



Facial Acne Multi-Class Classification Using the EfficientNetV2S Deep Learning Model

Aldi Yogie Pramono¹, Kusnawi, S.Kom., M.Eng.^{2*},

^{1,2} Department of Informatics, Amikom University Yogyakarta, Indonesia

Article Info	ABSTRACT
Article history: Submitted July 10, 2025 Accepted Published	Acne vulgaris is a common skin disorder that significantly affects adolescents and young adults, often impacting their self-esteem and quality of life. Accurate classification of acne types is essential for selecting appropriate treatment plans. This study proposes an automatic multi-class facial acne classification system using a Convolutional Neural Network (CNN) model based on the EfficientNetV2S architecture. The dataset consists of 4,673 annotated acne images categorized into five types: Blackheads, Whiteheads, Papules, Pustules, and Cyst. All images were preprocessed using normalization and data augmentation techniques. The model was trained using transfer learning from ImageNet weights and fine-tuned with custom classification layers. Experimental results on the separate test dataset demonstrate a high classification performance, achieving an overall accuracy of 97%, with macro average precision, recall, and F1-score all reaching 0.97. Notably, the model performs consistently across all acne categories, including the minority class. These findings indicate that the EfficientNetV2S-based CNN model is highly effective in classifying facial acne types and has strong potential as a decision support tool in dermatological diagnosis. This approach can assist dermatologists and telemedicine applications in providing timely and accurate acne classification from facial images.
Keywords: Acne classification; Deep learning; EfficientNetV2S; Convolutional Neural Network.	 
Corresponding Author: Kusnawi, S.Kom., M.eng., Department of Informatics, Amikom University Yogyakarta, Indonesia, Jl. Ring Road Utara, Condongcatur, Depok, Sleman, Yogyakarta 55281, Indonesia. Email: * khusnawi@amikom.ac.id	

1. INTRODUCTION

Skin diseases remain among the most prevalent health concerns worldwide, affecting individuals of various age groups and socioeconomic statuses. Acne vulgaris, a chronic inflammatory disorder of the pilosebaceous unit, is particularly common in adolescents and young adults. Beyond physical symptoms, acne significantly impacts psychological health, contributing to low self-esteem, anxiety, and depression (Martin & Udjulawa, 2024). In Indonesia, more than 60% of university students experience acne, highlighting the urgent need for scalable and efficient diagnostic solutions [1].

Conventional acne diagnosis largely depends on manual visual assessment conducted by dermatologists or general practitioners. Although commonly practiced, this method is inherently subjective, depending on the practitioner's expertise, focus, and interpretive judgment, which may result in inconsistent diagnoses and ineffective treatment plans [2], [3]. Furthermore, limited access to dermatologists—especially in remote or underserved areas—intensifies disparities in healthcare delivery [4]. This situation highlights the need for an automated, objective, and consistent acne classification system to support clinical decision-making processes.

In recent years, advancements in artificial intelligence (AI), particularly through Convolutional Neural Networks (CNNs), have provided new avenues for automated image-based diagnosis. CNNs are effective in medical image analysis due to their ability to learn spatial hierarchies and extract visual features [5], [6]. Several studies have reported classification accuracies exceeding 90% in dermatological applications [7]. However, much of the existing research focuses on limited disease categories such as melanoma or monkeypox and employs outdated architectures like VGG16 or AlexNet, which are less efficient in modern computing environments [8], [9], [10].

In Indonesia, multiple studies have begun exploring advanced CNN architectures such as MobileNetV2, ResNet, and EfficientNet by incorporating hyperparameter tuning and transfer learning to boost performance [11], [12], [13]. Although these developments mark progress, most experiments still rely on homogeneous

datasets that do not reflect real-world variability in image quality, lighting, and background [14]. Moreover, these studies often limit classification to just two or three acne categories, limiting their clinical applicability ([15], [16]).

To address these limitations, the present study employs EfficientNetV2-S, a state-of-the-art CNN architecture that utilizes compound scaling to simultaneously optimize network depth, width, and input resolution [17], [18]. The model is trained on a dataset consisting of real-world facial images, annotated by dermatology experts and augmented to reflect diverse image conditions. It is designed to classify acne into five clinically recognized categories: blackheads, whiteheads, papules, pustules, and cysts. Performance is evaluated using precision, recall, and F1-score for each class to obtain detailed performance insights [19]. Compared to previous studies, this research offers a broader classification scope, evaluates performance under heterogeneous imaging conditions, and benchmarks the model against established architectures such as MobileNet and ResNet [15], [16]. The proposed approach offers strong potential for integration into mobile health platforms and teledermatology systems, contributing to the advancement of AI-driven dermatological care in Indonesia.

