PROJECT REPORT

Skin Disease Classifier System

Group 22

Group Members

Liu Mohan (Leader)

Zhang Zhiyuan

Zhong Wenyu

Zhou Yuhao

Zhang Lixin

Submit Date:

Oct 31, 2024

Content

1. Executive Summary

This project develops an AI-powered diagnostic system for skin disease classification. By using advanced deep learning models in image classification area—ResNet50, DenseNet121, and VGG16, the system achieves a diagnostic accuracy range of 55-70% across 23 skin disease categories. And by further applying a hard-voting ensemble method, it combines model strengths, enhancing the precision and robustness of skin disease identification, notably improving classification performance for common conditions like acne, nail fungus, and melanoma.

Additionally, the system is deployed as an interactive web application, allowing healthcare providers and patients to upload images and receive real-time diagnostic feedback in a streamlined manner. The front-end is built using Vue.js for its flexibility and responsiveness, combined with Vite for fast-build development, which enhances the user experience. The interface uses Axios to handle asynchronous HTTP requests, enabling smooth data transmission between the front-end and the Django-based backend. To improve accessibility and foster trust, the system incorporates interpretable results, supporting clinicians in complex cases and providing patients with a preliminary self-assessment tool.

Future work includes broadening the dataset to ensure representation across skin tones, age groups, and rare conditions, as well as integrating multimodal data—such as patient history and clinical test results—for a more comprehensive diagnostic view. Incorporating a real-time feedback mechanism will also allow for continuous model retraining, improving diagnostic accuracy across diverse clinical applications. This AI-based diagnostic system sets a new standard in accessible, accurate dermatology tools, benefiting both medical professionals and patients worldwide.

2. Introduction

2.1 Project Background

The skin covers the entire human body, acting as a protective barrier for internal tissues and organs against external threats like mechanical injuries, chemical exposure, viruses, and immune responses. Additionally, it helps prevent lipid and moisture loss from the epidermis and dermis, which is crucial for maintaining the skin's barrier integrity[1] .

However, skin diseases are among the most common health issues and can occur at any age, affecting millions of people worldwide. According to a 2018 report by the British Skin Foundation, around 60% of the population in the UK experiences some form of skin condition[2] . In the United States, approximately 5.4 million new cases of skin cancer are recorded annually, with 1 in 5 Americans being diagnosed with malignant skin tumors at some point in their lifetime[3] . Skin diseases not only have a significant impact on physical health, but they also greatly affect patients' emotional well-being, social interactions, and mental health. For young people and young adults, skin conditions can diminish personal satisfaction and lead to issues such as pain, scarring, psychological distress, including feelings of depression and, in some cases, suicidal thoughts[4] .

Accurate and early diagnosis of skin diseases is essential for effective treatment, yet it remains challenging. Inexperienced doctors may rely on visual examinations, increasing the risk of misdiagnosis, especially when different diseases present similar symptoms. Patients, on the other hand, may struggle to identify symptoms and delay seeking medical attention, further exacerbating their condition. Additionally, individuals in rural areas face obstacles such as cost, time, and limited access to diagnostic centers, with few medical facilities for skin disease detection. These challenges highlight the need for advanced expert systems that can accurately classify common skin diseases and detect conditions like early-stage acne[5] .

Advancements in artificial intelligence (AI) and deep learning, along with the increasingly prominent visual representation of diseases, have opened up new possibilities in the field of medical imaging. AI systems have demonstrated significant potential in analyzing medical images and enhancing diagnostic accuracy across various domains. By applying these same techniques to dermatology, AI has the capability to revolutionize the diagnosis of skin diseases [6]. This project aims to use the concept and implementation of AI systems to develop a medical image recognition system for skin diseases that can assist doctors in diagnosis and enable patients to detect diseases early. By training the system on a large dataset of skin disease images, the model will learn to identify patterns and distinguish different diseases. In addition to providing a second opinion for doctors, the system enables patients to upload images of their skin conditions and receive preliminary diagnostic feedback.

2.2 Project Scope

The primary objective of this project is to develop a robust AI-powered system capable of accurately recognizing and classifying over 20 types of skin diseases based on our comprehensive training data, including acne, rosacea, and actinic keratosis. Previous studies have shown that skin disease detection accuracy using AI models ranges from 50% to 70% [7] . This project aims to enhance the detection accuracy to a minimum of 60%, thereby assisting target users such as dermatologists and general practitioners in making more precise diagnoses..

To achieve this, we implemented three advanced deep learning models—ResNet50, DenseNet121, and VGG16—recognized for their robustness in image recognition. By training these models on a comprehensive dataset of skin disease images, we can assess their effectiveness in identifying distinct features of various skin conditions. And to further improve accuracy and reliability, this project promote a hard-voting ensemble method combines the strengths of each model to enhance diagnostic performance.

The system has been developed with the objective of providing diagnostic support to medical practitioners and preliminary self-assessment to patients. For medical practitioners, the system serves as a decision-support tool, assisting in complex cases or providing secondary confirmation. Patients can upload images for preliminary analysis, which is especially valuable in rural areas with limited healthcare access.

The system offers real-time feedback, analysing uploaded images to provide an accurate preliminary diagnosis that guides users on the next steps, from self-care to seeking medical advice. Continuous updates with new data ensure the system remains effective as more cases are added.

2.3 Project Significance

2.3.1 Current Limitations

Current AI-driven skin disease diagnostic tools face several limitations that restrict their reliability, accuracy, and accessibility in medical applications. These limitations, including accuracy, explainability, and accessibility, hinder their use in diverse clinical environments and limit their effectiveness as diagnostic aids.

2.3.1.a Limited Diagnostic Accuracy

Most skin disease diagnostic tools in the market exhibit an accuracy range of 60-70%, which can lead to significant false-positive or false-negative results. A 2020 study by Karunanayake et al. found that their model achieved an accuracy of 65% when identifying common skin diseases, highlighting an industry-wide struggle to meet higher diagnostic benchmarks[7]. The current level of accuracy falls short in meeting the requirements of healthcare environments for its possibilities to result in misdiagnoses. Enhancing model accuracy is imperative for these systems to establish themselves as a dependable diagnostic tool.

2.3.1.b Lack of Model Explainability

The common use of complex 'black-box' models in skin disease diagnostics presents challenges, particularly for medical professionals who need clear insights into diagnostic outcomes. Frequently, AI models, even when integrated into user interfaces, lack explanations for each prediction, thereby impeding healthcare providers from interpreting or verifying results effectively. This lack of transparency can discourage practitioners from utilizing these tools in clinical settings where explainability and accountability are crucial.

2.3.1.c Accessibility and Usability Barriers

High-performing diagnostic models often require substantial computational resources or specialized training, rendering them inaccessible to individuals residing in rural or low-resource areas. Moreover, systems exhibit a lack of user-friendly designs, thereby limiting their usability for non-specialists and patients. This dearth of accessibility confines AI diagnostic tools to larger institutions, leaving those in remote regions devoid of access to these cutting-edge technologies. Addressing challenges pertaining to usability and accessibility is pivotal in facilitating widespread adoption and ensuring equitable access to advanced diagnostic support.

2.3.2 Proposed Solution and Innovations

To address these limitations, this project presents an AI-based skin disease diagnostic system that is transparent, accurate, and easy to access. It is designed to help both healthcare providers and patients. By combining several deep learning models in a hard-voting ensemble and building the system as a user-friendly web application, this project aims to set a new standard for AI diagnostic tools.

2.3.2.a Model Selection and Hard Voting Ensemble

This project utilizes ResNet50, DenseNet121, MobileNetv2 and VGG16, established deep learning models renowned for their strengths in image recognition, to improve the diagnostic accuracy of skin disease classification. Each model individually classifies the input images to identify different skin conditions. To enhance robustness, we employ a hard-voting ensemble method, where each model's prediction contributes equally to the final decision. This ensemble approach reduces individual model biases, thereby improving classification accuracy and reliability.

2.3.2.b Web Application for Real-Time Analysis

This project integrates its model with a web application that connects backend models with a user-friendly frontend. This application allows healthcare providers and patients to easily upload images, enabling real-time analysis to provide diagnostic feedback. The web interface streamlines user interaction with the intricate backend models, enhancing usability and accessibility, particularly for patients residing in remote or underserved areas.

2.3.2.c Scalability and Continuous Learning

Another focus of this project lies in its scalability and adaptability, enabling the model to seamlessly incorporate new images and data types. Continuous updates are implemented to ensure the model maintains its relevance and accuracy over time. Furthermore, a transparent diagnostic pipeline is established to provide healthcare providers with a clear understanding of the basis for each diagnosis, thereby fostering trust and accountability. This scalable and explainable approach guarantees that the system can effectively support diverse future healthcare applications.

3.Literature Review

3.1 Related Work

- Li et al.(2020) reviewed 45 studies from 2016 to 2020 that applied deep learning methods for skin disease diagnosis. They examined image datasets, data augmentation, and deep learning models, mostly CNNs. The study found that deep learning models, especially multi-model fusion methods, outperform dermatologists and other computer-aided systems. For instance, Sarkar et al. achieved a recognition accuracy of 99.5%, and Mahbod et al. achieved an AUC score of 97.5% using hybrid deep learning models[6] .

- According to Goceri's (2021) research, a modified MobileNet model with a hybrid loss function is used for diagnosing skin diseases. Five common skin conditions were investigated: seborrheic dermatitis, rosacea, hemangioma, psoriasis, and acne vulgaris. This method utilizes dilated convolutions to improve contextual information and LeakyReLU activation to prevent neuron death. The model was applied in a mobile app with a high accuracy of 94.76% in classifying these diseases. Specifically, hemangioma was classified with an accuracy of 98.62%, while seborrheic dermatitis showed the lowest accuracy at 90.40%[8] .

- The study by Li et al.(2021) focuses on using deep learning for skin disease diagnosis, particularly skin cancer, through image classification and segmentation methods. The researchers explored the effectiveness of CNNs in identifying various skin diseases like melanoma and basal cell carcinoma. By leveraging large publicly available datasets such as the PH2 and ISIC, they achieved top-1 and top-3 classification accuracies of 60% and 80.3% respectively. The results demonstrated that deep learning models outperform human specialists with significant advancements in certain diagnostic tasks[9] .

- In the research by Zhu et al.(2021), a deep learning framework was developed for diagnosing 14 types of skin diseases using dermoscopic images in a clinical setting. The method was based on the EfficientNet-b4 model, which was fine-tuned with a dataset of 13603 images. The study achieved an overall accuracy of 94.8%, with a sensitivity of 93.4% and a specificity of 95.0%. The model outperformed other popular CNN models with an area under the curve (AUC) of 0.985, and it showed comparable performance to board-certified dermatologists in an 8-class diagnostic task[10] .

3.2 Key Findings

- Deep learning models like MobileNet and EfficientNet have achieved high diagnostic accuracies, with MobileNet reaching 94.76% and EfficientNet-b4 reaching 94.8% for multiple skin diseases.

- Specific diseases such as hemangioma and psoriasis are classified with particularly high accuracy.

- Hybrid deep learning models and multi-model fusion techniques outperform traditional approaches.

