99.89%, compared to the other architectures [9].

Saidi et al. [10] utilized color digital imaging technology to detect early-stage caries lesions using several CNN variants, namely Vgg16, Vgg19, InceptionV3, and Resnet50. The results showed that the InceptionV3 model had the best performance with a training accuracy rate of 99.89% and a validation accuracy rate of 98.95%. The InceptionV3 model was far superior to the other models. This approach is very useful in e-health and IoT, enabling easy and effective remote patient care. Al Yassar et al. [11] developed a U-Net-based dental segmentation model with ResNet50, VGG19, and InceptionV3 backbones to improve dental caries classification. Consistent segmentation improved the classification performance of ResNeXt50, achieving a maximum accuracy of 79.17%, exceeding that of ResNet-50 and InceptionV3. However, this performance improvement was not statistically significant, requiring further exploration.

Furthermore, Fitria et al. [12] compared You Only Look Once (YOLO)v5 and YOLOv8 frameworks for detecting decayed, missing, and filled teeth using 294 augmented clinical images. Both models achieved high precision, recall, and mAP, with YOLOv8m at 90.6% slightly outperforming YOLOv5l at 90.4% but requiring longer training time. YOLOv5s, although less accurate, was the fastest and most suitable for real-time use. Moreover, YOLOv8 showed more stable training dynamics. These findings highlight that while both frameworks are effective for automated dental caries detection, practical deployment must balance accuracy and computational efficiency. Despite existing techniques achieving high accuracy, they remain inadequate for fully addressing dental caries classification challenges. Given the variety of strategies, such as dataset enhancement and parameter optimization, further research is needed to evaluate their limitations. Since individual algorithms possess distinct strengths, combining them through ensemble methods can yield superior performance by integrating predictions across models.

Muhajir et al. [13] showed that an ensemble approach combining ResNet152, MobileNetV2, and InceptionV3 models achieved the best performance in classifying students' facial emotions and confirmed that combining several models using the majority voting method could improve accuracy compared to using an Individual model. Desiani et al. [14] stated that the application of the majority voting ensemble method in cervical cancer classification, which combines SVM, Multi-layer perceptron (MLP), and K-Nearest Neighbors (KNN), successfully improved classification performance in accuracy by 1.72%, sensitivity by 0.74%, and specificity by 3.4% compared to the average results of an Individual model, thus proving that majority voting can provide more reliable results in early detection. Previous researchers have conducted several studies related to the application of ensemble models. Cao-Van K et al. [15], Kumari S et al. [16], Ilyas QM et al. [17], Khan AA et al. [18], and Manconi A et al. [19] state that ensemble learning has significant advantages in improving accuracy, robustness against complex data, and effectively handling class imbalance. Majority voting can exceed the performance of an Individual model; weighted methods produce more optimal predictions; and combining with data augmentation has proven efficient and improves classification performance.

Furthermore, Alsakar MY et al. [20] and Chandra TB et al. [21] declare that ensemble learning in medical imaging improves the accuracy and reliability of automated diagnosis. Bagging and majority voting ensembles have proven superior to single methods and

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

can support early diagnosis more robustly. In addition, Kang J et al. [22] and Mienye I et al. [23] agree that the ensemble approach to data classification can produce better performance than single classification models, especially when each classifier has high accuracy and provides significant prediction variation. These studies prove that ensemble classification strategies can improve classification performance.

Although numerous studies have reported high accuracy in dental caries detection, their performance remains strongly influenced by image modality, dataset characteristics, and model architecture. Research based on radiographic images generally achieves superior results due to clearer lesion contrast; however, such models often demonstrate limited generalizability when applied to intraoral photographs, which exhibit higher variability in illumination, texture, and acquisition conditions. In contrast, deep learning models trained on intraoral images frequently face challenges related to limited dataset size, class imbalance, and inconsistent annotation quality, which can lead to overfitting and suboptimal generalization performance. Furthermore, most existing studies adopt a single convolutional neural network (CNN) architecture, rendering their results highly sensitive to architectural bias and dataset-specific features. Although individual CNN models such as ResNet50 and VGG16 have achieved promising accuracies on caries datasets [24],[25], their effectiveness is often confined to controlled settings with homogeneous image modalities, balanced class distributions, and sufficiently large training samples. Single-model approaches are particularly vulnerable when applied to heterogeneous intraoral images affected by artifacts such as saliva reflections, motion blur, and inter-patient anatomical variability, as well as during cross-dataset evaluation, where performance degradation of 10–20% in F1-score has been reported [26]. These limitations highlight the necessity of an ensemble-based strategy that integrates multiple deep learning models through majority (hard) voting or weighted aggregation. By leveraging complementary feature representations from diverse architectures, ensemble methods can mitigate individual model biases, reduce variance, and enhance the robustness and reliability of classification outcomes, ultimately improving key performance metrics such as accuracy, precision, and recall for real-world dental caries detection systems.

Given the urgent need for a more robust, accurate, and generalizable method for the early detection of dental caries, this study focuses on developing a CNN-based image classification system. The proposed approach explores ensemble methods to integrate the advantages of various CNN architectures and significantly improve detection and classification performance, accuracy, and stability. This method integrated predictions from five Individual CNN models, ResNet152, MobileNetV2, InceptionV3, NasNet Mobile, and EfficientNetB5, which were selected based on their architectural advantages. ResNet152 can build intense CNN models with 152 layers without experiencing performance degradation due to vanishing gradients, residual learning, and identity shortcut connections. Enables more complex and representative feature extraction from image data [27]. MobileNetV2 is a lightweight and efficient deep learning architecture, specifically designed for resource-constrained devices such as smartphones and laptops. With its inverted residual and linear bottleneck innovations, this model can maintain high accuracy while accelerating training and inference. MobileNetV2 is also effective with limited training data. It can be fine-tuned for various classification tasks, making it ideal for medical and diagnostic applications that require efficient computation and accurate results [28]. InceptionV3 has a significant advantage in handling features on various scales and hierarchical structures in images, making it very effective for medical image analysis and complex classification tasks. Its architecture, which uses inception modules, enables parallel, multi-scale feature extraction, improving the model's ability to recognize subtle and coarse patterns. Although this model is relatively complex and requires considerable computational resources, InceptionV3 still provides a good balance between high accuracy and the ability to handle variations in object size in images, making it suitable for applications that require detailed and diverse feature recognition [29].

NasNet-Mobile excels because it uses Neural Architecture Search (NAS) to automatically find the optimal network architecture, resulting in high performance with computational efficiency suitable for mobile devices. This model can be fine-tuned for specific tasks such as disease recognition [30]. EfficientNetB5 is a CNN model that uses a structured scaling method to simultaneously increase width, depth, and image resolution, resulting in high performance with computational efficiency. This model is effective for disease detection. As a pre-trained model, EfficientNet-B5 easily adapts to specific tasks with fast training and optimal results [31]. Combining these five models is expected to improve the accuracy, reliability, and effectiveness of caries prediction based on clinical images. Thus, this research provides scientific and practical contributions to the development of modern diagnostic technology that is fast, accurate, and efficient.

To address the limitations of individual CNN-based classifiers, this study employs an ensemble approach by integrating five CNN architectures: ResNet152, MobileNetV2, InceptionV3, NASNetMobile, and EfficientNetB5. These models were deliberately

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

selected to represent diverse architectural characteristics and learning strategies. ResNet152 offers very deep residual learning capable of extracting complex hierarchical features, while MobileNetV2 and NASNetMobile provide lightweight architectures optimized for computational efficiency and stability. InceptionV3 excels at multi-scale feature extraction, which is crucial for identifying caries patterns of varying sizes, and EfficientNetB5 employs compound scaling to achieve high accuracy at a balanced computational cost.

The combination of these architectures is expected to enhance classification robustness by leveraging complementary strengths while mitigating individual weaknesses, such as overfitting or class bias. Furthermore, selecting five models ensures sufficient architectural diversity while maintaining computational feasibility and enables majority-based decision-making without tie-breaking in hard voting. This strategic design aims to improve accuracy, stability, and generalization capability in intraoral dental caries detection. By integrating heterogeneous architectures ranging from deep residual networks to efficient mobile designs, the ensemble leverages diverse inductive biases, decision boundaries, and feature-extraction strategies to mitigate overfitting and enhance robustness.

II. Method

Fig.1 illustrates the study's workflow, which involved Pre-processing clinical dental images, developing individual models and an ensemble Method, and analyzing the results. Each step is detailed in the following subsection.

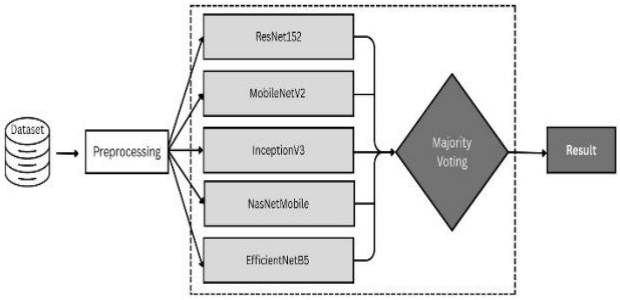


Fig. 1. Research Method

A. Dataset Collection

This stage involves collecting a high-quality, representative dataset of clinical dental images, sourced from previous research. The dataset was obtained from two main sources: clinical intraoral dental images collected through direct patient examinations at the Dental and Oral Hospital of Aceh, and a publicly documented dataset developed by Fitria et al.[11], [32] for use in model training and testing. The combined dataset was curated to ensure consistent labeling and image quality before being used for model training and evaluation. The dataset used in this study consists of clinical dental images grouped into two classes, namely caries and non-caries. This dataset was taken from a collection of clinical dental images of patients directly from patient examinations. The entire clinical dental image dataset was taken from previous research. The pre-processing process was carried out using tools to ensure optimal data quality. Images were selected based on lighting quality, object clarity, and image sharpness to ensure optimal feature extraction. Fig.2 shows a sample of clinical dental images of carious and normal teeth that will be used in the dataset. The dataset consists of 1,320 images, divided into caries and normal classes, with 660 images in each class.