2. RESEARCH METHODS

This study employed a quantitative experimental approach using facial acne image data, which were classified into five categories through a Convolutional Neural Network (CNN) model based on the EfficientNetV2-S architecture. To ensure the reliability and validity of the findings, the research was conducted through a structured and systematic workflow. The stages included data acquisition, preprocessing, image augmentation, dataset splitting, model training using EfficientNetV2-S, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. The complete research flow is illustrated in Figure 1.

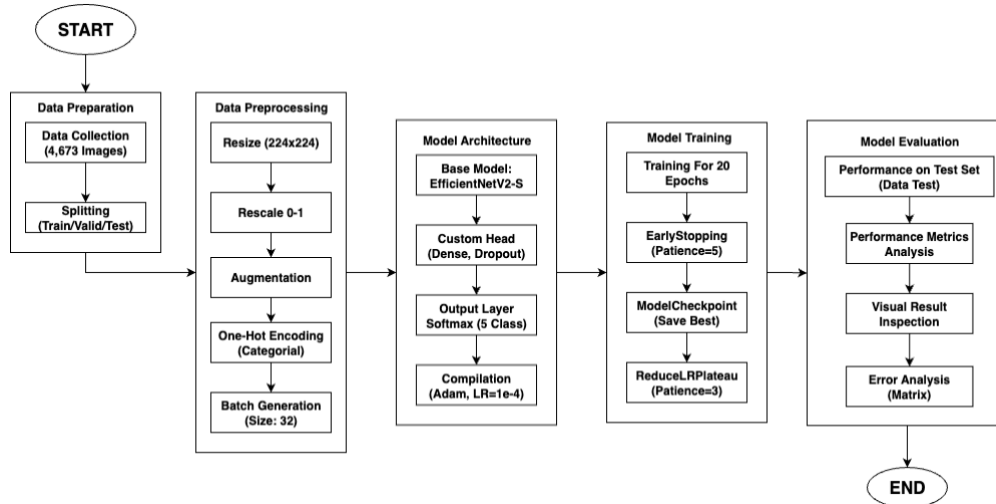


Figure 1. The Research Workflow

2.1 Data Collection

The dataset used in this study comprises a total of 4,673 facial images exhibiting various types of acne, categorized into five clinically recognized classes: blackheads, whiteheads, papules, pustules, and cysts. This dataset was obtained from a public repository on Kaggle, originally sourced from DermNet, and has been modified to suit the specific requirements of acne classification research based on digital facial images. It has also been utilized in previous studies involving deep learning models such as ResNet and MobileNet, enabling direct performance comparison with the EfficientNetV2-S architecture proposed in this work. The dataset reflects real-world variability in lighting, background, and skin tones, and all images have been annotated by dermatology experts. The data is split into three main subsets, including 2,834 images for training, 921 images for validation, and 918 images for testing, with the test set stored separately and used exclusively for final model evaluation in the multi-class classification task.

Table 1. Dataset Distribution

No.	Detail Dataset	Data		
		Training	Validation	Test
1	Facial Acne Dataset Total Images: 4,617 Class: Blackheads, Whiteheads, Papules, Pustules, Cysts	2,834	921	918



Figure 2. Sample images from the research dataset

2.2 Data Preprocessing

Data preprocessing is a fundamental stage in the deep learning modeling workflow to ensure the quality and consistency of the input data. This stage directly influences the model's ability to learn and generalize, a practice commonly adopted in various medical image classification studies [10], [13]. In this research, the preprocessing of acne image data consisted of three sequential steps: image resizing, pixel value normalization, and strategically applied data augmentation.

2.2.1 Image Resizing

Convolutional Neural Network (CNN) models, including the EfficientNetV2S architecture used in this study, require input data with uniform spatial dimensions [19]. The acne image dataset used contains images of various resolutions, thus necessitating size standardization. Every image in the training, validation, and test sets was resized to 224x224 pixels. This dimension was specifically chosen to match the pre-trained EfficientNetV2S base model's architecture, an approach that has proven effective in other skin disease classification studies [5], [13]. This standardization ensures that the tensor generated from each image has a consistent shape, allowing for efficient batch processing by the model.

2.2.2 Pixel Normalization

The next step is normalization, which aims to scale the range of pixel intensity values. Digital images typically have pixel values in the integer range of [0, 255]. This large range can slow down and disrupt the stability of the model's training process [4]. To address this, each pixel value in the images was normalized to the floating-point range of [0, 1] by dividing each pixel's value by 255.0. This rescaling process was applied to the entire dataset (training, validation, and test) and is a standard step in CNN implementations for image classification [16], [20]. Normalization ensures that all features (pixels) are on the same scale, thereby preventing disproportionate gradients and accelerating convergence during the optimization process.

