

Classification of Skin Diseases Using Transfer Learning with ResNet-50 Architecture and Data Preprocessing Using Real-ESRGAN and Wiener Filter

Micheila Jiemesha¹ and Daniel M. Wonohadidjojo¹

¹School of Information Technology, Universitas Ciputra Surabaya, Surabaya, Indonesia

Corresponding author: Daniel M. Wonohadidjojo (e-mail: daniel.m.w@ciputra.ac.id).

ABSTRACT The skin is a vital organ which serves as a barrier against external factors, yet it's highly susceptible to diseases. These diseases are often presented as lesions with similar appearances, making it difficult to be diagnosed and prone to human errors. To address this challenge, this study uses Deep Learning, particularly the ResNet-50 architecture using Transfer Learning, to classify skin diseases from lesion. In this study, data augmentation is implemented to increase dataset size, thus improving model performance and preventing overfitting. Data is then preprocessed using Real-ESRGAN to enhance resolution and the Wiener Filter to sharpen the features. Adam optimizer is used to further enhance the model's performance. Hyperparameter tuning is also implemented to optimize the model parameters, and dropout regularization is applied to enhance the model's ability to be able to accurately classify unseen data. The model managed to achieve a high accuracy of 99.09%, with 0.96 precision, 0.95 recall, and 0.95 F1-score. These results demonstrate the effectiveness of combining Real-ESRGAN and Wiener Filter with the ResNet-50 architecture and the Adam optimizer in developing a robust model for skin disease classification. This approach offers a promising tool for healthcare professionals which may help reduce human error in dermatological diagnosis.

KEYWORDS Real-ESRGAN, Skin lesion, Transfer Learning, Wiener Filter

I. INTRODUCTION

The skin is a sensitive part of the human body and is prone to diseases [1]. It plays a crucial role in bodily functions as a barrier against external agents such as wind, water, and sunlight [2]. Skin diseases rank fourth among illnesses that generally cause various human diseases which affects one-third of the global population [3]. Additionally, skin diseases can impact social life and mental health [4]. Therefore, skin health is essential.

One type of skin disease involves skin lesions, which often appear similar due to their nearly identical shapes. Based on data collected through interviews, physical examination is the most challenging diagnostic stage for most people. This can lead to misdiagnosis when performed visually by dermatology experts [5]. Moreover, further examination, such as a biopsy, is sometimes required, which incurs costs and requires technology not always available to all doctors.

Various methods have been used by researchers, each with its advantages and disadvantages. One frequently used method for classification is Convolutional Neural Network

(CNN), which has achieved significant success in image processing [6]. In a study [7], ResNet-50, a CNN method, was used along with the YOLOv5 detection. This model has shown higher classification accuracy due to deeper and more efficient feature extraction at higher levels [8]. The study demonstrated that the ResNet-50 model achieved excellent performance with 99.5% accuracy. Therefore, ResNet-50 will be used for classification in this research due to its proven accuracy.

However, several factors can affect model performance, such as data noise and imbalance, which can impact model accuracy or cause overfitting. Thus, data optimization and preprocessing are necessary to achieve more accurate and efficient results [9]. This includes hyperparameter tuning, which can enhance model accuracy, as these parameters significantly affect classification accuracy [10]. In this context, hyperparameter tuning involves modifying parameters such as the dropout value in dropout layers.

Initial data processing is conducted to improve the performance of the CNN model. For image data,

augmentation and digital image processing are used to enhance image quality. One common processing method is increasing image resolution using super-resolution techniques. Real-ESRGAN is a super-resolution technique that improves image resolution, enhancing model performance, as shown in a study [11], where Real-ESRGAN implementation resulted in a 5.5% mAP increase. Additionally, noise reduction in images can be performed using digital image processing technique such as the Wiener Filter, which has been proven to enhance classification model accuracy in a study [12].

II. THEORETICAL FRAMEWORK

A. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Network (CNN) is the most used Deep Learning method for classifying objects in images. CNN excels at extracting features from images, allowing for accurate classification. It consists of several sequential layers, as shown in Figure 1 [13]. The input layer receives the image data to be classified. Then, the feature extraction part extracts feature from these images, which are passed to the feature classification section to classify the images based on the obtained features, and finally, the classification results are output.

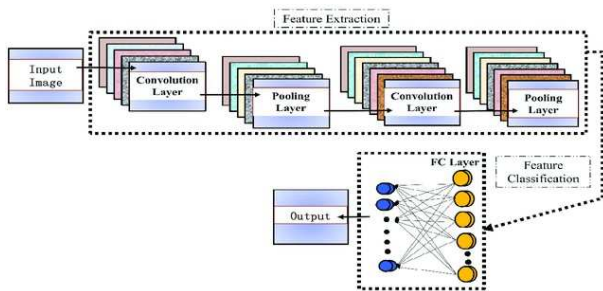


Figure 1. CNN Structure.

CNN has four types of layers: the convolutional layer, which primarily extracts features from the data; the pooling layer, which reduces the dimensionality of the feature map; the fully connected layer, which labels the input data and flattens the output; and the nonlinearity layer, which is an activation function that determines the output of the neural network, such as yes or no [14].

B. TRANSFER LEARNING

One method to improve the performance of a CNN model is to apply Transfer Learning, which involves using a model that has previously learned about one set of objects to apply that knowledge to learn about different objects [15]. Transfer Learning makes the data training process more efficient and enhances model development, leading to greater accuracy [16].

The Transfer Learning process begins with loading the data used to train the model for classification, typically followed by data augmentation to reduce overfitting. The

next step involves taking a pre-trained CNN model and training it on the desired data, but first, replacing some of the final layers, usually the fully connected layer, softmax layer, and classification layer. During this layer replacement process, dropout regularization can be applied by adding a dropout layer between the fully connected and softmax layers to reduce overfitting during model training. The model is then ready to be trained on the new data. After training, the model's performance can be tested and evaluated by examining its accuracy, precision, recall, and F1-score.

