DEEP LEARNING FINAL PROJECT

Facial Skin Type Classification

KENNETH SUNJAYA
2702273315
NOVELLINA EDYAWATI
2702228223
STEPANUS IMANUEL
2702355574

COMP6826001 - Deep Learning

SCHOOL OF COMPUTER SCIENCE
BINUS UNIVERSITY

1. Abstract

The use of skincare products is growing among various ages. Each individual needs to understand their facial skin type characteristics. In general, the skin types are divided into three categories: normal, dry, and oily. However, many individuals still lack adequate knowledge about their specific skin type, which often leads to the misuse of skincare products and results in various facial skin problems. Nowadays, improvements in image recognition using machine learning, particularly deep learning, have opened up new opportunities for classifying facial skin characteristics. In this study, we aim to build a model that classifies facial skin type using Convolutional Neural Network architectures. The EfficientNetB0 CNN architecture was used as the backbone for classifying skin type based on the input image in this study. This report study includes the preprocessing of the facial skin image dataset, preprocessing data, model training, and model evaluation. The dataset used consists of face images of various facial skin types collected from the Kaggle database. The model was trained and tested to classify the types of facial skin characteristics after going through several image preprocessing. The final model has successfully achieved an overall accuracy of 85.71% with a macro F1-score of 85.01%, indicating that the model can differentiate between dry, normal, and oily skin categories.

Keywords: facial skin classification, Convolutional Neural Network (CNN), EfficientNet, image recognition

2. Introduction

2.1 Background

In recent years, the use of skincare products has increased among people of all ages, from teenagers to adults to middle-aged individuals [1]. The use of skincare products varies depending on the characteristics of each individual's facial skin. The skin itself is an internal organ that plays an important role in an individual's appearance, especially the skin on the face [2]. Generally, skin types are divided into normal, dry, oily, and combination, where the characteristics of each skin type are influenced by moisture levels, oil production, and facial skin sensitivity [3]. Skin specialists emphasize the importance of understanding facial skin characteristics before applying skincare products directly to the skin [3]. However, many individuals still lack adequate knowledge about their specific skin type, which often leads to inappropriate product use and results in various facial skin problems [4].

Advancements in machine learning, particularly deep learning, have opened up new opportunities for classifying facial skin characteristics. Convolutional Neural Networks (CNNs) [5] are one of the models in deep learning that perform quite well for various image classification tasks, including image recognition for dermatological analysis. CNNs can classify images by automatically learning spatial features such as texture, glossiness, pore visibility, and dryness attributes that are highly relevant for skin analysis. Although they are quite effective, CNN architectures often require extensive parameter tuning and may have difficulty generalizing across different skin colors, lighting variations, and image quality commonly encountered in real-world settings.

Previous research has explored various approaches for facial skin classification, especially using Convolutional Neural Networks. A study by Dianthy et. al achieved an accuracy of 87.5% using four main categories — normal, oily, dry, and combination using CNN with MobileNetV2 [4]. A study conducted by Saiwaeo et al. [6] has evaluated various CNN architectures, including MobileNet-V2, EfficientNet-V2, InceptionV2, and ResNet-V1. The findings showed that EfficientNet-V2 performed the best, achieving an accuracy of 91.55% with an average loss of 22.74%.

EfficientNet, a transfer learning convolutional model developed through compound scaling, serves as a strong baseline for visual recognition tasks by achieving higher accuracy with fewer parameters compared to previous state-of-the-art architectures [7]. Using EfficientNet as a baseline has the potential to improve classification accuracy by adjusting the scale of network depth, width, and image resolution, making this architecture ideal for implementation as a facial skin type classification app.

This study aims to build a facial skin type classification model using CNNs, with EfficientNet as its baseline architecture. This research explores the effectiveness of deep convolutional feature extraction for skin-type recognition, evaluates the performance of various configurations, and analyzes the model's ability to generalize skin types under various image conditions. As a solution to overcome the limitations of older methods for analyzing skin types using modern deep learning approaches, it is hoped that this work can contribute to improving the automation of dermatological analysis and facilitate the personalization of skincare solutions for each individual.

