Group Project Development Final Report

Skin lesions Classification using Multi-model Deep Learning on the HAM10000 dataset

Members

Đặng Văn Tuấn	*BA11-096*
Trần Đăng An	*BI12-004*
Nguyễn Anh Duy	*BI12-127*
Đoàn Hữu Thành	*BI12-418*
Vũ Đức Thành	*BI12-419*

Supervisors

Nguyen Tu Anh
Dr. Nghiem Thi Phuong

White Neuro Co.LTd
ICT Labs
January, 2024

Abstract

The increasing prevalence and complexity of skin diseases pose significant challenges in healthcare, making the development of artificial intelligence (AI) tools for skin disease diagnosis a pressing need. With the advent of the 4th industrial revolution and the application of AI in healthcare, there is potential to improve the quality of diagnosis and overcome limitations in resources and geographical barriers. This research aims to explore the integration of AI and machine learning techniques to support dermatological disease diagnosis. The study focuses on analyzing skin lesion imaging data and employing deep learning techniques, including feature extraction, network refinement, and model optimization. Specifically, the research addresses the classification of psoriasis, atopic dermatitis, and skin cancer. Various techniques, such as image characterization, deep learning network refinement, and parameter optimization, are employed to enhance the accuracy and efficiency of the diagnostic process. The proposed methods are evaluated on both global open databases and skin lesion image databases. By harnessing the power of AI and machine learning, this research contributes to the advancement of dermatological disease diagnosis and paves the way for improved healthcare outcomes in the context of the 4th industrial revolution.

Acknowledgments

I would like to extend my sincere appreciation to my mentor, Nguyen Tu Anh, for their invaluable guidance and mentorship throughout the development of this paper on skin lesions. Their expertise, support, and dedication have played a crucial role in shaping the content and direction of this work.

Nguyen Tu Anh has consistently provided me with valuable insights, constructive feedback, and encouragement, which have greatly enhanced the quality of this research. Their deep understanding of the subject matter and their willingness to share their knowledge and expertise have been instrumental in my growth and development as a researcher.

I am truly grateful for Nguyen Tu Anh's patience, expertise, and mentorship. Their commitment to my success and their willingness to invest their time and effort in guiding me through this project have been invaluable. Their mentorship has not only strengthened my understanding of skin lesions but has also inspired me to pursue further research in this field.

I would like to express my deepest gratitude to Nguyen Tu Anh for their unwavering support, encouragement, and belief in my abilities. Their mentorship has been a source of motivation and inspiration, and I am fortunate to have had the opportunity to learn from their vast experience and knowledge.

Once again, I would like to thank Nguyen Tu Anh for their significant contributions to this paper and for their ongoing support throughout my academic and professional journey. Their mentorship has been instrumental in my growth and success, and I am truly grateful for their guidance and mentorship.

Table of contents

Abstract....1
Acknowledgments....2
Motivation....4
I. DATASET....5
1. Overview....5
2. Component....5
II. PROJECT'S OVERVIEW....6
1. Introduction....6
2. Exploit characteristics....7
3. Build models and refine....7
4. Review....7
III. APPLIED DEEP LEARNING MODELS DETAILS....7
1. ResNet-152....8
2. SE-ResNeXt-101....10
IV. TECHNIQUES TO IMPROVE MODEL PERFORMANCE....13
1. Dropout function....13
2. Image Augmentation....13
3. Batch Normalization....14
4. Fine-tuning model....15
V. TESTING AND EVALUATION....16
1. Related work....16
2. Our work and comparison....17
3. Discussion....19
VI. OUR PROJECT'S DRAWBACK....21
VII. BIBLIOGRAPHY....22

Motivation

The skin is the largest part of the body with a large environmental contact surface, so the number of patients suffering from skin diseases has been increasing in recent years and is very diverse and complex in terms of disease types. According to a 2006 WHO report, skin disease is the fourth leading disease in the global burden of disease, with billions suffering from it [1].

With the strong development of the 4.0 revolution, artificial intelligence has been applied very effectively in healthcare around the world as well as in Vietnam. Especially in the context of the current epidemic situation, going to the doctor with many difficulties is sometimes impossible. Many patients sent photos (taken with personal phones) to doctors for advice on diagnosis and treatment without directly seeing a doctor for a medical examination. Therefore, the research and application of the achievements of artificial intelligence (AI) technology to build tools to support doctors in improving the quality of skin disease diagnosis is an urgent requirement, in line with the development trend of the 4th industrial revolution. The combination of AI and the intelligence and experience of good experts will be an effective tool in diagnosing diseases, and solving many limitations in human resources, and difficulties in economic and social geography.

In AI, in addition to the importance of data, machine learning techniques are an integral element in any successful Artificial Intelligence project. A good machine-learning technique will help us get the best out of the data, helping machine learning-to be as effective as possible. The more difficult the problems using AI, the more important the role of data and engineering. In particular, the problem of classifying dermatological diseases is a difficult problem because of their small size, symptomatic diseases, and diverse and complex manifestations, especially when there are underlying diseases that have been treated.