Fig. 2. Sample dataset of oral cavity images for classification: (A–D) show cases of teeth with caries, while (E–F) show normal teeth.

The dataset was divided into three subsets with standardized proportions: training (80%), validation (10%), and test (10%). The dataset distribution is shown in Table 1.

Table 1. Dataset Distribution

Dataset	Training	Testing	Validation	Total
Normal	528	66	66	660
Caries	528	66	66	660
Total	**1056**	**132**	**132**	**1320**

B. Data Pre-Processing

The collected data must be processed before being used to train the model. A preprocessing step is performed to ensure optimal data quality, using tools that reduce data variation and improve model performance. Each image in the dataset has different sizes and dimensions, so cropping and resizing are

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

performed to ensure they have the same size and dimensions. The cropping process is done manually to avoid losing essential features in the image. Then each image is resized to 224×224 pixels. This size was chosen to align with the input architecture used in this study. The visualization of the cropping, resizing, and data augmentation process is shown in Fig 3. In addition to these two main stages, pre-processing may include data augmentation techniques, image cleaning, flipping, rotation, zooming, and normalization to improve the quality and diversity of the training data. All these stages are performed after the images are entered into the system, as part of the data preparation before being utilized by the machine learning model.

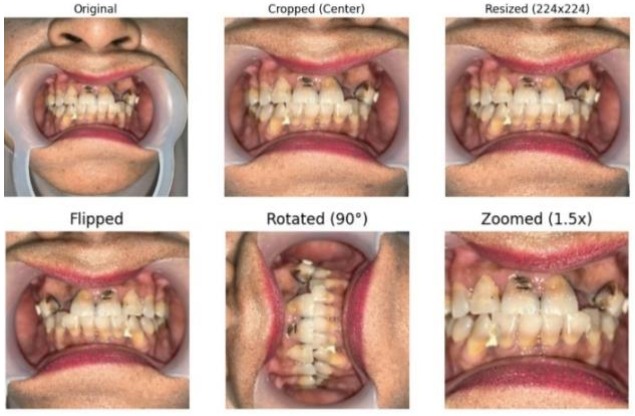


Fig. 3. The examples of image preprocessing and augmentation techniques applied to intraoral dental photographs.

All images in the dataset were standardized to a resolution of 224 × 224 pixels prior to training, aligning with the input requirements of ImageNet-pretrained CNN architectures such as ResNet152, EfficientNetB5, InceptionV3, MobileNetV2, and NasNet Mobile. This resizing process preserves essential diagnostic features, such as lesion edges and textures, in intraoral caries images while enabling efficient batch processing and transfer learning from large-scale natural image datasets.

Data augmentation techniques were systematically applied during training to enhance model generalization and mitigate overfitting, particularly given the limited size and inherent variability of medical imaging datasets. The augmentation pipeline included random rotations within ±15° to simulate varying camera angles during clinical intraoral photography, random zooming up to 20% to account for differences in proximity to teeth, and horizontal flipping with a 0.5 probability to reflect the bilateral symmetry of dental structures.

Additional augmentation strategies, such as brightness adjustments (±10%) and contrast enhancement, were incorporated to address common intraoral imaging artifacts, including uneven lighting from intraoral flashes and patient movement. These transformations were implemented in real time using libraries TensorFlow's ImageDataGenerator, effectively expanding the dataset by a factor of 8-10 per epoch and improving cross-validation performance by 5-8% in preliminary experiments. This preprocessing regimen ensures robust feature learning across diverse caries presentations, from early demineralization to advanced cavitation.

C. Experimental Setup

All experiments were conducted in Google Colaboratory (Colab), using a high-RAM NVIDIA T4 GPU runtime to facilitate efficient training of CNN ensembles on intraoral caries datasets. The software environment included Python 3.10 and the TensorFlow and Keras frameworks. Additional libraries included NumPy, OpenCV, Matplotlib, and Scikit-learn for structured model definition, transfer learning, and ensemble integration, ensuring compatibility with ImageNet-pre-trained weights for ResNet152, EfficientNetB5, InceptionV3, MobileNetV2, and NASNet Mobile. This configuration is sufficient to process batches of 32 images at 224×224 resolution while adjusting the memory-intensive ResNet152 backbone. The persistent Colab runtime supports thorough hyperparameter testing, with early stopping after 100 epochs per model. Key libraries include TensorFlow Addons for advanced metrics, Augmentations 1.4 for real-time data augmentation pipelines compatible with the displayed preprocessing (resizing, inversion, rotation, zoom), and scikit-learn 1.3 for stratified splitting and ensemble voting. Reproducibility is ensured through fixed random seeds across NumPy, TensorFlow, and Python, public dataset links, and hyperparameter tables such as the Adam optimizer with a learning rate of 1e-5 and the ReduceLROnPlateau scheduler.

D. Model Development

The model was built using five different classification architectures: ResNet152, MobileNetV2, NASNet Mobile, InceptionV3, and EfficientNetB5. The five models will be combined using the ensemble voting method to improve accuracy in detecting dental caries.

1. ResNet152

ResNet-152 is based on the concept of residual learning, which enables the effective training of very deep neural networks. Eq (1) represents the fundamental formulation of residual connections in the ResNet architecture. In this formulation, x denotes the input feature map. At the same time, $F(x)$ corresponds to a non-linear transformation typically composed of convolutional layers, batch normalization, and activation functions. The output (y) is obtained through element-wise addition between the transformed

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

features $F(x)$, and the original input feature map (x). The introduction of shortcut (skip) connections allows gradients to propagate directly through the network, mitigating the vanishing gradient problem and facilitating the optimization of deep architectures [33].

$$y = F(x) + x \quad (1)$$

This mechanism is known as residual learning, in which the network learns the residual mapping shown in Eq. (2).

$$F(x) = H(x) - x \quad (2)$$

instead of directly approximating the full mapping $H(x)$, resulting in Eq. (3).

$$H(x) = F(x) + x \quad (3)$$

Such a formulation mitigates the vanishing gradient problem, facilitates the training of very deep networks, and improves convergence stability and speed. When the dimensions of x and $F(x)$ are not identical, a linear projection, commonly implemented using a 1×1 convolution, is applied to align the dimensions before performing the addition. The ResNet architecture first demonstrated this concept with a 34-layer convolutional network, inserting residual connections every two layers. For deeper variants, a bottleneck design was adopted to improve computational efficiency. Each bottleneck block consists of three convolutional layers arranged as 1×1, 3×3, and 1×1 convolutions. The first 1×1 layer reduces the dimensionality of the feature maps, the 3×3 layer performs spatial feature extraction, and the final 1×1 layer restores the channel dimension. ResNet-152 is constructed by stacking bottleneck blocks into four stages, with the number of channels increasing progressively across them. This design enables the network to reach 152 layers while maintaining manageable computational complexity and strong representational capacity [34].

2. MobileNetV2

MobileNetV2 is a highly efficient convolutional neural network architecture designed specifically for mobile and embedded devices with limited computational resources. This model introduces an inverted residual structure combined with a linear bottleneck, which reduces computational cost while maintaining competitive accuracy. MobileNetV2 employs depthwise separable convolutions, which separate spatial and channel-wise filtering operations, thereby significantly reducing the number of parameters and computational complexity. This architecture is particularly suitable for tasks such as image classification and object detection on resource-constrained devices [35]. The core building block of MobileNetV2 is the inverted residual block. The inverted residual block can be expressed as Eq. (4), where x represents the input feature map with relatively few channels, and the function $Expand\ (x)$ performs channel expansion using a 1×1 convolution to capture richer feature representations. The operator $DW\ (\cdot)$ denotes a 3×3 depthwise convolution applied independently to each channel. Finally, $Proj\ (.)$ represents a 1×1 linear projection that reduces the channel dimension. The residual connection adds the transformed features to the original input, improving gradient flow, stabilizing training, and preserving essential information from earlier layers [36].

$$y = x + Proj(DW(Expand(x))) \quad (4)$$

3. InceptionV3

InceptionV3 is a deep learning architecture that combines multiple convolution filter sizes in a single layer through the Inception module. This design enables the network to efficiently capture features at various scales [33]. InceptionV3 is an extension of the Inception architecture that enhances computational efficiency through factorization and grid-size reduction techniques without compromising accuracy. The main Inception module can be expressed in Eq. (5), where x denotes the input feature map and y represents the output obtained by concatenating multiple convolutional branches along the channel dimension. The operation $1 \times 1(x)$performs dimensionality reduction while introducing additional non-linear transformations. The term $3 \times 3(1 \times 1(x))$extracts spatial features after channel compression, whereas $3 \times 3(3 \times 3(1 \times 1(x)))$corresponds to the factorization of a larger convolution (functionally equivalent to 5×5) into two consecutive 3×3 convolutions, expanding the receptive field with fewer parameters. Meanwhile, $1 \times 1(3 \times 3\,\text{pool}(x))$integrates contextual information from pooling and adjusts the channel dimension via 1×1 convolution. By concatenating these parallel branches, the module produces a computationally efficient multi-scale feature representation. Additionally, InceptionV3 incorporates asymmetric convolutions, auxiliary classifiers, and batch normalization to further stabilize training and improve classification performance on large-scale datasets such as ImageNet [37].

$$y = concat(1 \times 1(x), 3 \times 3(1 \times 1(x)), 3 \times 3(3 \times 3(1 \times 1(x))), 1 \times 1(3 \times 3 pool(x))) \quad (5)$$

4. NasNet Mobile

NASNetMobile is a convolutional neural network designed via Neural Architecture Search (NAS), in which an RNN-based controller automatically discovers optimal cell structures. Each cell is organized as a Directed Acyclic Graph (DAG) that combines operations, such as separable convolutions and pooling, to produce feature representations through channel-wise concatenation [38], [39]. NASNetMobile employs Neural Architecture Search (NAS) to automatically design optimal cell-based building

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

blocks. Each cell consists of $B = 5$blocks, where each block processes two previous hidden states $h_{i-2}$and h_{i-1}using two selected operations. The output of the block i can be expressed in Eq. (6), where $\oplus$denotes element-wise addition and op represents candidate operations such as 3×3 convolution, 5×5 separable convolution, 3×3 max pooling, 3×3 average pooling, skip connection, or none.

$$o_i = op_a(h_{i-2}, h_{i-1}) \oplus op_b(h_{i-2}, h_{i-1}), \quad (6)$$

The selection of operations is determined by an RNN-based controller that models the probability distribution as shown in Eq. (7), where θ denotes the controller parameters. The controller is optimized using the REINFORCE policy gradient method, with reward defined as $R = V_{valid} - 0.1 \cdot \text{Latency}$, balancing validation accuracy and computational efficiency.

$$P(\alpha_t \mid \alpha_{<t}; \theta) = \text{softmax}(\text{RNN}(\alpha_{<t}; \theta)), \quad (7)$$

The architecture stacks normal cells (stride = 1) and reduction cells (stride = 2) after an initial stem layer, followed by global average pooling and a softmax classifier. NASNetMobile stacks two types of cells: normal cells, which preserve spatial dimensions and are typically repeated six times, and reduction cells, which halve the spatial resolution using stride-2 separable convolutions while increasing channel depth [40].