- Large publicly available datasets like ISIC and PH2 enable CNN models to generalize better, with top-1 and top-3 classification accuracies reaching 60% and 80.3% respectively for skin cancer diagnosis.

3.3 Methodologies

- Convolutional Neural Networks: Used for image classification and segmentation of skin diseases.

- MobileNet Model (modified): Improves diagnostic accuracy and prevents neuron death through LeakyReLU activation and dilated convolutions.

- EfficientNet-b4 Model: Fine-tuned to achieve high diagnostic performance for multiple skin diseases.

- Hybrid Loss Functions: Enhance feature extraction and contextual information for better classification.

- Multi-Model Fusion Methods: Achieve higher recognition accuracy compared to individual models and traditional systems.

4.System Design

4.1 Application system

The application system is comprised of three principal modules: the external environment, the interactive interface, and the system itself. As illustrated in Figure X, the following section will analyse and present the flow.

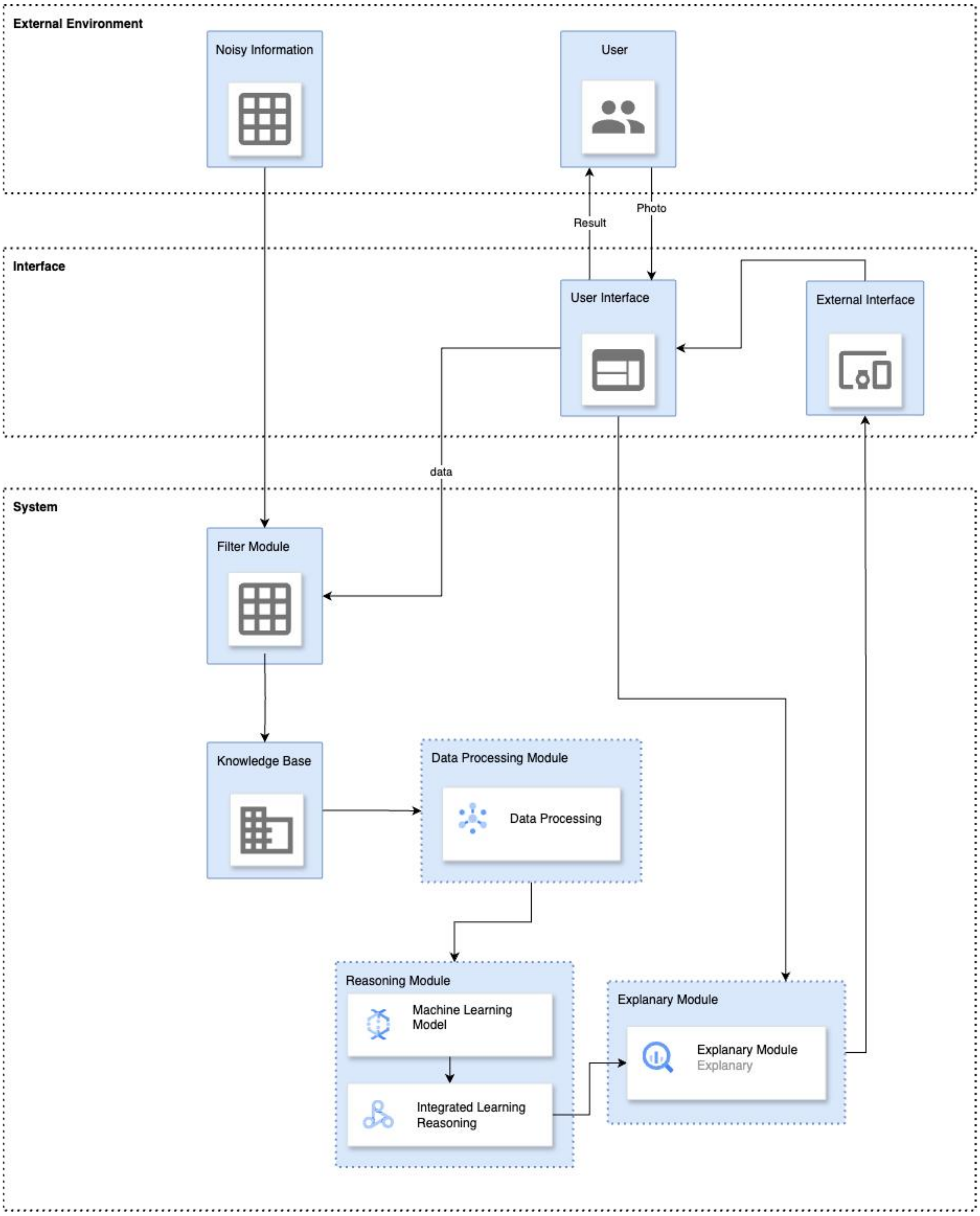

Figure 4-1. The Process of application system

The external environment has two segments, Noisy Information, which represents information to be collected externally about other skin conditions, and User, which includes skin patients, doctors, and ordinary people who want to observe their own skin conditions.

The second layer is the interaction layer, which is used to receive the information entering the system and to disseminate the information generated by the system to the user. The user then transfers the photographs to the user interface, which in turn receives the information emanating from the external interface and displays it to the user.

Regarding the system, it consists of five parts, and we describe the process separately by dividing it into the training process, and the application process.

- Training process

The Filiter Module is responsible for receiving all information and data from external sources and for organising and standardising them. This information and data is then stored in the Knowledge Base, where new data is regularly added to the data processing process, including data enhancement and feature extraction. The normalised data is then imported into the reasoning module for model training, where the newly trained models are evaluated and integrated, and finally dumped into the Explanary Module, representing the completion of training.

- Application process

The application process is considerably more straightforward, as the user interface accepts the data input by the user, imports it into the explanatory module, and then generates the results, which are then passed to the external interface.

The following section will concentrate on the two types of interface and the reasoning model.

4.2 User Interface Module

The module has been developed with the objective of providing a user-friendly interface for the submission of inputs and the subsequent display of results. The decision to utilise a web application was made on the grounds that it not only streamlines the design and development process, but also offers cross-platform compatibility. The web application was developed using the popular Vue.js framework, which is renowned for its flexibility, responsiveness and ease of use. The user interface permits the input of data in the form of image files. The files are then converted into PNG format, packaged into HTTP requests, and transmitted to a back-end server for subsequent processing. Subsequently, the model generates predictions, which are then conveyed back to the web application, where the type of skin disease is displayed.

4.3. External interface Module

The backend of this project, a RESTful interface has been defined for handling audio file uploads. This interface uses the POST method, allowing users to upload PNG files in multipart form data format. After successful upload, the interface will return the status of 201 Created and a success message. If an error occurs, detailed error information will be returned for the user to correct. The logic within the interface also provides scalability for future figure processing, allowing developers to conduct further analysis or feature extraction after file upload.

4.4 Reasoning Model Traning and Explanary Module

The process of training a model for pattern recognition comprises the following stages:

1. Data Processing
2. Model fine-tuning
3. Model training

4. Model evaluation

5. Model Ensemble (hard voting)

4.4.1Data Augmentation

The processing of image data is typically accomplished through the extraction of pixel information using the convolution neural network, the transformation of this information into features, and the subsequent generation of predictions. However, the majority of unenhanced images are designed to retain a greater degree of noise, blurring, or obscurity, which ultimately results in a deterioration of the visual quality. This can impede the recognition of crucial features, particularly in domains such as medical diagnosis, where precise accuracy is paramount[11] . Furthermore, computer vision algorithms may encounter challenges in accurately identifying or processing objects within an image, which could potentially lead to reduced accuracy and robustness in the analysis results.

So before image vectorisation, we need to enhance the image first, and our enhancement has the following main parts, which we will introduce in turn.

- Resize

All images are resized to a fixed size of 224x224. This standardises the size of the inputs, thereby avoiding computational problems caused by different image sizes. Furthermore, this is a common input size used in deep learning, which allows the model to better handle batch data.

- RandomHorizontalFlip

The image is randomly rotated 90 degrees either way to simulate different orientations. This operation helps to improve the robustness of the model by introducing perspective changes, increasing data diversity, and preventing the model from learning features from only a single orientation of the image.

- RandomRotation

The image is randomly rotated 90 degrees either way in order to simulate different orientations. This operation serves to enhance the robustness of the model by introducing perspective changes, increasing data diversity, and preventing the model from learning features from a single orientation of the image.

- ColorJitter

The image is randomly adjusted in terms of brightness, contrast and saturation in order to simulate changes in lighting conditions. This operation enables the model to adapt to disparate lighting conditions, thereby rendering it more resilient to fluctuations in chromaticity within the image.

4.4.2 Data Vectoring

Once the image enhancement process is complete, the enhanced image is tensorised as tensor, with the model training conducted within the PyTorch framework. This involves converting the image from PIL to PyTorch's Tensor and scaling the pixel values from the range of [0, 255] to the [0, 1] range. The necessity for data in a tensor format is a prerequisite for feeding image data into a neural network.

Subsequently, the transformed tensor is normalised using a predefined mean and standard deviation, which is typically applied to the mean and variance of the ImageNet dataset. This results in a more stable distribution of data values within the network for each channel, thereby accelerating the training process and preventing the issue of vanishing or exploding gradients.

4.4.3 Model Fine-tuning

In consideration of the computational resources available, four migration models were selected for fine-tuning and training purposes. These include Resnet50, Mobilenetv2, DenseNet121, and VGG16. The subsequent sections will present the details of the fine tune made to these models.

- Resnet50

The ResNet50 is a deep convolutional neural network that is widely used in computer vision tasks and belongs to the ResNet family. The ResNet50 addresses the issue of vanishing or exploding gradients, which is prevalent in deep networks, by incorporating skip connections or shortcut connections[12] . This enables deeper networks to be trained in an efficient manner.

Firstly, the pre-trained ResNet50 model was loaded and the weights trained on the ImageNet dataset (IMAGENET1K_V1) were utilised. This has enabled the model to learn a substantial number of fundamental visual features, including edges and textures. However, the new task may have a different number of categories, necessitating the removal of the original fully connected layer and its replacement with a custom classification layer. Firstly, the 2048-dimensional features output from ResNet50 are mapped to 512 dimensions via a fully connected layer. Subsequently, they are mapped to 23 dimensions, the 23 skin diseases within the dataset, via another fully connected layer.