This research focuses on analytical, deep learning/machine learning techniques. Technical materials automatically extract features from skin lesion imaging data, thinning techniques, and deep learning network refinement, including

- Extraction techniques that characterize images of skin lesions of psoriasis, atopic dermatitis, and skin cancer.
- Techniques to improve the quality of deep learning network models such as sparseness, and stratification
- Deep learning network fine-tuning techniques such as changing, optimizing parameters, and connection diagrams
- Testing and evaluating techniques on open databases in the world and skin lesion image databases

I. DATASET

1. Overview

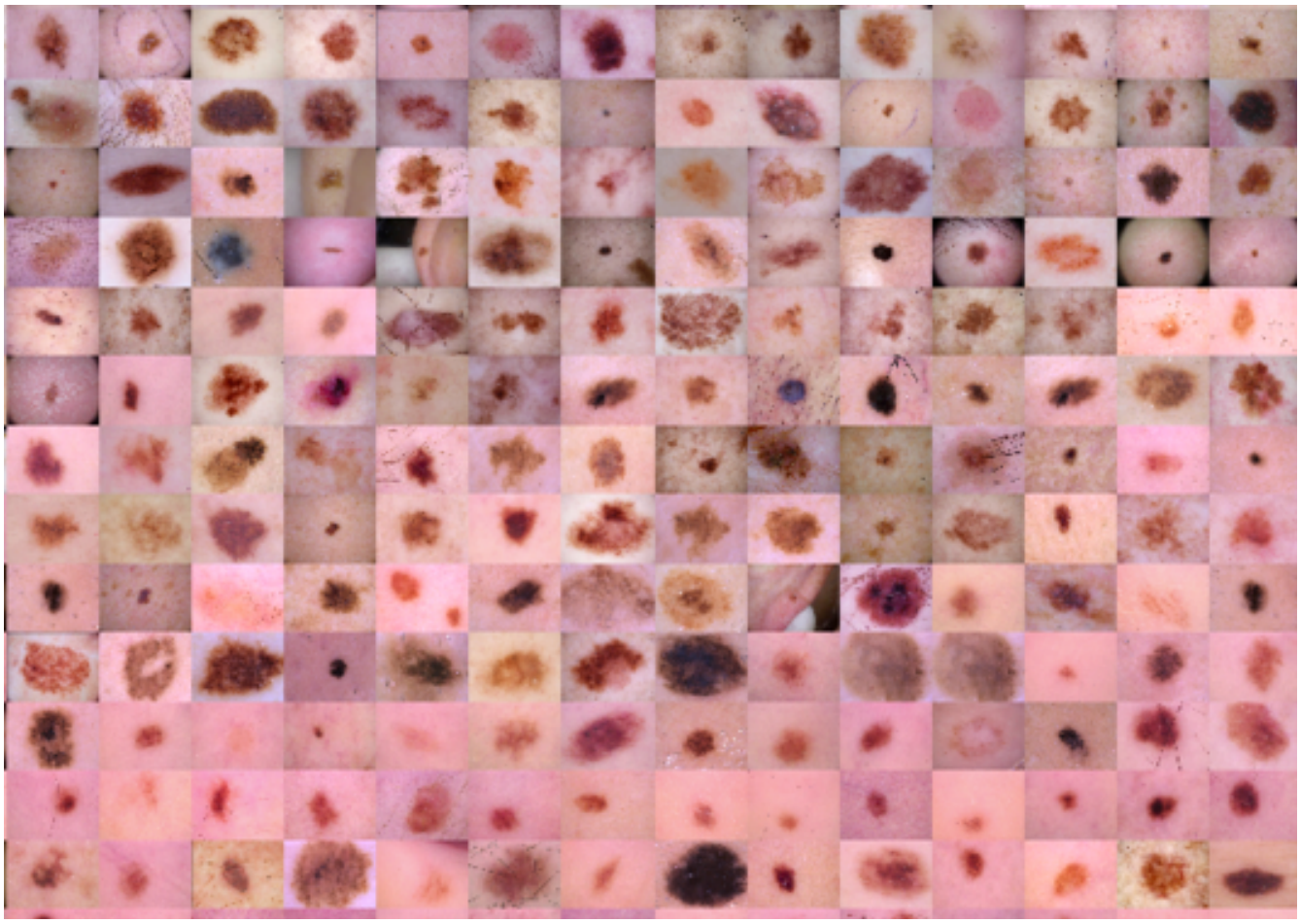

***Fig**. A few HAM10000 pictures*

HAM10000 (Human Against Machine with 10000 training images) is a dataset of about 10,000 close-up, high-quality dermatological images of skin lesions for research on Artificial Intelligence for the problem of identifying dermatological diseases. The dataset was collected from the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer clinic of Cliff Rosendahl in Queensland, Australia. The researchers used automation, machine learning, and manual evaluation to produce high-quality, consistent colors, middle lesions, and free-to-use. Specifics of the data collection process are found in their paper [2].

2. Component

The dataset consists of 7 disease labels denoted as follows:

- ak: Actinic Keratoses (Solar Keratoses) - Light Keratosis
- bcc: Basal cell carcinoma - Basal cell carcinoma
- bkl: Benign keratosis - Benign keratosis
- df: Dermatofibroma - Skin fibroids

- nv: Melanocytic nevus - Melanocyte microscope
- mel: Melanoma – Melanoma
- vasc: Vascular skin lesions

The images in the dataset were HAM10000 collected from patients presenting with skin-related symptoms and taken in high resolution. Each image is labeled with the corresponding type of skin pathology.

According to the author of HAM10000 [2], more than 50% of lesions are confirmed through histopathology, the actual label of the remaining cases is examination, monitoring (follow-up), expert consent (consensus), or in-vivo microscopy (confocal) identification. The data also includes lesions with multiple images (an area of damage taken with different images), which can be tracked with the accompanying metadata file. Not only can the lesion be observed and the corresponding disease label, but the metadata file also provides information about the method of identifying the disease (histo, follow-up, ...), the patient's gender, age, and the location of diseased skin.