C. RESNET-50

ResNet-50 is a CNN architecture that employs the concept of residual learning with skip connections. These skip connections help address the common issue of vanishing gradients in large architectures, where the gradients in a model diminish exponentially, hindering the learning process [17]. Skip connections allow information learned by previous layers to be directly passed to subsequent layers. This architecture was first introduced in [18]. As in Figure 2, ResNet-50 consists of main layers in the form of residual blocks, each containing two convolutional layers with a skip connection at the end of the block and uses global average pooling and a fully connected layer at the end of the architecture [19].

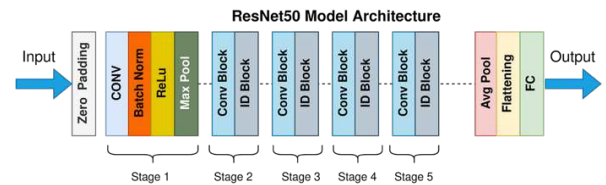


Figure 2. ResNet-50 Model Architecture.

D. REAL-ESRGAN

Real-ESRGAN is an image or video restoration model based on deep learning and uses a Generative Adversarial Network (GAN) to produce high-resolution images or videos from low-resolution ones [20]. Real-ESRGAN generates sharper, more detailed images and videos that are free from artifacts compared to other restoration algorithms [21]. Real-ESRGAN can be used for various applications, such as enhancing the resolution of old photos, restoring blurry videos, and improving the quality of images taken with mobile phone cameras.

E. WIENER FILTER

Wiener Filter, also known as the Minimum Mean Squared Error Filter, is a method in digital image processing designed to reduce noise in an image by combining the degradation function with the statistical characteristics of the noise [22]. The Wiener Filter works by minimizing the Mean Squared Error in the inverse filtering process to reduce noise. It is applied in the frequency domain using the following calculation formula [23].

F. ADAM OPTIMIZER

Adam is a stochastic optimization method that requires only first-order gradients with minimal memory usage, calculating adaptive learning rates for each parameter based on the first and second moment estimates of these gradients [24]. The Adam optimizer is a superior method that can enhance the training performance of a model more effectively than other optimizers like Stochastic Gradient Descent (SGD) and Stochastic Gradient Descent with Momentum (SGDM), especially when hyperparameter tuning is applied to the model [25].

G. DROPOUT REGULARIZATION

Dropout regularization is a technique commonly used to reduce overfitting in Deep Learning models. Overfitting occurs when a Deep Learning model learns the training data too closely, causing it to perform poorly on unseen data. Dropout regularization addresses this issue by randomly removing neurons during training, forcing the model to become more robust to new data.

Dropout works by randomly deactivating certain neurons (typically between 20% and 50% of all neurons) in each layer of the neural network during training [26]. These temporarily deactivated neurons are ignored, and their connections are not updated during that training step. This forces the remaining neurons to learn more independent features and prevents them from overly adapting to specific patterns in the training data [27].

III. PREVIOUS RESEARCH

Several previous studies have applied Deep Learning and Computer Vision to classify skin lesions. Some of the studies include the application of Deep Learning methods such as CNN with six types of architectures for classifying seven types of skin cancer, with the highest accuracy obtained using the modified InceptionV3 architecture [28]; the application of the CNN method with 8 Convolutional2D layers for binary skin cancer classification [29]; the Computer Vision method with the YOLOv5 algorithm combined with Deep Learning methods, namely Convolutional Neural Network (CNN) with ResNet-50 architecture, for seven types of skin cancer [7]; the classification of skin diseases with the use of several CNN architectures which are mostly state-of-the-art models such as VGG-16, MobileNet, ResNet-50, and many more [30]; and the use of VGG-16 model on skin disease classification [31]. These methods, of course, each have their own advantages and disadvantages.

In the study applying the Deep Learning method in the form of CNN with six types of architectures for the classification of seven types of skin cancer, it was found that among the six architectures, namely ResNet-50, VGG-16, DenseNet, MobileNet, InceptionV3, and Xception, the highest performing architecture was InceptionV3 [28]. In the application of the InceptionV3 architecture, several changes were made to the architecture, including a dense layer with the 'ReLU' activation function, a dropout layer, and a

softmax layer at the bottom layer, improvements to the parameter values, and modifications to the HAM10000 dataset used in this experiment for classifying skin lesions into cancer. In the InceptionV3 model, modifications were made to 9617 sample images for 30 epochs with a learning rate of 0.0001 and using the Adam optimizer. The results obtained from this experiment were an accuracy of 90%, with precision, recall, and F1-score values of 91%, 90%, and 90%, indicating the success of this experiment. However, in this experiment, no trials were conducted for initial processing using image processing techniques, and there was still an imbalance in the data set that likely affected the model's performance.

In the study applying the CNN method with 8 Convolutional2D layers for binary skin cancer classification by [29], the dataset was obtained from the ISIC 2018 challenge, which contains seven classes of skin lesions that were then reduced to five classes and divided into two categories: 841 non-melanocytic malignant data and 1356 benign data. Before conducting data training, preprocessing was carried out on the data by resizing the images to 100 x 75 and augmenting the data by rotating, enlarging, moving, and flipping the images.

The CNN model in [29] was trained for 100 epochs with a batch size of 48, with 80% training data, 20% test data, and 20% validation data from the training data. The model achieved an accuracy of 75%, with precision, recall, and F1-score values of 0.80, 0.82, and 0.81 for the benign class, and for the non-melanocytic malignant class were 0.64, 0.61, and 0.63. The results indicate that the model's performance was suboptimal. This could be due to the grouping of five skin cancer classes into two classes, resulting in a variety of lesions within a single class, making accurate classification challenging. The imbalanced precision, recall, and F1-score values between the two classes also indicate an imbalance in the data, with the benign class having more data, leading the model to classify this class more accurately.

In the classification of seven types of skin cancer using the Computer Vision method with the YOLOv5 algorithm combined with the CNN method with ResNet-50 architecture, the dataset used was a secondary dataset obtained from HAM10000, a publicly available dermoscopic data collection of seven categories of skin lesions [7]. During the process, a series of preprocessing steps were conducted on the image data. Firstly, data augmentation was performed to balance the dataset, reduce the risk of overfitting, and enrich the variety of information learned by the model. Then, noise reduction was performed to address disturbances in the images that could affect object segmentation. The final step was normalization to ensure consistency in the scale and distribution of pixel intensities, which can accelerate the object detection process and increase the model's stability against variations in lighting conditions and image quality. The experiment successfully achieved satisfactory results with precision of 99.0%, recall of 98.6%, DSC of 98.8%, accuracy of 99.5%, specificity of 99.8%, mAP from 0.0 to 0.5 of 98.3%, and mAP from 0.5 to 0.95 of 98.7%. This study

shows that ResNet-50 has excellent performance in classification.