5. EfficientNetB5:

EfficientNetB5 is one of the variants of the EfficientNet family, a deep learning architecture that systematically balances depth, width, and input resolution scalability using a compound scaling method. EfficientNetB5 is designed to achieve high accuracy with good computational efficiency, and it is ideal for deployment on various devices with different performance requirements [41]. This makes architecture offers improved accuracy compared to traditional convolutional models with fewer parameters and computational requirements. The key principle of EfficientNet is compound scaling, which balances network depth *(d),* width *(w),* and input resolution *(r)* simultaneously rather than scaling them independently. A single compound coefficient (ϕ), controls scaling and is applied to a baseline model, EfficientNet-B0. The dimensions are scaled as shown in Eq. 8:

$$d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi}, \quad (8)$$

subject to the constraint: $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$, where α, β, and $\gamma$are constants determined through grid search. This constraint ensures that for each increment of ϕ, the computational cost (FLOPs) increases by approximately 2^{ϕ}. The underlying intuition is that increasing input resolution requires proportional increases in depth to enlarge the receptive field and in width to capture more detailed feature representations. Compound scaling, therefore provides a systematic and efficient method for model scaling while maintaining computational balance [42].

All CNN architectures were initialized using ImageNet pre-trained weights and fine-tuned for binary classification (caries and normal). The convolutional base of each model was retained to leverage learned visual representations, while the final classification layers were replaced with a fully connected layer followed by a sigmoid activation function. Fine-tuning was applied to the top layers of each network to adapt the models to the dental caries dataset. All models were trained using the same hyperparameter configuration to ensure fair comparison. The Adam optimizer was used with a learning rate of 1×10^{-5}, a batch size of 32, and training for 100 epochs. Binary cross-entropy was employed as the loss function. The same training, validation, and testing splits, as well as identical hyperparameters, were consistently applied to all individual models to ensure a fair performance comparison.

Table 2. Hyperparameters in Training Model

Hyperparameter	Input
Batch Size	32
Learning Rate	10^{-5}
Optimizer	Adam
Epoch	100
Loss Function	Cross-Entropy

The model design and training were carried out on the Google Colab platform. The designed model was then trained and validated on a dataset that had undergone data transformation, with hyperparameter configurations as shown in Table 2, including a learning rate of 0.00001, the Adam optimizer with 100 epochs, and binary cross-entropy loss. Performance evaluation used accuracy, precision, recall, and F1-score metrics.

E. Ensemble Voting

Ensemble methods integrate multiple individual classifiers to form a new model with enhanced performance. One common approach is ensemble voting, which aggregates predictions from base models and determines the outcome by majority vote. This technique leverages the strengths of individual classifiers to improve overall decision-making [43]. This study developed an ensemble model by applying five classification architectures: ResNet152, MobileNetV2, NASNet Mobile, InceptionV3, and EfficientNetB5. The five models will be combined using model ensemble voting. Fig. 4 shows that the ensemble voting process can be run after obtaining results from several individual models. The results from these individual models will be processed in ensemble voting. Voting combines the predictions from several single classifications into a final

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

prediction. Hard and soft voting are the two main types of voting used in ensemble methods. Hard voting, or majority voting, determines the final class based on the majority of predictions from the base models, making it suitable for classification tasks with distinct mutually exclusive classes. In contrast, soft voting, or weighted voting, aggregates the probability scores from each model, computes their weighted average, and selects the class with the highest probability. Soft voting can be applied to classification and regression problems. According to Eq. (9), the class label determines the ultimate class $\acute{y}$ through majority (plurality) voting from each classifier $hj(x)$, where *j* represents the index for each class produced by an individual classifier.

$$\acute{y} = mode\,\{h1x, h2x, \ldots, hj(x)\} \quad (9)$$

In the soft voting strategy, all individual models were assigned equal weights, and the final prediction was obtained by averaging the predicted class probabilities produced by each base model. The class with the highest average probability was selected as the final output. In the hard voting strategy, the final class label was determined by majority voting across all base models' predictions. As an odd number of models was employed in the ensemble, tie conditions did not occur, ensuring an unambiguous final decision for each sample. In soft voting, each class's average probability is taken into account. Y denotes the final prediction, Pij is the predicted probability of membership in class i from the classifier for label class j, and Wj is the weighting parameter. Y can be calculated by using Eq. (10) as follows [44].

$$Y = argmax \sum WjPij, i \in \{0,1\}, [j = 1, \ldots m] \quad (10)$$

F. Model Evaluation

The model's performance was assessed using a confusion matrix to evaluate accuracy, precision, recall, and F1-score. A confusion matrix was utilized to evaluate the outcomes of this research. ResNet-152, MobileNetV2, InceptionV3, EfficientNetB05, and NASNet Mobile were evaluated using a confusion matrix, with accuracy, precision, recall, and F1-score as metrics. Furthermore, the ensemble method was evaluated using the same metrics to examine its effectiveness and the extent of its impact compared to the five Individual models. A receiver operating characteristic (ROC) curve was generated to assess the model's discriminatory power, and the area under the curve (AUC) was measured. Accuracy is the ratio of correct predictions to the total number of predictions made, calculated by dividing the total number of correct predictions (TP + TN) by the total number of predictions in the dataset. The accuracy is calculated by using Eq. (11) as follows.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

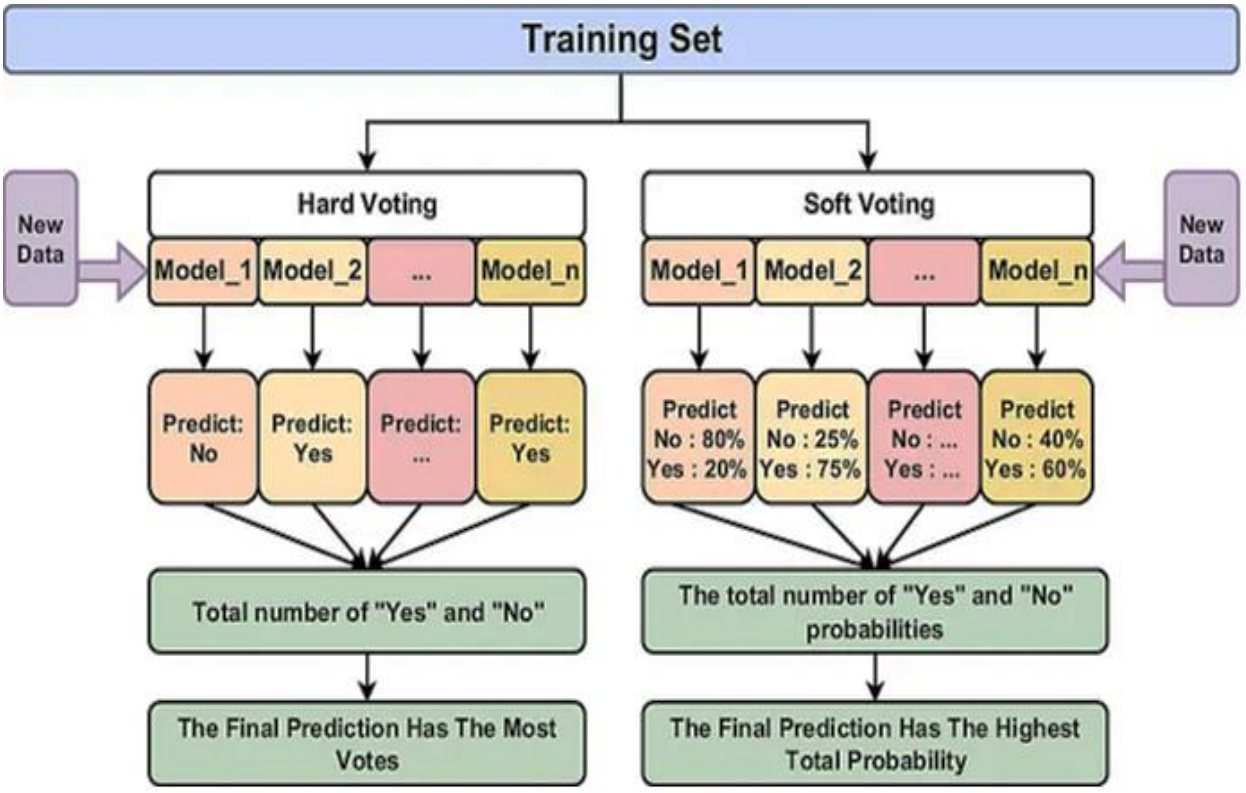


Fig. 4. The Training Set in Ensemble Voting [15]

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

Precision can be calculated by dividing the number of true positives (TP) by the total number of optimistic predictions, which includes true positives (TP) and false positives (FP). Precision indicates how accurately a model identifies positive examples among all positive predictions. The precision is calculated by using Eq. (12) as follows.