The HAM10000 dataset has been used in many artificial intelligence and deep learning studies, aiming to develop machine learning models to detect and classify skin diseases. These models can be used to aid in the diagnosis and treatment of skin diseases. Google Scholar [1] contains about 2.5K results when searching for the keyword "HAM10000 dataset". The studies include:

- Karar Ali et al. [3] proposed a convolutional neural network model for predicting dermatological disease using HAM10000 datasets to evaluate the model.
- Garg et al. [4] proposed a convolutional network model for skin disease diagnosis with the HAM10000 dataset to train and evaluate the effectiveness of the model.
- Faes et al. [5] also proposed a convolutional neural network model for skin disease diagnosis using this dataset to evaluate the effectiveness of the model.

HAM10000 is offered to participants of the ISIC-2018 classification challenge[6] hosted by the annual MICCAI conference in Granada, Spain. In 2019, the ISIC-2019 challenge kicked off with a usage dataset that already included ISIC-2018 and added more new data. Information about it will be presented in a later section. In short, HAM10000 is the most common, foundational dataset for dermatology machine learning problems.

II. PROJECT'S OVERVIEW

1. Introduction

This section presents a technical overview of the research conducted to develop techniques for extracting characteristics from skin lesion images data and network refinement after model building. The study was conducted on the dataset: HAM10000. The goal of this research is to improve the accuracy and efficiency of skin lesion analysis and diagnosis using advanced machine-learning techniques across multiple datasets. The study began by collecting and preprocessing skin lesion image data from the HAM10000 dataset. Each dataset contains a large number of images obtained from a variety of sources, including dermatology clinics and online archives. The images are carefully selected and annotated by professional dermatologists to ensure accurate, accurate labels for training and evaluation purposes. Preprocessing techniques, such as resizing, normalization, and noise reduction, are applied to normalize data and improve its quality.

[1]

2. Exploit characteristics

Feature extraction is an important step in identifying patterns of information and characteristics from images of skin lesions. Advanced techniques are used to automatically extract meaningful characteristics. Pre-trained convolutional neural networks (CNNs) on large-scale image datasets, such as ImageNet, are often used for this purpose. Pre-trained CNN networks serve as feature extractors, collecting relevant visual representations in a triggered form from intermediate layers. These features are then incorporated into subsequent classes or models for further analysis.

3. Build models and refine

After extracting and thinning the trait, deep learning models are built to classify skin lesions. Various models, built from deep neural networks (DNNs), are used for this purpose. Initially, a baseline model is trained using extracted characteristics and basic labels from each dataset. However, to optimize the performance of the model, fine-tuning techniques are applied. Refinement involves updating the parameters of a pre-trained network using a skin lesion dataset, allowing the model to adapt to the unique characteristics and nuances of each dataset, thereby improving its accuracy and reliability.

4. Review

The developed techniques are rigorously evaluated and validated using appropriate testing metrics and protocols. The dataset, namely HAM10000, is divided into training, validation, and testing subsets to evaluate the performance of the proposed methods. Common evaluation metrics such as accuracy, accuracy, reliability are used to quantify the model's performance on the dataset. Cross-validation techniques can also be used to ensure model generality across different subsets of the dataset.

In summary, this technical review highlights the research and development techniques used to extract characteristics from skin lesion imaging data and network fine-tuning after building models on the HAM10000 dataset. By using advanced deep learning methods and leveraging multiple datasets, the accuracy and efficiency of skin damage analysis can be significantly improved. The proposed techniques promise to assist dermatologists in accurately diagnosing skin lesions and assist in the early detection of skin diseases across different datasets.

III. APPLIED DEEP LEARNING MODELS DETAILS

We have used various advanced CNN architectures developed in recent years such as networks, boot networks, dense interconnect networks, and frameworks that support search architectures. To meet CNNs' relentless need for data, we used these pre-trained models on ImageNet, a large dataset of about 1.5 million nature scene images divided into 1000 classes. We refined these models on dermatology datasets to take advantage of the benefits of transfer learning. From the various CNN architectures tried for this task, we ended up choosing ResNet-152[16], SE-ResNeXt-101[17] and VGG16 [30] for better performance and the variety of different model architectures.

1. ResNet-152

In this section, the architecture of ResNet-152 is detailed.

The ResNet-152 architecture is a deep neural network designed for image analysis tasks, especially in the context of error diagnosis. The core component of ResNet-152 consists of stacked convolutional particles, allowing the network to efficiently extract deep nonlinear characteristics from the input image. In Figure 1, the convolutional process is achieved through the application of stacked convolutional particles. The input image undergoes a series of three-layer stacked convolutional transformations, producing tensor data with rich nonlinear characteristics. Unlike traditional neural networks that lose local features as depth increases due to pooled activity, ResNet-152 overcomes this limitation by using residual connections. These connections allow adding the input features of a certain layer to the output features, preserving important local information. It is important to note that the input and output features must be shaped the same for this shortening of identification to occur.

ResNet-152 is an ultra-deep network architecture, consisting of 152 convolution layers. Each convolutional layer performs three nonlinear transformations, thanks to the embedded three-layer stacked convolutional nucleus. Between each layer, average aggregation is applied to increase the number of feature channels while reducing the feature size. This process helps capture and retain relevant information across the network. To prevent over-equipping, the 0.5 ratio Dropout function is combined between each convolution layer. This function randomly removes neurons, effectively reducing their influence on the output of the network. A dropout rate of 0.5 shows that, on average, half of neurons are randomly removed during training. Finally, a fully connected layer will be connected to the network, with the number of output nodes corresponding to the number of categories in the classification task. This class synthesizes the characteristics extracted from the convolutional layers and produces the final classification result.