3. Literature Review

a. **Facial Skin Type**

A study by Baumann et al. [3] evaluated skin according to four basic parameters that more accurately characterize skin types. These parameters are dry or oily, sensitive or resistant, pigmented or non-pigmented, and wrinkled or unwrinkled. With those four parameters, consumers can easily define the most suitable topical treatments for their skin. In this study, we only do classification for dry, normal, and oily skin types.

b. **Convolutional Neural Network**

Convolutional Neural Network (CNN) is a deep learning approach that can highly abstract features and can identify objects efficiently [5]. CNN is based on the idea of weight sharing, which reduces the number of parameters required for training. Due to the less parameters, CNN can be trained smoothly and does not suffer from overfitting. CNNs are widely used in various domains such as image classification, object detection, face detection, speech recognition, etc [8].

c. EfficientNet

EfficientNet is a transfer learning model of CNN architectures introduced by Tan and Le (2019) [7] that proposes a new way to scale neural networks efficiently. EfficientNet introduces compound scaling, a method that uniformly scales all three dimensions based on a single coefficient. EfficientNet achieves significantly higher accuracy while using fewer parameters with less computation compared to previous CNN architectures. As a result, EfficientNet becomes a highly efficient model family that reaches state-of-the-art performance across multiple benchmarks with substantially reduced computational cost.

d. Lightweight Architectures

Marya et al. [4] developed an optimized facial skin type classification using a CNN with MobileNetV2, focusing on real-time smartphone applications. Their study demonstrates that MobileNetV2 offers a high accuracy of 87,5% showing a strong balance between accuracy and computational efficiency due to the efficient depthwise separable convolutions. The system successfully classified skin types while maintaining real-time performance, indicating the potential of lightweight CNN models for on-device inference. The study suggests focusing on expanding the dataset to include diverse demographics and integrating additional features, such as personalized skincare recommendations, to improve the system's capabilities and usability further.

Saiwaeo et al. [6] conducted a study using various CNN architectures such as MobileNet-V2, EfficientNet-V2, InceptionV2, and ResNet-V1 that were optimized and evaluated. EfficientNet-V2 reaches the best performance with an accuracy of 91.55% with an average loss of 22.74%. After conducting hyperparameter tuning, the model achieved an accuracy of 94.57% and a loss of 13.77%. The model performance was validated using 10-fold cross-validation and tested on unseen data, achieving an accuracy of 88.70% and a loss of 21.68%

Chiun et al. [9] proposed two different CNN architectures. The first one has two convolutional layers with two pooling layers and three fully connected layers. The other one has three convolutional layers, three pooling layers, and four fully connected layers. After comparing the result with LeNet-5, the CNN architecture that has three convolution layers, three pooling layers, and four fully connected layers, has the highest recognition rate. It serves as a baseline that builds a framework for detecting facial skin problems.

.

e. Limitations in Existing Research

- Imbalanced datasets present several drawbacks for various models, leading to biased model performance that favors the majority class.

- The data collection process focused primarily on specific skin areas, which introduced challenges in feature extraction and classification and normal and oily skin types.

4.Methodology

4.1 Dataset

This study uses the Oily, Dry, Normal Skin Types Dataset obtained from the Kaggle platform [Dataset Link]. This dataset is a collection of facial images or skin sections that have been grouped into three skin types: dry, normal, and oily. All images are organized into three main directories, namely train, valid, and test, where each directory has three subfolders according to class labels (dry, normal, and oily).

Initial experiments showed that the model performed very well on the training and validation subsets but achieved significantly lower performance on the test subset. This difference suggested a distribution mismatch between the test folder and other subsets such as lighting conditions and shooting angles. Because the dataset's creators did not provide documentation describing how the dataset was splitted, the exact source of the inconsistency could not be determined. However, this must be handled in the preprocessing part.

