Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 12(122)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2024.111101077

Convolutional neural network architectural models for multiclass classification of aesthetic facial skin disorders

Rismayani[1,2], Amil Ahmad Ilham[3*], Andani Achmad[1] and Muhammad Rifqy Yudhiestra Rachman[3]
Department of Electrical Engineering, Faculty of Engineering, Universitas Hasanuddin, Gowa, Indonesia[1]
Department of Information System, Dipa Makassar University, Makassar, Indonesia[2]
Department of Informatics, Faculty of Engineering, Universitas Hasanuddin, Gowa, Indonesia[3]

Received: 21-June-2024; Revised: 17-January-2025; Accepted: 19-January-2025

Abstract
Aesthetic facial skin disorders present a significant challenge in dermatology, often requiring accurate diagnosis for effective treatment. This study investigates the effectiveness of convolutional neural network (CNN) architectural models for multiclass classification of facial skin disorders, including oily skin, hyperpigmentation, acne, redness, blackheads, and normal skin types. The research aims to compare the performance of various CNN architectures, including EfficientNetB0, HRNet, DenseNet201, and ResNet50, in classifying facial skin image datasets into multiple categories of aesthetic skin disorders. Additionally, this study examines the impact of feature combination methods on classification accuracy by enhancing attribute representation. Specifically, it explores the integration of color moment (CM) for color features and Laplacian of Gaussian (LoG) for shape features with CNN architectural models. Performance evaluation and comparison are conducted between CNN architectures with and without feature combinations. The results demonstrate that the selected CNN architectures effectively classify various aesthetic facial skin disorders, achieving an accuracy rate of 0.95 using the ResNet50 architecture. Moreover, the findings highlight that feature combination techniques, particularly the integration of CNN architectures with CM and LoG, significantly enhance classification accuracy. This study emphasizes the potential of CNN architectural models to improve early diagnostic capabilities in dermatology. The integration of artificial intelligence (AI)-based technology in the classification of aesthetic facial skin disorders offers a promising avenue for more accurate and effective diagnosis and treatment, ultimately improving patient outcomes.

Keywords
Aesthetic facial skin disorders, Convolutional neural network, Multiclass classification, Feature combination methods, Artificial intelligence.

1.Introduction
Aesthetic disorders of facial skin, such as acne, rosacea, and hyperpigmentation, are common problems that interfere with self-confidence and quality of life [1, 2]. Accurate and timely diagnosis is essential for effective treatment and management of this condition [3]. However, conventional diagnostic techniques, which largely rely on a dermatologist's visual examination, can be subjective and prone to error [4, 5]. Artificial intelligence (AI) and deep learning techniques, especially convolutional neural network (CNN), have offered lucrative solutions for automated and objective diagnosis of skin disorders [6, 7]. CNN have proven very effective in image classification tasks, including medical image analysis [8, 9].

In addition, non-visual data information has been used to predict facial skin disorders with high accuracy using a model of artificial neural networks as the basic concept of deep learning [10].

Developing reliable and accurate AI-based diagnostic technology for aesthetic facial skin disorders faces several challenges. First, a large and diverse dataset with accurate diagnoses must be available to train a robust CNN model [11]. Second, analyzing and classifying images of facial skin disorders is very difficult due to their complexity and different variations, which are influenced by factors such as ethnicity, age, and lighting conditions [12–14]. Finally, interpretation and explanation of AI models are critical in medical applications, as clinicians need to understand how to make informed decisions about diagnosis and treatment [15, 16].

*Author for correspondence

This study investigated different CNN architectures for classifying different classes of aesthetic facial skin disorders. Our specific objectives are to compare the performance of EfficientNetB0, high-resolution network (HRNet), dense convolutional network with 201 layers (DenseNet201), and ResNet50 models when classifying facial skin image datasets categorized into different types of aesthetic skin disorders. The impact of laplacian of gaussian (LoG) and color moment (CM) methods on classification accuracy is analyzed to identify the most effective approach for improving early diagnosis. The performance of CNN models with and without feature combinations is evaluated to determine the best method.

This research contributes to AI-based dermatology, especially in classifying multiclass aesthetic facial skin disorders. This study comprehensively evaluates four state-of-the-art CNN architectures, EfficientNetB0, HRNet, DenseNet201, and ResNet50, for efficient and effective classification of aesthetic facial skin disorders. This study explores the impact of feature combination techniques, particularly CM and LoG, integrated with CNN architectures, on the accuracy of multiclass classification of aesthetic facial skin disorders. It emphasizes the effectiveness of these techniques in enhancing feature representation and improving model performance.

This study is organized into six main sections. Section 1 provides the introduction, while Section 2 presents the literature review. Section 3 outlines the materials and methods used in the study. The results are presented in Section 4, followed by a discussion in Section 5, which includes the study's implications, and limitations. Finally, Section 6 offers suggestions for future research and work.

2.Literature review

A literature review was conducted on recent research on the multiclass classification of facial skin problems using CNN architecture to identify research strengths and recent developments in this field.

Research conducted by Dai [16] suggested two types of CNN-based models for image aesthetic score prediction. The ensemble method significantly improves the aesthetic score prediction. Most activated feature maps correspond to the first fixation perspective, and Feature fusion slightly improves the accuracy of aesthetic judgment. Most activated feature maps correspond to the first fixation perspective, and Feature fusion slightly improves the accuracy of aesthetic judgment. The overall accuracy for the Ares model is 0.650. However, the ensemble model does not improve the F1 score. The comparison with the CUHK-PQ dataset is less clear regarding the quality criteria.

Li et al. [17] explained that AI improves dermatological diagnosis and treatment efficiency, aids in the rehabilitation of patients with skin tumors, and enhances diagnostic accuracy in dermatology through human-AI collaboration. AI recognizes subtle differences in lesion features, and Deep learning improves image classification and object detection. Feature learning finds representations for classification from raw data. The model refers to machine learning algorithms for dermatology, including artificial neural networks (ANN) and deep learning. It classifies skin lesions and assesses cancer risk. Thus, human-AI collaboration improves clinical decision-making performance. However, there are limitations in recognizing rare skin conditions, such as small sample size and inadequate quality manual annotation, a disconnect between AI algorithms and medical application scenarios, and Limitations in recognizing image artefacts such as tattoos. Furthermore, Accuracy is affected by image artefacts such as tattoos.

In their study, Sazzadul et al. [18] used CNN models to focus on skin disease detection and addressed eczema and psoriasis as common skin conditions. Furthermore, the Inception ResNetv2 architecture achieves the highest accuracy rate. Feature extraction using CNN architecture: Different skin diseases have unique shapes and colours. Eczema and psoriasis are the focus of feature identification. The study used five CNN architectures, namely VGG16, Inception v3, ResNet50, MobileNetv2 and ResNetv2. The context provided does not mention any limitations of the research paper. In their study, Dhar and Guha [19] focused on skin lesion detection using fuzzy logic and utilized CNN for dermoscopy image classification. Fuzzy rules improve the sharpness and accuracy of segmentation where the System shows significant performance on public datasets. CNN features are extracted in hidden layers, fully connected layers classify the extracted features. Filters detect patterns such as edges, shapes, and textures. CNN shows better performance in classification tasks. Then significant performance still needs to be compared with other learning algorithms.

In their study, Thwe and Yu [20] analyzed the challenges of hand gesture recognition in computer vision, focusing on skin color segmentation using the YCbCr color space with the aim of improving recognition accuracy in real-world conditions. Threshold-based skin color detection algorithms mainly affect lighting, various image capture devices, and objects such as human skin color and ethnicity. The current problem with YCbCr threshold-based segmentation is the inability to correctly recognize hands in situations where the background contains materials similar to human skin. But still need to use a mixed colour space model instead of just using the YCbC color space model, and use deep learning to learn attention segmentation on hands.

Chickaramanna and Thippeswamy [21] explained that computer vision and image processing techniques were applied in the Ayurveda domain to identify the dominant *prakriti*, age, and gender of the subject. 68-point facial landmarks are used for feature extraction with the geometric structure of the face analysed for dosha classification. But still requires well-trained practitioners for accurate identification. Small datasets limit the accuracy of decision trees and random forests, then rely on certain facial features for classification.

The research study conducted by Dahdouh et al. [22] proposed an innovative and intelligent system for real-time detection and classification of melanoma cancer lesions using the Internet of Things (IoT) connected to a Raspberry Pi camera and a deep learning model based on a CNN algorithm. The experimental results show the efficiency of the classification system, which reaches an accuracy level of 92%. However, further improvement can be achieved by developing new layers in deep learning algorithms and incorporating more specific and carefully curated datasets. In their study, Gadag and Palraj [23] explained that they adopted a UNet-based deep learning approach and integrated an attention mechanism, considering both low-level and high-level statistics. This mechanism is combined with feedback and skip connection modules, which helps to obtain robust quality without neglecting channel information.

Khomsi et al. [24] proposed a deep learning method based on CNN, which can be used in conjunction with thermographic images to determine the size and location of breast tumours. The model uses 32 filters for feature extraction, using thermal images to predict and estimate the size and location of breast tumours, whose architecture includes convolutional, pooling, and fully connected layers, tumour parameters are limited to size and location. Makhir et al. [25] classified cardiac ischemia using hybrid CNN models. Various machine learning models improve the baseline CNN model and focus on minimizing false negatives (FN) in ischemia detection. But there are no more specific research limitations.

In their study, Maquen-Niño et al. [26] focused on pneumonia detection using chest X-ray images and employed CNN for image classification. Three transfer models are compared: DenseNet, VGG19, and ResNet50. ResNet50 achieves the highest accuracy of 0.91. Limited dataset variation can affect model generalization. Accuracy decreased from 0.91 to 0.88 during evaluation. Then, many parameters in the pre-trained model are not visible to the user. The study by Benradi et al. [27] presented a CNN model for lung disease diagnosis, distinguishing between COVID-19, viral infections, and typical cases. Histogram equalization was used to enhance image contrast for better classification. CNN extracts feature from medical images, and Features are used for disease classification in chest X-rays.