2.2.3 Data Augmentation

Data scarcity is a common challenge in deep learning modeling that can lead to overfitting ([11]). To overcome this and to enhance the model's robustness against variations in real-world data, data augmentation techniques were applied. It is important to note that augmentation was applied only to the training set to enrich

data variety without introducing bias into the evaluation data, a crucial methodological principle [12]. The augmentation transformations were performed dynamically (on-the-fly), an approach also implemented in other acne classification studies to improve performance [21]. The applied transformations included:

- a. Random Rotation: Images were randomly rotated within a range of -15 to +15 degrees. This simulates capturing images from slightly different viewing angles.
- b. Random Zoom: Images were randomly zoomed in by up to 20%. This transformation helps the model recognize objects at various scales or distances.
- c. Horizontal Flip: Images were randomly flipped along the horizontal axis. For dermatological images, this is a logical transformation as acne can appear on either side of the face, thus creating realistic data variations.
- d. Brightness Adjustment: The brightness of the images was randomly modified, with a brightness factor ranging from 0.8 (darker) to 1.2 (brighter). This makes the model more tolerant to different lighting conditions during image capture.

The application of various augmentation techniques such as rotation and flipping has been proven effective for improving model generalization in skin disease classification tasks [5], [13]. By applying these augmentations, the model "sees" slightly different versions of the images in each training epoch. This effectively increases the size of the training dataset virtually and forces the model to learn features that are invariant to changes in position, scale, orientation, and lighting, ultimately reducing the risk of overfitting and improving the model's generalization ability on previously unseen data [18].

2.3 Model Architecture

This study employs the EfficientNetV2-S architecture for multi-class acne classification. EfficientNetV2-S is a convolutional neural network (CNN) that implements a compound scaling technique, which concurrently optimizes the network's depth, width, and resolution to enhance accuracy and computational efficiency. It also applies a progressive learning strategy, wherein input resolution and regularization are gradually increased during training, leading to better generalization and faster convergence [17], [18]. This combination makes the model well-suited for tasks involving high intra-class variability, such as dermatological image analysis.

The model was initialized using ImageNet-pretrained weights, leveraging generalized visual features learned from large-scale natural image datasets. To adapt to the acne classification task, the entire backbone was unfrozen and fine-tuned on a custom annotated dataset. All facial acne images were resized to $224 \times 224 \times 3$ and passed through the EfficientNetV2-S feature extractor. The original classification layers were replaced with a custom head comprising a GlobalAveragePooling2D layer, followed by two Dense layers with 512 and 128 neurons, respectively. Each Dense layer was followed by BatchNormalization and Dropout (rates of 0.4 and 0.3), promoting training stability and reducing overfitting.

The final output layer was a Dense layer with five units activated by a softmax function, allowing the model to classify images into five acne categories: blackheads, whiteheads, papules, pustules, and cysts. The full architecture consists of 21,056,101 parameters, of which 20,900,949 are trainable. This design reflects a balance between the power of a pretrained feature extractor and the specificity of a custom classifier. It follows current best practices in deep learning for medical imaging, particularly the use of transfer learning, architectural regularization, and modular output heads to improve accuracy, robustness, and generalizability [11], [21].

2.4 Model Training and Optimization

The model training process was configured and executed with a strategy aimed at achieving optimal performance while preventing overfitting. The model was compiled using the Adam optimizer with an initial learning rate of 1×10^{-4} , and the categorical crossentropy loss function, which is suitable for multi-class classification tasks as it measures the divergence between predicted probabilities and true class labels. Model performance was monitored during training using accuracy, precision, and recall, reflecting both overall and class-specific effectiveness in identifying acne categories [20], [22].

Training was conducted over a maximum of 20 epochs with a batch size of 32. To enhance efficiency and model robustness, several callback mechanisms were employed. The ModelCheckpoint callback was configured to save the model's weights only when the validation loss improved, ensuring retention of the best-performing model. To further mitigate overfitting, EarlyStopping was implemented with patience=5, halting training when no validation improvement was observed, and restoring the best weights recorded. Additionally, ReduceLROnPlateau dynamically adjusted the learning rate by a factor of 0.5 if the validation loss stagnated over three epochs, allowing the optimizer to make finer weight updates during later training stages [11], [13].

Model evaluation was carried out using a held-out test set that had not been exposed during training or validation. Metrics such as overall accuracy, per-class precision and recall, and a confusion matrix were used to assess the model's ability to distinguish among the five acne types: blackheads, whiteheads, papules, pustules,

and cysts. This evaluation protocol reflects standard practices in medical image analysis, emphasizing the importance of reliable classification in clinically relevant contexts [21].

2.5 Evaluation

Evaluation serves as a crucial stage after the model training process, aiming to assess the model's ability to produce accurate and reliable predictions. In this study, the model's classification performance is evaluated using four standard metrics: accuracy, precision, recall, and F1-score. These metrics are widely used in image classification tasks, particularly in the medical field, where the ability to distinguish between similar conditions is essential [13], [21].