In Figure, the overall architecture of ResNet-152 is depicted, showing the flow of information over the network. The combination of stacked convolutional particles, redundant connections, and elimination regularization contribute to enhancing the network's ability to understand complex representations and achieve high accuracy in image analysis tasks. The ResNet-152 architecture demonstrates the effectiveness of deep neural networks in capturing and leveraging rich nonlinear features for error diagnosis and other image analysis applications.

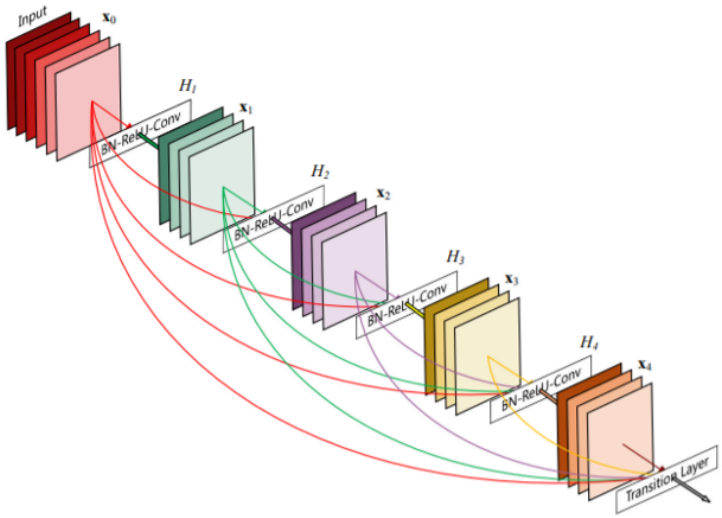

Fig. *The overall architecture of ResNet-152.*

Dynamics for the reconstruction of error characteristics and multiscale overlapping reception fields for insertion into the architecture of deep redundancy networks are introduced. The receiving field is the convolutional nucleus that implements the local perception of the corresponding input, the implementation of which is a weighted sum over a local region of the input. The size of the convolutional nucleus must be greater than 1 to have the effect of enhancing the cognitive field, such that the convolutional nucleus most commonly used for feature extraction cannot be 1. An even-sized convolutional kernel cannot ensure that the input feature map size and output feature map size do not change even if there is a symmetrically added buffer (for example, if the input is 4 × 4 and the convolutional particle size is 2 × 2 and the buffer on each side is 1, After sliding there will be a total of 5 outputs, which will not correspond to the input). Compared to larger convolutional nuclei, many layers are stacked on top of each other small convolutional particles have more activation functions, richer features, and greater lucidity. The convolutional activity comes with an activation function, and the use of more convolutional particles can make the decision function more discriminatory. Substituting a multilayer stacked convolutional nucleus for a large-sized convolutional nucleus involves parametric calculation. Table 1 shows a comparison between whether stacking is used or not for different types of networks or for the same type of network with different depths, including VGG-16, VGG-19, ResNet-50, and ResNet-152. As shown in Table 1, multilayer stacked convolutional nuclei have a greater number of parameters than large-sized convolutional nuclei and multi-layer convolutional nuclei, but the growth rate parameters are stable at less than 1%, the replacement of 7 × 7 convolutional nuclei with 3×3+3 × 3+3 ×3 helix nuclei in the VGG-19 network has a minimum parameter growth rate of 0.06%, and the replacement of 5 × 5 convolutions nuclei with 3 × 3 + 3 × 3 stacked helix nuclei in the ResNet-152 network have the largest parameter growth rate of 0.9%.

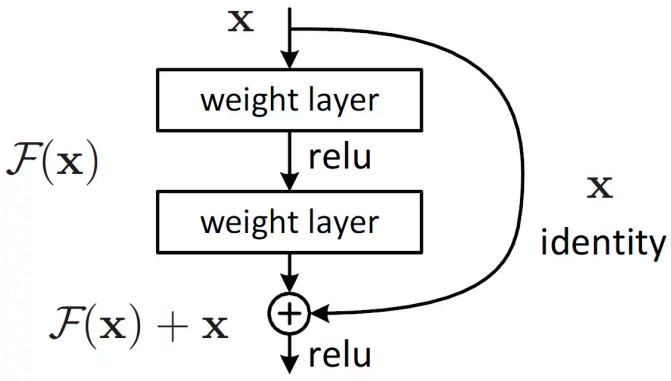

Fig. *The Residual Block architecture of ResNet-152.*

	$3 \times 3 + 3 \times 3$	5×5	$3 \times 3 + 3 \times 3 + 3 \times 3$	7×7
VGG-16	18,952,131	18,310,787	19,137,027	19,102,595
VGG-19	53,595,203	53,544,835	53,779,715	53,746,051
ResNet-50	762,691	714,703	1,234,093	1,197,084
ResNet-152	2,419,171	2,397,061	6,679,779	6,434,577

Table. *Compare the number of computational parameters of multi-layer stacked convolutional particles inserted into different networks.*

2. SE-ResNeXt-101

SE-ResNeXt-101 is a deep learning model that has demonstrated remarkable performance in various computer vision tasks, including image classification. It is an extension of the ResNeXt architecture, which combines the benefits of both residual connections and grouped convolutions. SE-ResNeXt-101 further introduces a mechanism called Squeeze-and-Excitation (SE) blocks, which enables the model to adaptively recalibrate channel-wise feature responses. This combination of ResNeXt architecture and SE blocks has proven to be highly effective in improving the representational power and discriminative capability of deep neural networks.