Often, aesthetic facial skin disorders are challenging to diagnose accurately through visual examination by a dermatologist alone, which can lead to delayed treatment or even misdiagnosis. Facial skin disorder classification, the performance of various CNN architectures and feature combination methods in the future, these data can help researchers and developers create more sophisticated and valuable AI models for dermatological diagnosis. Research that applies AI and CNN-based technology or with CNN architecture has been widely conducted. Still, facial skin disorders regarding aesthetics are not specific because they only focus on the disease. In contrast, before the appearance of skin disease, the disorder is first detected and can be treated appropriately so that it does not become a disease. In different cases, CNN is also widely applied with reasonable accuracy. As a result, this study is structured based on objectives, types of materials, previous theories adopted, and conclusions. This allows researchers to review and search, produce results, as asked, and study the most popular materials currently used to obtain higher accuracy results.

3.Materials and methods

This section displays the stages of research methodology: processing, feature extraction, training and testing of the model, and evaluation and indentation of the results. In addition, this stage also explains the hyperparameter, software and hardware requirement used in the study, which are described as follows:

3.1Experiment setup

The experiment setup used in this study is shown in *Table 1* and *Table 2*. *Table 1* is the hyperparameter of the experiment conducted in this study, where the loss value is cross entropy to measure the difference between the model prediction and the actual label. The optimizer uses Adam, a popular optimization algorithm, for its good performance. Furthermore, the learning rate used is 0.001, which controls how quickly the model adjusts its parameters during training. A smaller value indicates a minor step that can help the model find the global minimum. Furthermore, the epoch hyperparameter, set to 25, represents the number of iterations performed on the entire dataset; each epoch processes the entire dataset once. For batch size, the value is set to 10, which determines how many samples are used to update the model parameters in each training iteration; a smaller batch size results in faster model learning. *Table 2* shows the software and hardware configurations required to support the research process: central processing unit (CPU), graphics processing unit (GPU), operating system, RAM, disk, development environment, and programming language.

Table 1 Hyperparameter settings

Hyperparameter	**Value**
Loss	Cross Entropy Loss
Optimizer	Adam
Learning rate	0.001
epoch	25
Batch size	10

Table 2 Software and hardware requirement

Configuration Item	**Value**
CPU	Intel Core i9 13900K
GPU	NVIDIA GeForce RTX 4060 Ti
Operating System	Windows 11 (x64-based processor)
RAM	64GB
Disk	1.7 TB
Development Environment	Multiclass Classification
Programming Language	Python 3

3.2Research methodology

The research methodology section describes the block diagram shown in *Figure 1*, starting with the dataset collection process; the input dataset consists of photos of facial skin disorders. After that, the images are pre-processed, and feature extraction is carried out. This feature extraction combines techniques by utilizing the CNN architecture model with the CM method to recognize colors and shapes with LoG. After that, the dataset is split with a division of 80% training data and 20% testing. The models used for training and testing include CNN architectures such as EfficientNetB0, HRNet, DenseNet201, and ResNet50, which are compared. Subsequently, the results are evaluated and interpreted.

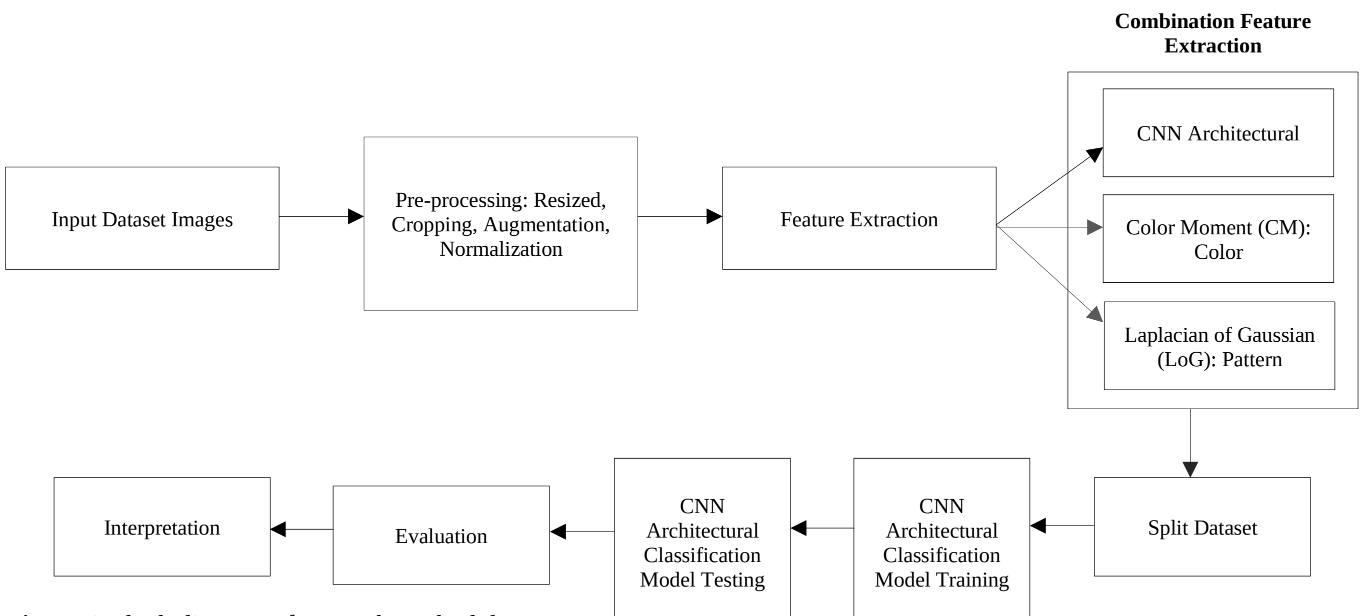


Figure 1 Block diagram of research methodology

3.3Data collection

The dataset used comprises public data sourced from DermNet on Kaggle and Roboflow Universe, as well as private data collected from respondents who consented to the data privacy statement. The dataset consists of 7,155 images of facial skin disorders, classified into six target categories: oily skin, hyperpigmentation, acne, redness, blackheads, and routine. The data is divided into 80% training data (5,724 images) and 20% test data (1,431 images). Each image has dimensions of 224×224×3 and is stored in JPEG format. Annotation is performed using a directory-based or folder-based labeling method, grouping images by folder corresponding to their class. The datasets were combined to enhance sample size, increase diversity (e.g., to represent different populations), balance class distribution, and improve model performance on underrepresented classes. *Table 3* illustrates the proportional distribution of the dataset.

Table 3 Dataset class distribution

Class	**Training**	**Testing**
Oily	1277	319
Hyperpigmentation	862	217
Acne	1019	254
Redness	1123	197
Blackheads	653	164
Normal	790	280
Total	**5724**	**1431**

Figure 2 shows sample images from the dataset, which consist of oily, hyperpigmentation, acne, redness, blackheads, and normal. This dataset image collection is used as input data processed by the CNN architecture model.

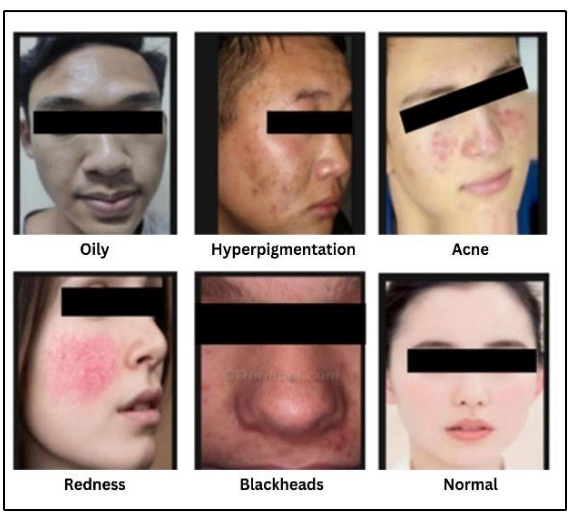


Figure 2 Sample dataset

3.4Pre-Processing

The following are the pre-processing stages carried out to improve data quality, performance, and reliability and to ensure that they can match the deep learning model used, such as the CNN architecture model.

1. Labelling is done to categorize each facial skin image into a particular aesthetic skin disorder class, such as (oily, hyperpigmentation, acne, redness, blackheads and normal).
2. Resize: The facial skin disorder images in the dataset may be of different sizes. Resizing ensures that all images have the same dimensions before being used to train the CNN architecture model.
3. Cropping: In this study, cropping was used to remove irrelevant parts of the image, such as hair or background, to focus on the areas of facial skin affected by disturbances.
4. Rotation and Flip: Rotating and flipping facial skin images can help CNN architecture models learn variations in the orientation and position of facial skin perturbations.
5. Normalization: Normalization is performed on facial skin images before they are used to train the CNN architecture model. This ensures that all pixels have the same range of values, which helps the model learn more efficiently.

Figure 3 shows the pre-processing steps, with labelling, resizing, cropping, rotation, and normalization being the crucial steps in image pre-processing for this study. These steps help prepare the data to fit the CNN architecture model, improve model performance, and reduce overfitting.