Each metric is derived from the values in the confusion matrix, which compares the predicted labels against the true labels. The core components are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Accuracy measures the overall correctness of the model across all classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision indicates the proportion of relevant instances among the retrieved results for a class.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall measures the model's ability to identify all actual relevant instances of a class.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score provides a balanced measure by combining both precision and recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

For this multi-class classification problem, the metrics of Precision, Recall, and F1-Score are calculated on a per-class basis (using a one-vs-rest approach), and the overall performance is presented using macro and weighted averages in the classification report.

The evaluation protocol was comprehensive. A confusion matrix was generated to visualize the distribution of predictions across all classes, allowing for the identification of commonly misclassified categories. A classification report was used to present the detailed per-class precision, recall, and F1-score. Finally, as part of the qualitative evaluation, a visual inspection of 25 randomly selected prediction results was conducted to illustrate how the model performs on real examples and to verify its ability to generalize..

3. RESULTS AND DISCUSSION

3.1 Research Results

This study presents an acne classification system using the EfficientNetV2-S architecture on a dataset of 4,673 annotated facial acne images. After preprocessing and data augmentation, the model was trained for 20 epochs. The training process proceeded stably, as evidenced by the performance graphs (Figures 3 and 4), indicating that the model learned effectively without overfitting.

Upon completion of training, the model's final performance was evaluated on a held-out test set to ensure an unbiased assessment. The model achieved a high overall accuracy of 97.29%. A detailed quantitative analysis is presented in the Classification Report for the test set (Figure 5), which reveals consistently high and balanced F1-Scores for all five acne classes, ranging from 0.97 to 0.98. The Confusion Matrix for the test set (Figure 6) further corroborates these findings, with the vast majority of predictions falling on the main diagonal, indicating minimal classification errors.