SE-ResNeXt-101 Model Architecture Summary :

- ResNeXt Architecture: SE-ResNeXt-101 builds upon the ResNeXt architecture, which is based on the residual network (ResNet) concept. ResNeXt introduces a cardinality parameter that controls the number of parallel paths or groups within each convolutional layer. This cardinality allows for increased diversity and complexity of feature representations.
- Squeeze-and-Excitation (SE) Blocks: SE-ResNeXt-101 incorporates SE blocks into its architecture. An SE block consists of two steps: squeeze and excitation. In the squeeze step, global average pooling is applied to the feature maps, resulting in a channel-wise descriptor. This descriptor captures the importance of each channel in terms of its contribution to the overall feature representation. In the excitation step, a fully connected network with a sigmoid activation function is used to model channel-wise dependencies and generate a set of channel-wise scaling factors.
- Channel-Wise Recalibration: The SE blocks in SE-ResNeXt-101 enable the model to adaptively recalibrate the channel-wise feature responses. This recalibration allows the network to assign different weights or importance to different channels, focusing on the most informative and discriminative features. By emphasizing important channels and suppressing less relevant ones, SE-ResNeXt-101 enhances the discriminative power of the network and improves its representation learning capability.
- Training and Inference: SE-ResNeXt-101 is trained using large-scale image datasets with supervised learning techniques, such as stochastic gradient descent (SGD) or adaptive optimization algorithms like Adam. During inference, the model takes an input image and passes it through a series of convolutional layers, SE blocks, and pooling operations to extract hierarchical and context-aware features. The final prediction is obtained using a classifier, such as a fully connected layer or a softmax layer, which maps the learned features to the corresponding class probabilities.
- Performance and Applications: SE-ResNeXt-101 has achieved state-of-the-art performance on various image classification benchmarks, demonstrating its effectiveness in capturing intricate

features and improving the overall accuracy of deep learning models. It has been widely used in applications such as object recognition, scene understanding, and medical image analysis, where precise and robust classification is essential.

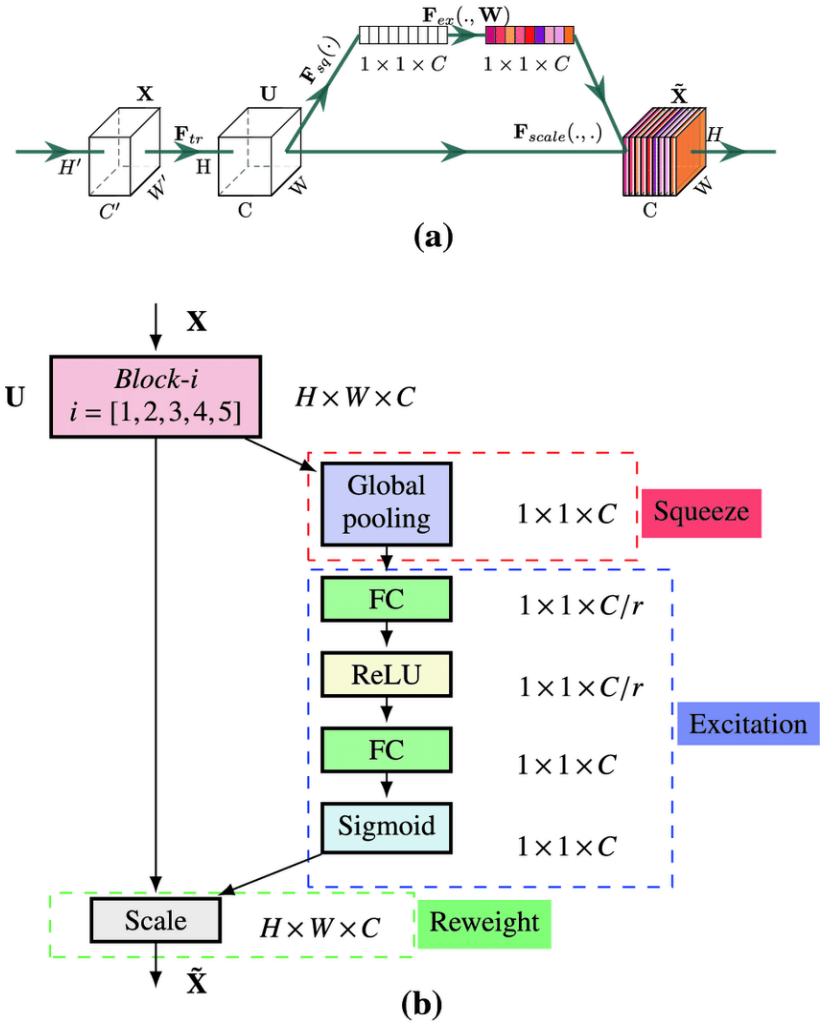

***Fig**. The model's architecture of* SE-ResNeXt-101

The Squeeze-and-Excitation (SE) block consists of two key steps: squeeze and excitation. In the squeeze step, global average pooling is applied to the feature maps, which results in a channel-wise descriptor. Global average pooling computes the average of each channel across spatial locations, reducing the spatial dimensions to 1x1. The resulting descriptor captures the importance of each channel in terms of its contribution to the overall feature representation. In the excitation step, the channel-wise descriptor is fed into a fully connected network, typically comprising one or more fully connected layers followed by non-linear activation functions. This network aims to model channel-wise dependencies and generate a set of channel-wise scaling factors. These scaling factors represent the importance or relevance of each channel and are used to recalibrate the feature responses. The recalibration process allows the network to assign different weights or importance to different channels based on their importance for the task at hand. By emphasizing important channels and suppressing less relevant ones, the SE block enhances the discriminative power of the network. This adaptability enables the model to focus on the most informative and discriminative features, enhancing its representation learning capability.