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

Recall measures the extent to which a model can correctly identify positive cases or values, referring to the ratio between correct optimistic predictions and the overall positive data. The recall is calculated by using Eq. (13) as follows.

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

The F1 Score is a metric that balances precision and recall, providing a comprehensive overview of model performance in an easy-to-understand manner. The F1-score is calculated by using Eq. (14) as follows [45].

$$F1 - score = \frac{2 \times (precision \times recall)}{(precision + recall)} \quad (14)$$

III. Result

A. Individual Model

Fig. 5 illustrates the results of training individual models using predetermined hyperparameters, namely ResNet152, MobileNetV2, InceptionV3, NasNet Mobile, and EfficientNetB5. The ResNet152 curve shows a stable increase in training accuracy and a fairly consistent decrease in loss. However, there are indications of mild overfitting due to differences in performance between the training and validation data. MobileNetV2 produced a relatively stable training curve, where the training and validation accuracy were balanced without indicating significant overfitting. The loss also decreased consistently in both datasets. Fig. 5. Performance comparison of five CNN architectures on the image classification task. Each subfigure shows training (blue) and validation (orange) accuracy and loss across epochs for: (a,b) ResNet152, (c,d) MobileNetV2, (e,f) InceptionV3, (g,h) NasNet Mobile, and (i,j) EfficientNetB5. Each model is shown in a graph of accuracy and loss during training and validation, illustrating the trend of increasing accuracy and decreasing loss for each architecture over the epochs. Moreover, the results of training individual models InceptionV3 and NasNet Mobile using predetermined hyperparameters. InceptionV3 shows a training curve with good accuracy improvement in training and validation, and a consistent decrease in loss, although the validation accuracy is slightly lower than the training accuracy. NasNet Mobile showed a training curve, with training accuracy increasing from 0.50 to 0.70, but validation accuracy stagnated at 0.62-0.66 after the 20th epoch. This condition indicates overfitting because, even though the training loss decreased consistently, the validation loss remained stable around 0.65–0.70. Further, the results of training individual models EfficientB5 using predetermined hyperparameters. EfficientNetB5 displayed an excellent training curve, increasing training accuracy from 0.53 to over 0.72, while validation accuracy remained relatively stable at around 0.70. There were no indications of overfitting, as the validation loss consistently decreased and was even lower than the training loss.

Meanwhile, Fig. 6 shows the confusion matrices for the Individual models, namely ResNet152, MobileNetV2, InceptionV3, NASNet Mobile, and EfficientB5. In the testing phase, ResNet152 achieved an accuracy of 78.79% with a precision of 79.69% and a recall of 77.27%. The confusion matrix results show that the model still produces errors with 13 negative data incorrectly predicted as positive (false positive) and 15 positive data incorrectly predicted as negative (false negative), indicating that although ResNet152 can generalize quite well, the model is still less sensitive to detecting positive data. Furthermore, the MobileNetV2 model achieved an accuracy of 81.82% with precision, recall, and F1-score all at the same value, namely 81.82%. The confusion matrix shows a balanced distribution of errors, with 12 false positives and false negatives. These results indicate that MobileNetV2 achieves stable, balanced performance in detecting both classes without strong bias.

Besides, InceptionV3 achieved an accuracy of 81.06% in testing, with precision 79.71% and recall 83.33%. Based on the confusion matrix, the model produced 14 false positives and 11 false negatives. This pattern shows that InceptionV3 is more sensitive to positive classes, although it tends to produce excessively optimistic predictions, which lowers precision. However, NasNet Mobile, in testing, achieved the lowest accuracy of 74.24%, with precision 71.05% and recall 81.82%. The confusion matrix shows a relatively high number of errors with 22 false positives and 12 false negatives. It indicates that NasNetMobile is more biased towards the positive class, often misprediction negative data. Despite this, EfficientNetB5 achieved the highest performance among all models, with an accuracy of 84.85%, precision of 80.26%, recall of 92.42%, and an F1-score of 85.96%. The confusion matrix shows only five false negatives and 15 false positives, meaning the model is highly reliable in detecting positive cases with a low error rate. Table 3 presents the test results for the single model, highlighting performance variations across the architectures used. EfficientNetB5 outperforms other models, achieving a balanced combination of accuracy, precision, and sensitivity. This model can detect almost all positive cases well, reducing the possibility of classification errors in caries data. MobileNetV2 also showed competitive

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

performance with stable and consistent results across all evaluation metrics. InceptionV3 performed exceptionally well and tended to be more sensitive in detecting positive cases, although its precision was slightly lower. ResNet152 showed relatively moderate performance, with fairly reliable classification capabilities but still limited in detecting all positive cases. Meanwhile, despite its fairly good sensitivity, NasNet Mobile had the lowest performance and the lowest accuracy. Overall, these results show that each model has its own characteristics and advantages, but EfficientNetB5 can be considered the optimal single model in this study.

B. Ensemble Voting

The ensemble voting model was built by combining five Individual models, namely ResNet152, MobileNetV2, InceptionV3, NasNetMobile, and EfficientNetB5. Each model was trained independently on the same dataset and then tested on a test set of 132 images. All models

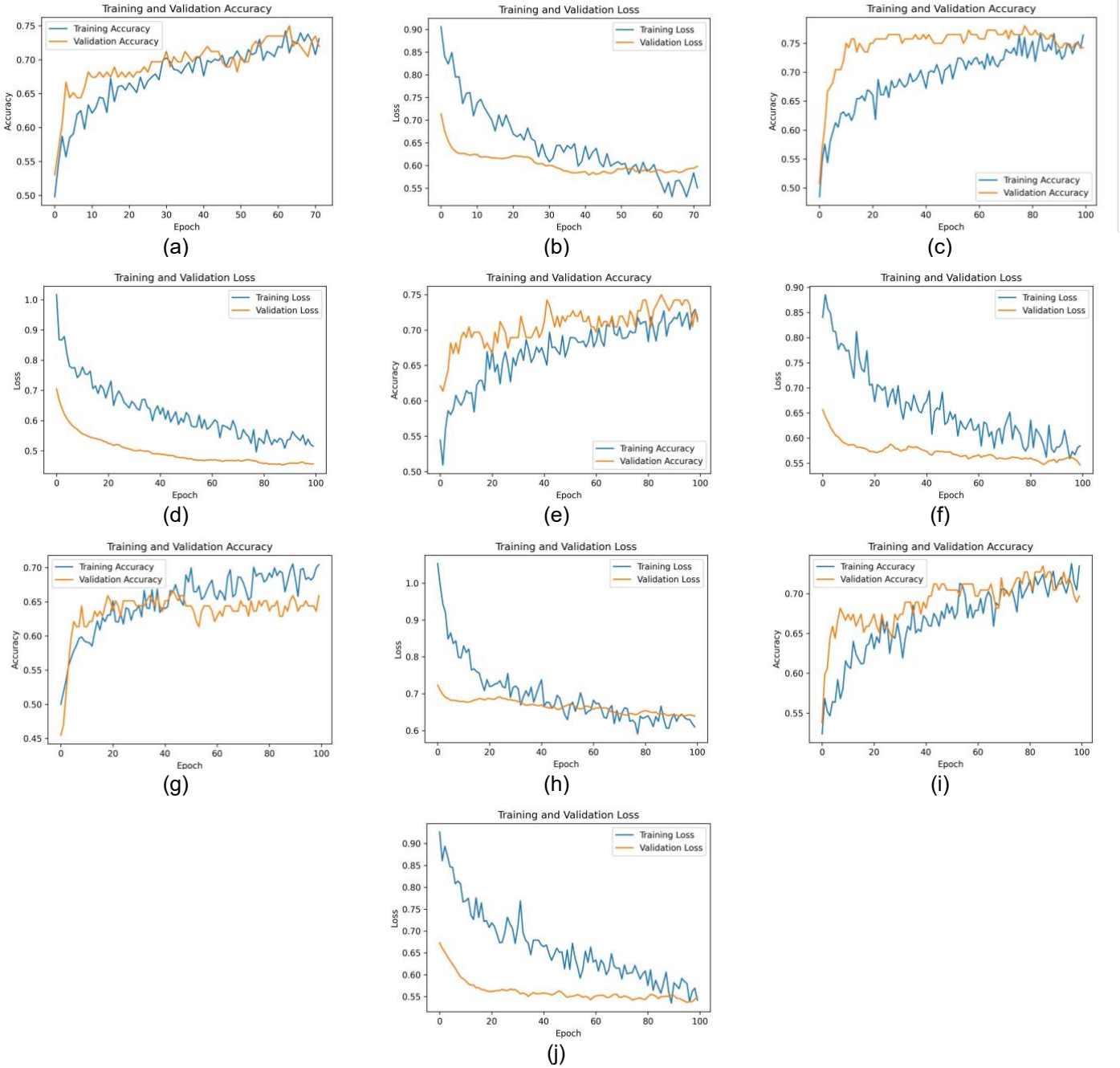
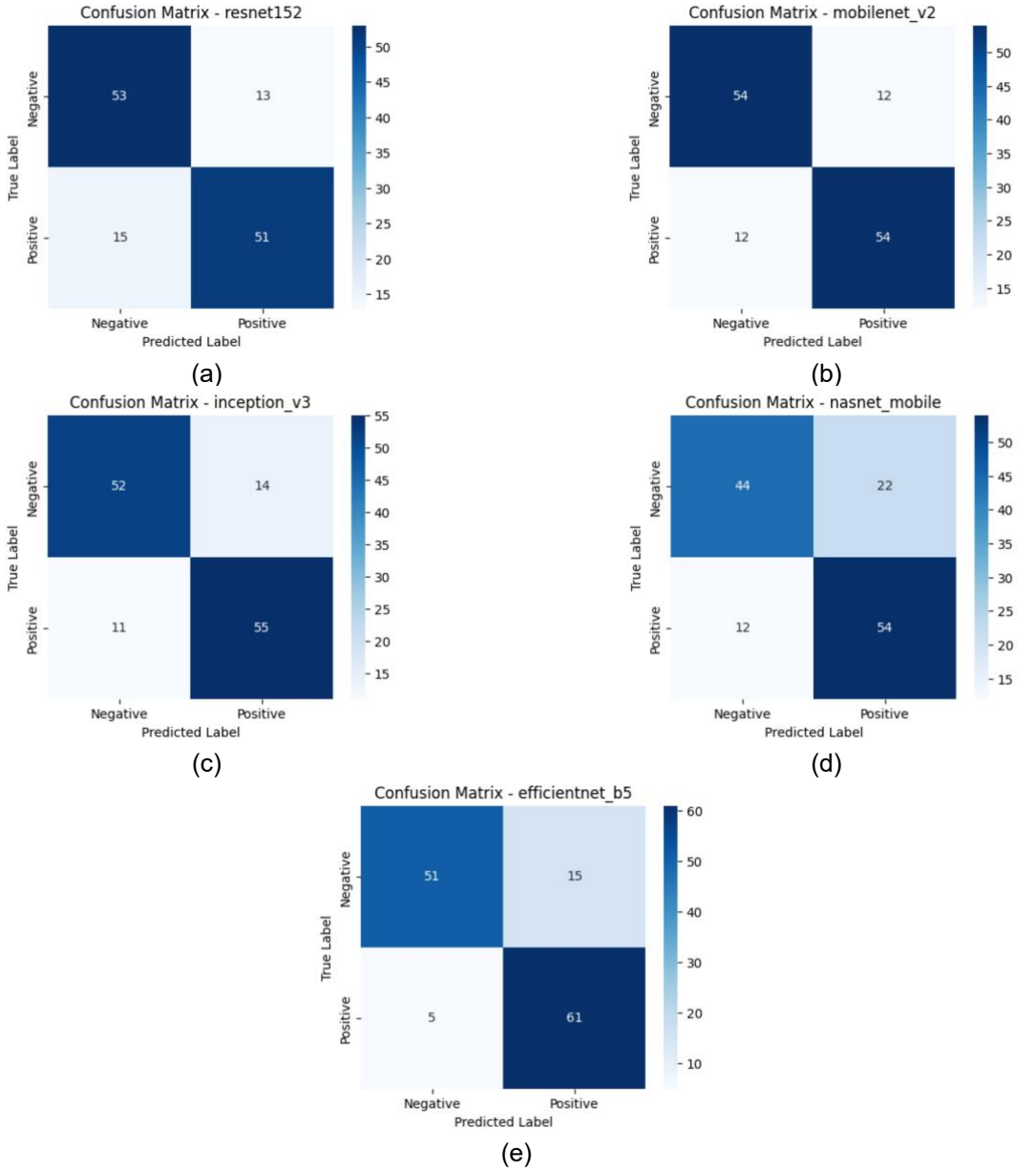