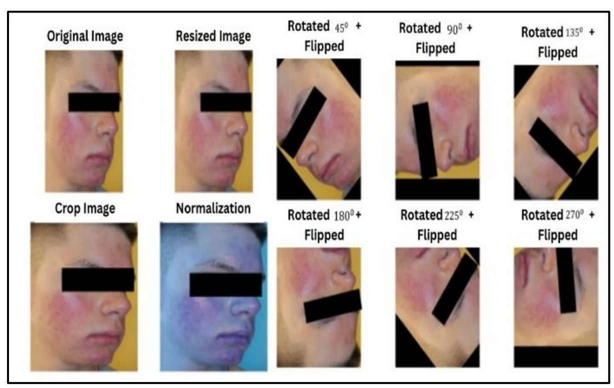


Figure 3 Pre-processing stages

3.5Feature extraction

This study utilizes feature extraction techniques to improve the representation of facial skin image data before classification using the CNN model architecture. The feature extraction techniques used include CM to extract colour information and LoG to extract texture information. This technique is expected to improve model performance in distinguishing various aesthetic facial skin disorders.

3.5.1EfficientNetB0

EfficientNeB0 optimizes computational resources using a co-scaling approach, simultaneously scaling the input image resolution, network depth, and channel width, whose architecture is shown in *Figure 4*. The formulas are presented in Equations 1, 2, and 3 as follows:

a. Input Resolution

$$R = \alpha^i \times 224 \quad (1)$$

b. Network Depth

$$D = \alpha^i \times 1 \quad (2)$$

c. Channel width

$$W = \alpha^i \times 1 \quad (3)$$

Where:

- α is a scale factor (usually around 1.2).
- i is an index indicating the variation of EfficientNetB0.

MBConv, a special convolutional block designed for efficiency, is used by every EfficientNetB0 block. MBConv consists of:

a. Expansion: Increase the representation capacity by increasing the channel width.
b. Depth convolution: Perform convolution with a small kernel to reduce computational cost.
c. Squeeze: Reduces computational costs by reducing the channel width.

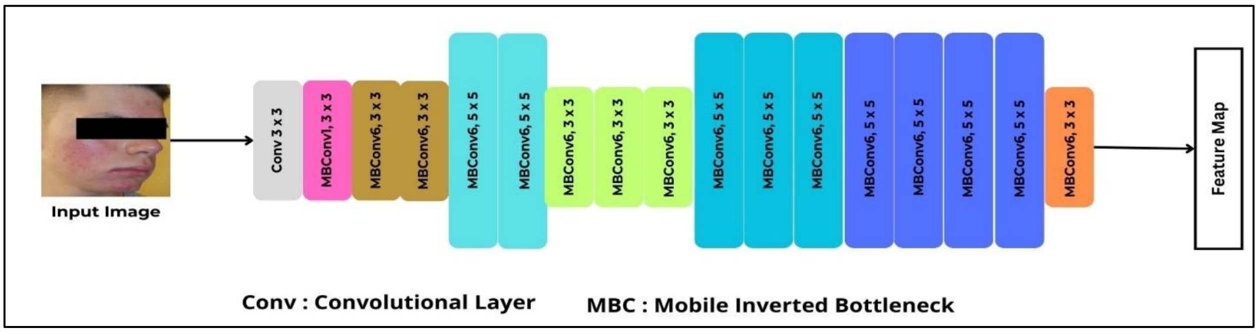


Figure 4 Feature extraction architecture using convolutional layers and MBConv

3.5.2High resolution network (HRNet)

HRNet does not have a formula that governs its entire architecture. This architecture is more complex, and maintaining high resolution throughout the process involves many convolutional layers, pooling layers, and unique connections, whose architecture is shown in *Figure 5*.

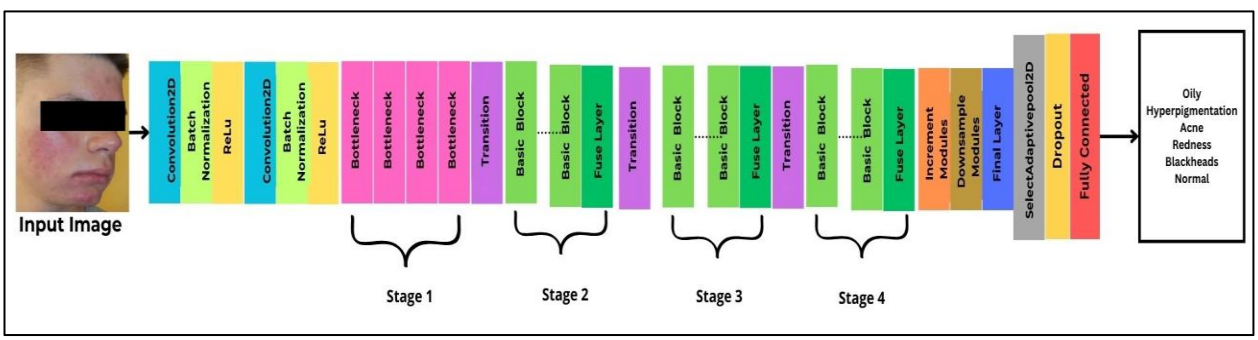


Figure 5 HRNet architecture for facial skin disorder classification

3.5.3DenseNet201

DenseNet201 compare its performance and advantages with other CNN models in classifying aesthetic facial skin disorders. The DenseNet201 architecture is more complex and consists of many convolutional, pooling, and dense layers rather than a single formula governing its entire structure. Although there is no specific formula for dense connectivity, it can be thought of as the sum of the current layer inputs and the outputs of all previous layers, whose architecture is shown in *Figure 6.*

DenseNet201 consists of dense blocks, which consist of multiple convolutional layers. The outputs of all previous layers are connected to the inputs of the next layer through dense connections between each block. Dense blocks consist of a combination of convolutional layers, activation functions, and normalization techniques.

Transition layers are used between dense blocks to control the size of map features and reduce the number of features. The transition layer usually consists of a 1×1 convolution and pooling layers. The growth rate (k) represents the number of new features added by each layer within a dense block. For DenseNet201, the growth rate is set to k=32.

DenseNet201 consists of 201 layers divided into four dense blocks. A specific formula does not determine the number of layers, but the designed architecture determines the number. For the majority of its layers, DenseNet201 utilizes the rectified linear unit (ReLU) activation function, as described in Equation 4.

$$ReLU(x) = \max(0, x). \quad (4)$$

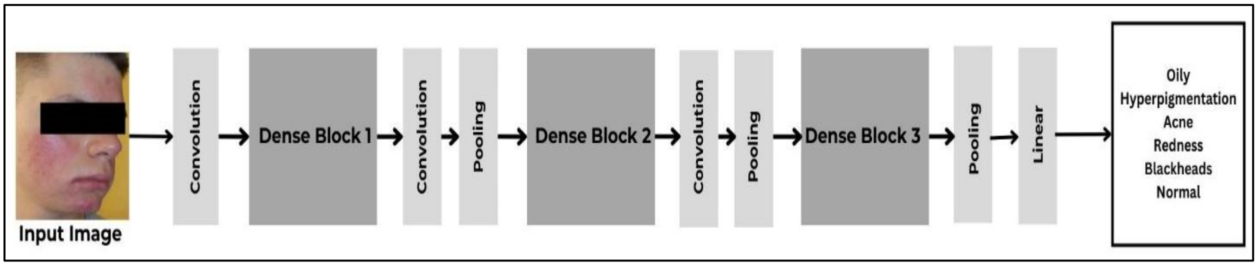


Figure 6 DenseNet201 architecture for facial skin disorder classification

3.5.4ResNet50

ResNet50 was utilized to evaluate its performance and effectiveness in classifying aesthetic facial skin disorders when compared with various CNN model architectures. ResNet50 leverages residual connections to facilitate training and enhance representation capabilities by enabling direct data flow through the network. Its architecture is illustrated in *Figure 7*. The output of the residual layer is defined in Equation 5 as follows:

$$y = F(x) + x \quad (5)$$

Where:

- x is input to the residual layer.
- $F(x)$ is output of a convolutional layer (or series of layers) in the layer.
- y is the output of the residual layer.

ResNet50 consists of residual blocks, each with multiple convolutional layers. Each block uses a residual connection to add the output of each convolutional layer to the input of the block. The residual block requires a combination of convolution, activation, and normalization, and there is no unique formula. Convolutional layers with different kernel sizes are used in ResNet50. Convolution formula is shown in Equation 6 as follow:

$$y = conv(x, filter) \quad (6)$$

Where:

- x is input to the convolutional layer.
- $filter$ is convolution kernel.
- y is output of the convolutional layer.

ResNet50 utilizes pooling layers, including max pooling and average pooling, to extract essential features from feature maps. These operations are defined in Equations 7 and 8 as follows:

a. Max Pooling

$$y = \max(x, kernel) \quad (7)$$

b. Average Pooling

$$y = avg(x, kernel) \quad (8)$$

Where:

- x is input to the pooling layer.
- $kernel$ is kernel pooling.
- y is output of the pooling layer.

For most layers, ResNet50 uses the ReLU activation function. $ReLU(x) = \max(0, x)$. ResNet50 consists of fifty layers divided into various residual blocks. A specific formula does not determine the number of layers, but the designed architecture determines the number.

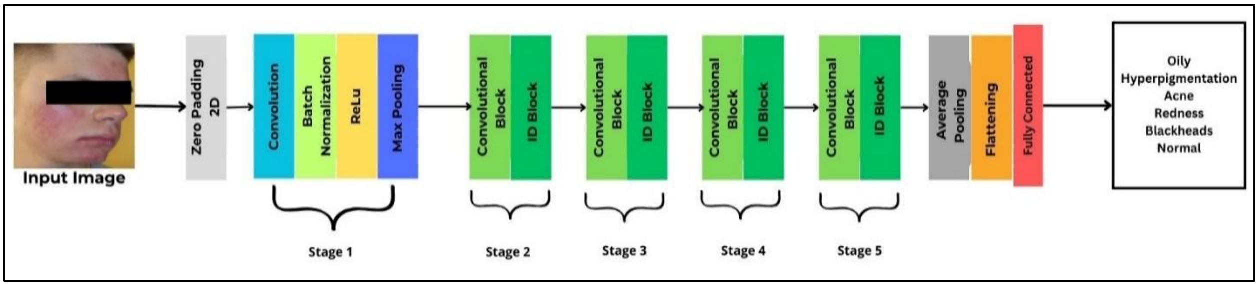


Figure 7 ResNet50 architecture for facial skin disorder classification

3.5.5Color moment

Colour is one of the characteristics of facial skin disorders that uses the CM method with a graphical representation of the frequency distribution of intensity values in an image. CM is used in image processing as they calculate the frequency distribution of pixel intensity values in an image, as illustrated in Equations 9, 10, and 11.