A study where 4 CNN architectures were compared using the Fruit 360 dataset also has proven that ResNet-50 has a better performance than other CNN models, which in this case it's being compared to VGG-16, MobileNet, and SC-3 [32]. The study shows how that ResNet-50 has the highest performance compared to the other 3 models, with a second to lowest training time. This shows that although the ResNet-50 gives the highest performance, it's also considered faster than other models.

Another study has compared multiple models towards skin disease classifications. In the study, state-of-the-art models such as VGG-16, VGG-19, MobileNet, ResNet-50, InceptionV3, Inception-ResNetV2, Xception, DenseNet-121, DenseNet-169, DenseNet-201 and NASNet Mobile are compared using the same case which is the skin disease classification [30]. As a result, the MobileNet managed to obtain the highest accuracy of 94.1%. Here, we can see an opportunity to improve other model's performance by implementing preprocessing methods and adjusting the model which may result in a higher accuracy than the state-of-the-art model's accuracy presented in the study.

Although high accuracy may indicate a good performance, it cannot guarantee that the model will perform well towards real-life scenarios. Further validation is needed to ensure the model's robustness and that there is no overfitting. The study of using VGG-16 on skin disease classification shows that validation is very important since overfitting may occur even after using a state-of-the-art model such as VGG-16 [31]. The findings in the study indicate that the model has a high accuracy in skin disease classification, achieving 98% accuracy. However, the model shows signs of overfitting, as reflected by the validation accuracy of 88%. To handle this overfitting problem, the model may need regularization, or the dataset may need to be preprocessed.

In a study [11] that uses Real-ESRGAN to preprocess the data, it is shown that data preprocessing using Real-ESRGAN has proven to be able to increase the model's performance, which in this case is YOLOv5. Before any preprocessing was conducted, the model managed to achieve the highest mAP of 89.2%, whereas after preprocessing the data using Real-ESRGAN, the mAP increased to 94.7%. This shows an increase of 5.5% in the mAP which is a major improvement.

Wiener Filter also has been proven to improve a CNN model's performance compared to other image preprocessing methods such as CLAHE and unsharp masking. This can be seen in a study where a MobileNetV2 model is trained using data preprocessed with Wiener Filter which obtains a higher overall performance than using other preprocessing methods [12].

Other than data preprocessing, some adjustments can also be made to the model's layers such as dropout regularization. This method was used in research where the model experienced overfitting [33]. After dropout was used, the

model no longer had overfitting and has a decent performance. This proves how dropout regularization can prevent overfitting, yet more adjustments must be made to maximize the model's performance.

To further optimize the model's performance, a model optimizer was used, specifically Adam optimizer which is a state-of-the-art optimizer. Compared to other state-of-the-art optimizers, Adam can give a higher performance which can be seen in research where 10 state-of-the-art optimizers are compared and Adam optimizer managed to achieve the highest accuracy of 99% for a brain tumor segmentation [34].

Therefore, in this study the ResNet-50, a variant of the Residual Network (ResNet) with 50 layers that has proven to have high performance based on previous research, is proposed. However, ResNet-50 has shortcomings, including its inability to handle data imbalance, susceptibility to overfitting, and trouble in detecting small objects [35]. Therefore, dataset balancing is applied, followed with data augmentation to handle the imbalanced data. Preprocessing using Real-ESRGAN and Wiener Filter is implemented to enhance the features of skin diseases to make it easier for the model to detect. Not only that, but dropout regularization is also applied to avoid overfitting during data training where the dropout value is tuned to achieve maximum performance since hyperparameter tuning can significantly affect classification accuracy [10]. Lastly, Adam optimizer is also used to further increase the model's performance. Overall, the adjustments on the ResNet-50 model are implemented to increase the model's performance without any overfitting to ensure the model's ability to handle real-data.

IV. RESEARCH METHOD

A. DATA COLLECTION

In this study, five secondary datasets sourced from Kaggle and Roboflow websites were used. They provide various facial skin diseases and examples of healthy skin. Dataset A is a collection of facial segmentation images [36], and Dataset B is a collection of images of 8 types of skin diseases caused by bacterial, viral, fungal, and parasitic infections [37]. Additionally, Dataset C contains 23 types of skin diseases from Dermnet [38], Dataset D is a collection of images of 4 skin types [39], and Dataset E is a collection of images of 5 types of facial skin diseases [40].

Dataset A consists entirely of secondary data obtained from the Kaggle website [36], containing segmented facial images. The facial images data from Dataset A along with images from Dataset D, sourced from the Roboflow website [39], which contains a variety of facial skin types, is sorted and cropped to only take the healthy skin areas of the face to produce the healthy skin class. Dataset B, sourced from Kaggle, contains various skin diseases caused by infections [37]. From this, the tinea ringworm class was selected. Additionally, from Dataset C, sourced from Kaggle by [38], containing skin disease images from Dermnet, two classes, psoriasis and eczema, were obtained. There is also Dataset E from the Roboflow website [40], containing 5 types of skin diseases, of which only two types, acne and rosacea, were

used. Ultimately, the processed data from Dataset A, Dataset B, Dataset C, Dataset D, and Dataset E were selected, resulting in 6 classes: 5 classes of skin disease lesions which are tinea ringworm, psoriasis, eczema, acne, and rosacea, and 1 class of healthy skin without disease, each containing 100 images. The information and classes of each dataset are presented in Table 1.