As a complement, a qualitative evaluation was performed by visually inspecting the model's predictions on 25 random test samples (Figure 7), which confirmed the model's high fidelity and ability to generalize to real-world diagnostic scenarios.

```

Epoch 1/20
89/89 0s 12s/step - accuracy: 0.2530 - loss: 2.2831 - precision: 0.2747 - recall: 0.1690
Epoch 1: val_loss improved from inf to 1.60759, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 1598s 16s/step - accuracy: 0.2535 - loss: 2.2811 - precision: 0.2753 - recall: 0.1694 - val_accuracy: 0.3735 - val_loss: 1.6076 - val_precision: 1.0000
Epoch 2/20
89/89 0s 692ms/step - accuracy: 0.3959 - loss: 1.7223 - precision: 0.4411 - recall: 0.3000
Epoch 2: val_loss improved from 1.60759 to 1.44448, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 72s 806ms/step - accuracy: 0.3962 - loss: 1.7217 - precision: 0.4414 - recall: 0.3003 - val_accuracy: 0.4354 - val_loss: 1.4445 - val_precision: 0.8059
Epoch 3/20
89/89 0s 730ms/step - accuracy: 0.4891 - loss: 1.3707 - precision: 0.5672 - recall: 0.4176
Epoch 3: val_loss improved from 1.44448 to 1.18989, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 74s 828ms/step - accuracy: 0.4893 - loss: 1.3702 - precision: 0.5675 - recall: 0.4179 - val_accuracy: 0.5494 - val_loss: 1.1899 - val_precision: 0.7979
Epoch 4/20
89/89 0s 716ms/step - accuracy: 0.5824 - loss: 1.1234 - precision: 0.6461 - recall: 0.5002
Epoch 4: val_loss improved from 1.18989 to 0.80951, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 80s 811ms/step - accuracy: 0.5828 - loss: 1.1227 - precision: 0.6466 - recall: 0.5007 - val_accuracy: 0.7014 - val_loss: 0.8095 - val_precision: 0.8632
Epoch 5/20
89/89 0s 729ms/step - accuracy: 0.6660 - loss: 0.8903 - precision: 0.7274 - recall: 0.5995
Epoch 5: val_loss improved from 0.80951 to 0.60493, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 73s 822ms/step - accuracy: 0.6661 - loss: 0.8901 - precision: 0.7275 - recall: 0.5998 - val_accuracy: 0.7774 - val_loss: 0.6049 - val_precision: 0.8689
Epoch 6/20
89/89 0s 722ms/step - accuracy: 0.7468 - loss: 0.6976 - precision: 0.8034 - recall: 0.6800
Epoch 6: val_loss improved from 0.60493 to 0.52217, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 76s 844ms/step - accuracy: 0.7470 - loss: 0.6974 - precision: 0.8035 - recall: 0.6803 - val_accuracy: 0.8154 - val_loss: 0.5222 - val_precision: 0.8854
Epoch 7/20
89/89 0s 741ms/step - accuracy: 0.7972 - loss: 0.5483 - precision: 0.8431 - recall: 0.7491
Epoch 7: val_loss improved from 0.52217 to 0.44582, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 75s 839ms/step - accuracy: 0.7973 - loss: 0.5480 - precision: 0.8432 - recall: 0.7493 - val_accuracy: 0.8534 - val_loss: 0.4458 - val_precision: 0.8911
Epoch 8/20
89/89 0s 715ms/step - accuracy: 0.8403 - loss: 0.4373 - precision: 0.8802 - recall: 0.8118
Epoch 8: val_loss improved from 0.44582 to 0.34829, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 72s 808ms/step - accuracy: 0.8403 - loss: 0.4372 - precision: 0.8803 - recall: 0.8118 - val_accuracy: 0.8936 - val_loss: 0.3483 - val_precision: 0.9063
Epoch 9/20
89/89 0s 714ms/step - accuracy: 0.8694 - loss: 0.3608 - precision: 0.8926 - recall: 0.8473
Epoch 9: val_loss improved from 0.34829 to 0.31319, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 72s 805ms/step - accuracy: 0.8695 - loss: 0.3605 - precision: 0.8927 - recall: 0.8475 - val_accuracy: 0.8979 - val_loss: 0.3132 - val_precision: 0.9223
Epoch 10/20
89/89 0s 707ms/step - accuracy: 0.9019 - loss: 0.3004 - precision: 0.9186 - recall: 0.8800
Epoch 10: val_loss improved from 0.31319 to 0.25945, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 72s 804ms/step - accuracy: 0.9019 - loss: 0.3002 - precision: 0.9186 - recall: 0.8801 - val_accuracy: 0.9197 - val_loss: 0.2595 - val_precision: 0.9394
Epoch 11/20
89/89 0s 717ms/step - accuracy: 0.9221 - loss: 0.2343 - precision: 0.9331 - recall: 0.9076
Epoch 11: val_loss did not improve from 0.25945
89/89 69s 771ms/step - accuracy: 0.9221 - loss: 0.2343 - precision: 0.9331 - recall: 0.9076 - val_accuracy: 0.9240 - val_loss: 0.3065 - val_precision: 0.9330
Epoch 12/20
89/89 0s 651ms/step - accuracy: 0.9421 - loss: 0.1858 - precision: 0.9517 - recall: 0.9304
Epoch 12: val_loss improved from 0.25945 to 0.22983, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 67s 758ms/step - accuracy: 0.9421 - loss: 0.1857 - precision: 0.9517 - recall: 0.9305 - val_accuracy: 0.9370 - val_loss: 0.2298 - val_precision: 0.9491
Epoch 13/20
89/89 0s 654ms/step - accuracy: 0.9529 - loss: 0.1485 - precision: 0.9645 - recall: 0.9431
Epoch 13: val_loss improved from 0.22983 to 0.18140, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 81s 743ms/step - accuracy: 0.9528 - loss: 0.1485 - precision: 0.9644 - recall: 0.9431 - val_accuracy: 0.9457 - val_loss: 0.1814 - val_precision: 0.9507
Epoch 14/20
89/89 0s 650ms/step - accuracy: 0.9534 - loss: 0.1576 - precision: 0.9571 - recall: 0.9440
Epoch 14: val_loss improved from 0.18140 to 0.17160, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 82s 739ms/step - accuracy: 0.9534 - loss: 0.1574 - precision: 0.9571 - recall: 0.9441 - val_accuracy: 0.9533 - val_loss: 0.1716 - val_precision: 0.9562
Epoch 15/20
89/89 0s 715ms/step - accuracy: 0.9665 - loss: 0.1036 - precision: 0.9735 - recall: 0.9615
Epoch 15: val_loss improved from 0.17160 to 0.16618, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 73s 811ms/step - accuracy: 0.9665 - loss: 0.1037 - precision: 0.9734 - recall: 0.9614 - val_accuracy: 0.9479 - val_loss: 0.1661 - val_precision: 0.9540
Epoch 16/20
89/89 0s 718ms/step - accuracy: 0.9663 - loss: 0.1058 - precision: 0.9715 - recall: 0.9621
Epoch 16: val_loss improved from 0.16618 to 0.14067, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 72s 808ms/step - accuracy: 0.9663 - loss: 0.1058 - precision: 0.9715 - recall: 0.9622 - val_accuracy: 0.9577 - val_loss: 0.1407 - val_precision: 0.9638
Epoch 17/20
89/89 0s 723ms/step - accuracy: 0.9719 - loss: 0.0980 - precision: 0.9739 - recall: 0.9687
Epoch 17: val_loss did not improve from 0.14067
89/89 70s 782ms/step - accuracy: 0.9719 - loss: 0.0980 - precision: 0.9739 - recall: 0.9687 - val_accuracy: 0.9533 - val_loss: 0.1740 - val_precision: 0.9560
Epoch 18/20
89/89 0s 664ms/step - accuracy: 0.9768 - loss: 0.0754 - precision: 0.9784 - recall: 0.9738
Epoch 18: val_loss did not improve from 0.14067
89/89 77s 719ms/step - accuracy: 0.9768 - loss: 0.0754 - precision: 0.9784 - recall: 0.9738 - val_accuracy: 0.9609 - val_loss: 0.1517 - val_precision: 0.9650
Epoch 19/20
89/89 0s 662ms/step - accuracy: 0.9815 - loss: 0.0663 - precision: 0.9843 - recall: 0.9801
Epoch 19: val_loss did not improve from 0.14067
Epoch 19: ReduceLROnPlateau reducing learning rate to 4.99999873689376e-05.
89/89 64s 716ms/step - accuracy: 0.9814 - loss: 0.0663 - precision: 0.9842 - recall: 0.9800 - val_accuracy: 0.9446 - val_loss: 0.1921 - val_precision: 0.9527
Epoch 20/20
89/89 0s 656ms/step - accuracy: 0.9824 - loss: 0.0637 - precision: 0.9836 - recall: 0.9775
Epoch 20: val_loss improved from 0.14067 to 0.09370, saving model to /content/drive/MyDrive/amikom/Data/Acne Dataset/best1_model.keras
89/89 69s 776ms/step - accuracy: 0.9824 - loss: 0.0637 - precision: 0.9836 - recall: 0.9775 - val_accuracy: 0.9729 - val_loss: 0.0937 - val_precision: 0.9780
Restoring model weights from the end of the best epoch: 20.