The SE-ResNeXt-101 deep learning model, with its combination of ResNeXt architecture and SE blocks, represents a significant advancement in the field of computer vision. Its ability to adaptively recalibrate channel-wise feature responses has demonstrated superior performance in image classification tasks, making it a valuable tool for researchers and practitioners in the scientific community.

3. VGG16

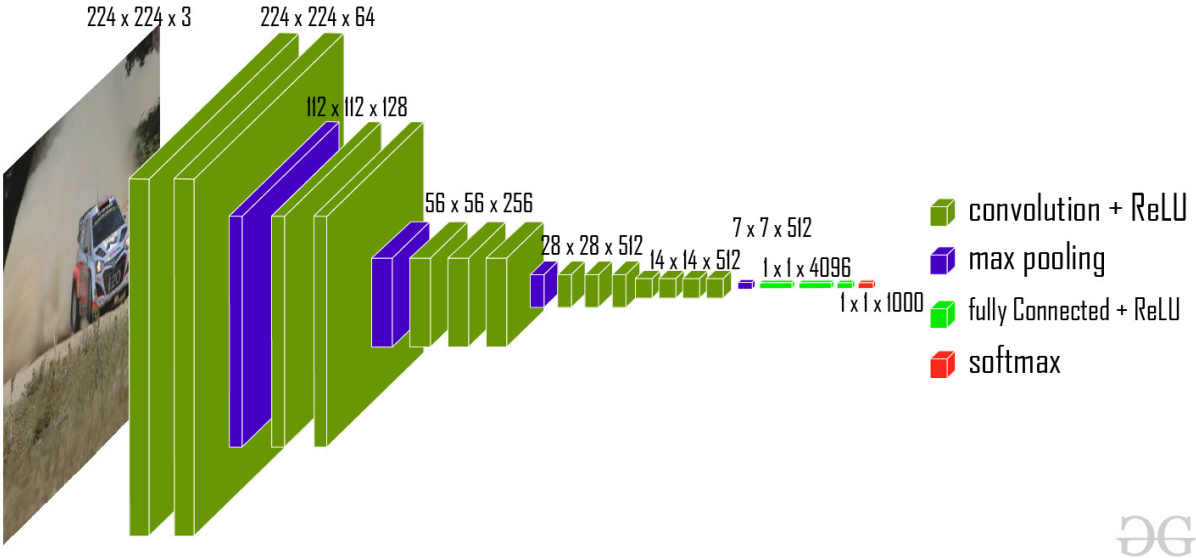

***Fig**. Network structure diagram of VGG16*

VGG16 is a convolutional neural network model developed by the Visual Geometry Group (VGG) of the University of Oxford and the winner of the 2014 ILSVRC object identification algorithm23. The critical work of VGG16 is to demonstrate that extending the depth of the network can improve the performance of the network in certain situations. Compared with the classic AlexNet, VGG16's improvement lies in the use of multiple 3×3 convolution cores to replace the larger convolution cores (11×11, 7×7, 5×5), which can broaden the depth of the network to improve the network performance effectively, and the use of smaller convolution cores can also reduce the number of network parameters. The VGG16 network model comprises 13 convolutional layers, three fully connected layers and five pooling layers.

VGG16 continues the characteristics of the classical network's simple structure, expands the network's depth through the flexible use of 3 × 3 convolution, and successfully improves network performance. However, the VGG16 model also has some drawbacks in the application. First, the fully connected layer has many parameters, which occupy much memory and consume many computing resources, making the VGG16 model encounter obstacles in the front-end deployment. Secondly, the network model structure is single, and its performance is weak compared with some sophisticated advanced networks. Moreover, VGG16 lacks an effective method to prevent gradients' disappearance and problems such as slow convergence speed and gradient explosion are likely to occur in the model's training. Aiming at the defects of the VGG16, this paper improved the VGG16 model by drawing on advanced network models such as ResNet, SqueezeNet, and DenseNet. The improved VGG16 model consists of 14 convolutional layers, five BN layers, six pooling layers, and one fully connected layer.

The network structures of VGG16 and Improved VGG16 are shown in Fig.