4.2 Preprocessing

To address the distribution mismatch observed in the original dataset split, all images from all three directories were first merged into a single dataset grouped by their respective class labels. The unified pool was then randomly re-split into new training, validation, and testing subsets following 80-10-10 ratio.

To improve the model's generalization capability and mitigate overfitting, several augmentation strategies were applied during the training process:

- Image resizing to 224 x 224 pixels (as the requirement of EfficientNet-B0 input)
- Random vertical flip with probability = 0.6
- Conversion to tensor format
- Standard ImageNet normalization using mean and standard deviation values:

$$p = [0.485,\ 0.456,\ 0.406],\ \sigma = [0.229,\ 0.224,\ 0.225]$$

For validation and testing subsets, only resizing and normalization were applied. This ensures consistent image preprocessing while preventing augmentation bias during model evaluation.

4.3 Model Architecture

In this study, the EfficientNetB0 Convolutional Neural Network (CNN) architecture is used as the backbone model for classifying skin type based on input image. EfficientNet is a family of CNN models proposed to improve classification accuracy while maintaining computational efficiency through a systematic compound scaling that uniformly scales network depth, width, and input resolution. It uses the inverted bottleneck residual block of MobileNetV2, in addition to squeeze-and-excitation modules (SE Modules). Among the EfficientNet family, the B0 model serves as the baseline model. This model offers an optimal balance between complexity and performance.

4.3.1 Compound Scaling Principle

Unlike traditional CNN architecture that scale a single dimension such as increasing depth or width independently, EfficientNet introduces compound model scaling where network depth, width, and input resolution are scaled simultaneously using fixed coefficients. This approach affects the capacity of the network to grow proportionally across all dimensions which prevents the performance saturation that is caused by imbalance scaling. EfficientNet-B0 is the baseline architecture discovered through neural architecture search under constrained computational budgets. The compound scaling process makes the EfficientNet models achieve higher accuracy even with fewer parameters and floating-point operations compared to conventional architectures.

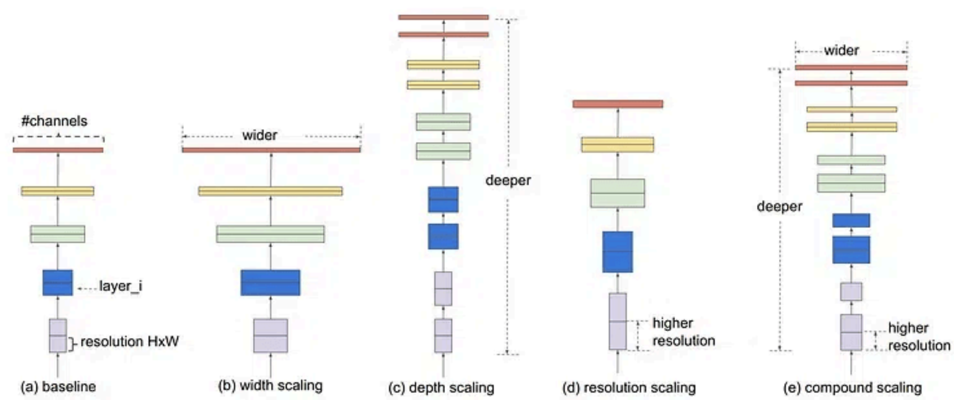


Fig 1. Compound Scaling in EfficientNet

4.3.2 EfficientNetB0 Architecture Design

The EfficientNetB0 architecture consists of three main components: A stem layer, a sequence of feature extraction blocks, and a classification head. The stem layer consists of an initial 3x3 convolutional layer with a stride of 2, followed by batch normalization and a Swish activation function. This stage performs early spatial downsampling while capturing low-level visual patterns such as edges and color gradients, which are essential for texture-based image analysis.