```

Figure 3. Model Training Process Log Over 20 Epochs

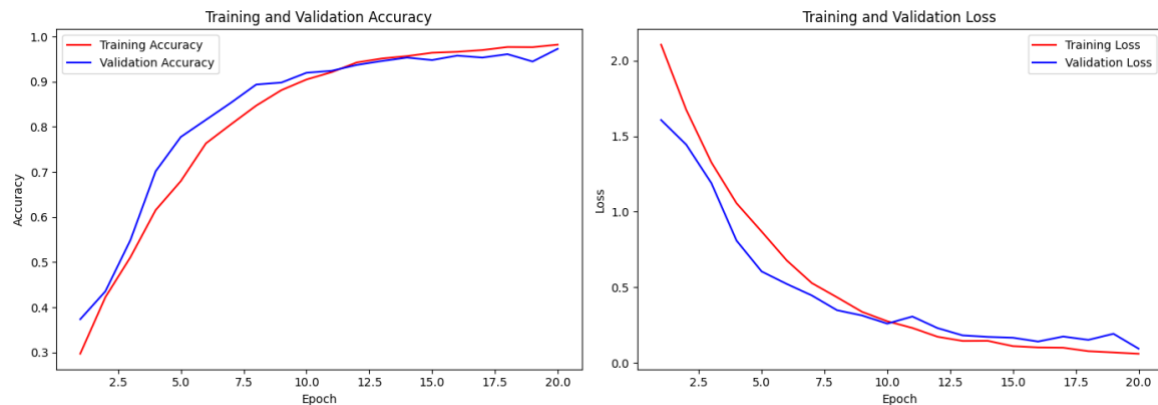


Figure 4. Accuracy (left) and Loss (right) Graphs During the Training Process

29/29 250s 9s/step

Classification Report (Test Set):

	precision	recall	f1-score	support
Blackheads	0.97	0.98	0.98	265
Cyst	0.96	0.98	0.97	189
Papules	0.98	0.95	0.96	202
Pustules	0.96	0.98	0.97	205
Whiteheads	1.00	0.98	0.99	57
accuracy			0.97	918
macro avg	0.97	0.97	0.97	918
weighted avg	0.97	0.97	0.97	918