1. First Moment (Mean): (9)

a. $\mu(R) = \frac{\Sigma(R(i))}{N}$

b. $\mu(R) = \frac{\Sigma(G(i))}{N}$

c. $\mu(R) = \frac{\Sigma(B(i))}{N}$

2. The Second Moment (Variants): (10)

a. $\sigma^2(R) = \frac{\Sigma((R(i)-\mu(R))^2)}{N}$

b. $\sigma^2(G) = \frac{\Sigma((G(i)-\mu(G))^2)}{N}$

c. $\sigma^2(B) = \frac{\Sigma((B(i)-\mu(B))^2)}{N}$

3.The Third Moment (Skewness): (11)

a. $\gamma(R) = \frac{\Sigma((R(i)-\mu(R))^3)}{[N\times(\sigma^2(R))^{\frac{3}{2}}]}$

b. $\gamma(G) = \frac{\Sigma((G(i)-\mu(G))^3)}{[N\times(\sigma^2(G))^{\frac{3}{2}}]}$

c. $\gamma(B) = \frac{\Sigma((B(i)-\mu(B))^3)}{[N\times(\sigma^2(B))^{\frac{3}{2}}]}$

where:

- $R(i), G(i), B(i)$ is the colour intensity values at the i pixel for the red (R), green (G), and blue (B) colour channels.
- N is the total number of pixels in the image.
- μ(R), μ(G), μ(B) is average colour for the red (R), green (G), and blue (B) colour channels.
- σ^2(R), σ^2(G), σ^2(B) is colour variance for red (R), green (G), and blue (B) colour channels.
- γ(R), γ(G), γ(B) is colour skewness for the red (R), green (G), and blue (B) colour channels.
- Σ is the count symbol indicates the number of all pixels in the image.

3.5.6Laplacian of gaussian

For shape features, the LoG operator is used to identify corners and edges in the image. This operator works by convolving the image with a Gaussian kernel and then applying the Laplacian, as shown in Equation 12.

$$LoG(x, y) = (x^2 + y^2 - 2\sigma^2) \times \exp\left(-\frac{(x^2 + y^2)}{(2\sigma^2)}\right) \quad (12)$$

where:

- LoG(x, y) is LoG value at coordinates (x, y),
- σ is Gaussian kernel standard deviation,
- x, y is coordinates of points in the image.

3.5.7Extraction feature combination technique

Figure 8 is a flow diagram of the feature extraction combination process that displays the process stages by utilizing the CNN architecture model's automatic feature extraction feature with the CM and LoG methods for color and pattern features.

3.6Models training dan testing

This study uses CNN architecture models such as EfficientNetB0, HRNet, DenseNet201 and ResNet50 to train and test datasets. The CNN architecture is chosen for several reasons, including the computational efficiency of EfficientNetB0, which is designed to achieve a balance between performance and computation. Furthermore, HRNet can process high-resolution images, and DenseNet201 and ResNet50 can learn complex features.

The training and testing process of CNN architecture models involves systematic steps to train and evaluate model performance. This study uses a comprehensive approach to compare different CNN architectures and feature combination techniques to identify the most effective solution for aesthetic facial skin disorder classification.

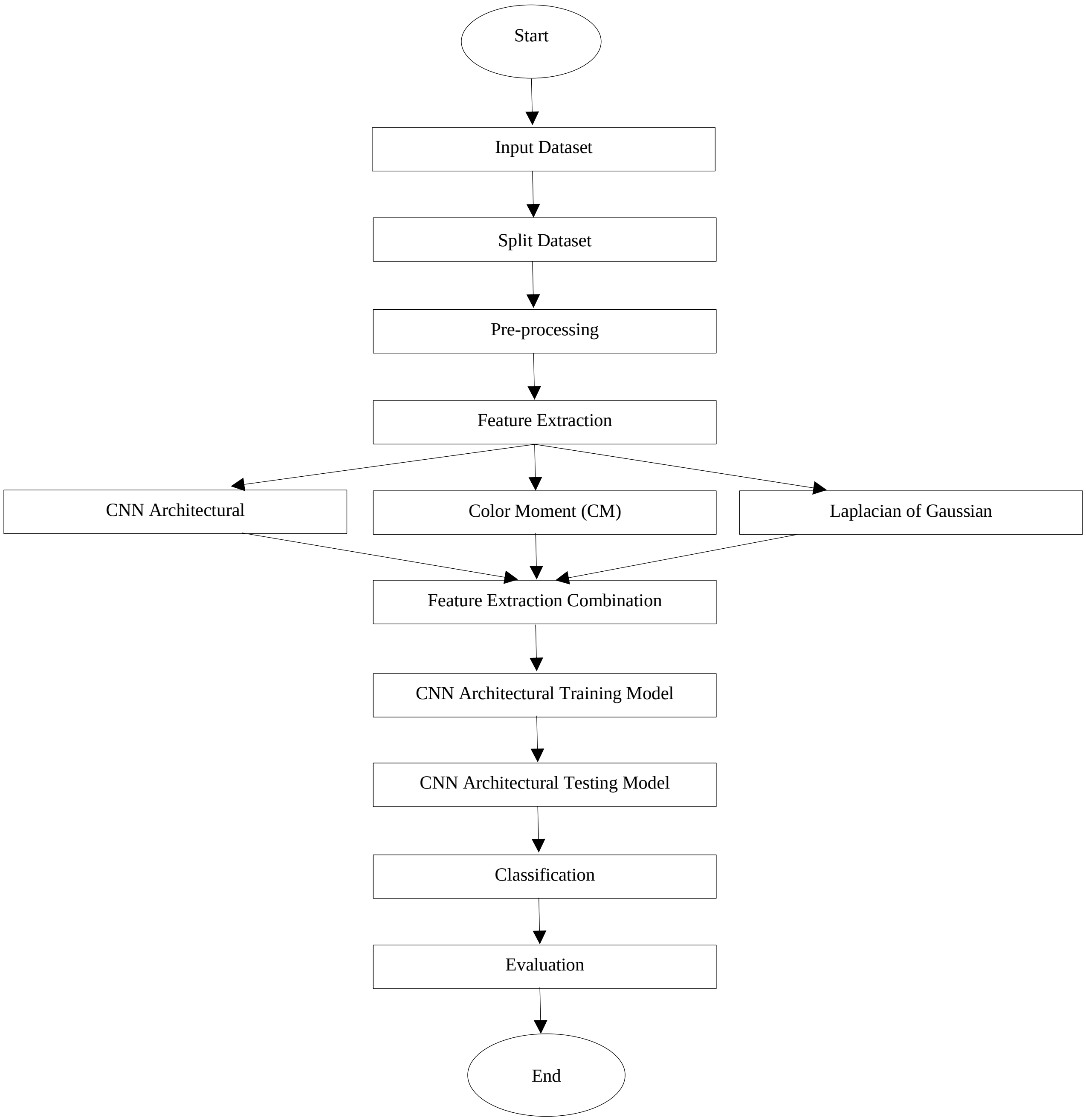


Figure 8 Flow diagram of feature extraction and combination

Algorithm I. EfficientNetB0-CM-LoG combination

1. Input Image Dataset
2. Split the dataset into training and testing with a ratio of 80:20
3. Preprocessing stage is performed on image dataset
4. Define an EfficientNetB0 model with a combination of CM and LoG features:
   a. Extraction with EfficientNetB0: The image is fed into the efficientNetB0 model, which processes and produces the EfficientnetB0 feature vector.
   b. Extraction with CM:
      - Colour Channel Separation: splitting an image into individual colour channels. Channel separation

divides an image into three channels (Red, Green, and Blue). The result is a channel consisting of three arrays or matrices, each representing the R, G, and B channels.
   - Feature Vector Initialization
   - Iterating Through Colour Channels: iterate through each previously separated colour channel (R, G, and B).
   - Color Moment Calculation for Each Channel
   - Feature Merger: The moment list includes the mean, standard deviation, and slope for each colour channel. After the loop completes, the moments will contain nine values (three for each channel: mean, standard, deviation).

   c. Extraction with LoG:
   - Convert image to Grayscale
   - LoG transform on grayscale images.
   - Displays an image that emphasizes edges and details based on the spatial scale specified by the parameters in the Gaussian filter.
   - Calculates the average pixel intensity of a filtered image.
   - Average Extraction of features
   - Displays information about the texture and details in the image.

5. Combine EfficientNetB0, Color Moments, and LoG features using torch cat.
6. The results of the feature combination are entered into the fully connected layer for classification.
7. Classification model initialization
   model = EfficientNetB0Combination ()
8. Input the testing data into the initialized model to complete the classification process.
9. Classification process by model.
10. Classification result output.
11. Calculate classification report and displays confusion matrix.