TABLE I
DATASET INFORMATIONS

Dataset	Description	Classes
A	A collection of facial segmentation images	Healthy Skin
B	Various skin diseases caused by infections	Tinea Ringworm
C	Skin disease images from Dermnet	Psoriasis and Eczema
D	4 types of facial skin types	Healthy Skin
E	5 types of skin diseases	Acne and Rosacea

B. DATA PREPROCESSING

Before using the data to train the CNN model, preprocessing was conducted to improve data quality. First, the data is cleaned by selecting 100 clean images from each class which results in 600 images since there are 6 classes. After that, the data is cropped to only show the region of interest. Then, data augmentation was performed by randomly rotating, flipping, and shearing the images. This aims to improve the classification accuracy and reduce overfitting, allowing the model to accurately classify not only the training data but also unseen data. From this data augmentation, 400 images per class were obtained, resulting in 2400 images. Next, digital image processing was applied to the images using the Real-ESRGAN method to enhance the data quality by increasing the image resolution and quality using the library provided in a GitHub Repository [41]. After that, Wiener Filter is applied to reduce noise in the images and sharpen existing features. By implementing these two methods, the noises in the image data have been reduced and the quality has improved thereby increasing the CNN model's accuracy. The final step in dataset processing was dividing the dataset into 70% training data, 15% validation data, and 15% testing data. The preprocessed data is then passed onto the model to be trained. The system architecture diagram of the process can be seen in Figure 3.

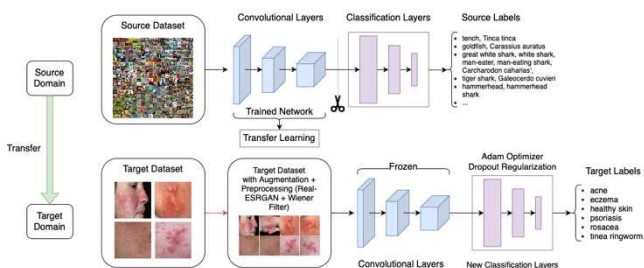


Figure 3. System Architecture Diagram.

C. MODEL TRAINING

The model training in this study utilized the Convolutional Neural Network (CNN) architecture, specifically ResNet-50, which has proven to be effective. Using a well-established architecture can accelerate model development and ensure good performance. This architecture also has lower model complexity compared to the InceptionV3 model used by [28]. To meet the specific needs of the study, such as reducing overfitting and achieving high accuracy, some of the model's final layers were modified, including the fully connected, softmax, and classification layers. These modifications aim to tailor the model to the dataset and classification tasks at hand.

Additionally, dropout regularization was applied by adding a dropout layer, with values determined through hyperparameter tuning, to reduce the risk of overfitting [42]. This technique effectively helps prevent the model from relying too heavily on certain subsets of the training data, thus enhancing the model's ability to generalize to new data.

The model was trained with 100 epochs to ensure the stability of the model's accuracy and loss, 32 batch size that's the most optimal batch size after some experiments, Adam optimizer, and a learning rate of 0.001 as it is the default learning rate of Adam optimizer. This configuration was chosen after a series of experiments to ensure the model achieves good convergence without significant overfitting or underfitting.

D. MODEL EVALUATION

After the training is complete, the results are evaluated using a confusion matrix, which is then used to obtain several evaluation metrics, including accuracy, precision, recall, and F1-score, to comprehensively and objectively assess the model's performance. This is done by looking at the numbers of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), by using the formula for each metric. This analysis helps identify areas that may need further improvement and establish next steps for better model development. The higher the values of accuracy, precision, recall, and F1-score, the better the model's performance in classification.

V. RESULT AND DISCUSSION

A. RESULT

After conducting an experiment using the ResNet-50 architecture model on data consisting of 5 classes of facial skin diseases and 1 class of healthy skin, high accuracy was achieved with stable model performance without overfitting due to dropout regularization. Additionally, the application of preprocessing data with Real-ESRGAN and Wiener Filter followed by hyperparameter tuning and the use of Adam optimizer significantly improved the classification accuracy.

Originally, without any data preprocessing or model adjustments and testing using 30 epochs, the model has only managed to achieve an accuracy score of 0.8937 with

overfitting which can be seen from the accuracy graph in Figure 4 and loss graph in Figure 5.

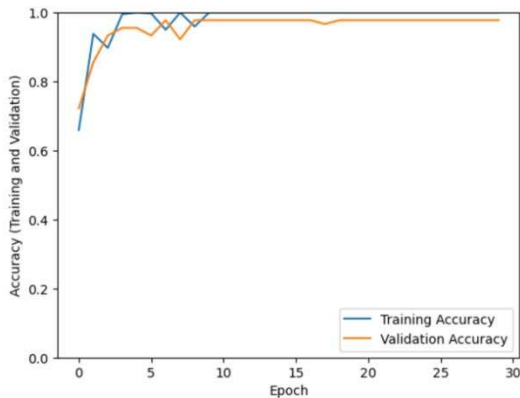


Figure 4. ResNet-50 Model Training and Validation Accuracy Graph Without Any Adjustments or Preprocessing.

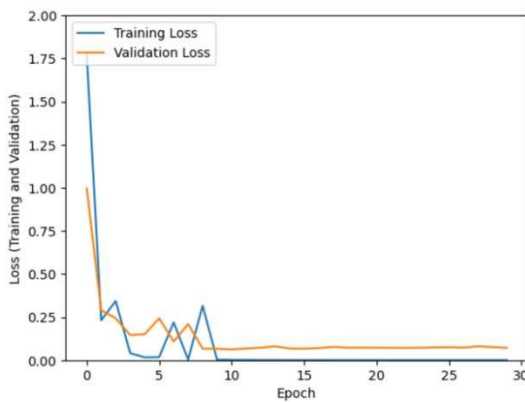


Figure 5. ResNet-50 Model Training and Validation Loss Graph Without Any Adjustments or Preprocessing.

But after conducting data augmentation, the model managed to achieve a high accuracy of 0.978. But despite having high accuracy, the model still experienced overfitting which indicates that it will not be as accurate towards real data. This can be seen by the constant difference between the training and validation accuracy graph as shown in Figure 6 and the loss graph in Figure 7.

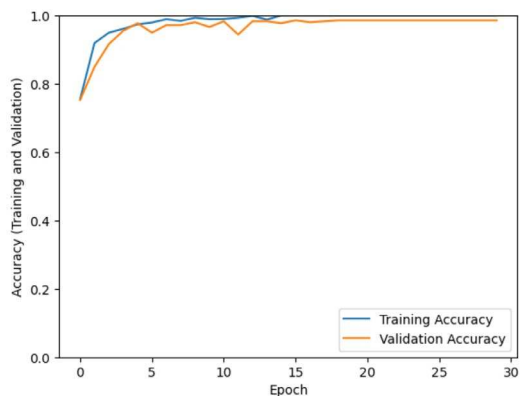


Figure 6. ResNet-50 Model Training and Validation Accuracy Graph with Only Data Augmentation.