Figure 5. Classification Report on Test Set

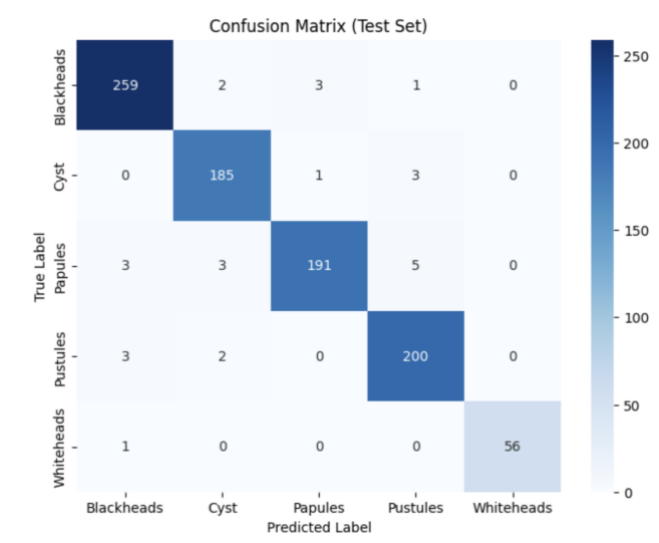


Figure 6. Confusion Matrix on Test Set



Figure 7. Example of Model Prediction Results on 25 Random Samples

3.2 Artikel Quality

3.2.1 Comparison with Other Research

To demonstrate the quality and scientific contribution of this research, the results obtained were compared with several previous relevant studies. The accuracy of 97.29% achieved in this study shows highly competitive performance and places it at the forefront of the dermatological image classification field. For comparison, a previous study by Pangestu & Kusriani (2024) using the ResNet50 architecture reported an accuracy of 94.00% [23], while Junayed et al. (2023) with a deeper Deep ResNet architecture achieved 95.89% [24]. Even when compared to a classic architecture like VGG16, which was reported by Lubis et al. (2023) to only achieve 82.00% [25], the superiority of the EfficientNetV2-S model in this study is evident.

This success can be attributed to several factors. First, the EfficientNetV2-S architecture is inherently more efficient and powerful. Second, the implementation of a comprehensive data augmentation strategy significantly improved the model's robustness. These results confirm that the selection of an advanced architecture combined with appropriate regularization techniques is a crucial factor for achieving peak performance on complex and highly variable medical image classification tasks.

4. CONCLUSION

This study successfully applied EfficientNetV2-S for multiclass acne classification, achieving 97.29% accuracy, outperforming established models such as ResNet50 and VGG16. Key strengths include the use of transfer learning with fine-tuning, robust data augmentation, and a well-structured training process. Nonetheless, limitations exist in terms of dataset diversity, as all samples originated from a single collection. For broader applicability, future work should involve validation on external datasets and implementation of model compression methods such as quantization or pruning. This would enable deployment on resource-constrained devices, offering scalable, AI-powered preliminary acne diagnosis accessible to the general public.

REFERENCE

- [1] A. ; K. Chandra I., "Analisis faktor penyebab jerawat pada mahasiswa berdasarkan kebiasaan hidup," *Jurnal Kesehatan Masyarakat*, vol. 14, no. 1, pp. 55–62, 2022.
- [2] L. ; S. Hakim Z.; Handhajani, H., "Klasifikasi citra pigmen kanker kulit menggunakan CNN," *Jurnal RESTI*, vol. 5, no. 5, pp. 887–894, 2021, [Online]. Available: <https://jurnal.iaii.or.id/index.php/RESTI/article/view/3001>
- [3] I. ; D. Fathurrahman M., "CNN untuk klasifikasi penyakit kulit digital," *Jurnal Infotek Hamzanwadi*, vol. 8, no. 1, pp. 11–20, 2025, [Online]. Available: <https://e-journal.hamzanwadi.ac.id/index.php/infotek/article/view/28655>
- [4] U. ; R. Wirantasa W.; Wona, M. M. A., "Deteksi penyakit kulit dengan CNN," *JRAMI – Univ. Indraprasta PGRI*, vol. 6, no. 2, pp. 44–50, 2025, [Online]. Available: <https://jim.unindra.ac.id/index.php/jrami/article/view/9210>
- [5] A. M. ; S. Manurung I.; Anugrah, A., "Deep learning for skin disease classification using EfficientNet-B1," *JAIC*, vol. 6, no. 1, pp. 67–74, 2025, [Online]. Available: <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/9100>
- [6] A. ; K. Chandra I. *et al.*, "CNN untuk klasifikasi penyakit kulit digital," *JAIC*, vol. 5, no. 1, pp. 887–894, 2024, [Online]. Available: <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/9621>
- [7] A. S. ; R. Sindhu C. K., "CNN-based skin disease detection," *IEEE Xplore Conference Proceedings*, 2024, [Online]. Available: <https://ieeexplore.ieee.org/document/11034901>
- [8] S. S. ; H. Kabir M. A.; Moz, S. H., "MPCNN: A novel approach for monkeypox detection," *Indonesian Journal of Artificial Intelligence Research*, 2024, [Online]. Available: <https://www.researchgate.net/publication/384560164>
- [9] M. I. F. ; F. Nuzula C., "Optimasi CNN untuk kanker kulit dengan Bayesian," *ILKOMNIKA*, vol. 8, no. 1, pp. 25–32, 2025, [Online]. Available: <http://journal.unublitar.ac.id/ilkomnika/index.php/ilkomnika/article/view/690>
- [10] H. A. Faudyta, J. T. Sinaga, and E. R. Subhiyakti, "Implementation of MobileNet Architecture for Skin Cancer Disease Classification," *Journal of Applied Informatics and Computing*, vol. 8, no. 2, pp. 589–597, 2024, doi: 10.30871/jaic.v8i2.8771.
- [11] M. I. ; S. Zulfa H., "Hyperparameter tuning pada MobileNetV2 untuk deteksi penyakit kulit," *SINTA: Sistem Informasi dan Teknologi Akademik*, vol. 4, no. 2, pp. 114–120, 2025, [Online]. Available: <http://jurnalsinta.id/index.php/sinta/article/view/43>
- [12] A. Fathir, R. Januar, J. Indra, D. S. Kusumaningrum, and S. Faisal, "Application of Convolutional Neural Network (CNN) Algorithm with ResNet-101 Architecture for Monkey Pox Detection in Human," vol. 9, no. 3, pp. 1006–1012, 2025.