The main improvement methods of the improved VGG16 model are as follows. Remove the F6 and F7 fully connected layers of VGG16. Add Conv6 and Global Average Pooling Layer(GAP). The two fully connected layers FC6 and FC7, in the VGG16, willfully connect each neuron with all the neurons in the previous layer, thus generating a considerable number of parameters and occupying many computing resources. Therefore, these two fully connected layers need to be discarded. GAP is a new idea proposed by M. Lin et al. (2014), which can replace the fully connected layer, and it has been proved by experiments that GAP can reduce the number of parameters, the amount of calculation, and the amount of overfitting in the model24. GAP can calculate the mean value of the pixel points in each feature map, output a feature point, fuse these feature points into feature vectors, and input them to the Sofmax layer, thus reducing the number of parameters, the amount of calculation, and the over-fitting. Besides, GAP can output a feature graph for each category, which directly endows features with real meaning and connects each category and feature graph more intuitively. As more and more researchers have confirmed GAP's function, many advanced network models, such as GoogLeNet, ResNet, SqueezeNet, and DenseNet, have introduced a GAP. SqueezeNet25 added a convolutional layer with a convolution kernel size of 1×1 before the GAP to balance input and output channel size. This operation again reduced the number of parameters and computations in the model and significantly accelerated the speed. Therefore, in this paper, a convolutional layer Conv6 with a convolution kernel size of 1×1 was placed in front of the added GAP to optimize the model further. The number of filters on the model performance was analyzed by setting various filters (128/256/512) during model construction. [30]

IV. TECHNIQUES TO IMPROVE MODEL PERFORMANCE

Model quality improvement techniques such as dropout function, images augmentations, batch normalization and model fine-tuning are introduced in this section.

1. Dropout function

In recent years, deep learning has emerged as a powerful paradigm in the field of artificial intelligence, revolutionizing various fields such as computer vision, natural language processing, and speech recognition. However, deep neural networks tend to be over-learned, making them unable to generalize well to unseen data. To address this challenge, the dropout function has attracted considerable attention and has become an integral component of deep learning architecture. In this article, we provide a comprehensive review of the dropout function, its fundamentals, and its diverse applications in deep learning. We delve into the theoretical basis of dropouts and clarify its role in preventing overarming by randomly eliminating units during both the training and inference phases. Furthermore, we discuss different strategies for incorporating dropout into different types of neural network architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer models. We highlight the benefits of dropouts in improving model generalization, enhancing regularization, reducing model complexity, and minimizing the impact of input noise.

2. Image Augmentation

Image augmentation is a widely used technique in the field of computer vision and image processing, which plays a crucial role in enhancing the performance of machine learning models. It involves applying a variety of transformations to existing images in order to create new, synthetic training data. The augmented images retain the same semantic content as the original ones, but possess variations in terms of appearance, orientation, scale, and other properties. This technique has gained significant popularity in scientific research, particularly in the domain of deep learning, for its ability to alleviate issues such as overfitting, data scarcity, and generalization challenges.

Image Augmentation Techniques:

- Rotation: Rotating an image by a certain degree to simulate different viewing angles and orientations.
- Translation: Shifting an image horizontally or vertically to simulate changes in position or perspective.
- Scaling: Resizing an image to a larger or smaller size, replicating the effect of zooming in or out.
- Flipping: Mirroring an image horizontally or vertically, allowing models to learn from both orientations.
- Noise Injection: Adding random noise to an image to enhance its robustness against noise in real-world scenarios.
- Brightness and Contrast Adjustment: Modifying the brightness and contrast levels of an image to account for varying lighting conditions.
- Cropping: Selecting a region of interest from an image and discarding the rest, focusing the model's attention on specific features.
- Gaussian Blur: Applying a blurring effect to an image, simulating out-of-focus or hazy environments.
- Color Channel Shifting: Altering the intensity or balance of color channels in an image, enabling models to handle variations in color composition.
- Elastic Distortion: Deforming an image using elastic transformations, introducing local spatial variations to the data.

By applying these image augmentation techniques, researchers can significantly expand their training datasets and improve the generalization capability of deep learning models. These augmented images provide diversity and variability, enabling models to learn robust and invariant representations, ultimately enhancing the overall performance and reliability of scientific studies in computer vision and image analysis.

3. Batch Normalization

Batch normalization is a widely used technique in deep learning that has significantly contributed to the success of convolutional neural networks (CNNs) in various computer vision tasks, including image classification. It aims to address the internal covariate shift problem by normalizing the activations of each layer across a mini-batch of training examples. By doing so, batch normalization helps to stabilize and accelerate the training process, leading to improved convergence and enhanced performance of deep learning models.

Batch Normalization Technique:

- Normalization: Batch normalization applies a normalization step to the activations of each layer in a neural network. It calculates the mean and standard deviation of the activations within a mini-batch and normalizes the activations to have zero mean and unit variance. This normalization step helps in reducing the internal covariate shift, which refers to the change in the distribution of layer inputs during training.
- Scale and Shift: After normalization, batch normalization introduces two learnable parameters, scale and shift, for each normalized activation. These parameters allow the model to learn the optimal scale and shift factors, effectively enabling the network to adapt and preserve the representational capacity of the original network.
- Training and Inference: During training, batch normalization computes the mean and standard deviation of the mini-batch and uses them to normalize the activations. The scale and shift parameters are updated through backpropagation. However, during inference, the mean and standard deviation of the entire training set are used instead to normalize the activations, ensuring consistency between training and inference.
- Benefits and Impact: Batch normalization offers several benefits in deep learning. It accelerates training by reducing the dependence of gradients on the scale of the parameters, allowing for larger learning rates. It also mitigates the vanishing or exploding gradient problem, making it easier to train deeper networks. Additionally, batch normalization acts as a form of normalization between layers, making the network more robust to changes in input distribution and improving generalization performance.

By incorporating batch normalization into deep learning models, researchers have achieved notable improvements in various image classification tasks. Its ability to address the internal covariate shift and stabilize the training process has made it a fundamental technique in the field of deep learning, contributing to the advancement of computer vision research.