Fig. 5. Training and validation accuracy and loss curves for five CNN architectures: (a,b) ResNet152, (c,d) MobileNetV2, (e,f) InceptionV3, (g,h) NasNet Mobile, and (i,j) EfficientNetB5.

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343



(a) (b) (c) (d) (e)

Fig. 6. Confusion matrix of classification prediction results from five individual CNN architectures: (a) ResNet152, (b) MobileNetV2, (c) InceptionV3, (d) NasNet Mobile, and (e) EfficientNetB5.

trained and tested separately are combined into a hard-voting ensemble model. The majority voting mechanism was used to determine the final prediction based on the majority vote across models. The test results showed a significant performance improvement compared to an individual model. Based on Fig. 7, the confusion matrix shows that out of 132 test data in hard voting, 55 normal data were correctly predicted as true negatives, only 11 normal data were incorrectly predicted as caries false positives, 62 caries data were successfully detected as true positives, and only four caries data were misclassified as normal false negatives. the hard voting ensemble method's accuracy was 88.64%, with a precision of 84.93%, a recall of 93.94%, and an F1-score of 89.21%. In soft voting, 56 negative data were correctly predicted as true negatives, while 10 negative data were misclassified as positive false positives. Meanwhile, the accuracy is 87.12%, the precision is 85.51%, the recall is 89.39%, and the F1-score is 87.41%. On the other hand, 59 positive data were correctly identified as true positives, while seven positive data were misclassified as negative false negatives. The frequency distributions of the predicted classes from hard and soft voting are also illustrated in Fig. 7. These results show a lower classification error rate than any individual

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

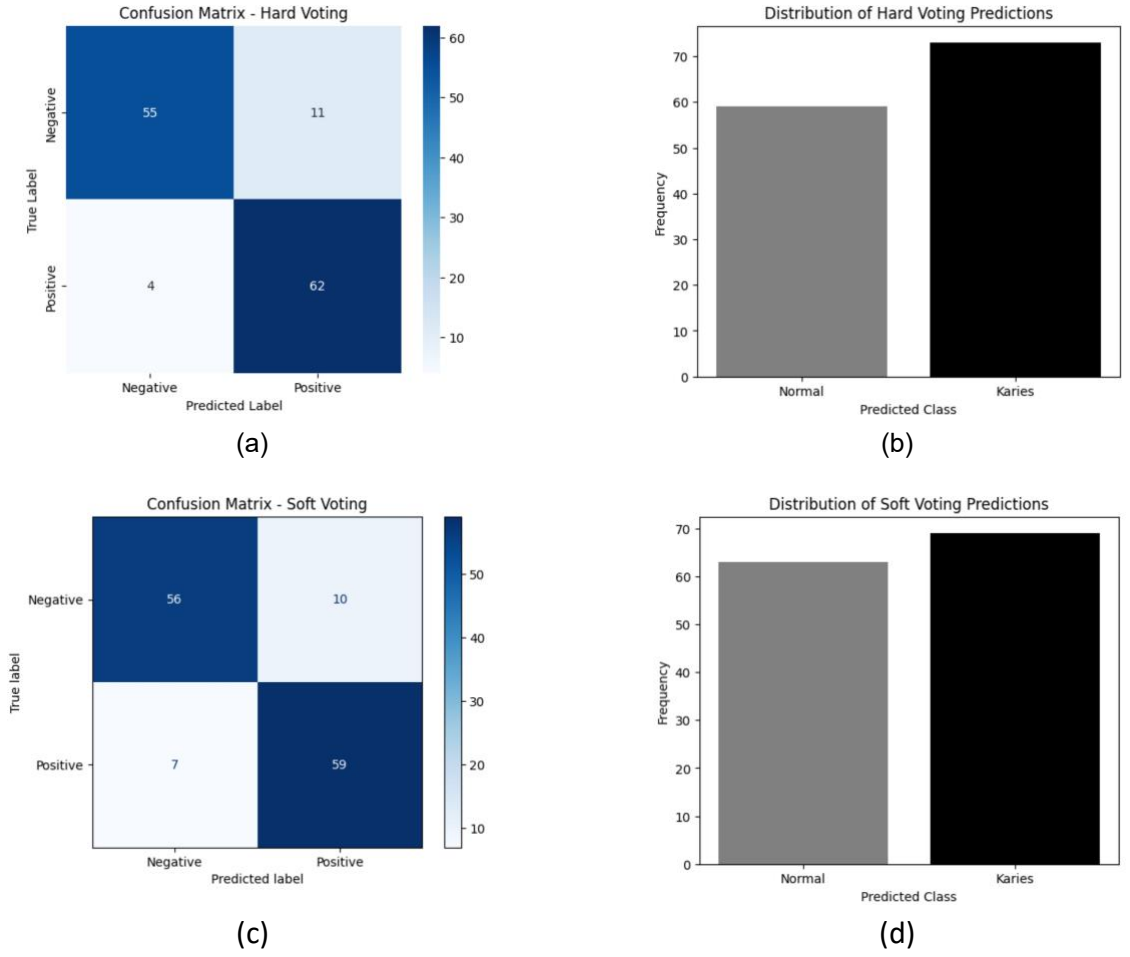


Fig. 7. Confusion matrices and prediction distributions for (a,b) hard voting and (c,d) soft voting ensemble methods.

Table 3. Individual Model Testing Results

Model	Accuracy	Precision	Recall	F1-Score
InceptionV3	0.8106	0.7971	0.8333	0.8148
ResNet152	0.7879	0.7969	0.7727	0.7846
MobileNetV2	0.8182	0.8182	0.8182	0.8182
NasNetMobile	0.7424	0.7105	0.8182	0.7606
EfficientNetB5	0.8485	0.8026	0.9242	0.8592

model. In addition, the high recall value indicates that the hard voting ensemble effectively detects caries cases, an important aspect of early diagnosis systems. The ROC and Precision–Recall curves demonstrate that ensemble methods consistently achieve higher AUC values compared to individual models. A higher ROC-AUC indicates improved discriminative power, while a higher PR-AUC reflects better performance under class imbalance, which is crucial for reliable caries detection. Fig. 8 shows the ROC and PR curves for the hard-voting ensemble, which produced an AUC of 0.89, and the soft-voting ensemble, which produced an AUC of 0.93. It indicates excellent ability to distinguish between normal and caries classes. Meanwhile, the Precision-Recall curve shows precision remains stable at various recall levels, thereby reinforcing the reliability of this method in classification.