The core feature extraction stage is built using Mobile Inverted Bottleneck Convolution (MBConv) blocks, which were adopted due to their efficiency and strong representational capability. Each of the MBConv blocks includes:

- An expansion phase that increases the number of channels
- A depthwise separable convolution that reduces computational complexity
- A Squeeze-and-Excitation module that adaptively recalibrates channel-wise feature responses
- A projection layer that compresses the feature maps into a lower-dimensional space.

Residual connections are used when the input and output dimensions are compatible. This might improve gradient flow and training stability. By stacking multiple MBConv blocks with varying kernel sizes, expansion ratios, and strides, the EfficientNet-B0 can extract hierarchical features ranging from local textures to high-level semantic representations.

Following feature extraction, a 1x1 convolutional layer is used to refine channel dimensionality. Global average pooling is then used to aggregate spatial information into a compact feature vector. The final layer is a fully connected classification head with a softmax activation function.

4.4 Training Setup

For the skin image classification task, the EfficientNet-B0 was selected as the backbone. This model offers strong accuracy and efficiency due to its compound scaling strategy. This network was initialized with pretrained ImageNet weights, consistent with dermatology-related EfficientNet applications where transfer learning improves feature extraction for skin textures. The final classification layer was replaced by a fully connected layer of size 3, indicating three classes of skin types.

The training was conducted in a PyTorch environment using the following settings:

Optimizer	Stochastic Gradient Descent (SGD)
	Learning rate: 3 x 10-3
	Momentum: 0.9
	Weight Decay: 1 x 10-4
LR Scheduler	StepLR
	Step Size: 15 epochs
	Decay factor (gamma): **0.1**

Loss Function	Cross-entropy loss
Batch Size	64
Epochs	50 with early stopping
Early Stopping Procedure	Patience: 10 epochs
	Minimum improvement threshold: 0.0
Hardware	CUDA - NVIDIA GeForce RTX 3070 8GB

The training loop collected accuracy, precision, recall, and F1-macro at each epoch.

To prevent overfitting, an early stopping mechanism was applied based on the validation loss data with patience = 10 and minimum improvement threshold = 0.0. The model state with the lowest validation loss was saved as the final best checkpoint.

4.5 Evaluation metrics

To assess the performance of the EfficientNet-B0 model for this multi-class classification, several combinations of evaluation metrics were used. These metrics measure different aspects of predictive performance, especially in the presence of class imbalance.

- Accuracy

Accuracy measures the proportion of correctly classified samples among all predictions. On the other hand, accuracy alone can be misleading when the class distributions are not u uniform. It is defined as:

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

- Precision

Precision evaluates the model's ability to correctly identify positive predictions for each class, indicating how reliable the predicted labels are. Macro-averaged precision is used to ensure equal weighting across all classes

$$Precision = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FP_i}$$

- Recall

Recall measures the model's ability to correctly retrieve all relevant samples for each class. Macro-averaged recall provides a class-balanced assessment.

$$Recall = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FN_i}$$

- F1-Score

The F1-Score is the mean of the combination between precision and recall and provides a balanced measure when the dataset is imbalanced. The macro-averaged form is computed as:

$$F1 = \frac{1}{K}\sum_{i=1}^{K}\frac{2 \cdot Precision_i \cdot Recall_i}{Precision_i + Recall_i}$$

- Confusion Matrix

The confusion matrix provides a detailed breakdown of model predictions by comparing true labels with predicted labels for each class. It enables the identification of misclassification patterns, revealing which classes are more challenging for the model

5. Implementation & Results

5.1 System Details

The proposed skin-type classification system was developed and evaluated using a deep learning pipeline built on the PyTorch framework. All experiments were executed on a workstation equipped with an NVIDIA GeForce RTX 3070 GPU to accelerate the training process. The system architecture integrates a pre-trained EfficientNet-B0 model, customized preprocessing modules, and an end-to-end training and evaluation pipeline. The software environment includes Python 3.11.4, PyTorch, Torchvision, and CUDA 13.0.