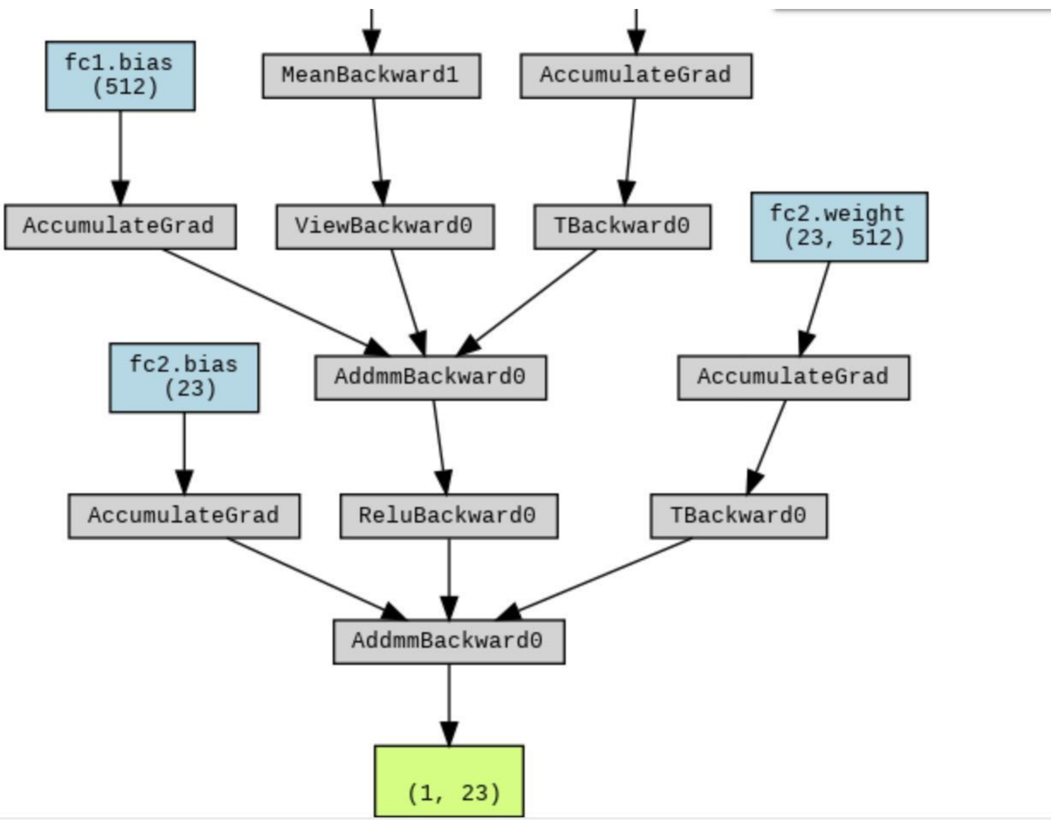

Figure 4-2.Fully Connected Layer based on Resnet50 after fine-tuning

- DenseNet121

DenseNet121 is a highly efficient deep learning model known for its densely connected convolutional layers [13] . DenseNet121 leverages feature reuse and improves gradient flow, making it effective for image classification tasks.

In this implementation, similarly to the ResNet50, the DenseNet121 model pre-trained on the ImageNet dataset is utilized to extract essential visual features like textures and shapes. The original classification layer of DenseNet121 is removed, allowing for custom classification suited to specific tasks. The 1024-dimensional output from the base model is then passed through a fully connected layer that reduces the feature dimension to 512, followed by another fully connected layer that maps the output to 23 dimensions. Finally, the output is projected down to 23 dimensions, corresponding to the various skin disease categories in the dataset.

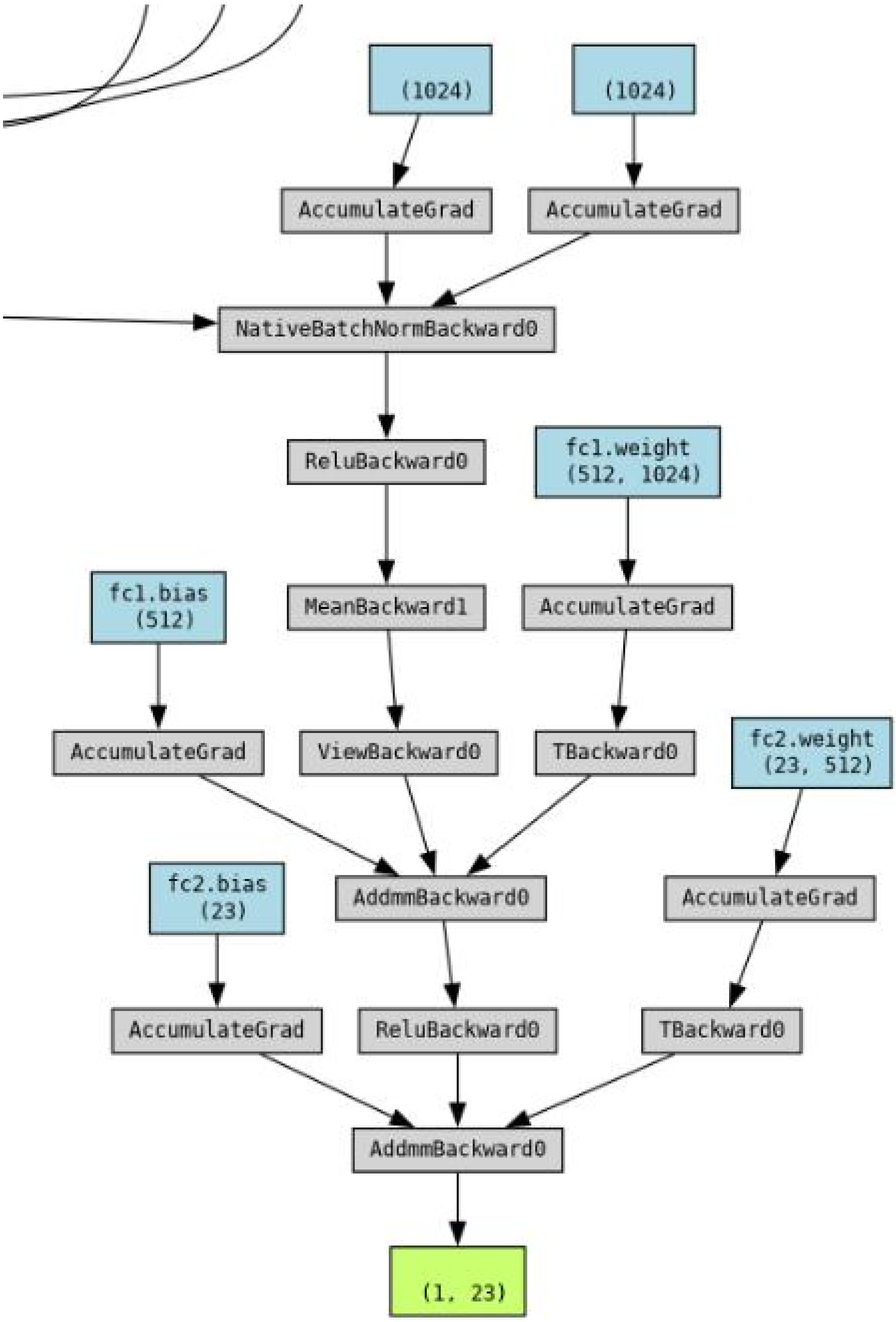

Figure 4-3.Fully Connected Layer based on DenseNet121 after fine-tuning

- VGG16

VGG16 is also a well-established deep convolutional neural network known for its simplicity and effectiveness in image classification tasks [14] . VGG16 utilizes a series of convolutional layers followed by pooling layers, which progressively reduce the spatial dimensions of the input while increasing the depth of feature representation.

Similarly to the ResNet50, the pre-trained VGG16 model is loaded, utilizing weights learned from the ImageNet dataset. This model effectively captures essential visual features such as shapes and textures. The last fully connected layer of VGG16 is removed to allow for custom classification tailored to specific tasks. The model outputs 4096 features after the convolutional and pooling layers, which are then fed into a fully connected layer that reduces the dimensionality to 512.

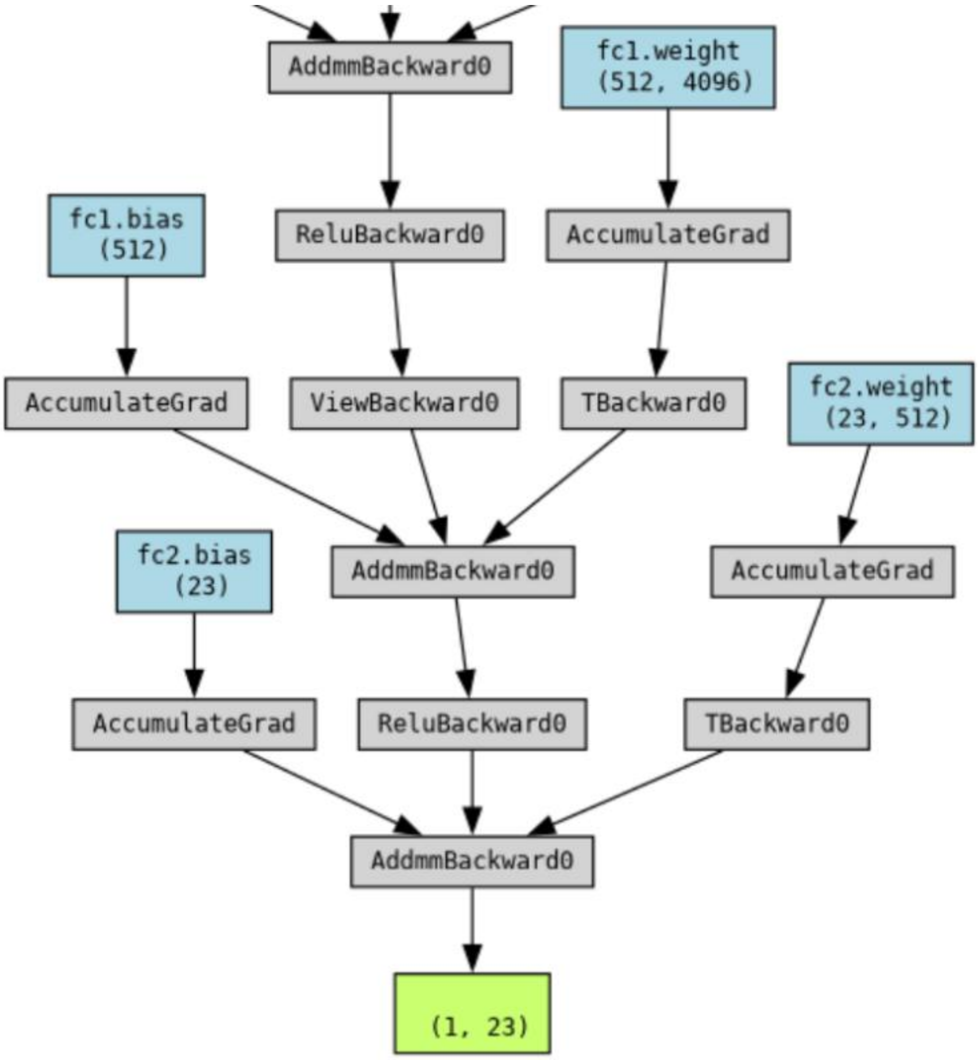

Figure 4-4 .Fully Connected Layer based on VGG16 after fine-tuning

- MobileNetV2

MobileNetV2 [15] is an efficient deep neural network architecture designed for mobile and low-resource devices. It builds on MobileNetV1 by introducing an inverted residual structure and linear bottlenecks, which reduce memory usage and computation while maintaining accuracy. MobileNetV2 is widely used in mobile vision applications due to its speed and efficiency.