Algorithm II. HRNet-CM-LoG combination

1. Input Image Dataset
2. Split the dataset into training and testing with a ratio of 80:20
3. Preprocessing stage is performed on image dataset
4. Define an HRNet model with a combination of CM and LoG features:
   a. Extraction with HRNet: The model processes the image and produces a feature vector (HRnet_features), which is a numerical representation of the image main characteristics.
   b. Extraction with CM:
   - Colour Channel Separation: splitting an image into individual colour channels. Channel separation divides an image into three channels (Red, Green, and Blue). The result is a channel consisting of three arrays or matrices, each representing the R, G, and B channels.
   - Feature Vector Initialization
   - Iterating Through Colour Channels: iterate through each previously separated colour channel (R, G, and B).
   - Color Moment Calculation for Each Channel
   - Feature Merger: The moment list includes the mean, standard deviation, and slope for each colour channel. After the loop completes, the moments will contain nine values (three for each channel: mean, standard, deviation).

   c. Extraction with LoG:
   - Convert image to Grayscale
   - LoG transform on grayscale images.
   - Displays an image that emphasizes edges and details based on the spatial scale specified by the parameters in the Gaussian filter.
   - Calculates the average pixel intensity of a filtered image.

- Average Extraction of features
- Displays information about the texture and details in the image.

5. Combine HRNet, Color Moments, and LoG features using torch cat.
6. The results of the feature combination are entered into the fully connected layer for classification.
7. Classification model initialization
   model = HRNetCombination()
8. Input the testing data into the initialized model to complete the classification process.
9. Classification process by model.
10. Classification result output.

Calculate classification report and displays confusion matrix.

Algorithm III. DenseNet201-CM-LoG combination

1. Input Image Dataset
2. Split the dataset into training and testing with a ratio of 80:20
3. Preprocessing stage is performed on image dataset
4. Define an DenseNet201 model with a combination of CM and LoG features:
   a. Extraction with DenseNet201: The model processes the image and produces a feature vector (densenet_features). This vector is a numerical representation of the main characteristics of the image.
   b. Extraction with CM:
      - Colour Channel Separation: splitting an image into individual colour channels. Channel separation divides an image into three channels (Red, Green, and Blue). The result is a channel consisting of three arrays or matrices, each representing the R, G, and B channels.
      - Feature Vector Initialization
      - Iterating Through Colour Channels: iterate through each previously separated colour channel (R, G, and B).
      - Color Moment Calculation for Each Channel
      - Feature Merger: The moment list includes the mean, standard deviation, and slope for each colour channel. After the loop completes, the moments will contain nine values (three for each channel: mean, standard, deviation).
   c. Extraction with LoG:
      - Convert image to Grayscale
      - LoG transform on grayscale images.
      - Displays an image that emphasizes edges and details based on the spatial scale specified by the parameters in the Gaussian filter.
      - Calculates the average pixel intensity of a filtered image.
      - Average Extraction of features
      - Displays information about the texture and details in the image.
5. Combine DenseNet201, Color Moments, and LoG features using torch cat.
6. The results of the feature combination are entered into the fully connected layer for classification.
7. Classification model initialization
   model = DenseNet201Combination ()
8. Input the testing data into the initialized model to complete the classification process.
9. Classification process by model.
10. Classification result output.
11. Calculate classification report and displays confusion matrix.

Algorithm IV. ResNet50-CM-LoG combination

1. Input Image Dataset
2. Split the dataset into training and testing with a ratio of 80:20
3. Preprocessing stage is performed on image dataset
4. Define an ResNet50 model with a combination of CM and LoG features:

a. Extraction with ResNet50: The model processes the image and produces a feature vector (resnet_features). This vector is a numerical representation of the main characteristics of the image.

b. Extraction with CM:
- Colour Channel Separation: splitting an image into individual colour channels. Channel separation divides an image into three channels (Red, Green, and Blue). The result is a channel consisting of three arrays or matrices, each representing the R, G, and B channels.
- Feature Vector Initialization
- Iterating Through Colour Channels: iterate through each previously separated colour channel (R, G, and B).
- Color Moment Calculation for Each Channel
- Feature Merger: The moment list includes the mean, standard deviation, and slope for each colour channel. After the loop completes, the moments will contain nine values (three for each channel: mean, standard, deviation).

c. Extraction with LoG:
- Convert image to Grayscale
- LoG transform on grayscale images.
- Displays an image that emphasizes edges and details based on the spatial scale specified by the parameters in the Gaussian filter.
- Calculates the average pixel intensity of a filtered image.
- Average Extraction of features
- Displays information about the texture and details in the image.

5. Combine ResNet50, Color Moments, and LoG features using torch cat.
6. The results of the feature combination are entered into the fully connected layer for classification.
7. Classification model initialization
   model = ResNet50Combination ()
8. Input the testing data into the initialized model to complete the classification process.
9. Classification process by model.
10. Classification result output.
11. Calculate classification report and displays confusion matrix.

3.7Evaluation

After training, the performance of the CNN architecture models is thoroughly evaluated with independent test data. The calculation of performance metrics such as accuracy in Equation 14, precision in Equation 15, recall in Equation 16, and F1-Score in Equation 17 for each CNN architecture (EfficientNetB0, HRNet, DenseNet201, and ResNet50) is used in the analysis of the evaluation results, both with and without the integration of CM and LoG features. The evaluation results focus on finding the most effective models and feature extraction techniques for classifying aesthetic facial skin disorders and how this will impact the development of AI-based diagnostic technologies. The formula as shown in Equations from 13 to 16:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (13)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (14)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (15)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (16)$$

Where:
- Accuracy is calculating the percentage of correct predictions out of all predictions made by the model.
- Precision is shows how accurate the model is in predicting the positive class.
- Recall is measuring the proportion of correct optimistic predictions from all positive examples. Recall shows how well the model finds all positive examples.
- F1-Score is the harmonic mean of precision and recall, which provides a balance between precision and recall.
- TP is a true positive prediction
- TN is a true negative prediction
- FP is a false positive prediction
- FN is a false negative prediction.

3.8Interpretation

The CNN model architecture provides predictions and insights into the most influential factors in determining predictions and analyzing models that can identify the most critical features in classifying aesthetic facial skin disorders; the analysis is carried out on the most influential features in displaying the results of multiclass classification of aesthetic facial skin disorders. Furthermore, *Table 4* shows the target of the multiclass classification of aesthetic facial skin disorders.

4.Result

This section presents the performance evaluation results of four CNN architectures (EfficientNetB0, HRNet, DenseNet201, and ResNet50) with and without applying the feature combination techniques, CM and LoG. The evaluation was conducted using separate testing data never used during training. The performance metrics used include accuracy, precision, recall, and F1-Score. *Figure 9* illustrates the system architecture showcasing the feature extraction combination process. The system begins with an input dataset of facial skin disorder images, followed by a pre-processing stage. Each CNN architecture model is integrated with color and shape features using the CM and LoG methods. Subsequently, training and testing are conducted with the CNN architecture models, producing multiclass classification results, including oily, hyperpigmentation, acne, redness, blackheads, and normal. This study evaluates different CNN architectures and feature extraction techniques to identify the most effective approach for facial skin disorder classification, emphasizing the optimization of classification performance through these techniques.

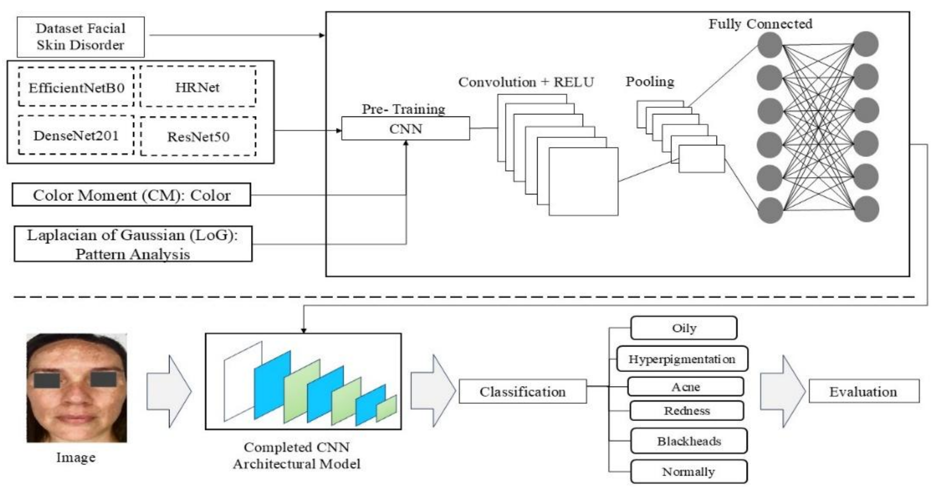


Figure 9 System architecture for facial skin disorder classification using CNN and feature extraction techniques

Table 4 displays the multiclass target of facial skin disorders related to aesthetics. The target class consists of six facial skin disorders that are most often experienced by someone. Each multiclass target is given an identifier with the variables D1, D2, D3, D4, D5, and D6. Then, each class is also given a value from 0 to 5.

Table 4 Target class labels and corresponding variable values

Target Class	Variable	Value
Oily	D1	0
Hyperpigmentation	D2	1
Acne	D3	2
Redness	D4	3
Blackheads	D5	4
Normal	D6	5

4.1Experimentation with CNN architecture model without feature extraction combination

This experiment, shown in *Table 5*, evaluates the performance of four CNN architectures, namely EfficientNetB0, HRNet, DenseNet201, and ResNet50, without using additional feature extraction techniques. The goal is to establish a performance baseline and understand how each architecture plays a role in classifying aesthetic facial skin disorders. The ability of each architecture to extract features directly from raw images can be separated by removing CM and LoG. The results of this experiment provide an essential reference combination the performance of models after the incorporation of feature extraction methods. This allows us to measure how the incorporation of color and texture information impacts the classification accuracy. Based on the evaluation results, EfficientNetB0 obtained the highest accuracy with a rate of 0.92, and HRNet had the lowest rate of 0.84.