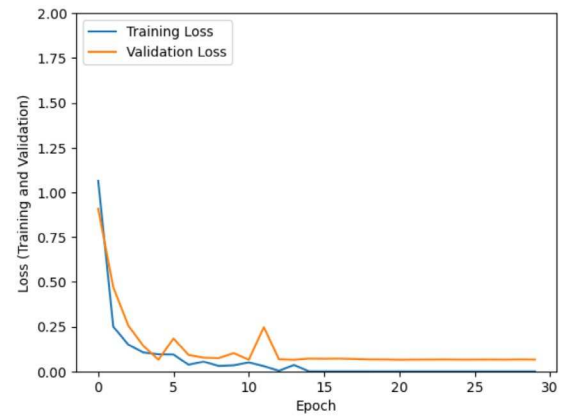


Figure 7. ResNet-50 Model Training and Validation Loss Graph with Only Data Augmentation

Then, to compare the performance of ResNet-50, an ablation study was conducted to towards ResNet-50 model and the other state-of-the-art CNN models, including VGG16, InceptionV3, and MobileNet. The results of each model can be seen in Table 2.

TABLE II
MODEL COMPARISON TOWARDS THE CLASSIFICATION 5 CLASSES OF FACIAL SKIN DISEASE

Model	Accuracy	Precision	Recall	F1-Score
ResNet-50	0.978	0.99	0.99	0.99
MobileNet	0.7975	0.79	0.78	0.78
VGG-16	0.9765	0.97	0.97	0.97
InceptionV3	0.639	0.69	0.65	0.66

From Table 2, ResNet-50 has proven to achieve the highest accuracy with an almost perfect precision, recall, and F1-score. Hence, further experiments are conducted using the model. Table 3 shows the accuracy percentages of the ResNet-50 model, where the accuracy increases as more advancements towards the model are made.

TABLE III
RESNET-50 MODEL'S PERFORMANCE WITH IMPROVEMENTS

Model Improvement	Accuracy	Precision	Recall	F1-Score
Dropout 0.5	0.9634	0.98	0.98	0.98
Dropout 0.5 + Adam	0.9663	0.98	0.98	0.98
Dropout 0.5 + Adam + Real-ESRGAN	0.9811	0.98	0.98	0.98
Dropout 0.5 + Adam + Wiener Filter	0.9703	0.99	0.99	0.99
Dropout 0.5 + Adam + Real-ESRGAN + Wiener Filter	0.9859	0.97	0.97	0.97

After obtaining a high accuracy from the implementation of dropout regularization along with Adam optimizer, Real-ESRGAN, and Wiener Filter, the training epochs was increased from 30 to 100 epochs to make sure that the

model's accuracy and loss is stable. Other than that, hyperparameter tuning was done, specifically in the value of the dropout layer and batch size, to further improve the model's performance while maintaining its robustness since the dropout layer can reduce overfitting and batch size may improve the model's accuracy.

TABLE IV
RESNET-50 MODEL'S PERFORMANCE WITH DROPOUT VALUE TUNING

Dropout Value	Accuracy	Precision	Recall	F1-Score
0.5	0.9745	0.95	0.95	0.95
0.6	0.9809	0.97	0.97	0.97
0.7	0.9823	0.96	0.96	0.96
0.8	0.9909	0.96	0.95	0.95
0.9	0.9781	0.94	0.94	0.94

TABLE V
RESNET-50 MODEL'S PERFORMANCE WITH BATCH SIZE TUNING

Batch Size	Accuracy	Precision	Recall	F1-Score
32	0.9909	0.96	0.95	0.95
64	0.9882	0.97	0.97	0.97
128	0.9724	0.96	0.96	0.96

The results of the accuracy, precision, recall, and F1-score of the ResNet-50 model that has undergone data preprocessing, Adam optimizer, dropout regularization, and hyperparameter tuning according to the tested dropout values can be seen in Table 4 and batch size in Table 5. In that table, the dropout value that provides the best model performance is 0.8 with 32 batch size, with an accuracy of 0.9909. The model's performance can also be seen in the training and validation accuracy graph in Figure 8, loss graph in Figure 9, and confusion matrix in Figure 10.

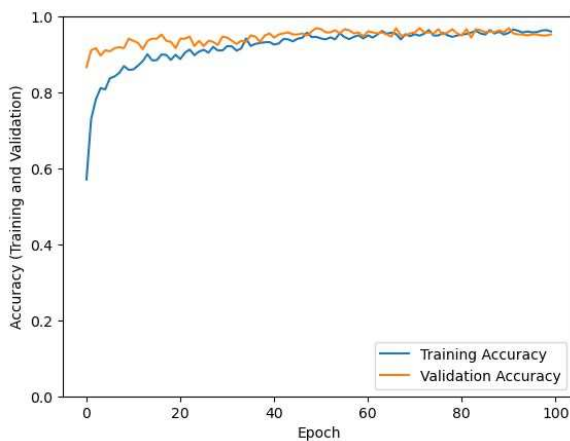


Figure 8. ResNet-50 Model Training and Validation Accuracy Graph with Preprocessing, Optimizer, 0.8 Dropout Value, and 100 Epochs.

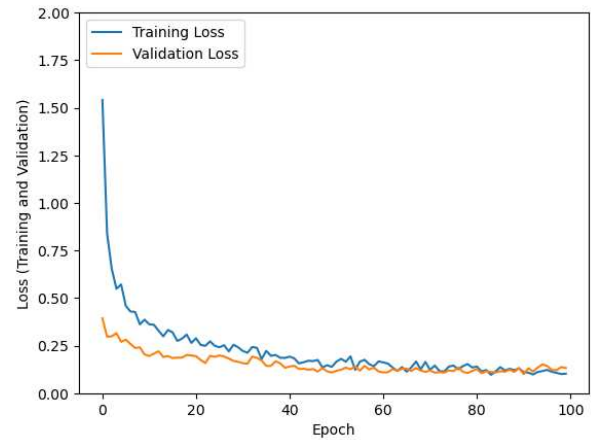


Figure 9. ResNet-50 Model Training and Validation Loss Graph with Preprocessing, Optimizer, 0.8 Dropout Value, and 100 Epochs.