C. Performance Comparison of Individual Models and Ensemble Model

Based on the evaluation, each deep learning model demonstrated varying performance across accuracy, precision, recall, and F1-score. EfficientNetB5 achieved the best results in individual models, showing high accuracy and recall with a low error rate. MobileNetV2 produced consistent and balanced outcomes across all metrics, while InceptionV3 performed reasonably well but with a higher false positive rate. Meanwhile, ResNet152 showed slightly lower performance. However, despite a relatively high recall, NasNet Mobile had the weakest results with low accuracy and precision. These findings indicate that

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

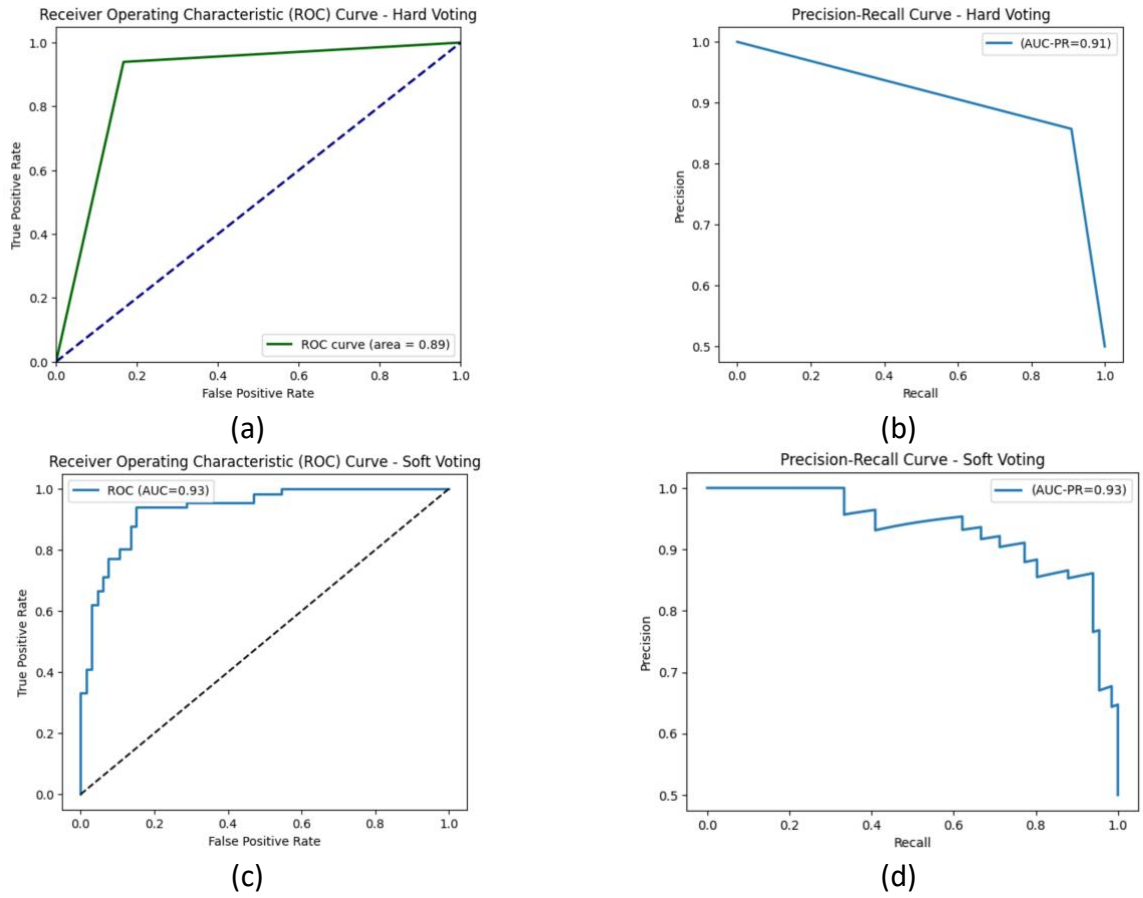


Fig. 8. Comparison of ROC and Precision-Recall curves for Vard Voting (a,b) and Soft Voting (c,d) ensemble methods.

each model has distinct strengths and weaknesses, and individually they do not yet provide optimal classification performance. To address these limitations, this study applied ensemble voting, namely hard voting and soft voting techniques, to combine multiple model predictions to achieve more accurate results. Table 4 shows the comparison of the individual model and ensemble voting. The comparison results between individual models and the ensemble approach show that hard and soft voting methods improve performance compared to all individual models. The hard voting method generally produces the best performance, with higher accuracy, precision, sensitivity, and F1-score than each model. It confirms that combining predictions from several models can reduce the weaknesses of each architecture, making the classification system more stable and reliable. Soft voting also shows competitive performance and is slightly below hard voting. Although it does not exceed the results of hard voting, this method is still superior to most single models, especially in terms of the balance between precision and sensitivity. The difference in results between hard and soft voting shows that the decision aggregation strategy affects final performance, with hard voting being more effective in this study.

Table 4. Comparison of Individual Model and Ensemble Voting Results

Model	Accuracy	Precision	Recall	F1-Score
Inception V3	0.8106	0.7971	0.8333	0.8148
ResNet152	0.7879	0.7969	0.7727	0.7846
MobileNetV2	0.8182	0.8182	0.8182	0.8182
NasNetMobile	0.7424	0.7105	0.8182	0.7606
EfficientNetB5	0.8485	0.8026	0.9242	0.8592
Hard Voting	**0.8864**	**0.8493**	**0.9394**	**0.8921**
Soft Voting	**0.8712**	**0.8551**	**0.8939**	**0.8741**

The results in the table confirm the ensemble approach's superiority over individual models. Combining the strengths of various CNN architectures, such as EfficientNetB5, InceptionV3, and MobileNetV2, the ensemble voting method produces predictions that are more accurate, consistent, sensitive, balanced, and reliable than those produced by individual models.

To assess whether the performance improvement from the ensemble method is statistically significant,

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

McNemar's test was used to compare the classification outcomes of the hard voting ensemble and the best-performing individual model, EfficientNetB5, on the same test dataset. The test used paired predictions, focusing on samples where the two models produced different classifications. The obtained p-value was below the 0.05 significance threshold, indicating a statistically significant difference between the two methods. This result confirms that the observed performance improvement in the ensemble is unlikely to be due to random variation and reflects a genuine enhancement in classification capability.

IV. Discussion

A. Deep Interpretation of Results

The experimental results show clear differences in behavior among individual CNN models, as shown in Table 3. EfficientNetB5 achieved the highest recall of 0.9242, indicating effective detection of most caries cases in the test set. This high recall indicates that EfficientNetB5 is more sensitive to various intraoral image patterns. However, the lower precision of 0.8026 indicates that a significant proportion of healthy teeth were misclassified as caries. These results show that EfficientNetB5 prioritizes sensitivity over specificity. On the other hand, MobileNetV2 shows balanced performance, with a precision and recall of 0.8182. These results reflect stable classification without noticeable bias towards either class. However, the overall accuracy and F1 score of 0.8182 are lower than those of EfficientNetB5. InceptionV3 shows a higher recall (0.8333) than precision (0.7971), indicating a tendency to favor positive predictions. NASNetMobile achieved the lowest accuracy (0.7424) and precision (0.7105). However, it achieved a high recall score of 0.8182. These results confirm that each model has unique strengths and limitations, and no single architecture achieves optimal performance across all evaluation metrics. The ensemble results in Table 4 show that combining these heterogeneous models significantly improves classification performance. The Hard Voting ensemble achieves the highest recall of 0.9394, outperforming all individual models, including EfficientNetB5. At the same time, its precision increases to 0.8493, which is much higher than that of EfficientNetB5. This improvement indicates that the voting mechanism effectively reduces false positives from highly sensitive models while preserving their ability to detect caries. As a result, the Hard Voting ensemble achieves the highest F1 score of 0.8921, reflecting more balanced and reliable classification performance. On the other hand, the Soft Voting ensemble achieved the highest precision of 0.8551, though its recall was slightly lower at 0.8939 than Hard Voting. These results indicate that Soft Voting produces more conservative predictions, reducing false positives while potentially missing some caries cases. These findings show that the ensemble strategy effectively integrates the complementary characteristics of individual models, leading to more stable performance.

Table 5. Performance comparison with related studies

Author (Year)	Method	Image Type	Accuracy
Park et al. (2022) [5]	CNN with segmentation	Intraoral photographs	83.0%
Duong et al. (2021) [6]	SVM	Smartphone images	88.76%
Moral et al. (2021) [8]	Inception CNN	Bitewing radiographs	73.3%
This study (2025)	Hard Voting (5 CNNs)	Clinical Intraoral Photographs	88.64%

B. Comparison with Existing Literature

To assess the competitiveness of the proposed method relative to previous methods, the ensemble results were compared with those of related studies on dental caries detection, as summarized in Table 5. The proposed hard voting ensemble achieved an accuracy of 88.64%, which is comparable to the SVM-based approach reported by Duong et al. [6] (88.76%) using smartphone images. Compared with Park et al. [5], who used a single CNN for tooth surface segmentation in intraoral photos and achieved an accuracy of 83%, the proposed method shows a clear performance improvement. The results of this study suggest that ensemble learning may be more effective than single-model segmentation-based approaches for intraoral image analysis. The lower accuracy reported by Moral et al. [8] (73.3%) is likely due to differences in imaging modality and problem formulation. Bitewing X-ray images reveal caries with distinctive visual features, and adding lesion severity levels increases the complexity of classification. Overall, this comparison shows that the proposed ensemble approach achieves competitive performance with noninvasive intraoral clinical photographs, supporting its feasibility for routine dental screening.

C. Limitations of the Study

While the results of this study are promising, several limitations must be acknowledged. First, the dataset comprised 1,320 images sourced from a limited regional set. Although the dataset was balanced between caries and normal classes, it may not capture broader population variability, including differences in tooth morphology, imaging equipment, or lighting conditions. Second, from a methodological standpoint,

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

the ensemble approach increases computational complexity and inference time because multiple convolutional neural network models must be evaluated for each prediction. This limitation may impede implementation in real-time clinical environments or settings with constrained resources. Additionally, ensemble models operate as black-box systems, which complicates the interpretation of the reasoning behind specific predictions. Third, model evaluation was conducted using a curated dataset with relatively high image quality. The robustness of the system under real-world conditions, including images affected by severe shadows, blurring, saliva reflections, or suboptimal angles, was not thoroughly investigated. These factors could influence performance in practical clinical applications.