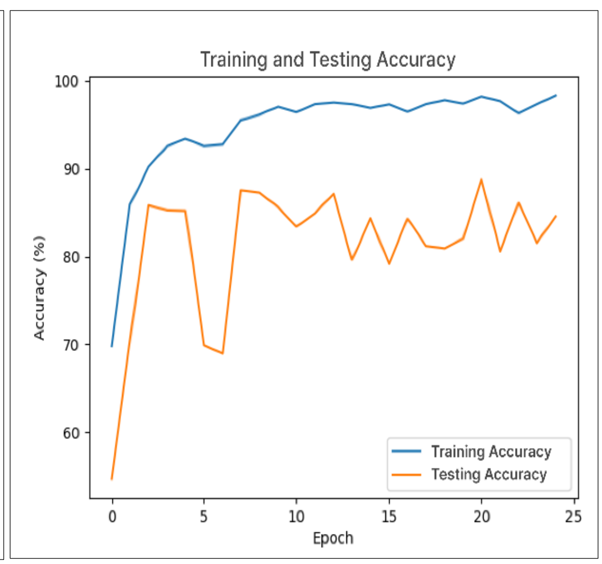
While HRNet is successful in tasks that require a lot of spatial detail, facial skin abnormalities do not necessarily require very high resolution to be classified correctly. HRNet is too complex for the dataset used, making it prone to overfitting. At the same time, EfficientNetB0, which is more computationally efficient, has been quite successful in extracting the essential features needed for classification. Element of randomness in deep learning model training. Despite using the same seed, there is still variation in model performance due to different weight initializations. *Figure 10* shows the training and testing accuracy graphs of the CNN architecture model without a combination of feature extraction wherein part (a) the efficientNetB0 architecture produces a testing accuracy of 0.92 with a training accuracy of 0.99, (b) HRNet produces a training accuracy of 0.99 and a testing accuracy of 0.84, (c) DenseNet201 produces a training accuracy of 0.99 and testing accuracy of 0.89 and (d) ResNet50 obtains a training accuracy of 0.99 with a testing accuracy of 0.91.

Table 5 CNN architecture models result without feature extraction combination

Model	Matrix	Class D1	D2	D3	D4	D5	D6	Accuracy	Micro Average ROC-AUC
M1	Precision	0.91	0.93	0.87	0.94	0.93	0.94	0.92	0.95
	Recall	1.00	0.76	0.87	1.00	0.87	1.00		
	F1 Score	0.95	0.84	0.87	0.97	0.90	0.97		
	ROC-AUC	0.97	0.94	0.95	0.95	0.94	0.96		
M2	Precision	0.80	0.92	0.84	0.84	0.81	0.85	0.84	0.90
	Recall	0.92	0.72	0.88	0.85	0.78	0.82		
	F1 Score	0.85	0.81	0.86	0.85	0.79	0.84		
	ROC-AUC	0.92	0.85	0.92	0.91	0.88	0.89		
M3	Precision	0.95	0.90	0.93	0.90	0.87	0.82	0.89	0.93
	Recall	0.92	0.81	0.87	0.88	0.89	0.96		
	F1 Score	0.93	0.85	0.90	0.89	0.88	0.88		
	ROC-AUC	0.95	0.90	0.93	0.93	0.94	0.95		
M4	Precision	0.96	0.91	0.95	0.94	0.79	0.89	0.91	0.94
	Recall	0.95	0.88	0.86	0.88	0.93	0.94		
	F1 Score	0.95	0.90	0.90	0.91	0.85	0.91		
	ROC-AUC	0.97	0.93	0.92	0.94	0.95	0.95		

M1 = EfficientNetB0; M2 = HRNet; M3 = DenseNet201; M4 = ResNet50

(a)



(b)

Figure 10 (a)EffcientNetB0, (b)HRNet, (c)DenseNet201 and (d)ResNet50 training and testing accuracy without feature extraction combination

4.2Experimentation with CNN architecture model and feature extraction combination

This experiment investigates how the incorporation of Log and CM feature extraction techniques impacts the performance of four CNN architectures in facial skin disorder classification. The goal of incorporating color CM and shape LoG information into the raw image representation is to improve the model ability to distinguish between different types of facial skin disorders, the CM and LoG feature extraction techniques. To measure the performance improvements brought about by the addition of richer feature information, the results of this experiment are compared with the results of an experiment that did not involve feature extraction. A thorough analysis of these results reveals whether the improvements in accuracy and generalization of the classification model offset the increase in computational complexity caused by feature integration. Based on the evaluation results shown in *Table 6*, ResNet50 has the highest accuracy rate of 0.95, and HRNet has the lowest accuracy rate of 0.88.

The ResNet50 architecture is more effective in extracting relevant features for the multiclass classification of aesthetic facial skin disorders. Although HRNet is designed for spatial details, the most important features for classification are still global or more straightforward to detect by the ResNet50 architecture. Next, the way ResNet50 and HRNet interact with the LoG and CM features are different. ResNet50 uses color and shape information better when using the classification process.

Table 6 CNN architecture models result with feature extraction combination

Model	**Matrix**	**Class**						**Accuracy**	**Micro Average ROC-AUC**
		D1	**D2**	**D3**	**D4**	**D5**	**D6**		
M1	Precision	0.97	0.85	0.94	0.92	0.85	0.92	0.92	0.95
	Recall	0.95	0.94	0.86	0.92	0.85	0.95		
	F1 Score	0.96	0.89	0.90	0.92	0.86	0.93		
	ROC-AUC	0.97	0.95	0.93	0.95	0.92	0.96		
M2	Precision	0.88	1.00	0.67	1.00	1.00	0.86	0.88	0.92
	Recall	0.93	0.69	1.00	0.93	0.77	1.00		
	F1 Score	0.90	0.81	0.80	0.97	0.87	0.92		
	ROC-AUC	0.93	0.84	0.95	0.94	0.91	0.93		
M3	Precision	1.00	0.92	0.92	0.82	0.95	0.92	0.92	0.94
	Recall	0.88	0.92	0.73	0.93	1.00	1.00		
	F1 Score	0.93	0.92	0.81	0.88	0.97	0.96		
	ROC-AUC	0.92	0.95	0.90	0.94	0.95	0.97		

Model	Matrix	Class						Accuracy	Micro Average ROC-AUC
		D1	D2	D3	D4	D5	D6		
M4	Precision	1.00	1.00	0.83	1.00	1.00	0.94	0.95	0.95
	Recall	1.00	1.00	1.00	0.79	0.95	0.94		
	F1 Score	1.00	1.00	0.93	0.88	0.98	0.94		
	ROC-AUC	0.96	0.96	0.94	0.93	0.94	0.96		

Information:M1 = EfficientNetB0, M2 = HRNet, M3 = DenseNet201, M4 = ResNet50

Figure 11 shows the graph of training and testing accuracy with a combination of feature extraction. Each CNN architecture model for EfficientNetB0, HRNet, DenseNet201, and ResNet50 produces a training accuracy of 0.99. Then, sequentially, for the highest accuracy, the rates are 0.92, 0.88, 0.92, and 0.95.

Figure 12 shows a recapitulation diagram of the overall results of the accuracy comparison of the CNN architecture model with a combination of feature extraction from CM and LoG. The chart shows that the highest accuracy is obtained with a rate of 0.95 on the ResNet50 architecture.