Similarly to the ResNet50, the MobileNetV2 model pre-trained on the ImageNet dataset was loaded to represent essential visual features such as textures and shapes. The original classification layer was replaced according to our task. The extracted 1280-dimensional features from MobileNetV2 are projected down to 512 dimensions via a fully connected layer, followed by a final mapping to 23 dimensions, representing the skin disease categories in the dataset.

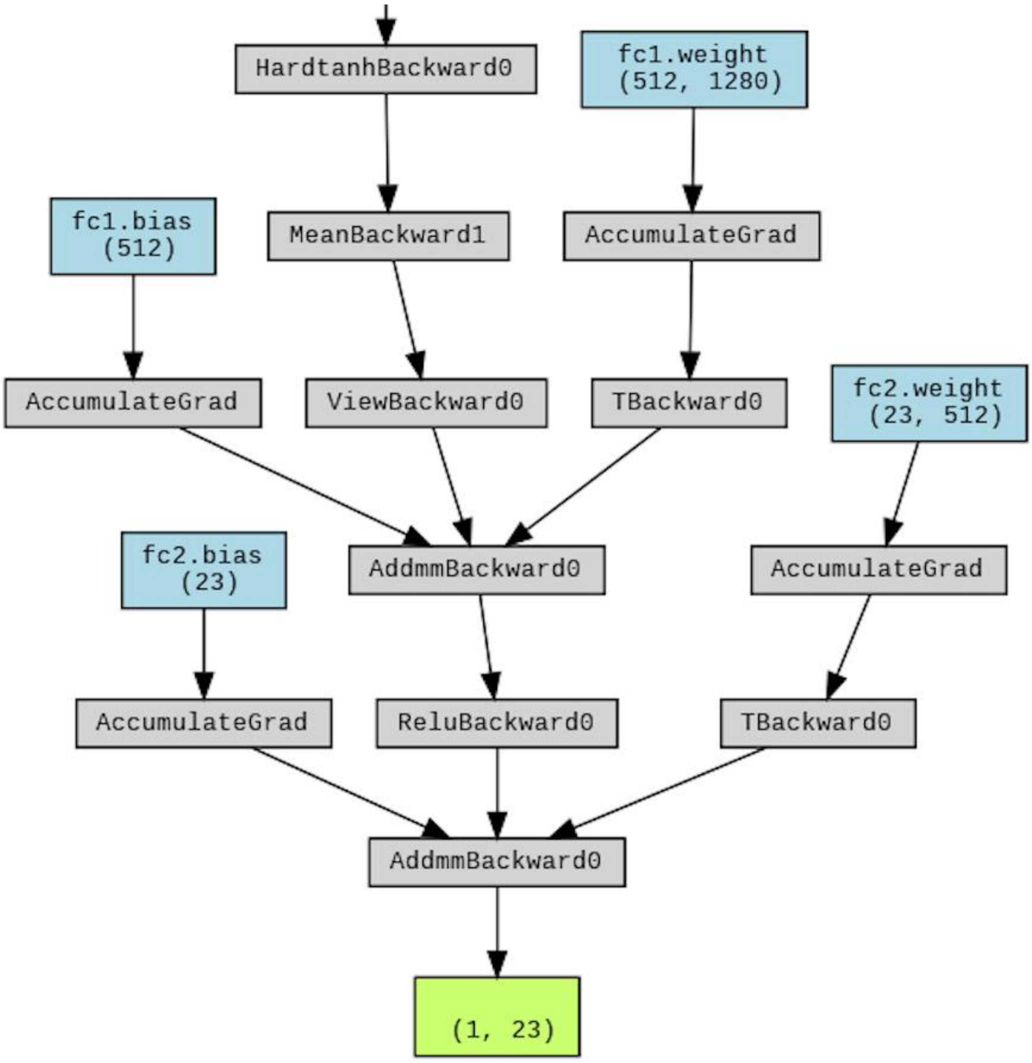

Figure 4-5.Fully Connected Layer based on MobileNetV2 after fine-tuning

4.4.4 Model Training

Following the fine-tuning of each model, training will commence. It should be noted that training methods vary between models, with differences evident in settings for batch size, epoch, learning rate and other hyperparameters. These will be introduced in turn, along with additional training details.

- Resnet50

In order to train the ResNet50 model, we employed the cross-entropy loss function and the Adam optimiser for the purpose of updating the model weights. During the training process, we set 50 epochs, as well as a batch size of 5. The model then performs forward propagation within each epoch, and optimises the model parameters to minimise the loss by backpropagation. Following the conclusion of each training epoch, the performance of the model is assessed by disabling the validation phase of gradient computation. Concurrently, the learning rate is adjusted in real-time according to the validation loss, employing the **ReduceLROnPlateau** learning rate scheduler to avert convergence issues or overfitting resulting from an excessively high learning rate [16] .

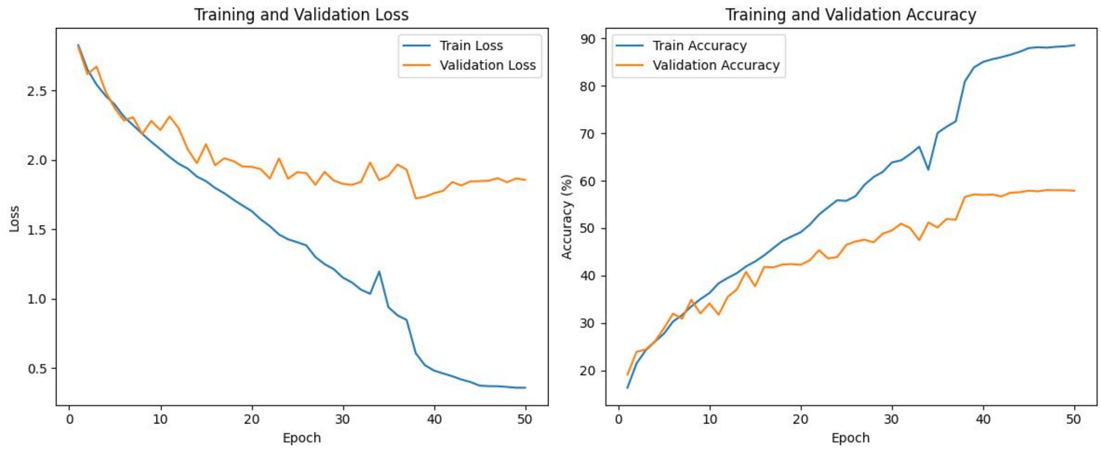

Figure 4-6.Loss and accuracy changes during model training for Resnet50

- DenseNet121

To train the Custom DenseNet121 model, we utilized the cross-entropy loss function and the Adam optimizer. The training was conducted over 50 epochs with a batch size of 5. We adjusted the initial learning rate from 0.001 to 0.0009 for improved convergence stability, using the ReduceLROnPlateau scheduler to dynamically modify the learning rate based on validation loss.

Early stopping was implemented with a patience of 10 epochs to prevent overfitting, monitoring validation loss to halt training if no improvement was observed.

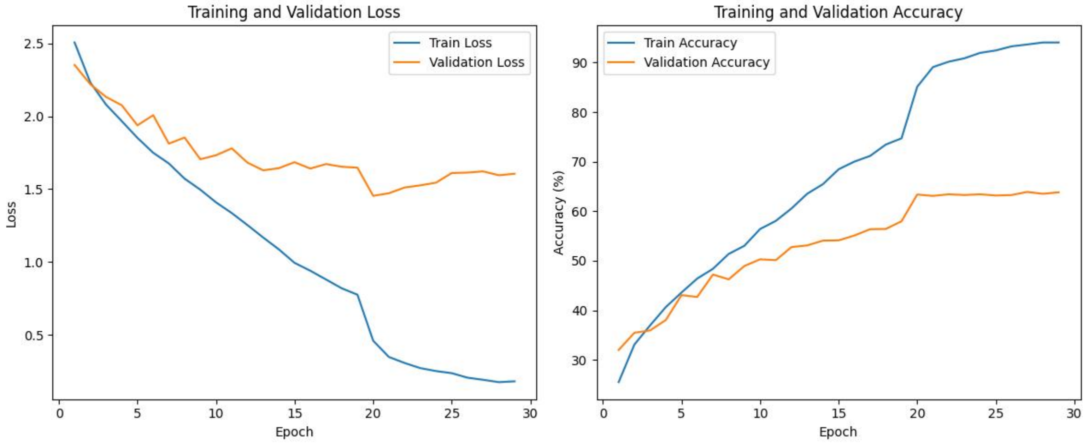

Figure 4-7.Loss and accuracy changes during model training for Custom DenseNet121

- VGG16

For the VGG16 model, we followed a similar training protocol, maintaining the same loss function and optimizer. However, due to the larger number of parameters, we found that a learning rate of 0.001 could cause gradient explosion, impeding convergence. To address this, we set the initial learning rate to 1e-4. The ReduceLROnPlateau scheduler and early stopping with a patience of 10 epochs were also applied to monitor validation loss and mitigate overfitting.

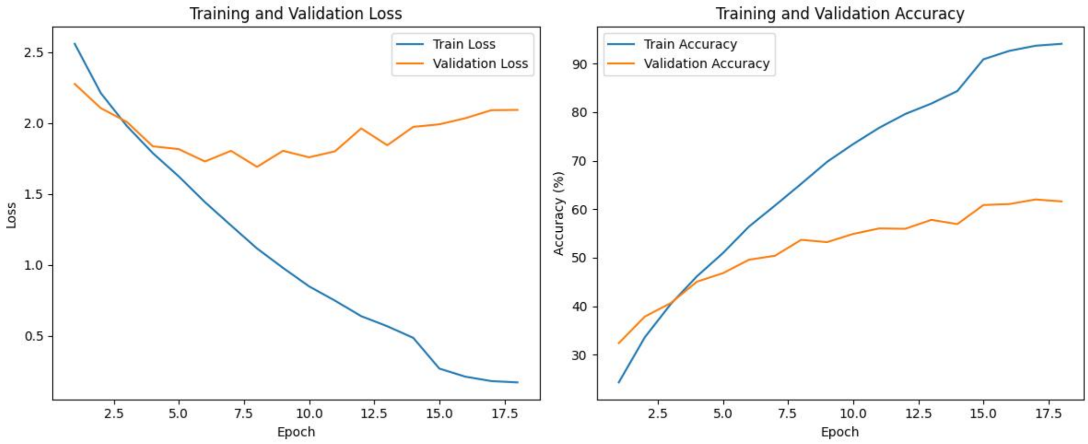

Figure 4-8.Loss and accuracy changes during model training for VGG16

- MobileNetV2

As to the MobileNetV2 model, the ross-entropy loss function and the Adam optimizer were also adopted, with a learning rate of 0.001. The training contained 50 epochs in total and batch size was set to 5. ReduceLROnPlateau scheduler was applied, but early stopping wasn't.