5.2 Experiments

A fixed random seed of 42 was used to ensure deterministic sampling. After the preprocessing stage, the newly generated training subset was analyzed using frequency counts to verify that the sampling process preserved the natural imbalance present in the original dataset.

The EfficientNet-B0 was selected as the backbone due to its strong accuracy-efficiency trade-off. Pretrained ImageNet1K weights were used for transfer learning. The final classifier layer was modified from a 1000-class output to a 3-class fully connected layer, corresponding to the skin-type labels. All remaining weights in the backbone were fine-tuned during training.

The training was conducted for 50 epochs with an early-stopping mechanism (patience = 10) to prevent overfitting. The training and validation losses, accuracy, precision, recall, and F1-score were recorded at each epoch. The model weights for the lowest validation loss were stored as the final model. After training, the model was evaluated on the test set. Predictions were compared against ground-truth labels to compute performance metrics.

5.3 Visualization

To provide deeper insight into the learning behavior of the EfficientNet-B0 model during the training process, several visualizations were generated. These visualizations were created to illustrate the convergence properties of the model, highlight class distribution patterns, and evaluate prediction reliability across different skin-type categories.

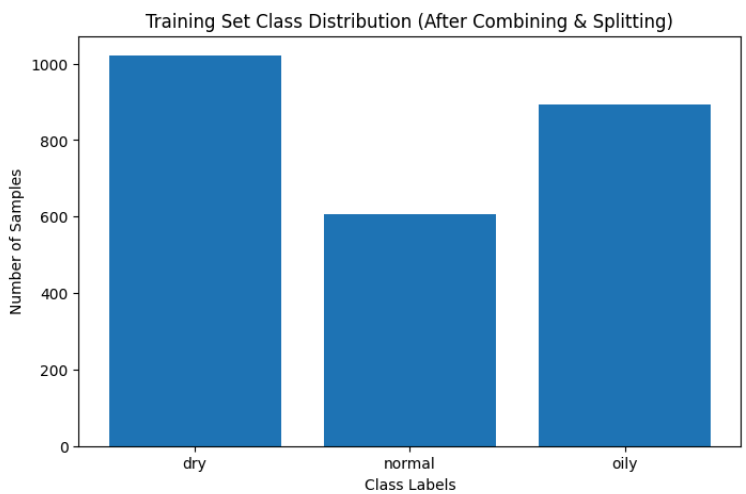


Fig 2. Class Distribution Chart

Fig 2 illustrates the distribution of samples across the three skin-type categories after combining all original data and applying the final 80/10/10 stratified split. The plot highlights a moderate class imbalance, where the normal class contains fewer samples compared to dry and oily. These imbalance can influence the learning dynamics of the model and motivates the use of macro-averaged evaluation metrics.

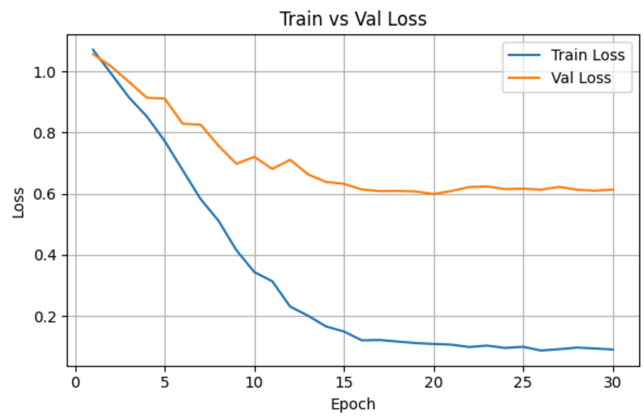


Fig 3. Training and Validation Loss Curves

Fig 3 illustrates the training and validation loss curves across epochs. The training loss demonstrates a consistent downward trend, while the validation loss stabilizes after several epochs, indicating that the model has successfully converged without significant overfitting.