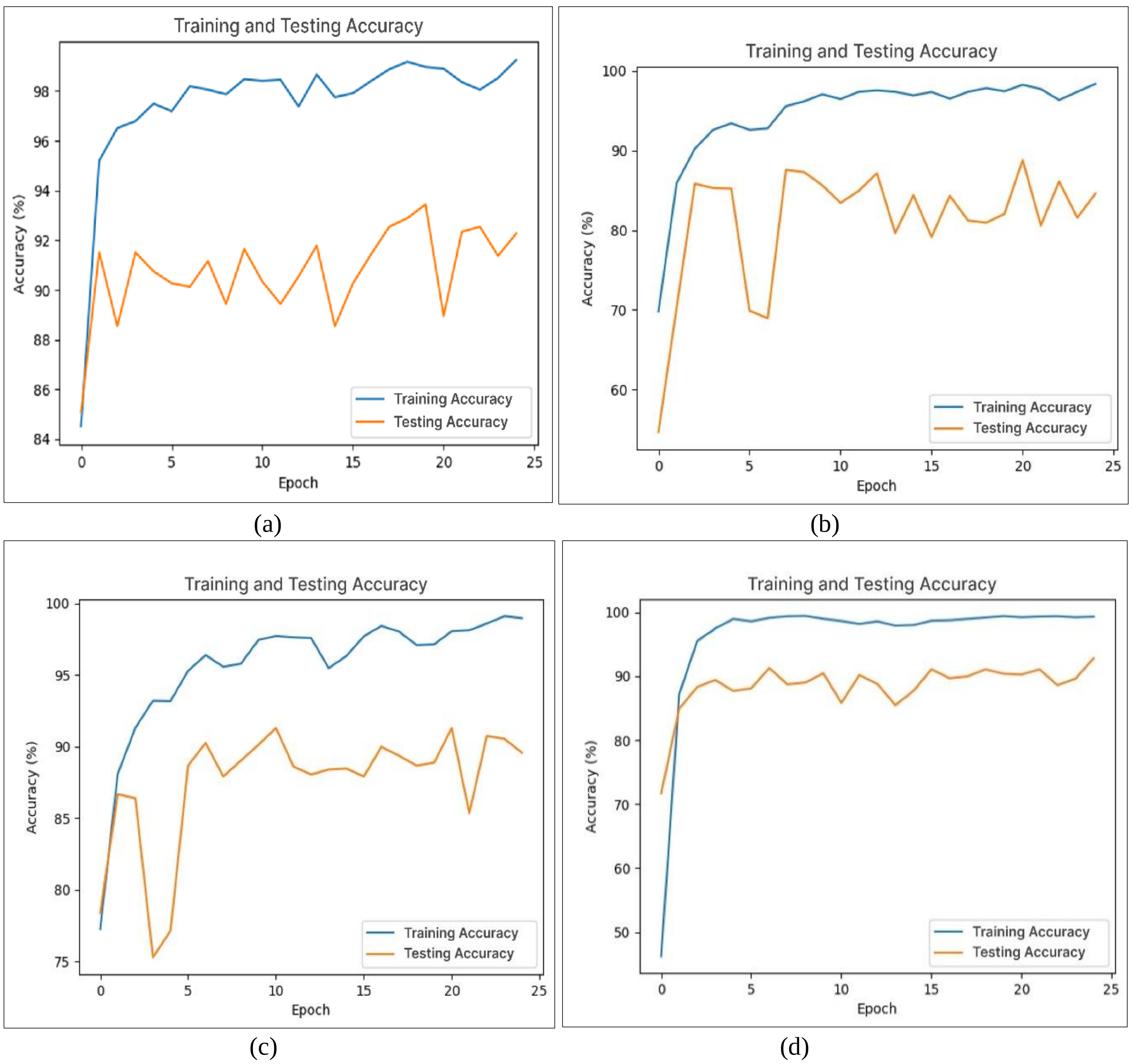


Figure 11 (a)EffcientNetB0, (b)HRNet, (c)DenseNet201 and (d)ResNet50 training and testing accuracy with feature extraction combination

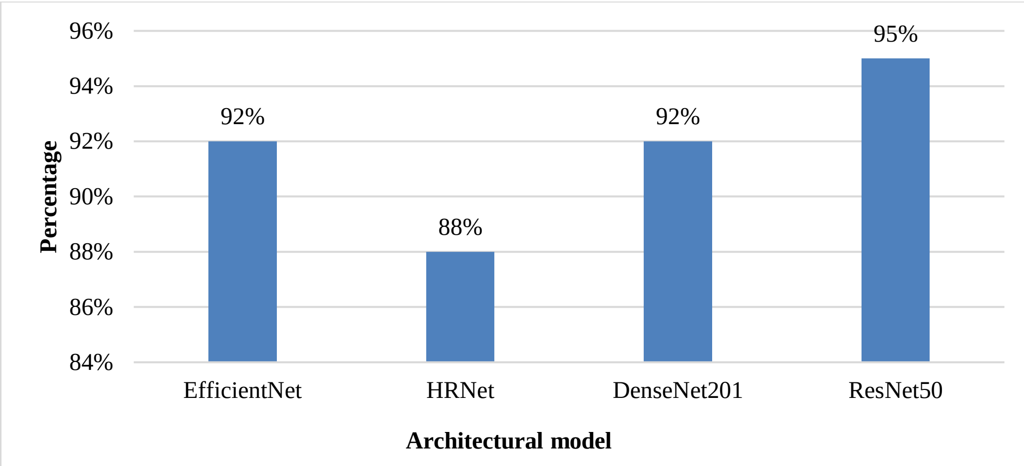


Figure 12 Comparison of CNN architectural models based on classification accuracy

5.Discussion

Based on the presented results, this section discusses how the research findings impact the development of AI-based diagnostic systems for aesthetic facial skin disorders. It also examines how various CNN architectures work, how feature extraction methods affect them, and the constraints of the methods used. Performance evaluation of four CNN architectures (EfficientNetB0, HRNet, DenseNet201, and ResNet50) in categorizing aesthetic facial skin disorders, both with and without the incorporation of feature extraction techniques. An analysis of the components contributing to the differences in model performance is conducted, highlighting how these findings impact the development of AI-based diagnostic systems. *Table 7* shows a comparison of the accuracy of the architecture model without combination and using a combination of feature extraction, where after the combination of feature extraction, there is an increase in accuracy results on the HRNet, DenseNet201, and ResNet50 architectures. However, on EfficientNetB0, the accuracy results remain the same is 0.92. EfficientNetB0 can already implicitly extract colour and shape information from raw images. There is no point in adding CM and LoG features as the EfficientNetB0 architecture already captures this information. In other words, the features extracted by CM and LoG are redundant, as they are already represented in the features learned by EfficientNetB0.

Table 7 Comparison of model accuracy results

Architectural model without combination extraction feature	Accuracy	Average ROC-AUC	Architectural model with combined feature extraction	Accuracy	Average ROC-AUC
EfficientNetB0	0.92	0.95	EfficientNetB0-CM-LoG	0.92	0.95
HRNet	0.84	0.90	HRNet-CM-LoG	0.88	0.92
DenseNet201	0.89	0.93	DenseNet201-CM-LoG	0.92	0.94
ResNet50	0.91	0.94	ResNet50-CM-LoG	0.95	0.95

Several previous studies have investigated skin problems, focusing on skin diseases in specific body regions using CNN-based models for classification. For example, Superficial Spreading Melanoma, Nodular Melanoma, and Lentigo Maligna Melanoma were classified with an accuracy of 0.88 [28]. Similarly, ResNet50 was employed for Melanoma, Basal Cell Carcinoma, and Squamous Cell Carcinoma classification, achieving accuracies of 0.83 [29] and 0.67 [30], respectively. The classification of low-risk and high-risk melanomas using DenseNet121 yielded an accuracy of 0.84 [31], while AlexNet achieved an accuracy of 0.79 [32]. Additionally, MobileNet, VGG19, ResNet, EfficientNet, Inception, and DenseNet were used to classify Acne, Eczema, Psoriasis, Dermatitis, and Skin Infections, resulting in an accuracy of 0.90 [33].

The findings of this study have significant implications for the development process of AI-based diagnostic systems for aesthetic facial skin disorders. It is possible to create accurate and reliable diagnostic systems by combining appropriate feature extraction techniques, such as CM and LoG, with efficient CNN architectures, such as EfficientNetB0.

In particular, ResNet50 ability to classify facial skin disorders shows the importance of choosing the right architecture and integrating features well. This study provides empirical evidence to support the development of AI-based diagnostic tools that can help dermatologists diagnose skin conditions more quickly and accurately, improving the quality of patient care and treatment outcomes. In addition, these findings pave the way for future research focused on model optimization, the development of more sophisticated feature extraction methods, and validation on more extensive and different datasets.

Limitations

Several limitations of this study need to be considered. First and foremost, the size of the dataset used is tiny, which increases the risk of overfitting and reduces the reliability and generalization of the model. A more extensive and more diverse dataset would improve the reliability and generalization of the model. Second, research into alternative feature extraction methods may improve classification performance, although LoG and CM methods were used. Third, this study only examined a few CNN architectures; further research could provide additional information. The last, this study only classifies aesthetic facial skin disorders. Adding other types of skin disorders and combining them with other clinical data will increase the clinical value of this system. These limitations point to further research that can improve the reliability and generalization of the AI-based facial skin disorder diagnosis system.

A complete list of abbreviations is listed in *Appendix I*.

6.Conclusion and future work

This study evaluates the performance of four CNN architectures, namely EfficientNetB0, HRNet, DenseNet201, and ResNet50, in multiclass classification of aesthetic facial skin disorders, both with and without LoG and CM feature extraction techniques. The results show that ResNet50 consistently has the best performance, both with and without feature integration. The testing accuracy is high, with a ratio before the combination of feature extraction of 0.91. After the combination, it obtained an accuracy increase of 0.95 and became the highest accuracy value of the architecture model used. These findings suggest that proper selection of CNN architecture is crucial and that performance improvement through proper feature integration is possible, although the performance improvement on EfficientNetB0 is not significant. Furthermore, ResNet50 shows the best performance.

Future work will involve collecting more extensive and diverse datasets, which will be a top priority in improving model generalization and reducing the risk of overfitting. More sophisticated feature extraction techniques, beyond CM and LoG, will be explored to improve classification performance and expand the system coverage to cover different types of skin disorders. Integration with other clinical data will increase the clinical value of the system.

Acknowledgment

The author expresses gratitude for the utilization of the Cloud Computing and Information Systems Laboratory facilities at the Informatics Department, Universitas Hasanuddin, Gowa, Indonesia.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

The dataset used in this study includes public datasets from the following sources:

1. **Roboflow Universe:**
   - Skin Problem Detection Relabel Clean (https://universe.roboflow.com/parin-kittipongdaja-vwmn3/skin-problem-detection-relabel-clean)
   - Acne Detection (https://universe.roboflow.com/kritsakorn/acne-kbm0q)
   - Hyperpigmentation and Wrinkles (https://universe.roboflow.com/upeksha/hyperpigmentation-and-wrinkles-zun1j)
2. **DermNet on Kaggle:**
   - DermNet Dataset (https://www.kaggle.com/datasets/shubhamgoel27/dermnet)

Primary Data: Due to privacy and ethical considerations, detailed datasets containing sensitive information cannot be shared publicly. However, anonymized data used for statistical analysis and interpretation in this study may be made available to researchers upon reasonable request to the authors, provided they meet the criteria for access to confidential data and follow institutional and ethical guidelines.

Author's contribution statement

Rismayani: Conceptualization, Investigation, Data Curation, Writing - original draft, Analysis and editing. **Amil Ahmad Ilham:** Conceptualization, Supervision, Investigation, Data Curation, Writing - initial draft, Writing - review and editing. **Andani Achmad:** Data collection, Supervision, Writing - review, Analysis and Interpretation of results. **Muhammad Rifqy Yudhiestra Rachman:** Data Processing, Data Collection, Training, Testing.