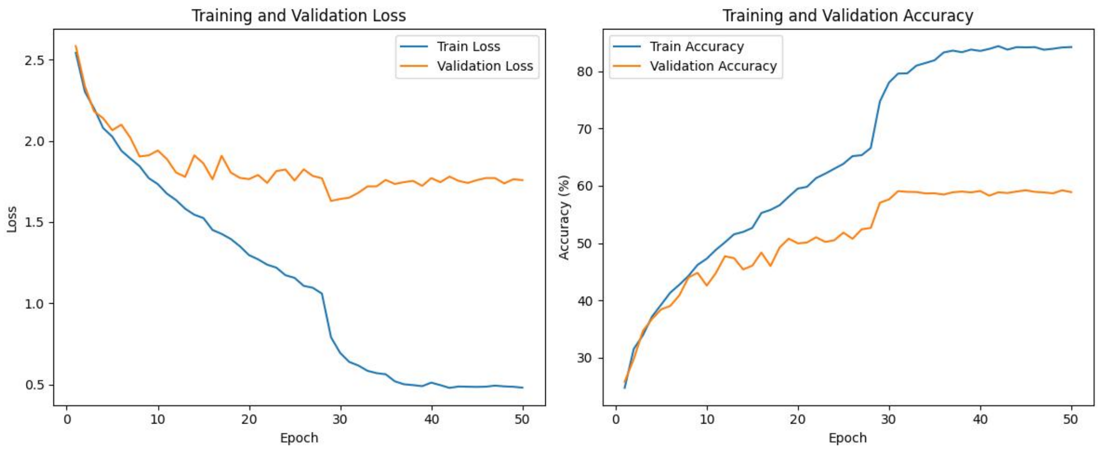

Figure 4-9.Loss and accuracy changes during model training for MobileNetV2

4.4.5 Model Evaluation

Following the training of the model, a sequential evaluation is conducted to ascertain its performance and effectiveness.

- Resnet50

Category	Precision	Recall	F1-score	Support
Acne and Rosacea Photos	0.82	0.89	0.86	312
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	0.64	0.57	0.6	288
Atopic Dermatitis Photos	0.58	0.49	0.53	123
Bullous Disease Photos	0.41	0.43	0.42	113
Cellulitis Impetigo and other Bacterial Infections	0.35	0.36	0.35	73
Eczema Photos	0.59	0.61	0.6	309
Exanthems and Drug Eruptions	0.39	0.5	0.44	101
Hair Loss Photos Alopecia and other Hair Diseases	0.51	0.55	0.53	60
Herpes HPV and other STDs Photos	0.42	0.41	0.42	102
Light Diseases and Disorders of Pigmentation	0.43	0.44	0.44	143
Lupus and other Connective Tissue diseases	0.4	0.34	0.37	105

Melanoma Skin Cancer Nevi and Moles	0.62	0.61	0.61	116
Nail Fungus and other Nail Disease	0.79	0.78	0.78	261
Poison Ivy Photos and other Contact Dermatitis	0.39	0.37	0.38	65
Psoriasis pictures Lichen Planus and related diseases	0.54	0.54	0.54	352
Scabies Lyme Disease and other Infestations and Bites	0.49	0.41	0.44	108
Seborrheic Keratoses and other Benign Tumors	0.56	0.64	0.6	343
Systemic Disease	0.47	0.53	0.5	152
Tinea Ringworm Candidiasis and other Fungal Infections	0.67	0.62	0.64	325
Urticaria Hives	0.47	0.45	0.46	53
Vascular Tumors	0.59	0.5	0.54	121
Vasculitis Photos	0.47	0.48	0.47	105
Warts Molluscum and other Viral Infections	0.59	0.58	0.58	272
macro avg	0.53	0.53	0.53	4002

Table 4-1. Evaluation of Resnet 50

According to *Table 4-1,* the results of the assessment demonstrate that the Resnet 50 exhibits lower precision, recall, and F1-score for categories including Cellulitis Impetigo and other Bacterial Infections, Bullous Disease Photos, Poison Ivy Photos, and other Contact Dermatitis, Exanthems and Drug Eruptions, Lupus and other Connective Tissue Diseases, and other Infectious Diseases. These findings indicate that the model may have limited capacity to accurately represent the characteristics of these diseases, which could result in an increased likelihood of misclassification.

However, the model exhibits high precision and recall in classification, particularly in the identification of melanoma, acne, and nail fungus. This illustrates the reliability of the model in the identification of specific skin diseases.

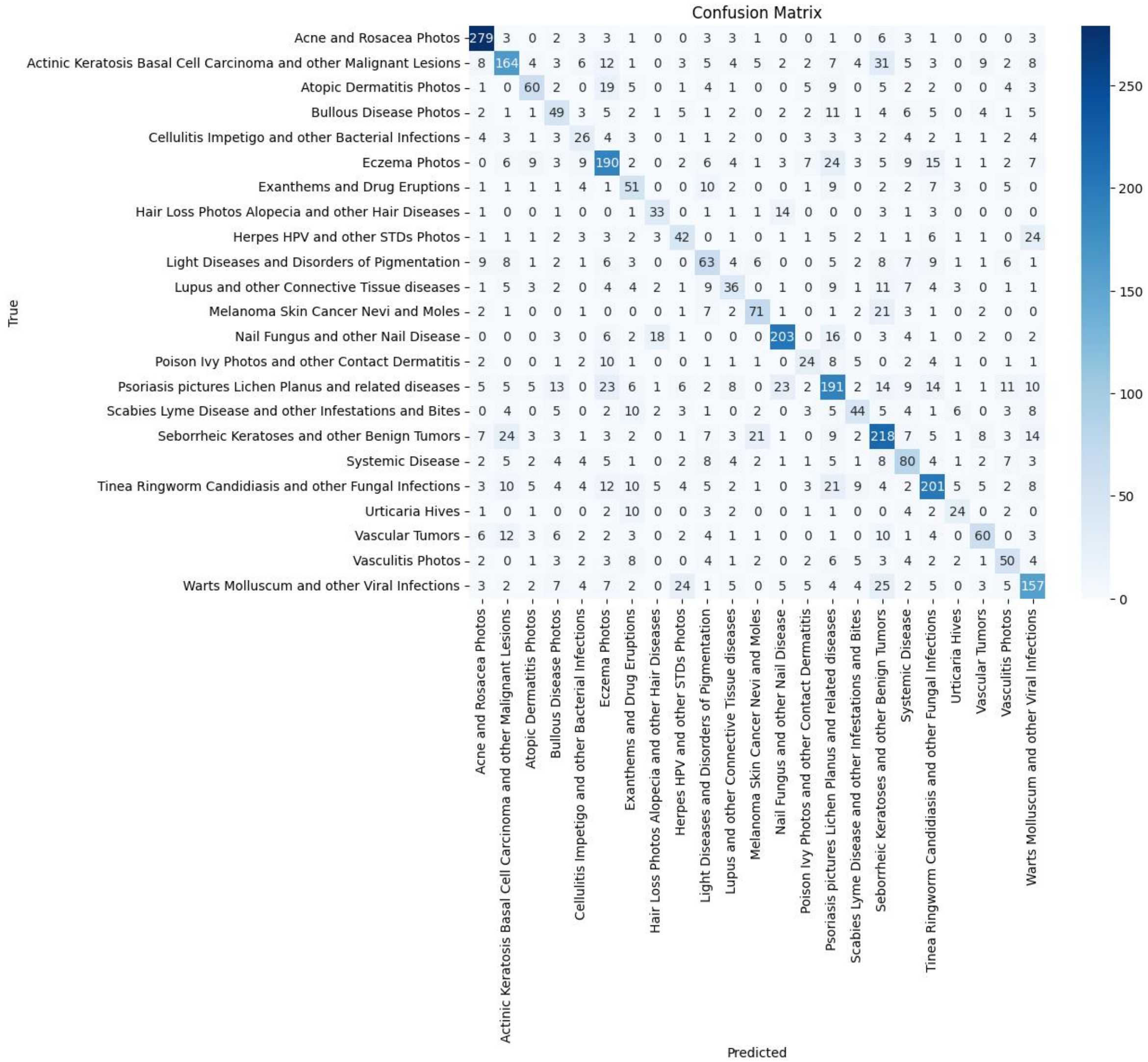

Figure 4-10. Confusion Matrix of ResNet50

- DenseNet121

Category	Precision	Recall	F1-score	Support
Acne and Rosacea Photos	0.84	0.94	0.89	312
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	0.71	0.68	0.69	288
Atopic Dermatitis Photos	0.63	0.51	0.57	123
Bullous Disease Photos	0.49	0.51	0.50	113
Cellulitis Impetigo and other Bacterial Infections	0.48	0.40	0.44	73
Eczema Photos	0.67	0.63	0.65	309
Exanthems and Drug Eruptions	0.43	0.55	0.48	101
Hair Loss Photos Alopecia and other Hair Diseases	0.58	0.67	0.62	60
Herpes HPV and other STDs Photos	0.52	0.48	0.50	102

Light Diseases and Disorders of Pigmentation	0.56	0.59	0.57	143
Lupus and other Connective Tissue diseases	0.62	0.44	0.51	105
Melanoma Skin Cancer Nevi and Moles	0.66	0.72	0.69	116
Nail Fungus and other Nail Disease	0.80	0.79	0.80	261
Poison Ivy Photos and other Contact Dermatitis	0.53	0.45	0.48	65
Psoriasis pictures Lichen Planus and related diseases	0.62	0.61	0.61	352
Scabies Lyme Disease and other Infestations and Bites	0.52	0.54	0.53	108
Seborrheic Keratoses and other Benign Tumors	0.66	0.71	0.68	343
Systemic Disease	0.50	0.59	0.54	152
Tinea Ringworm Candidiasis and other Fungal Infections	0.73	0.69	0.71	325
Urticaria Hives	0.65	0.64	0.65	53
Vascular Tumors	0.70	0.61	0.65	121
Vasculitis Photos	0.59	0.60	0.59	105
Warts Molluscum and other Viral Infections	0.65	0.67	0.66	272
macro avg	0.61	0.61	0.61	4002

Table 4-2. Evaluation of DenseNet121

According to *Table 4-2,* the Custom DenseNet121 model displays varying precisions and recalls across different categories. The model effectively captures the characteristics of skin diseases with distinct features, especially achieving a precision of 0.84 and a recall of 0.94 in identifying Acne and Rosacea Photos. In complex categories like "Lupus and other Connective Tissue Diseases", it achieves a precision of 0.62 compared to ResNet50's 0.40, resulting in an improvement of 0.22, indicating DenseNet121's strong ability to capture features of such complex diseases. Additionally, the precision for Bullous Disease Photos increased by 0.08, while for Cellulitis Impetigo and other Bacterial Infections, it improved by 0.13, reflecting enhancements in DenseNet121's performance.

However, in some categories, the performance differences between DenseNet121 and ResNet50 are less significant. For instance, in "Psoriasis" and "Warts and other Viral Infections", they demonstrates comparable capabilities.

Overall, while DenseNet121 shows significant strengths in classifying certain diseases compared to ResNet50, further optimization for specific categories remains crucial for enhancing the model's overall diagnostic reliability.