- [13] L. Hakim, Z. Sari, and Handhajani, "Klasifikasi Citra Pigmen Kanker Kulit Menggunakan Convolutional Neural Network," *Jurnal RESTI*, vol. 5, no. 2, pp. 379–385, 2021, doi: 10.29207/resti.v5i2.3001.
- [14] S. H. ; M. Pusadan A.; Yazdi, M., "Pattern recognition untuk klasifikasi penyakit kulit," *Jurnal Informatika dan Komputer Nasional*, vol. 4, no. 2, pp. 20–27, 2025, [Online]. Available: <https://jurnal.unw.ac.id/index.php/IKN/article/download/3563/2540>
- [15] R. I. ; T. Susanto T., "Klasifikasi penyakit cacar dengan CNN AlexNet," *Jurnal Algoritme*, vol. 21, no. 4, pp. 55–63, 2024, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/algoritme/article/view/9045>
- [16] M. G. Febrian, B. A. Utomo, and G. Muhammad, "Implementasi CNN untuk klasifikasi penyakit kulit," *Semnasa AMIKOM Solo*, 2024, [Online]. Available: <https://ojs.amikomsolo.ac.id/index.php/semnasa/article/view/692>
- [17] R. Mekala, "Cloud-enhanced skin cancer diagnosis using EfficientNet-based CNN," *IJ Biomedical Signal Processing and AI*, 2024, [Online]. Available: <https://www.researchgate.net/publication/392598079>
- [18] A. ; S. Yadav S., "Skin disease classification using EfficientNet," *Springer Nature – Lecture Notes in Computer Science*, 2024, [Online]. Available: <https://link.springer.com/article/10.1007/s00521-024-09543-z>
- [19] K. V. ; P. Prakusa T. D., "Analisis klasifikasi penyakit kulit wajah dengan CNN," *Semnasa AMIKOM Solo*, 2024, [Online]. Available: <https://ojs.amikomsolo.ac.id/index.php/semnasa/article/view/527>
- [20] A. Chandra and I. Kotimah, "Analisis faktor penyebab jerawat pada mahasiswa berdasarkan kebiasaan hidup," *Jurnal Kesehatan Masyarakat*, vol. 14, no. 1, pp. 55–62, 2022.
- [21] F. A. Farid, S. Rahman, and J. Uddin, "DLI-Net: Deep learning framework for acne detection," *J Imaging*, vol. 10, no. 4, p. 82, 2024, [Online]. Available: <https://www.mdpi.com/2313-433X/10/4/82>
- [22] I. Fathurrahman and M. Djamaluddin, "CNN untuk klasifikasi penyakit kulit digital," *Jurnal Infotek Hamzanwadi*, vol. 8, no. 1, pp. 11–20, 2025, [Online]. Available: <https://e-journal.hamzanwadi.ac.id/index.php/infotek/article/view/28655>
- [23] H. A. ; K. Pangestu, "Peningkatan akurasi ResNet50 dalam klasifikasi penyakit kulit melalui augmentasi dan regularisasi," *TEMATIK*, vol. 11, no. 1, pp. 65–71, 2024, [Online]. Available: <https://jurnal.itpln.ac.id/index.php/tematik/article/view/1876>
- [24] M. ; I. Junayed M. T.; Hossain, M. A., "Deep residual learning for skin lesion classification using ensemble CNNs," *Biomed Signal Process Control*, vol. 82, p. 104593, 2023, [Online]. Available: <https://doi.org/10.1016/j.bspc.2022.104593>
- [25] C. ; Y. Lubis D.; Tarumanagara, U., "Klasifikasi penyakit kulit menggunakan CNN dengan arsitektur VGG16," *Jurnal Saintek*, vol. 8, no. 1, pp. 55–61, 2023.