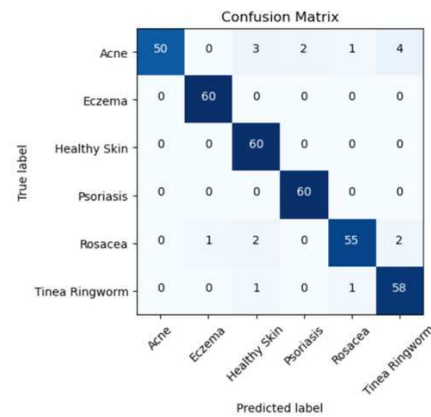


Figure 10. ResNet-50 Model Confusion Matrix with Preprocessing, Optimizer, 0.8 Dropout Value, and 100 Epochs.



Figure 11. ResNet-50 Model Prediction Result on Skin Disease Pictures Sourced from Google.

The model performance is tested using images that are sourced from Google. They consist of different types of skin disease that the model was trained to classify. Figure 11 shows image examples of 5 skin diseases. The Figure shows that out of 5 images, the model managed to correctly classify all 5 of them. Other than testing on those images, a cross validation was also conducted using the K-Fold Cross Validation method with $k=5$, which resulted in a 98.33% accuracy in majority vote of the test. Additionally, the model achieved a high precision of 0.98 which indicates that false positives are minimal, 0.98 recall which reflects the model's ability to capture relevant cases without missing many true positives, and a F1-score of 0.98 which represents a balance between precision and recall. The confusion matrix can be seen in Figure 12.



Figure 12. K-Fold Cross Validation on ResNet-50 Model's Confusion Matrix with Preprocessing, Optimizer, 0.8 Dropout Value, and 100 Epochs with $k=5$.

B. DISCUSSION

The accuracy and loss graphs in Figure 8 and Figure 9, indicate that the ResNet-50 model, with the implementation of dropout regularization, Adam optimizer, Real-ESRGAN, and Wiener Filter, was successful in the data training process, with significantly increasing accuracy and decreasing loss over iterations. This demonstrates the model's effectiveness in learning from the data and reducing prediction errors.

The model developed in this study achieved an accuracy of 99.09%, indicating that most predictions made by the model are correct. This result shows that the model performs better than the previous research's ResNet-50 model without any additional adjustments and preprocessing [30], with a 37.69% higher accuracy.

The use of preprocessing with Real-ESRGAN and Wiener Filter has also contributed in increasing the accuracy from 97.8% to 99.09% and dropout regularization plays a role in reducing the overfitting significantly compared to the study of skin disease classification using VGG-16 where there is an overfitting with a 10% difference in training and validation accuracy [31] while in this experiment the

difference between the test accuracy of the model with the cross validation is only 0.76%.

Additionally, the model reached a precision of 0.96, indicating that most of the positive predictions made by the model are accurate. The model also achieved a recall of 0.95, showing that it successfully identified most of the actual positive conditions in the data. Finally, the model achieved an F1-score of 0.95, which is the harmonic mean of precision and recall, providing a balanced assessment of both metrics with no overfitting in the accuracy and loss graph that can be seen in Figure 8 and Figure 9. Overall, based on these evaluation metrics, the ResNet-50 model demonstrates excellent performance in the given classification task, proving that the application of dropout regularization along with the Adam optimizer, Real-ESRGAN, Wiener Filter, and hyperparameter tuning can enhance model performance. However, the model used has limitations due to the data still containing noise in the Healthy Skin, Eczema, and Rosacea classes. This noise can affect the model's performance during the classification process, where the model's ability to accurately distinguish between classes may be impaired.

From the confusion matrix in Figure 10, we can see where the model tends to make mistakes in classifying 'Acne' and 'Rosacea' as other skin conditions. This indicates that there are some skin conditions that are more difficult for the model to predict accurately. In such cases, additional training data, different preprocessing techniques, or model parameter adjustments might be needed for those specific skin conditions. Additionally, the 'Acne' and 'Rosacea' classes may contain noise that has not yet been filtered appropriately, which could be impacting the model's ability to classify it accurately. Addressing this noise through data cleaning or enhanced preprocessing could further improve the model's performance for this class. To gain a deeper understanding of the problem, additional experimentation and analysis are required to get the best result.

Based on the presented evaluation metrics, the ResNet-50 model with preprocessing, optimizer, and regularization demonstrates very good performance. This has also been proven in Figure 11 where the model managed to correctly classify all skin disease images sourced from Google which shows that the model is capable classifying real images accurately. This is also supported with the K-Fold Cross Validation which shows a high accuracy of 98.33% in the test accuracy of the majority vote, not so far from the model's test accuracy of 99.09%, and 0.98 in precision, recall, and F1-score with the confusion matrix can be seen in Figure 12.

However, there is always room for improvement in the model. The area of improvement includes precision and recall, model generalization, specific classification error, and the model's F1-score. Although the model's precision and recall are already quite high, there might still be some false positives or false negatives. Increasing precision means reducing the number of false positives, cases where the model predicts positive but is negative, while increasing recall means reducing the number of false negatives, cases where the model predicts negative but is positive. While the

model has high accuracy on training data, it is important to ensure that it also performs well on data it has never seen before. This can be checked by conducting cross-validation or using separate validation and test datasets. This may be applied for further research.

Although the F1-score is already quite high, improving this metric could indicate a better balance between precision and recall. Improving one metric might lead to a decrease in another, such as between recall and precision. Therefore, it is important to consider these trade-offs when optimizing the model.

Overall, the ResNet-50 model along with preprocessing using Real-ESRGAN and Wiener Filter, Adam optimizer, and dropout regularization with hyperparameter tuning produces superior results compared to previous studies, with an accuracy of 99.09%, 0.96 precision, 0.95 recall, F1-score of 0.95, and cross validation result of 98.33% which is improvement over previous research. These findings open opportunities for further research on preprocessing, data handling, and model parameter adjustments by future researchers.