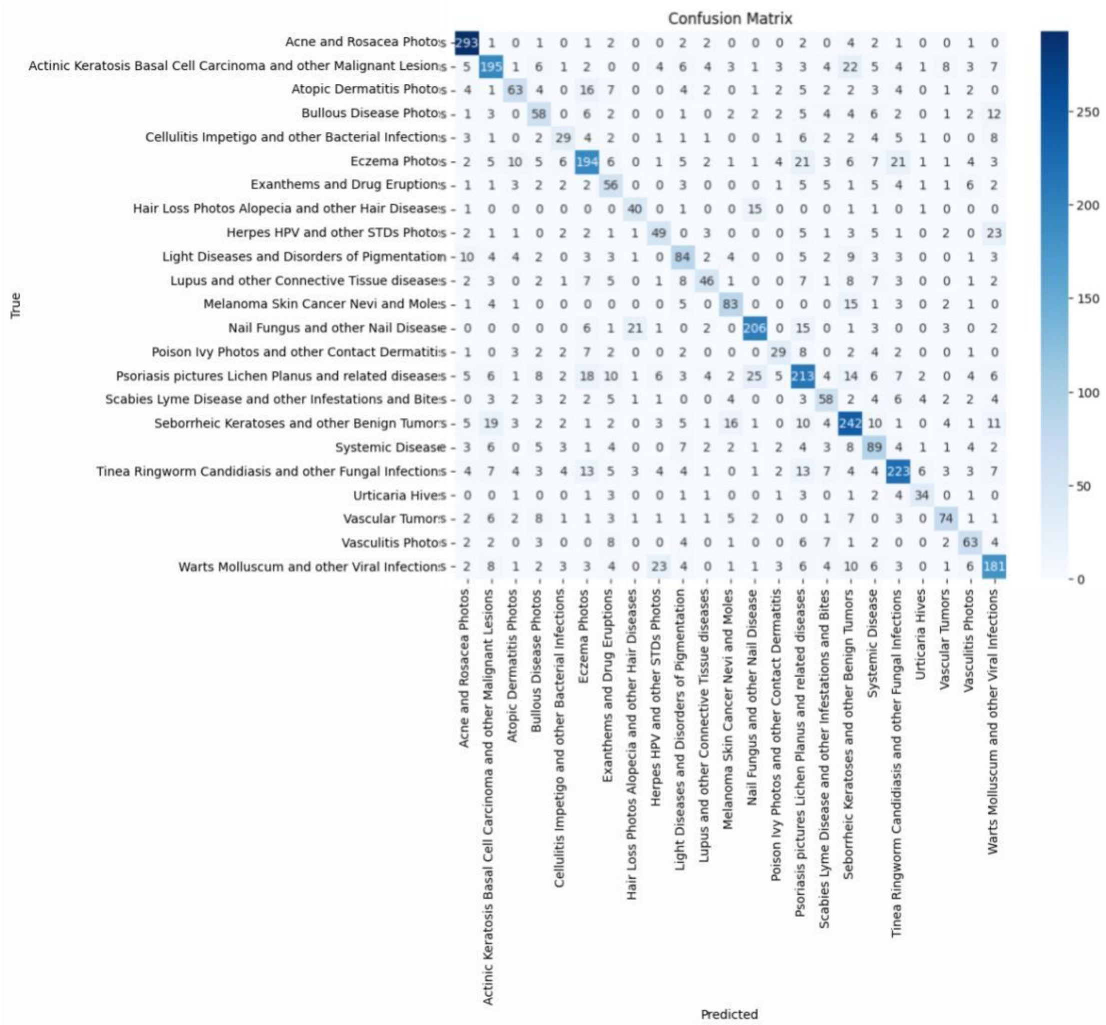

Figure 4-11. Confusion Matrix of DenseNet121

- VGG16

Category	Precision	Recall	F1-score	Support
Acne and Rosacea Photos	0.78	0.96	0.86	312
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	0.62	0.69	0.65	288
Atopic Dermatitis Photos	0.53	0.48	0.50	123
Bullous Disease Photos	0.54	0.49	0.51	113
Cellulitis Impetigo and other Bacterial Infections	0.38	0.32	0.34	73
Eczema Photos	0.64	0.64	0.64	309
Exanthems and Drug Eruptions	0.48	0.44	0.46	101
Hair Loss Photos Alopecia and other Hair Diseases	0.49	0.57	0.53	60
Herpes HPV and other STDs Photos	0.47	0.44	0.46	102
Light Diseases and Disorders of Pigmentation	0.51	0.59	0.54	143

Lupus and other Connective Tissue diseases	0.47	0.45	0.46	105
Melanoma Skin Cancer Nevi and Moles	0.70	0.69	0.70	116
Nail Fungus and other Nail Disease	0.80	0.76	0.78	261
Poison Ivy Photos and other Contact Dermatitis	0.48	0.43	0.46	65
Psoriasis pictures Lichen Planus and related diseases	0.62	0.55	0.58	352
Scabies Lyme Disease and other Infestations and Bites	0.61	0.46	0.53	108
Seborrheic Keratoses and other Benign Tumors	0.62	0.69	0.65	343
Systemic Disease	0.52	0.46	0.49	152
Tinea Ringworm Candidiasis and other Fungal Infections	0.64	0.63	0.64	325
Urticaria Hives	0.59	0.64	0.61	53
Vascular Tumors	0.62	0.61	0.62	121
Vasculitis Photos	0.50	0.51	0.51	105
Warts Molluscum and other Viral Infections	0.59	0.58	0.59	272
macro avg	0.57	0.57	0.57	4002

Table 4-3. Evaluation of VGG16

When comparing performances of models, we find that the accuracy of VGG16 lies between that of DenseNet121 and ResNet50. VGG16's performance metrics are lower than DenseNet121's in most of the categories. In the categories where it performs relatively well, e.g., VGG16's accuracy is higher than DenseNet121's in the category of “Exanthems and Drug Eruptions”, it still has lower recall. This suggests that although it can identify samples in this category, the number of samples identified is relatively small, which may lead to some samples being missed.

Overall, compared to VGG16, DenseNet121 is more capable in pattern recognition for the dataset, especially when dealing with complex or diverse skin diseases. The weaker performance of VGG16 may be attributed to the fact that it has a higher parameter count and greater model complexity, which are more suitable for larger datasets.

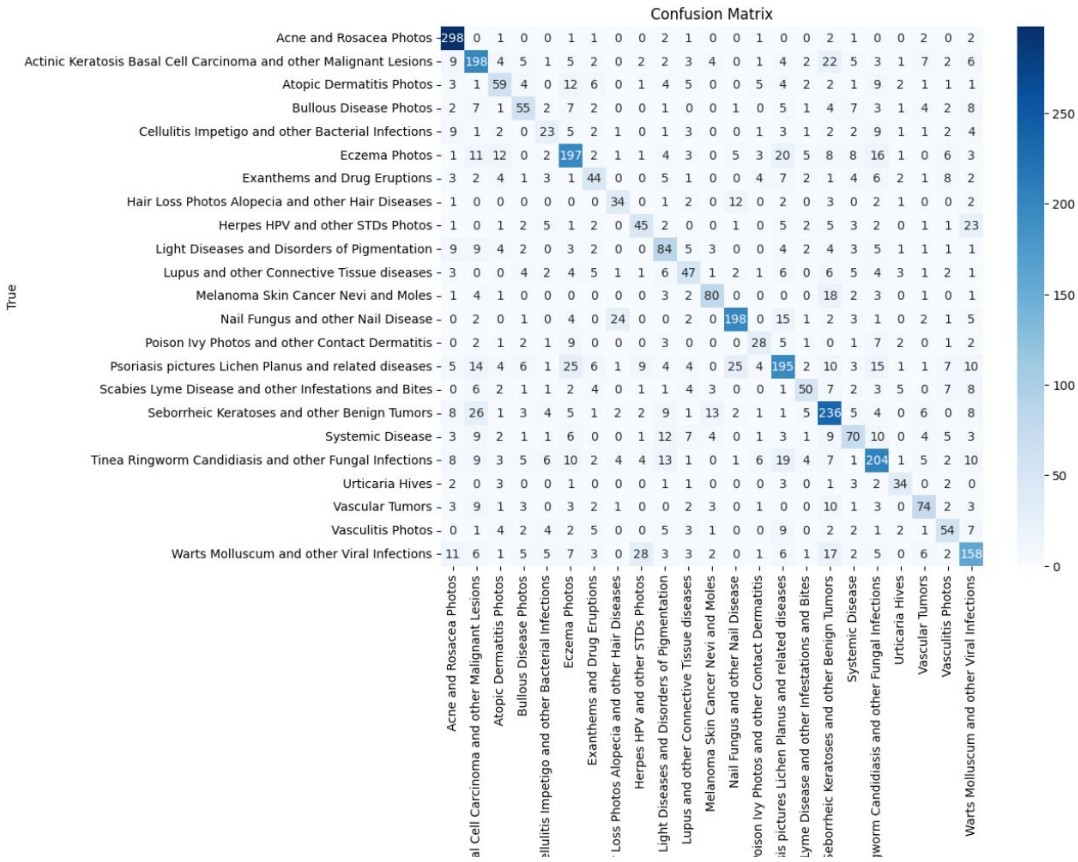

Figure 4-12. Confusion Matrix of VGG16

- MobileNetV2

Category	Precision	Recall	F1-score	Support
Acne and Rosacea Photos	0.80	0.91	0.85	312
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	0.65	0.61	0.63	288
Atopic Dermatitis Photos	0.51	0.49	0.50	123
Bullous Disease Photos	0.44	0.39	0.42	113
Cellulitis Impetigo and other Bacterial Infections	0.35	0.26	0.30	73
Eczema Photos	0.59	0.66	0.62	309
Exanthems and Drug Eruptions	0.51	0.54	0.53	101
Hair Loss Photos Alopecia and other Hair Diseases	0.51	0.57	0.54	60
Herpes HPV and other STDs Photos	0.53	0.42	0.47	102
Light Diseases and Disorders of Pigmentation	0.41	0.47	0.44	143
Lupus and other Connective Tissue diseases	0.41	0.35	0.38	105
Melanoma Skin Cancer Nevi and Moles	0.70	0.60	0.65	116
Nail Fungus and other Nail Disease	0.81	0.75	0.78	261
Poison Ivy Photos and other Contact Dermatitis	0.38	0.32	0.35	65

Psoriasis pictures Lichen Planus and related diseases	0.58	0.55	0.56	352
Scabies Lyme Disease and other Infestations and Bites	0.35	0.42	0.38	108
Seborrheic Keratoses and other Benign Tumors	0.61	0.60	0.61	343
Systemic Disease	0.46	0.48	0.47	152
Tinea Ringworm Candidiasis and other Fungal Infections	0.62	0.64	0.63	325
Urticaria Hives	0.67	0.57	0.61	53
Vascular Tumors	0.60	0.56	0.58	121
Vasculitis Photos	0.49	0.54	0.51	105
Warts Molluscum and other Viral Infections	0.60	0.60	0.60	272
macro avg	0.55	0.54	0.54	4002

Table 4-4. Evaluation of MobileNetV2

According to *Table 4-4*, MobileNetV2's performance indicates lower precision, recall, and F1-scores for categories like "Cellulitis Impetigo and other Bacterial Infections," "Bullous Disease Photos," "Poison Ivy Photos and other Contact Dermatitis," and "Lupus and other Connective Tissue Diseases." This implies the model may face challenges in correctly identifying these particular skin conditions with a higher chance of misclassification.