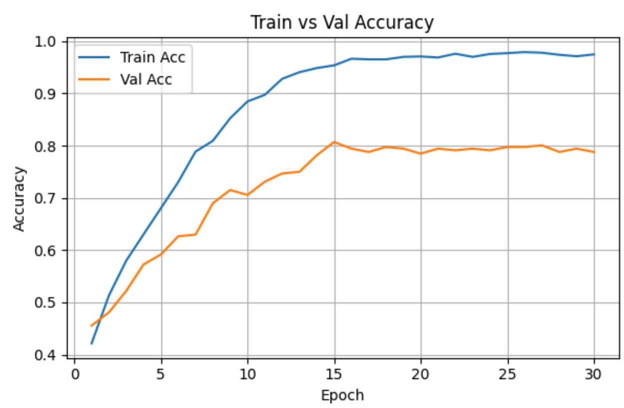


Fig 4. Training and Validation Accuracy Curves

Fig 4 presents the training and validation accuracy curves, showing a substantial performance increase in the early epochs followed by gradual saturation. The gap between training and validation accuracy provides insight into generalization capability.

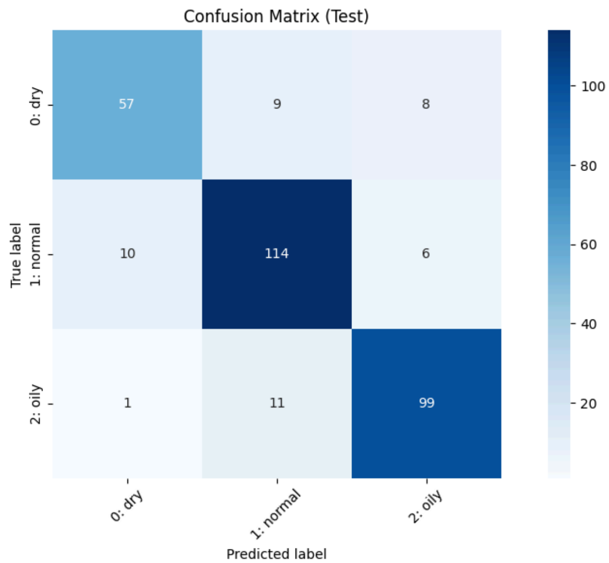


Fig 5. Confusion Matrix

The confusion matrix in Fig 5 was constructed using the final test predictions. This visualization provides a clear overview of class-wise-performance, highlighting correctly classified samples as well as the misclassification that occurs between each categories. These insights are essential for interpreting error patterns and potential model or pipeline improvements.

5.4 Result Obtained

Class	Precision	Recall	F1-Score	Support
Dry (0)	0.8382	0.7703	0.8028	74
Normal (1)	0.8507	0.8769	0.8636	130
Oily (2)	0.8761	0.8919	0.8839	111
Accuracy	-	-	0.8571	315
Macro-Avg	0.8550	0.8464	0.8501	315
Weighted-Avg	0.8567	0.8571	0.8565	315

Table 1. Classification Report

The model achieved an overall test accuracy of 85.71%, indicating reliable generalization on unseen data. Table 1 summarizes the precision, recall, and F1-score for each class. All three categories obtained F1-scores above 0.80, with only the oily skin class achieving the highest performance (F1-score = 0.8839). These results indicate that the model is capable of learning discriminative features across different skin conditions despite the visual similarity among classes.

Looking back at Fig 5 which reveals the confusion matrix for the test dataset, the diagonal values indicate correct predictions (Dry: 57, Normal: 114, Oily: 99). Misclassification primarily occurred between dry and normal. This also likely occurred between normal and oily classes.

6.Discussion & Limitations

6.1 Analysis of Performance

The performance of the EfficientNet-B0 model is analyzed by examining not only the numerical results, but also the underlying behavioral trends reflected in the evaluation metrics and confusion matrix. We will be focusing on interpreting those results, identifying performance patterns, and understanding the model's operational strength and limitations.