VI. CONCLUSION

This study discusses the application of Convolutional Neural Networks (CNN) using transfer learning. The transfer learning process involves using a pre-trained CNN model. In this study, the ResNet-50 architecture is used along with data preprocessing using Real-ESRGAN and Wiener Filter. Before training, data processing is performed through data augmentation by implementing random rotation, random flipping, and random shearing, increasing the data to reduce overfitting and enhance the model's ability to classify unseen images. The training process is optimized using Adam, and dropout regularization with hyperparameter tuning is applied to further reduce overfitting and increase model performance. The model's performance is evaluated using four metrics: accuracy, precision, recall, and F1-score.

Based on the results discussed, it can be concluded that using CNN with transfer learning produces a model with highly satisfactory classification capabilities. The implementation of Real-ESRGAN and Wiener Filter has proven to significantly improve the model's performance from 96.63% accuracy, only after using dropout regularization and Adam optimizer, to 99.09% accuracy after using Real-ESRGAN and Wiener Filter in data preprocessing which is further improved. Overall, the CNN method with the ResNet-50 architecture along with data preprocessing using Real-ESRGAN and Wiener Filter, Adam optimizer, dropout regularization and hyperparameter tuning, has successfully achieved high performance in classifying facial skin diseases with an accuracy of 99.09%. With 0.96 precision, 0.95 recall, and 0.95 F1-score, the use of ResNet-50 can be concluded to provide excellent results.

For future research, it is recommended to experiment with newer CNN architectures, increase training data, and explore different preprocessing techniques. Additionally, parameter

adjustments using an algorithm can be conducted to achieve optimal results. The model's generalization ability can also be tested on different objects or datasets. It is important to consider the trade-off between accuracy and interpretability, as more complex models may improve accuracy but make the model difficult for others to interpret and understand.

AUTHORS CONTRIBUTION

Micheila Jiemesha: Conceptualization, Methodology, Research, Investigation, Formal Analysis, Resources, Software, Visualization, Original Draft Writing.

Daniel Martomanggolo Wonohadidjojo: Supervision, Validation, Original Draft Writing Preparation, Review Writing & Editing.

COPYRIGHT



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

REFERENCES

- [1] R. Lestari, "Hubungan Sanitasi Lingkungan Dengan Gejala Penyakit Kulit Di Wilayah Kerja Puskesmas Sukamenanti Kabupaten Pasaman Barat," *nthn*, vol. 17, no. 1, pp. 14–23, Apr. 2022, doi: 10.59963/nthn.v17i1.98.
- [2] K. Merin, M. Shaji, and R. Kameswaran, "A review on sun exposure and skin diseases," *Indian J Dermatol*, vol. 67, no. 5, p. 625, 2022, doi: 10.4103/ij.d.ij.d_1092_20.
- [3] C. Flohr and R. Hay, "Putting the burden of skin diseases on the global map," *Br J Dermatol*, vol. 184, no. 2, pp. 189–190, Feb. 2021, doi: 10.1111/bjd.19704.
- [4] Y. Ozcan, M. A. Sungur, B. Y. Özcan, Y. Eyüp, and E. Özlü, "The Psychosocial Impact of Chronic Facial Dermatoses in Adults," *Dermatol Pract Concept*, p. e2023029, Jan. 2023, doi: 10.5826/dpc.1301a29.
- [5] K. Sreekala, N. Rajkumar, R. Sugumar, K. V. D. Sagar, R. Shobrani, K. P. Krishnamoorthy, A. K. Saini, H. Palivela, A. Yeshitla, and V. Kumar, "Skin Diseases Classification Using Hybrid AI Based Localization Approach," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–7, Aug. 2022, doi: 10.1155/2022/6138490.
- [6] Y. Chen, "Research on Convolutional Neural Network Image Recognition Algorithm Based on Computer Big Data," *J. Phys.: Conf. Ser.*, vol. 1744, no. 2, p. 022096, Feb. 2021, doi: 10.1088/1742-6596/1744/2/022096.
- [7] M. Elshahawy, A. Elnemr, M. Oproescu, A-G. Schiopu, A. Elgarayhi, M. M. Elmogy, and M. Sallah, "Early Melanoma Detection Based on a Hybrid YOLOv5 and ResNet Technique," *Diagnostics*, vol. 13, no. 17, p. 2804, Aug. 2023, doi: 10.3390/diagnostics13172804.
- [8] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," *Applied Sciences*, vol. 12, no. 18, p. 8972, Sep. 2022, doi: 10.3390/app12188972.
- [9] M. Krichen, "Convolutional Neural Networks: A Survey," *Computers*, vol. 12, no. 8, p. 151, Jul. 2023, doi: 10.3390/computers12080151.
- [10] S. Mezzah and A. Tari, "Practical hyperparameters tuning of convolutional neural networks for EEG emotional features classification," *Intelligent Systems with Applications*, vol. 18, p. 200212, May 2023, doi: 10.1016/j.iswa.2023.200212.
- [11] F. Akhyar, L. Novamizanti, and T. Riantiami, "Sistem Inspeksi Cacat pada Permukaan Kayu menggunakan Model Deteksi Obyek YOLOv5," *ELKOMIKA*, vol. 10, no. 4, pp. 990–1005, Oct. 2022, doi: 10.26760/elkomika.v10i4.990.