D. Practical and Theoretical Implications

From a practical standpoint, the Hard Voting ensemble, which achieves a recall of 93.94%, is well-suited for deployment as a screening tool in dental clinics. High sensitivity is essential for early caries detection, as false negatives may delay intervention and allow disease progression. In contrast, the Soft Voting ensemble, which emphasizes precision, is more suitable in contexts of reducing false positives and avoiding unnecessary follow-up examinations is prioritized. From a theoretical perspective, the present study reinforces the effectiveness of ensemble learning for medical image classification. The results indicate that integrating heterogeneous convolutional neural network (CNN) architectures can reduce individual model bias, decrease variance, and enhance overall classification stability. These findings support ensemble strategies as reliable solutions for complex diagnostic tasks. Future research should prioritize improving interpretability through explainable artificial intelligence (AI) methods and reducing computational costs to enable clinical implementation.

V. Conclusion

This study proves that the application of the ensemble hard voting method by combining five CNN models, namely ResNet152, MobileNetV2, InceptionV3, Nas Net Mobile, and EfficientNetB05, successfully improves the performance of the dental caries detection system. The evaluation results show that the ensemble model has higher accuracy, precision, recall, and F1-score than the individual model. The ensemble majority hard voting proved highly effective by achieving the highest accuracy of 88.64% and a good balance between precision and recall. It confirms that combining various architectures can utilize each model's strengths, overcome Individual models' weaknesses, improve system stability, and strengthen the ability to generalize to new data. Thus, the research objective of improving the performance of the dental caries detection system was achieved through the ensemble learning approach. The main contribution of this research is to provide empirical evidence that the voting method can be an effective solution in overcoming the limitations of Individual models in medical image classification. With superior results, ensemble hard voting has the potential to serve as a fast, accurate, non-invasive diagnostic tool in image-based caries detection systems. For further research, it is recommended to explore other methods, such as soft voting or stacking ensembles, as well as the use of larger and more varied datasets, so that the system can be more robust and improve its generalization capabilities for clinical applications.

Future research should explore more advanced ensemble strategies, such as weighted soft voting and stacking-based ensemble methods with meta-learners, in conjunction with larger, more diverse datasets to further improve classification performance, generalization, and clinical applicability.

Acknowledgment

This research was supported by Direktorat Penelitian dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia, year 2025, with number 189/UN11.L1/PG.01.03/DPPM/2025. We would like to express our gratitude to all dental and medical experts at Dental and Oral, Regional General Hospital, Dr. Zaineol Abidin (RSUDZA), who assisted, helped, and participated throughout the research process.

Funding

Direktorat Penelitian dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia, year 2025, supported this research with number 189/UN11.L1/PG.01.03/DPPM/2025.

Data Availability

No datasets were generated or analyzed during the current study.

Author Contribution

Oktiana M conceptualized and secured funding for the study, administered the project, and supervised all stages. Rizkiah P managed data curation, investigation, methodology, software development, and visualization, and drafted the original manuscript. Saddami K performed formal analysis and validation. Fitria M provided the necessary resources. Oktiana M, Saddami

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

K, Arnia F, Walidainy H, and Yunida Y contributed to manuscript review and editing. All authors reviewed and approved the final manuscript, taking responsibility for its accuracy and integrity.

Declarations

Ethical Approval

This study, which utilizes a combined dataset from direct patient interactions at the Dental and Oral Hospital of Aceh and a pre-existing dataset on caries from a prior study by Maya Fitria and team, received ethical approval from the competent ethics committee at the Dental and Oral Hospital of Aceh. All methods were carried out in accordance with relevant ethical guidelines and regulations.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] K. C. Li *et al.*, "Detection of Tooth Position by YOLOv4 and Various Dental Problems Based on CNN With Bitewing Radiograph," *IEEE Access*, vol. 12, no. 2023, pp. 11822–11835, 2024, doi: 10.1109/ACCESS.2023.3348788.

[2] V. P. Mathur and J. K. Dhillon, "Dental Caries: A Disease Which Needs Attention," *Indian J. Pediatr.*, vol. 85, no. 3, pp. 202–206, 2018, doi: 10.1007/s12098-017-2381-6.

[3] R. Alsubhi, H. Alsharif, H. Kadi, and M. Barashi, "Dental Crowding Prediction from Occlusal View Images Using Deep Learning," *2024 IEEE Int. Conf. Autom. Control Intell. Syst. I2CACIS 2024 - Proc.*, no. June, pp. 95–100, 2024, doi: 10.1109/I2CACIS61270.2024.10649861.

[4] I. D. S. Chen, C. M. Yang, M. J. Chen, M. C. Chen, R. M. Weng, and C. H. Yeh, "Deep Learning-Based Recognition of Periodontitis and Dental Caries in Dental X-ray Images," *Bioengineering*, vol. 10, no. 8, pp. 1–13, 2023, doi: 10.3390/bioengineering10080911.

[5] E. Y. Park, H. Cho, S. Kang, S. Jeong, and E. K. Kim, "Caries detection with tooth surface segmentation on intraoral photographic images using deep learning," *BMC Oral Health*, vol. 22, no. 1, pp. 1–9, 2022, doi: 10.1186/s12903-022-02589-1.

[6] D. L. Duong, Q. D. N. Nguyen, M. S. Tong, M. T. Vu, J. D. Lim, and R. F. Kuo, "Proof-of-concept study on an automatic computational system in detecting and classifying occlusal caries lesions from smartphone color images of unrestored extracted teeth," *Diagnostics*, vol. 11, no. 7, 2021, doi: 10.3390/diagnostics11071136.

[7] L. Zheng, H. Wang, L. Mei, Q. Chen, Y. Zhang, and H. Zhang, "Artificial intelligence in digital cariology: a new tool for the diagnosis of deep caries and pulpitis using convolutional neural networks," *Ann. Transl. Med.*, vol. 9, no. 9, pp. 763–763, 2021, doi: 10.21037/atm-21-119.

[8] M. Moran, M. Faria, G. Giraldi, L. Bastos, L. Oliveira, and A. Conci, "Classification of approximal caries in bitewing radiographs using convolutional neural networks," *Sensors*, vol. 21, no. 15, pp. 1–12, 2021, doi: 10.3390/s21155192.

[9] R. Krishnaveni and R. Sudarmani, "Classification of Dental Disease through various CNN Techniques and Performance Analysis," *Proc. 2024 Int. Conf. Sci. Technol. Eng. Manag. ICSTEM 2024*, 2024, doi: 10.1109/ICSTEM61137.2024.10560940.

[10] D. Saini, R. Jain, and A. Thakur, "Dental Caries early detection using Convolutional Neural Network for Tele dentistry," in *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, 2021. doi: 10.1109/ICACCS51430.2021.9442001.

[11] M. K. Al Yassar *et al.*, "The Role of U-Net Segmentation for Enhancing Deep Learning-based Dental Caries Classification," *Indones. J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 2, pp. 253–269, 2025, doi: 10.35882/ijeeemi.v7i2.75.

[12] M. Fitria, Y. Elma, M. Oktiana, K. Saddami, and R. Novita, "The Deep Learning Model for Decayed - Missing - Filled Teeth Detection: A Comparison Between Yolo V5 And Yolo V8," vol. 10, no. 03, pp. 335–349, 2024.

[13] M. Muhajir, K. Muchtar, M. Oktiana, and A. Bintang, "Students' emotion classification system through an ensemble approach," *SINERGI*, vol. 28, p. 413, 2024, doi: 10.22441/sinergi.2024.2.020.

[14] A. Desiani *et al.*, "Majority Voting as Ensemble Classifier for Cervical Cancer Classification," *Sci. Technol. Indones.*, vol. 8, no. 1, 2023, doi: 10.26554/sti.2023.8.1.84-92.

[15] K. Cao-Van, T. C. Minh, L. G. Minh, T. T. B. Quyen, and H. M. Tan, "Soft-Voting Ensemble Model: An Efficient Learning Approach for Predictive Prostate Cancer Risk," *Vietnam J. Comput. Sci.*, vol. 11, no. 4, pp. 531–552, Nov. 2024, doi: 10.1142/S2196888824500155.

[16] S. Kumari, D. Kumar, and M. Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier," *Int. J. Cogn. Comput. Eng.*, vol. 2, 2021, doi:

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

10.1016/j.ijcce.2021.01.001.

[17] Q. M. Ilyas and M. Ahmad, "An Enhanced Ensemble Diagnosis of Cervical Cancer: A Pursuit of Machine Intelligence Towards Sustainable Health," *IEEE Access*, vol. 9, pp. 12374–12388, 2021, doi: 10.1109/ACCESS.2021.3049165.

[18] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," 2024. doi: 10.1016/j.eswa.2023.122778.

[19] A. Manconi, G. Armano, M. Gnocchi, and L. Milanesi, "A Soft-Voting Ensemble Classifier for Detecting Patients Affected by COVID-19," *Appl. Sci.*, vol. 12, no. 15, 2022, doi: 10.3390/app12157554.

[20] Y. M. Alsakar, N. Elazab, N. Nader, W. Mohamed, M. Ezzat, and M. Elmogy, "Multi-label dental disorder diagnosis based on MobileNetV2 and swin transformer using bagging ensemble classifier," *Sci. Rep.*, vol. 14, no. 1, p. 25193, 2024, doi: 10.1038/s41598-024-73297-9.

[21] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble," *Expert Syst. Appl.*, vol. 165, Mar. 2021, doi: 10.1016/j.eswa.2020.113909.

[22] J. Kang, Z. Ullah, and J. Gwak, "Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, 2021, doi: 10.3390/s21062222.