Overall, the model demonstrates strong classification capability across all three skin type categories, achieving an overall test accuracy of 85.71% and balanced macro-averaged precision, recall, and F1-scores. The confusion matrix reveals that the model is particularly reliable in identifying normal and oily skin types, with high true-positive counts and low misclassification rates. In contrast, the dry class shows comparatively higher confusion with neighboring categories, indicating that its intra-class variability or visual similarity to the oily class may limit separability.

The oily class obtains the highest recall (0.8919), suggesting that the model is highly sensitive to features characteristic of oily skin. Meanwhile, the dry class exhibits the lowest recall (0.7703), showing that some dry-skin samples are misclassified likely due to overlaps in texture or lighting induced artifacts. Nevertheless, the model maintains strong precision across all categories.

Despite the strong performance of the EfficientNet-B0 model, several challenges were encountered throughout the study. First, the dataset exhibited class imbalance, particularly with fewer samples in the normal skin category. This imbalance introduced bias during training and required careful handling through stratified splitting and data augmentation strategies.

Second, the image qualities are not consistent, such as lighting, shadows, and noise–which introduces difficulties for reliable feature extraction. These inconsistencies reduced model robustness and necessitated extensive preprocessing, including normalization and augmentation to ensure stable convergence.

Third, the model displayed limitations in classifying visually similar categories (e.g., dry and normal), as reflected in the confusion matrix. And finally, computational constraints imposed restrictions on hyperparameter exploration and model scaling. While EfficientNet is designed for efficiency, deeper variants (B3, B4, B7) could not be evaluated due to training time and hardware limitations.

The selection of EfficientNet-B0 involves several trade-offs balancing accuracy, computational cost, and model complexity. Although larger EfficientNet variants typically achieve higher accuracy, the B0 architecture offers the balance between performance and efficiency. Additionally, using a lightweight model accelerates inference, making real-time applications more responsive. However, this comes at the cost of reduced representational capacity compared to deeper networks. Consequently, while the adopted model performs well for general classification, it may not capture finer-grained distinctions as effectively as more complex architectures.

7.Conclusion & Future Work

In this study, an EfficientNet-B0 based model was developed for skin type classification using a dataset consisting of three categories: dry, normal, and oily. The dataset was reconstructed through a re-pooling and re-splitting process to address distribution inconsistencies found in the original partition. The model was trained using transfer learning from ImageNet weights combined with data augmentation to improve generalization.

Accuracy, precision, recall, and F1-score were monitored throughout the training process while early stopping was applied to prevent overfitting. The final model achieved an overall test accuracy of 0.8571, with each class F1-scores are 0.8028 (Dry), 0.8636 (Normal), 0.8839 (Oily). These results indicate that the model is effective in distinguishing normal and oil skin types while classification of dry skin remains more challenging. Overall, the experiment demonstrates that EfficientNet-B0 provides a strong accuracy-efficiency trade-off for dermatological applications and validates the feasibility of applying deep convolutional architectures to automated facial skin type assessment with decent performance.

Future work can focus on integrating clinical metadata such as age, gender, sensitivity level, or skin history into a multimodal model. Since skin type is not influenced only by the visual appearance but also by biological and lifestyle variables, combining metadata with image can improve classification reliability. Also, the exploration of more advanced architectures such as EfficientNetV2 or Vision Transformers may capture finer skin texture that could be improving the classification of challenging classes such as dry skin. Another important direction is expanding and improving the quality of the dataset. A richer representative dataset will enhance the model’s generalization in real-world use.

8.References

[1] X. Kuang, C. Lin, Y. Fu, Y. Wang, J. Gong, Y. Chen, Y. Liu, and F. Yi, "A comprehensive classification and analysis of oily sensitive facial skin: a cross-sectional study of young Chinese women," *Scientific Reports*, vol. 15, Art. no. 1633, Jan. 2025, doi:10.1038/s41598-024-85000-z.