- [12] E. Avsar, "Effects of Image Preprocessing on the Performance of Convolutional Neural Networks for Pneumonia Detection," in 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Kocaeli, Turkey: IEEE, Aug. 2021, pp. 1–5. doi: 10.1109/INISTA52262.2021.9548351.
- [13] S. Ali, J. Li, Y. Pei, M. S. Aslam, Z. Shaikat, and M. Azeem, "An Effective and Improved CNN-ELM Classifier for Handwritten Digits Recognition and Classification," *Symmetry*, vol. 12, no. 10, p. 1742, Oct. 2020, doi: 10.3390/sym12101742.
- [14] P. Purwono, A. Ma'arif, W. Rahmانيar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, "Understanding of Convolutional Neural Network (CNN): A Review," *IJRCS*, vol. 2, no. 4, pp. 739–748, Jan. 2022, doi: 10.31763/ijrsc.v2i4.888.
- [15] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabdullah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.
- [16] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *J Big Data*, vol. 9, no. 102, pp. 1–19, Oct. 2022, doi: 10.1186/s40537-022-00652-w.
- [17] S. A. Agrawal, V. D. Rewaskar, R. A. Agrawal, S. S. Chaudhari, Y. Patil, and N. S. Agrawal, "Advancements in NSFV Content Detection: A Comprehensive Review of ResNet-50 Based Approaches," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 4, pp. 41–45, Sep. 2023.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 10, 2015, arXiv: arXiv:1512.03385. Accessed: Aug. 05, 2024. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [19] E. Suherman, B. Rahman, D. Hindarto, and H. Santoso, "Implementation of ResNet-50 on End-to-End Object Detection (DETR) on Objects," *Sinkron*, vol. 8, no. 2, pp. 1085–1096, Apr. 2023, doi: 10.33395/sinkron.v8i2.12378.
- [20] Ş. B. Çetin, "Real-ESRGAN: A deep learning approach for general image restoration and its application to aerial images," *Arsej*, vol. 3, no. 2, pp. 90–99, Dec. 2023.
- [21] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," Aug. 17, 2021, arXiv: arXiv:2107.10833. Accessed: Aug. 05, 2024. [Online]. Available: <http://arxiv.org/abs/2107.10833>
- [22] D. S. Nurjanah, D. Suhaedi, and E. Harahap, "Denoising Restorasi Citra Digital Menggunakan Filter Wiener," *Journal Matematika UNISBA*, vol. 15, no. 1, pp. 1–6, May 2016.
- [23] N. Heryana and R. Mayasari, "Implementasi Nose Removal Menggunakan Wiener Filter untuk Perbaikan Citra Digital," *Syntax J. Inf.*, vol. 5, no. 2, pp. 159–164, Dec. 2016, doi: 10.35706/syji.v5i2.709.
- [24] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 29, 2017, arXiv: arXiv:1412.6980. Accessed: Aug. 05, 2024. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [25] N. Gkouti, P. Malakasiotis, S. Toumpis, and I. Androustopoulos, "Should I try multiple optimizers when fine-tuning pre-trained Transformers for NLP tasks? Should I tune their hyperparameters?" Feb. 10, 2024, arXiv: arXiv:2402.06948. Accessed: Aug. 05, 2024. [Online]. Available: <http://arxiv.org/abs/2402.06948>
- [26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jun. 2014.
- [27] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," Jul. 03, 2012, arXiv: arXiv:1207.0580. Accessed: Aug. 05, 2024. [Online]. Available: <http://arxiv.org/abs/1207.0580>
- [28] M. S. Akter, H. Shahriar, S. Sneha, and A. Cuzzocrea, "Multi-class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network," in 2022 IEEE International Conference on Big Data (Big Data), Mar. 2023, pp. 5404–5413. doi: 10.1109/BigData55660.2022.10020302.
- [29] L. Hakim, Z. Sari, and H. Handhajani, "Klasifikasi Citra Pigmen Kanker Kulit Menggunakan Convolutional Neural Network," *RESTI*, vol. 5, no. 2, pp. 379–385, Apr. 2021, doi: 10.29207/resti.v5i2.3001.
- [30] J. S. Velasco, J. V. Catipon, E. G. Monilar, V. M. Amon, G. C. Virrey, and L. K. S. Tolentino, "Classification of Skin Disease Using Transfer Learning in Convolutional Neural Networks," 2023, doi: 10.48550/ARXIV.2304.02852.
- [31] Nurkhasanah and Murinto, "Klasifikasi Penyakit Kulit Wajah Menggunakan Metode Convolutional Neural Network," *Sainteks*, vol. 18, no. 2, p. 183, Feb. 2022, doi: 10.30595/sainteks.v18i2.13188.
- [32] K. Liu, "Comparison of different Convolutional Neural Network models on Fruit 360 Dataset", *HSET*, vol. 34, pp. 85–94, Feb. 2023, doi: 10.54097/hset.v34i.5385.
- [33] B. Jabir and N. Falih, "Dropout, a basic and effective regularization method for a deep learning model: a case study," *IJECS*, vol. 24, no. 2, p. 1009, Nov. 2021, doi: 10.11591/ijeecs.v24.i2.pp1009-1016.
- [34] M. Yaqub et al., "State-of-the-Art CNN Optimizer for Brain Tumor Segmentation in Magnetic Resonance Images," *Brain Sciences*, vol. 10, no. 7, p. 427, Jul. 2020, doi: 10.3390/brainsci10070427.
- [35] A. Stateczny, G. Uday Kiran, G. Bindu, K. Ravi Chythanya, and K. Ayyappa Swamy, "Spiral Search Grasshopper Features Selection with VGG19-ResNet50 for Remote Sensing Object Detection," *Remote Sensing*, vol. 14, no. 21, p. 5398, Oct. 2022, doi: 10.3390/rs14215398.
- [36] K. Roman, "Face Dataset – Segmentation." 2023. [Online]. Available: <https://www.kaggle.com/datasets/tapakah68/face-segmentation>
- [37] S. Biswas19, "Skin-Disease-Dataset." 2023. [Online]. Available: <https://www.kaggle.com/datasets/subirbiswas19/skin-disease-dataset>
- [38] S. Goel and B. Hall, "Dermnet." 2020. [Online]. Available: <https://www.kaggle.com/datasets/shubhamgoel27/dermnet/data>
- [39] theaskin, "Skin types2 Dataset." Jan. 2024. [Online]. Available: <https://universe.roboflow.com/theaskin/skin-types2>
- [40] facial skin dataset, "facial skin diseases Dataset." Apr. 2024. [Online]. Available: <https://universe.roboflow.com/facial-skin-dataset/facial-skin-diseases>
- [41] bycloudai, 2021, "Real-ESRGAN-Windows," Version 0.1.1, [Source code]. <https://github.com/bycloudai/Real-ESRGAN-Windows>
- [42] I. Salehin and D.-K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *Electronics*, vol. 12, no. 14, p. 3106, Jul. 2023, doi: 10.3390/electronics12143106.