[23] I. D. Mienye and Y. Sun, *A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects*, vol. 10, no. September. IEEE, 2022, pp. 99129–99149. doi: 10.1109/ACCESS.2022.3207287.

[24] S. Lee, S. il Oh, J. Jo, S. Kang, Y. Shin, and J. won Park, "Deep learning for early dental caries detection in bitewing radiographs," *Sci. Rep.*, vol. 11, no. 1, pp. 1–8, 2021, doi: 10.1038/s41598-021-96368-7.

[25] I. Lasri, N. El-Marzouki, A. Riadsolh, and M. Elbelkacemi, "Automated Detection of Dental Caries from Oral Images using Deep Convolutional Neural Networks," *Int. J. online Biomed. Eng.*, vol. 19, no. 18, pp. 53–70, 2023, doi: 10.3991/ijoe.v19i18.45133.

[26] R. A. Welikala *et al.*, "Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer," *IEEE Access*, vol. 8, pp. 132677–132693, 2020, doi: 10.1109/ACCESS.2020.3010180.

[27] T. Kim, K. Oh, J. Kim, Y. Lee, and J. Choi, "Development of ResNet152 UNet++-Based Segmentation Algorithm for the Tympanic Membrane and Affected Areas," *IEEE Access*, vol. 11, no. May, pp. 56225–56234, 2023, doi: 10.1109/ACCESS.2023.3281693.

[28] M. Akay *et al.*, "Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model," vol. 2, pp. 104–110, 2021, doi: 10.1109/OJEMB.2021.3066097.

[29] A. Sharma and R. Parvathi, "Enhancing Cervical Cancer Classification: Through a Hybrid Deep Learning Approach Integrating DenseNet201 and InceptionV3," *IEEE Access*, vol. 13, no. December 2024, pp. 9868–9878, 2025, doi: 10.1109/ACCESS.2025.3527677.

[30] F. A. Shah, M. A. Khan, M. Sharif, U. Tariq, and A. Khan, "A Cascaded Design of Best Features Selection for Fruit Diseases Recognition A Cascaded Design of Best Features Selection for Fruit Diseases Recognition," no. March, 2023, doi: 10.32604/cmc.2022.019490.

[31] M. M. Shahriar Maswood, T. Hussain, M. B. Khan, M. T. Islam, and A. G. Alharbi, "CNN Based Detection of the Severity of Diabetic Retinopathy from the Fundus Photography using EfficientNet-B5," *11th Annu. IEEE Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2020*, no. November, pp. 147–150, 2020, doi: 10.1109/IEMCON51383.2020.9284944.

[32] M. Fitria *et al.*, "Development of Intraoral Clinical Image Dataset for Deep Learning Caries Detection," in *Proceeding - 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering: Sustainable Development for Smart Innovation System, COSITE 2023*, 2023. doi: 10.1109/COSITE60233.2023.10249428.

[33] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," 2022. doi: 10.3390/app12188972.

[34] D. Popescu, A. Stanciulescu, M. D. Pomohaci, and L. Ichim, "Decision Support System for Liver Lesion Segmentation Based on Advanced Convolutional Neural Network Architectures," *Bioengineering*, vol. 9, no. 9, 2022, doi: 10.3390/bioengineering9090467.

[35] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-Stage CNN Architecture for Face Mask Detection," in *2021 6th International Conference for Convergence in Technology, I2CT 2021*, 2021. doi: 10.1109/I2CT51068.2021.9418207.

[36] M. Ardi, "MobileNetV2 Paper Walkthrough: The Smarter Tiny Giant | Towards Data Science."

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

Accessed: Feb. 07, 2026. [Online]. Available: https://towardsdatascience.com/mobilenetv2-paper-walkthrough-the-smarter-tiny-giant/

[37] L. Xu, S. S. Teoh, and H. Ibrahim, "A deep learning approach for electric motor fault diagnosis based on modified InceptionV3," *Sci. Rep.*, vol. 14, no. 1, pp. 1–15, 2024, doi: 10.1038/s41598-024-63086-9.

[38] P. Bhardwaj *et al.*, "Advanced CNN models in gastric cancer diagnosis: enhancing endoscopic image analysis with deep transfer learning," *Front. Oncol.*, vol. 14, no. September, pp. 1–21, 2024, doi: 10.3389/fonc.2024.1431912.

[39] M. Tan *et al.*, "Mnasnet: Platform-aware neural architecture search for mobile," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 2815–2823, 2019, doi: 10.1109/CVPR.2019.00293.

[40] A. O. Adedoja, P. A. Owolawi, T. Mapayi, and C. Tu, "Intelligent Mobile Plant Disease Diagnostic System Using NASNet-Mobile Deep Learning," *IAENG Int. J. Comput. Sci.*, vol. 49, no. 1, pp. 216–231, 2022.

[41] Abdullahi Lawal Rukuna *et al.*, "Enhancing Monkeypox Detection with Efficientnet-B5 And Image Augmentation Fusion Technique," *Int. J. Sci. Res. Sci. Technol.*, vol. 11, no. 6, pp. 646–661, 2024, doi: 10.32628/ijsrst241161119.

[42] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.

[43] D. Turimov and W. Kim, "Enhancing Sarcopenia Prediction Through an Ensemble Learning Approach: Addressing Class Imbalance for Improved Clinical Diagnosis," *Mathematics*, vol. 13, no. 1, p. 26, 2025, doi: 10.3390/math13010026.

[44] A. Mahabub, "A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers," *SN Appl. Sci.*, vol. 2, no. 4, pp. 1–9, 2020, doi: 10.1007/s42452-020-2326-y.

[45] S. N. Sindhu and R. S. Prasad, "Dental Caries Detection Using Neural Turing Machines (NTM) and High Intensity Color Detection (NTM-HICD) Model," *Rev. d'Intelligence Artif.*, vol. 38, no. 2, pp. 671–679, 2024, doi: 10.18280/ria.380231.

Author Biography

Putri Rizkiah received a Bachelor of Engineering (S.T.) in Electrical Engineering from Syiah Kuala University in Indonesia in 2010. She is currently pursuing a master's degree in electrical engineering at the Faculty of Engineering, Syiah Kuala University. Her research centers on the application of artificial intelligence, with a particular emphasis on deep learning and machine learning, to address challenges in electrical engineering, biomedical engineering, and digital health. Her thesis develops a dental caries detection system that uses ensemble learning to improve the accuracy and reliability of early diagnosis from clinical intraoral images. Her broader research interests include medical image processing, the development of computer-aided diagnostic tools, and pattern recognition.

Maulisa Oktiana received her Bachelor's degree in Electrical Engineering (S.T.) from Syiah Kuala University (USK) in 2013 and completed her Ph.D. in Electrical and Computer Engineering at the same institution in 2020. She was awarded a scholarship by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia under the PMDSU scheme. In November 2018, she participated in an academic exchange at Chiba University. She is currently a lecturer in the Department of Electrical Engineering and Computer Science at Syiah Kuala University. Her research interests focus on image processing, biometrics, pattern recognition, deep learning, machine learning, and artificial intelligence.

Khairun Saddami obtained a Bachelor's degree in 2015 from Syiah Kuala University, Indonesia. He received his PhD in Electrical and Computer Engineering from Syiah Kuala University, Indonesia, in 2020, where he was awarded a scholarship from the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia under the Scheme of Pendidikan Magister Menuju Doktor Untuk Sarjana Unggul (PMDSU). He has been with the Multimedia and Signal Processing Research Group (MUSIG) at Syiah Kuala University since 2020. He also acts as a reviewer in reputable journals such as IEEE Access and ACM Computer Survey. He is a member of IEEE and the International Association for Pattern Recognition (IAPR). His research interests are in document image analysis, deep learning, biometric and biomedical image processing, and pattern recognition.

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343

Maya Fitria began her career as a lecturer and member of the Department of Electrical Engineering and Computer Science at Syiah Kuala University in 2017 and has been there ever since. She earned her Bachelor's degree in Computer Science from the University of Indonesia (UI) in 2012. In 2013, she continued her studies at the Department of Computer Engineering at the University of Duisburg-Essen, Germany, specializing in Interactive Systems and Visualization. During her studies, she received support from the DAAD-LPSDM Aceh Scholarship. She completed her master's degree in 2016, earning a Master of Science in Computer Engineering. Her research interests are in Human-Centered AI and IoT for Diagnostic and Interactive Systems.

Fitri Arnia received a B. Eng degree from Universitas Sumatera Utara (USU) Medan. She completed her master's and doctoral degrees at the University of New South Wales (UNSW), Sydney, Australia, and Tokyo Metropolitan University, Japan. She is a professor of multimedia signal processing in the Department of Electrical Engineering, Faculty of Engineering, Syiah Kuala University. She was a visiting researcher at Tokyo Metropolitan University in Japan and at Suleyman Demirel University in Isparta, Turkey. She has been awarded the AusAID Scholarship, the Hashiya Scholarship, and the Latvian Government's Latvian Fellowship for Research. Since 2008, she has served as the Editor-in-Chief for Jurnal Rekayasa Elektrika (Accredited by RISTEKDIKTI). Her research interests include signal, image, and multimedia information processing. She is a member of PII, IEEE, ACM, and APSIPA.

Yunida received the B.Eng degree in Electrical Engineering from Syiah Kuala University, Banda Aceh, Indonesia, in 2013. Then she received her PhD in Electrical and Computer Engineering from Syiah Kuala University in 2020 through the "Magister Program of Education Leading to Doctoral for Excellent Graduates (PMDSU)" scholarship from the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia. She is currently a lecturer at the Department of Electrical and Computer Engineering, Faculty of Engineering, Syiah Kuala University, Banda Aceh, Aceh, Indonesia. Since 2016, she has published about six articles in Scopus-Indexed Journals. Her research interests include digital communications, wireless communications, and information theory.

Manuscript received 14 November 2025; Revised 10 February 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1343