However, MobileNetV2 shows better performance with higher precision and recall in categories such as "Acne and Rosacea Photos," "Nail Fungus and other Nail Disease," and "Melanoma Skin Cancer Nevi and Moles." This suggests MobileNetV2 is more effective at capturing the characteristics of these conditions, which can support more accurate identification in a diagnosis.

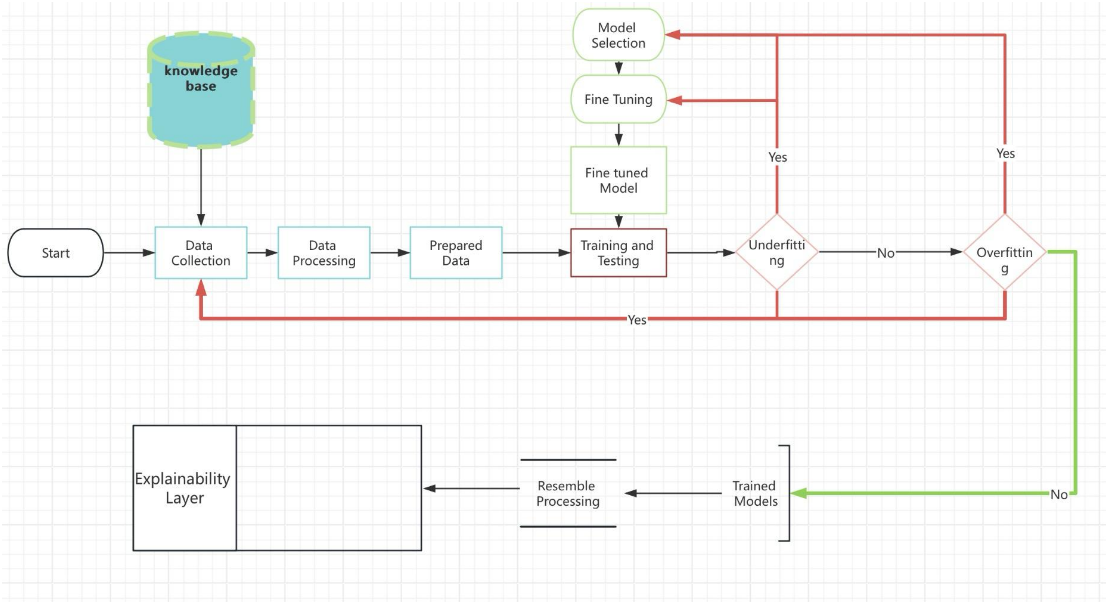

Figure 4-13.The training process of Models

4.4.6 Model Ensemble (hard voting)

In the preceding section, we trained distinct deep learning models and evaluated them. While the overall accuracy is comparable for all three models, for a specific classification, the accuracy of different models for a particular skin disease varies. In other words, the predictive probability of different models for a single skin disease exhibits a considerable degree of variance. To enhance the stability of this prediction, we integrated multiple models to develop an integrated model for hard voting.

In the "hard voting" phase, the selected models vote on the classification results. This involves aggregating the predictions of each model and selecting the prediction that receives the most votes (or supports the most models) as the final decision. As *Figure 4-14,* to illustrate, if three models predicted classification A, then classification A is the final decision in this case.

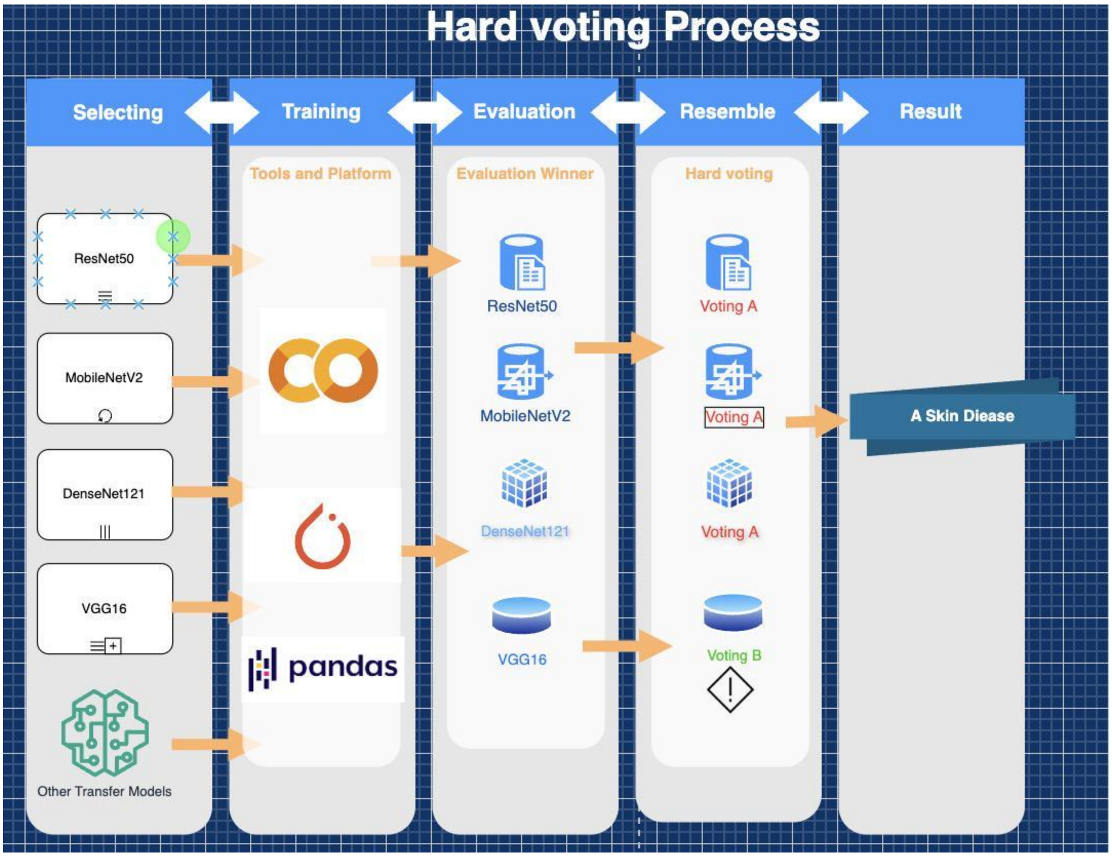

Figure 4-14. Hard voting Process

We briefly evaluated the integrated model for hard voting and obtained an accuracy of 62.3%, which is not too high in terms of accuracy but its variance in the accuracy of prediction for different types of skin diseases has been reduced, which improves the stability to some extent.

Diease Class	Accuracy (%)
Acne and Rosacea Photos, Accuracy	94.23
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions,	61.46
Atopic Dermatitis Photos,	52.03
Bullous Disease Photos,	42.48
Cellulitis Impetigo and other Bacterial Infections,	32.88
Eczema Photos,	55.34
Exanthems and Drug Eruptions,	43.56

Hair Loss Photos Alopecia and other Hair Diseases,	56.67
Herpes HPV and other STDs Photos,	51.96
Light Diseases and Disorders of Pigmentation,	54.55
Lupus and other Connective Tissue diseases,	40
Melanoma Skin Cancer Nevi and Moles,	62.07
Nail Fungus and other Nail Disease,	76.63
Poison Ivy Photos and other Contact Dermatitis,	35.38
Psoriasis pictures Lichen Planus and related diseases,	46.02
Scabies Lyme Disease and other Infestations and Bites,	37.04
Seborrheic Keratoses and other Benign Tumors,	77.55
Systemic Disease,	57.89
Tinea Ringworm Candidiasis and other Fungal Infections,	67.38
Urticaria Hives,	62.26
Vascular Tumors,	43.8
Vasculitis Photos,	50.48
Warts Molluscum and other Viral Infections,	51.1

Table 4-5.Single Evaluation of Resemble

5.Web application Development

5.1 Initiation

Based on the previous experiments and analyses, the corresponding web system was designed to enable users to upload pictures of skin diseases and obtain the recognition results in web pages. This chapter describes the development and implementation process of the system and provides a detailed description of the front and back end respectively.

5.2 Front-End Development

The user interface of our deception detection system is an interactive web application built with Vue 3 and Vite. The front-end collects user input, sends it to the back end as an HTTP request, gets a response, and presents it in a clear and readable way.

5.2.1 Tools

Vue.js

Vue is a progressive JavaScript framework for building user interfaces. It is known for its flexibility, reactivity, and ease of use. It builds on top of standard HTML, CSS, and JavaScript and provides a declarative, component-based programming model that helps efficiently develop user interfaces.

Vite

Vite is a modern, fast-build tool designed to improve the development experience, especially for front-end projects. It provides a rapid, lightweight development server with hot module replacement (HMR) and fast production builds. Vite works out of the box with Vue 3 and significantly improves the efficiency of our web development process.

PrimeVue

PrimeVue is a UI component library for Vue.js that provides a wide array of ready-made components designed to enhance the speed and quality of front-end development. It is highly customizable and suitable for both simple and complex applications. It also allows easy integration with Vue's Composition API. We used its FileUpload component for uploading audio files and the ConfirmDialog component for displaying results in our web application.

Axios

Axios is a popular JavaScript library used for making HTTP requests from both the browser and Node.js environments. It provides an easy-to-use API for sending asynchronous requests (GET, POST, PUT, DELETE, etc.) to interact with RESTful APIs. Axios supports features like request and response interceptors, automatic JSON transformation, and error handling. It is adopted in our system to achieve communication between front-end and back-end servers.

5.2.2 Application Design

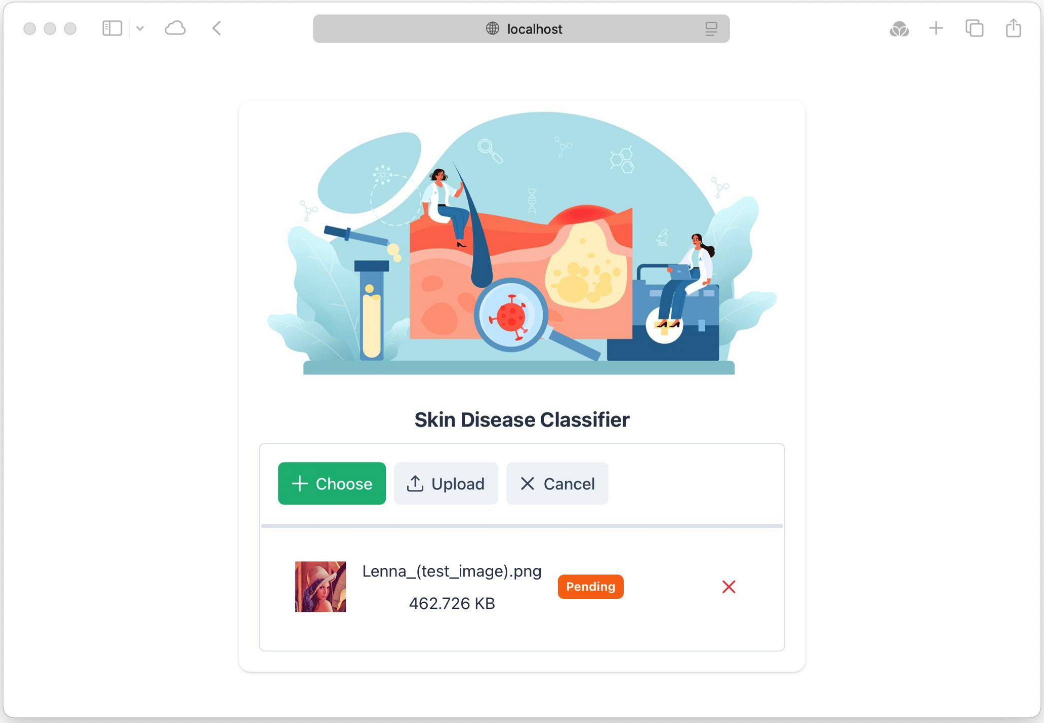

Figure 5-1. Web Application Screenshot

Our web application is quite simple and easy to use. By simply clicking the "Choose" button in the frame, the user can upload an image file, which will be sent to the back-end server. After submitting, the machine-learning model in the back-end will generate an outcome with a piece of description attached, and then the result is displayed on the web page.