References

[1] Yang TT, Lan CC. Impacts of skin disorders associated with facial discoloration on quality of life: novel insights explaining discordance between life quality scores and willingness to pay. Journal of Cosmetic Dermatology. 2022; 21(7):3053-8.

[2] Shekhar N, Eeshaan R, Tripathy DM, Siddharth M, Bhavni O. Pigmented purpuric dermatoses: a review. Pigment International. 2024; 11(1):1-11.

[3] Nandraj JO, Gajanan KA, Karande KM. A concise review on contemporary and novel treatments addressing the prevention and control of hyperpigmentation. Asian Journal of Pharmaceutical Research and Development. 2024; 12(2):19-27.

[4] Darwish E, Abdelgawad W, Makhlouf M, Abdalla H, Nassar YM, El-tohamy MH, et al. A mobile-based deep learning system for skin disease diagnosis. In intelligent methods, systems, and applications 2024 (pp. 39-44). IEEE.

[5] Raman R, Kumar V, Saini D, Rabadiya D, Patre S, Meenakshi R. Machine learning-driven approaches for dermatological disease diagnosis. In international conference on data science and network security 2024 (pp. 1-5). IEEE.

[6] Malik SG, Jamil SS, Aziz A, Ullah S, Ullah I, Abohashrh M. High-precision skin disease diagnosis through deep learning on dermoscopic images. Bioengineering. 2024; 11(9):1-24.

[7] Pugazhenthi V, Naik SK, Joshi AD, Manerkar SS, Nagvekar VU, Naik KP, et al. Skin disease detection and classification. International Journal of Advanced Engineering Research and Science. 2019; 6(5):396-400.

[8] Wu X, Feng Y, Xu H, Lin Z, Chen T, Li S, et al. CTransCNN: combining transformer and CNN in multilabel medical image classification. Knowledge-Based Systems. 2023; 281:1-15.

[9] Azad R, Kazerouni A, Heidari M, Aghdam EK, Molaei A, Jia Y, et al. Advances in medical image analysis with vision transformers: a comprehensive review. Medical Image Analysis. 2024; 91:103000.

[10] Ilham AA, Achmad A, Rachman MR. Facial skin disorder prediction based on non-visual information using ANN model. In 7th international conference on informatics and computational sciences 2024 (pp. 450-5). IEEE.

[11] Wang Y, Li D, Li L, Sun R, Wang S. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model. Heliyon. 2024; 10(15):1-11.

[12] Wu ZH, Zhao S, Peng Y, He X, Zhao X, Huang K, et al. Studies on different CNN algorithms for face skin disease classification based on clinical images. IEEE Access. 2019; 7:66505-11.

[13] Pawshe V, Bhagwat A, Yelai S, Irkar S, Kornule P. Face skin disease classification based on images. Journal of Emerging Technologies and Innovative Research. 2020; 7(6): 384-7.

[14] Srinivasu PN, Sivasai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors. 2021; 21(8):1-27.

[15] Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Applied Sciences. 2021; 11(11):1-23.

[16] Dai Y. Building CNN-based models for image aesthetic score prediction using an ensemble. Journal of Imaging. 2023; 9(2):1-14.

[17] Li Z, Koban KC, Schenck TL, Giunta RE, Li Q, Sun Y. Artificial intelligence in dermatology image analysis: current developments and future trends. Journal of Clinical Medicine. 2022; 11(22):1-14.

[18] Sazzadul IPM, Mahjabin FS, Bulbul AM, Zihadur RM, Kabir HAB, Shamim KM. Deep learning-based skin disease detection using convolutional neural networks (CNN). In the fourth industrial revolution and beyond: select proceedings of IC4IR+ 2023 (pp. 551-64). Singapore: Springer Nature Singapore.

[19] Dhar P, Guha S. Skin lesion detection using fuzzy approach and classification with CNN. International Journal of Engineering and Manufacturing. 2021; 11(1):11-8.

[20] Thwe PM, Yu MT. Analysis on skin colour model using adaptive threshold values for hand segmentation. International Journal of Image, Graphics and Signal Processing 2019; 11(9):25-33.

[21] Chickaramanna SG, Thippeswamy VS. Identification and classification of prakriti of human using facial features. IAES International Journal of Artificial Intelligence. 2024; 13(2):2093-101.

[22] Dahdouh Y, Boudhiranouar A, Ahmed M. Embedded artificial intelligence system using deep learning and raspberrypi for the detection and classification of melanoma. IAES International Journal of Artificial Intelligence. 2024; 13(1): 1104-11.

[23] Gadag S, Palraj P. Hybrid channel and spatial attention-UNet for skin lesion segmentation. IAES International Journal of Artificial Intelligence (IJ-AI). 2024; 13(1):1077-89.

[24] Khomsi Z, El FM, Bellarbi L. CNN-based approach for non-invasive estimation of breast tumor size and location using thermographic images. International Journal of Online & Biomedical Engineering. 2024; 20(4):160-75.

[25] Makhir A, El YMH, Alaoui LB. Comprehensive cardiac ischemia classification using hybrid CNN-based models. International Journal of Online & Biomedical Engineering. 2024; 20(3):154-65.

[26] Maquen-niño GL, Nuñez-fernandez JG, Taquila-falderon FY, Adrianzén-olano I, De-la-cruz-VdV P, Carrión-barco G. Classification model using transfer learning for the detection of pneumonia in chest X-ray images. International Journal of Online & Biomedical Engineering. 2024; 20(5):150-61.

[27] Benradi H, Bouganssa I, Chater A, Lasfar A. Discriminative approach lung diseases and COVID-19 from chest X-Ray images using convolutional neural networks: a promising approach for accurate

diagnosis. International Journal of Online & Biomedical Engineering. 2023; 19(14):131-41.
[28] Allugunti VR. A machine learning model for skin disease classification using convolution neural network. International Journal of Computing, Programming and Database Management. 2022; 3(1):141-7.
[29] Bonechi S, Bianchini M, Bongini P, Ciano G, Giacomini G, Rosai R, et al. Fusion of visual and anamnestic data for the classification of skin lesions with deep learning. In new trends in image analysis and processing–ICIAP Trento, Italy, 2019 (pp. 211-9). Springer International Publishing.
[30] Pacheco AG, Krohling RA. The impact of patient clinical information on automated skin cancer detection. Computers in Biology and Medicine. 2020; 116:103545.
[31] Chin YP, Hou ZY, Lee MY, Chu HM, Wang HH, Lin YT, et al. A patient-oriented, general-practitioner-level, deep-learning-based cutaneous pigmented lesion risk classifier on a smartphone. British Journal of Dermatology. 2020; 182(6):1498-500.
[32] Sriwong K, Bunrit S, Kerdprasop K, Kerdprasop N. Dermatological classification using deep learning of skin image and patient background knowledge. International Journal of Machine Learning and Computing. 2019; 9(6):862-7.
[33] Nath S, Das GS, Saha S. Deep learning-based common skin disease image classification. Journal of Intelligent & Fuzzy Systems. 2023; 44(5):7483-99.

Rismayani is a doctoral student in Electrical Engineering with a concentration in Informatics at Universitas Hasanuddin, Faculty of Engineering, Indonesia. She is also currently an Associate Professor at Dipa Makassar University, Department of Information Systems, Indonesia. Her fields of interest include Artificial Intelligence, Deep Learning, and Machine Learning.
Email: rismayani22d@student.unhas.ac.id

Amil Ahmad Ilham is Professor at the Department of Informatics, Faculty of Engineering, Universitas Hasanuddin, Indonesia, and has been a lecturer since 1998. He received his bachelor's degree in Electrical Engineering from Universitas Hasanuddin, Indonesia, in 1997 and his Master of Information Technology degree from The University of Newcastle, Australia, in 2003. Then, he received his doctoral degree from Kyushu University, Japan, in 2011. He is currently the Vice Dean for Academic and Student Affairs at the Faculty of Engineering, Universitas Hasanuddin. His research interests include various fields in Information Systems, Big Data, Cloud Computing, and Computer Systems. Additionally, he also has interests in image processing, clustering and similarity analysis, sentiment analysis, machine learning, deep learning, and artificial intelligence.
Email: amil@unhas.ac.id

Andani Achmad is a Professor at the Department of Electrical Engineering, Faculty of Engineering, Universitas Hasanuddin, Indonesia, and has been a lecturer since 1987. He received his bachelor's degree in 1986, master's degree in 2000, and doctoral degree in 2010 from Universitas Hasanuddin in Indonesia. He is currently the Head of the Electrical Engineering Doctoral Study Program at the Faculty of Engineering, Universitas Hasanuddin. His research interests are primarily in the field of Information and Communication Technology. Additionally, he also has interests in Artificial Intelligence, Image Processing, Machine Learning, Deep Learning, and the Internet of Things.
Email: andani@unhas.ac.id

Muhammad Rifqy Yudhiestra Rachman is a Bachelor student at the Universitas Hasanuddin in department of informatics. He fields of interest are Artificial Intelligence, Deep Learning, Mobile Android, Cloud Computing and Machine Learning.

Email: rachmanmry20d@student.unhas.ac.id

Appendix I

S. No.	Abbreviation	Description
1	AI	Artificial Intelligence
2	CNN	Convolutional Neural Network
3	CPU	Central Processing Unit
4	CM	Color Moment
5	DenseNet201	Dense Convolutional Network with 201 Layers
6	EfficientNetB0	Efficient Neural Network B0
7	FP	False Positive
8	FN	False Negative
9	GPU	Graphics Processing Unit
10	HRNet	High-Resolution Network
11	LoG	Laplacian of Gaussian
12	ReLU	Rectified Linear Unit
13	ResNet50	Residual Network 50 layers
14	TP	True Positive
15	TN	True Negative