[2] S. Saidah, Y. N. Fuadah, F. Alia, N. Ibrahim, R. Magdalena, and S. Rizal, "Facial skin type classification based on microscopic images using convolutional neural network (CNN)," in *Proc. 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics*, Lecture Notes in Electrical Engineering, vol. 746, Singapore: Springer, 2021, pp. 75–83, doi: 10.1007/978-981-33-6926-9_7.

[3] L. Baumann, "Understanding and treating various skin types: the Baumann skin type indicator," *Dermatologic Clinics*, vol. 26, pp. 359–373, 2008, doi: 10.1016/j.det.2008.03.007.

[4] D. Marya, C. M. S. Ramadhani, and R. Ardiansyah, "Development of an Optimized Facial Skin Type Classification System using CNN MobileNetV2 for Real-Time Smartphone Applications," *Journal of Telecommunication Network (Jurnal Jaringan Telekomunikasi)*, vol. 15, no. 1, 2025.

[5] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network – a deep learning approach," *Procedia Computer Science*, vol. 132, pp. 679–688, 2018

[6] S. Saiwaeo, S. Arwatchananukul, L. Mungmai, W. Preedalikit, and N. Aunsri, "Human skin type classification using image processing and deep learning approaches," *Heliyon*, vol. 9, no. 11, p. e21176, Oct. 2023, doi: 10.1016/j.heliyon.2023.e21176.

[7] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, May 2019, doi: 10.48550/arXiv.1905.11946.

[8] Wallach, I., Dzamba, M., and Heifets, A. (2015) "AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery." arXiv preprint arXiv:1510.02855

[9] C. -L. Chin, M. -C. Chin, T. -Y. Tsai and W. -E. Chen, "Facial skin image classification system using Convolutional Neural Networks deep learning algorithm," 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, Japan, 2018, pp. 51-55, doi: 10.1109/ICAwST.2018.8517246.

9. Appendix

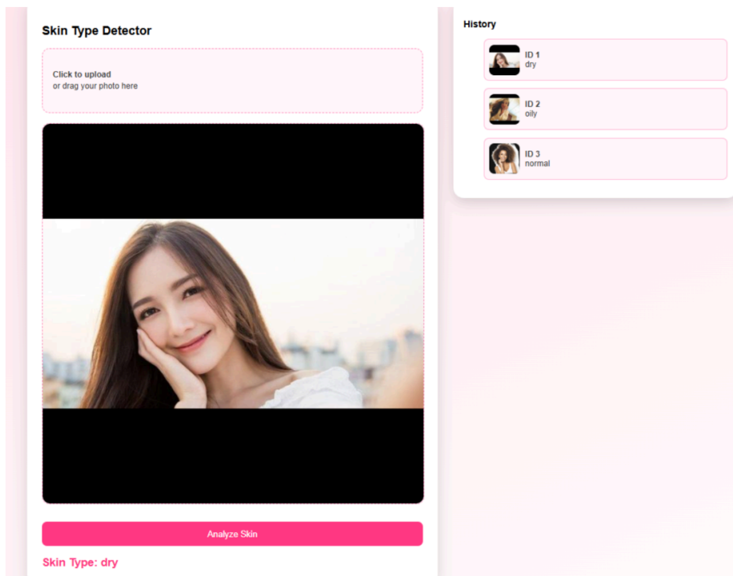
Team Contribution Statement

Kenneth Sunjaya: Formal analysis, Data Curation, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing.

Novellina Edyawati: Conceptualization, Resources, Investigation, Writing - original draft, Writing - review & editing.

Stepanus Imanuel: Formal analysis, Investigation, Supervision, Writing - original draft, Writing - review & editing, Software.

Screen Shots



Code snippets:
https://github.com/stepanusimnuel/skin-type-classification

Video:
https://drive.google.com/file/d/17xA3wT_xsh_L8E6TmtuAdxgDz6TnMIDG/view?usp=sharing