5.3 Back-End Development zwy

5.3.1 System Architecture

The back-end is designed using the Django framework, which adopts the MVC design pattern (Model-View-Controller), which effectively reduces the coupling between modules and facilitates code refactoring afterwards.

5.3.2 Database Design

The system defines an image model to manage the image files uploaded by users. The model definition contains the file uploaded by the user and the corresponding upload time, which will facilitate the further management of the file.

In the database the model corresponds to a table where the primary key is a unique identifier while other fields are used to store the image file path and upload time. And these fields allow the system to efficiently manage and retrieve image data.

For example, when the user uploads an image file, the system will generate a record in the database, including the path of the image file and its upload time.

```
sqlite> SELECT * FROM api_skinimage ORDER BY uploaded_at DESC;
5|images/disease3_eMyjDSh.jpg|2024-10-25 09:21:21.322608
4|images/123.jpg|2024-10-25 09:21:17.764805
3|images/disease3.jpg|2024-10-25 09:20:45.924501
```

Figure 5-2. image record

5.3.3 Api Design

The API design of this project consists of three modules: Serializer, URL and View.These modules together constitute the complete process of audio file processing and ensures efficient and reliable data exchange between the front-end and back-end. The following diagram shows the relationship between these three modules and their data flow.

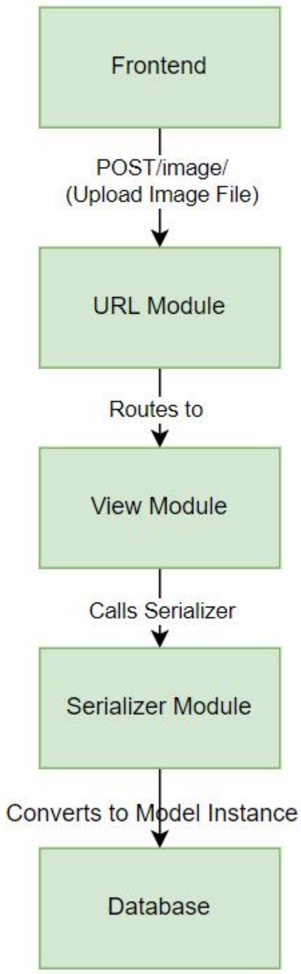

Figure 5-3. API Design

5.3.3.a Serializer Module

The serializer SkinImageSerializer is defined in this module. Its function is to serialize and deserialize data. When the API needs to return image file information, the deserializer will automatically convert the model data to JSON format by defining fields such as id, image and uploaded_at. When the user uploads an image file through the API, the serializer will validate the input data, and after validation, it will convert the data into a model instance and then save it to the database.

5.3.3.b URL Module

This module mainly defines the routing of the API related to skin disease image files, and uses the routing function of Django REST Framework in the process of realization, which makes the management and access of API more convenient and effective.

The module sets up an image folder to store the image files uploaded by users. The module sets up the image folder to store the image files uploaded by users and includes the URL patterns generated by the router in the list of urlpatterns. When a user accesses the URL of an image file, it will be routed to the corresponding module to handle the request.

5.3.3.c View Module

This module is responsible for recognizing the skin disease image uploaded by the user and returning the corresponding results.

When receiving the request, the module will call serializer to validate the uploaded data. If the validation is successful, the system saves the image file to the database and generates the corresponding record. After the uploaded skin disease image is loaded, the module will process the data accordingly and use the previously trained model to make the prediction, after which the prediction results will be sent back to the front-end.

6.Challenges and Future Work

6.1 Challenge

6.1.1 Limitation of Data and Features

Despite the extraction of two gigabytes of pircture data, comprising about 2000 pieces of pircture, the resulting dataset remains relatively modest in scale for a personal computer.

Further data is required to facilitate the training of the inference engine, although this presents a significant challenge for the current team. Furthermore, the considerable number of images labelled with skin diseases must be extracted from more hospital-like locations. Furthermore, the process of feature extraction presents a significant challenge. Individuals possess a range of skin colours and features, which can impede the accurate detection of skin disease characteristics. For instance, the engine may erroneously identify a skin disease prevalent in individuals with white skin, even if the individual does not exhibit the disease. This could lead to inaccurate assessments. Addressing this issue and ensuring the engine extracts features with greater precision is crucial.

6.1.2 Problem for the accuracy and generation of reasoning engine

Although the Hard Voting Integration Model has demonstrated more consistent performance on the test set, it remains uncertain whether this accuracy can be generalized to other datasets, as well as to environments with significant distractions and to these specific environments where the user picture is not clear, which may be unstable. Moreover, it would be beneficial to ascertain whether the model's accuracy can be enhanced, for instance, to 80%. This would serve to bolster the user's confidence in the model's judgement. Finally, with regard to the interpretability of the model, given that this project employs all deep learning models, we currently encounter significant challenges in elucidating the internal workings of these deep multi-layer neural networks. This is a crucial aspect in the medical domain, where the user may not have confidence in the machine's judgement due to a lack of sufficient evidence, even when the accuracy is at an exceedingly high level.

6.2 Future Work

6.2.1 Enhancing Model Performance and Robustness

To push the boundaries of diagnostic accuracy, additional architectures and most up-to-date model present promising avenues for model development. Testing those up-to-date models in isolation and as part of an ensemble can help identify the most effective architecture for dermatological image recognition.

Furthermore, incorporating weighted or soft voting ensemble methods may improve adaptability. In these methods, models that excel in specific disease categories may contribute more heavily to the final decision, enhancing robustness across diverse conditions.

6.2.2 Expanding the Dataset and Diversity

The accuracy and generalizability of skin disease recognition models depend on diverse, representative training data. Current datasets didn't encompass all skin tones, age groups, or regions, potentially leading to biases. To address this, future efforts will focus on collaborating with healthcare institutions to build required and comprehensive dataset. Expanding data collection to include a broader range of skin tones, backgrounds, and ages will help the model learn demographic variations. Collaborating with dermatology experts will ensure that the data accurately reflects a global population, enhancing both applicability and fairness.

6.2.3 Multimodal Diagnostic Integration

As the complexity of skin diseases varies, incorporating multimodal data could greatly enhance diagnostic precision. Integrating patient history, symptoms, and additional diagnostic data (such as previous treatments, associated symptoms, or biopsies) can add valuable context to image-based diagnosis, providing a holistic view of the patient’s condition. Future versions of the system could incorporate these elements by linking image data with text-based health records from medical consultations. In addition, the fusion of modalities such as clinical test results, symptom descriptions, and environmental factors could allow the system to make more informed diagnoses in complex cases, which has the potential to improve the system’s capacity to weigh and interpret the significance of each modality, ultimately supporting more nuanced diagnostic outputs.

6.2.4 Real-Time Data Collection and Model Updating

To keep diagnostic capabilities current, a real-time feedback loop could be integrated for continuous learning and adaptation. By collecting verified diagnostic outcomes and real-world feedback, the model could retrain with the latest user data, maintaining relevance in clinical settings. Implementing a feedback-driven retraining pipeline would allow healthcare providers to contribute anonymized patient data with consent, adding to the system's knowledge base. Incremental training techniques would help model to integrate new data efficiently without full retraining, ensuring diagnostic accuracy over time.

Reference

[1] Hu Z. and Yu C. S., "Functional research and development of skin barrier", *Chinese Journal of Clinicians*, vol. 7, no. 7, 2013, pp. 3101–3103.

[2] British Skin Foundation., May 2018, [online] Available: http://www.britishskinfoundation.org.uk.

[3] A. Esteva, et al., "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, vol. 542, no. 7639, pp. 115-118, Feb. 2017.

[4] M. E, J. Newton, S.-B. S, S.-B. S, R. T and F. A, "The quality of life in acne: a comparison with general medical conditions using generic questionnaires", *British Journal of Dermatology*, pp. 672-676, 1999.

[5] Nawar, Anika, et al. "Skin disease recognition: a machine vision based approach." *2021 7th International conference on advanced computing and communication systems (ICACCS).* Vol. 1. IEEE, 2021.

[6] Li, Ling-Fang, et al. "Deep learning in skin disease image recognition: A review." *IEEE Access*, vol. 8, 2020, pp. 208264–208280.

[7] K Karunanayake, R.K.M.S. et al.“CURETO: Skin Diseases Detection Using Image Processing And CNN.” *2020 14th International Conference on Innovations in Information Technology (IIT)* , 2020, pp. 1–6.

[8] Goceri, Evgin. "Diagnosis of skin diseases in the era of deep learning and mobile technology." *Computers in Biology and Medicine*, vol. 134, 2021, 104458.

[9] Li, Hongfeng, et al. "Skin disease diagnosis with deep learning: A review." *Neurocomputing*, vol. 464, 2021, pp. 364–393.

[10] Zhu, Chen-Yu, et al. "A deep learning based framework for diagnosing multiple skin diseases in a clinical environment." *Frontiers in Medicine*, vol. 8, 2021, 626369.

[11] Maharana, Kiran, Surajit Mondal, and Bhushankumar Nemade. "A review: Data pre-processing and data augmentation techniques." *Global Transitions Proceedings* 3.1 (2022): 91-99.

[12] Mukti, Ishrat Zahan, and Dipayan Biswas. "Transfer learning based plant diseases detection using ResNet50." *2019 4th International conference on electrical information and communication technology (EICT)*. IEEE, 2019.

[13] Huang, Gao, et al. "Densely Connected Convolutional Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261-2269.

[14] Simonyan, Karen, and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *International Conference on Learning Representations (ICLR)*, 2015.

[15] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[16] Al-Kababji, Ayman, Faycal Bensaali, and Sarada Prasad Dakua. "Scheduling techniques for liver segmentation: Reducelronplateau vs onecyclelr." *International Conference on Intelligent Systems and Pattern Recognition*. Cham: Springer International Publishing, 2022.