4. Fine-tuning model

Model refinement refers to the process of modifying and adapting a pre-trained deep learning model to a new task or domain. Refinement involves using a pre-trained model, which has been trained on a large dataset and further trained on a dataset dedicated to the smaller task.

Here's an overview of the concept of fine-tuning change and how to use it in deep learning:

- Pre-trained models: Deep learning models, especially large-scale models such as Transformative Neural Networks (CNNs) or Transformer models, are often pre-trained on large datasets, such as ImageNet or large text sets. These pre-trained models learn common features that can be transferred to different tasks or domains.
- Transferable learning: Refinement promotes transferable learning, in which knowledge gained from previous training is transferred to a new task. Instead of training deep learning models from scratch on a small dataset, pre-trained models are used as starting points and then refined to tailor them to a specific task or domain.
- Task-specific adaptation: During fine-tuning, the weights of the pre-trained model will be updated using a smaller task-specific dataset. The final classes or a subset of classes in the model are often replaced or modified to fit the target task. The model is then

trained on the new dataset and the weights are updated to optimize performance for the specific target task.

- Benefits of Refinement: Changing fine-tuning offers several benefits in deep learning. It allows deep models to be trained efficiently even when the task-specific dataset is limited, as it leverages the knowledge learned from the pre-training stage. Refinement also leads to faster convergence and better generalization, as the pre-trained model that has learned common features can transfer to the new task.
- Application Tuning: Tuning is widely used across many different areas of deep learning. In computer vision, pre-trained CNN models are often fine-tuned for tasks such as image classification, object detection, or semantic segmentation. In natural language processing, pre-trained Transformer models such as BERT or GPT are fine-tuned for tasks such as sentiment analysis, answering questions, or creating text. Refinements can also be applied to other areas such as speech recognition, recommendation systems, and reinforcement learning.

Refinement change provides a practical and effective way to apply deep learning to new tasks or areas by leveraging pre-trained models. It enables efficient use of limited data, faster convergence, and improved performance of the target task.

V. TESTING AND EVALUATION

1. Related work

Automatic diagnosis of skin diseases has received a lot of attention from researchers for a long time. However, most of these studies are limited to binary or tertiary classification even when a large number of classes are available. The importance of early detection of melanoma is understandable because of the increasing risk it poses to patient survival with each passing day. However, there are thousands of other skin diseases that may not be as fatal as melanoma but have a tremendous impact on a patient's quality of life. DL is extremely capable of taking on hundreds of classes simultaneously, as evidenced by our results. We believe this is the right time to harness the full potential of Deep Learning and begin conducting truly impactful research that can truly translate into an industry-standard solution for automated skin disease diagnosis on a larger scale. These solutions can have far-reaching societal impact by not only helping dermatologists make diagnoses in clinical settings but also providing economical and effective initial screening for poor patients in both developed and developing countries. Another consideration about the application of DL in dermatology is that many researchers use private or public datasets with their own choice of training/test division (although taken at random) and number of layers. For this reason, there is little in common and sometimes nothing in common to compare different classification methods – as noted by Brinker et al. [20]. This incomparable problem can be solved by collecting and maintaining a large standardized dataset that is publicly available with clearly specified training/test separations and standard performance metrics for benchmarking. Although some public datasets, such as the HAM10000 dataset, do offer this training/pre-test separation, their size is usually small and tasks are usually limited to binary or tertiary classification. Any study of such small datasets cannot be reliably generalized, and although the results can be published, they cannot be used as a stepping stone for practical applications of AI in real-world diagnostics. On the other hand, large public datasets often have a lot of noise, low-resolution images, or blurred stamps. Important useful information needed for a detailed classification of seemingly similar

diseases will be lost in such low-resolution images or watermarks. In addition, non-visual metadata, such as medical history, is generally not available in medical imaging datasets. However, this additional information may be key to an accurate and definitive diagnosis. We were able to use a disease classification approach for the HAM10000 dataset and improve our outcomes by 2.5% (refer to Table A1). If multi-model datasets are managed and publicly available, AI can certainly leverage additional information to improve its classification performance. While understanding and interpreting the results of any AI-based classification, it is important to recognize that accuracy or even sensitivity and specificity may not describe the complete picture of a model's performance. That's why the Performance Characteristics Curve (ROC) of the sub-receive zone (AUC) is also reported along with other performance metrics. From an AI perspective, we can argue that achieving an average sensitivity of about 80% with an average false positive rate of 1.6% (Table 3, Exp-2) for the 23-ary classification task using low-resolution datasets and low-resolution watermarks is a difficult one. reasonable achievement. However, the actual performance of any AI-based classifier can vary significantly in an actual clinical setting as Navarrete-Dechent et al noted. [22]. They found that the classifier developed by Han et al. [21] did not generalize well when presenting data from the archive on demographics differently than the data used to train the classifier. For a dermatologist, it's definitely a cause for concern. However, Han et al. advocated in their response [21] that a classifier should not be evaluated on the basis of sensitivity and specificity alone. ROC curves show the classifier's true ability to perform below multiple operating points or thresholds while making diagnostic predictions for a given image. Changing this threshold from 0 to 1 on the model's output can change the balance between sensitivity and specificity while yielding different accuracy. Thus, a higher AUC value ensures that the model is capable of accurately predicting a given disease, e.g. melanoma, with minimal chance of classifying any other disease as that particular disorder.

Model	Top-1 Accuracy (%)	Top-5 Accuracy (%)	AUC (%)
Resnet-152	60.82 ± 0.51	82.16 ± 0.43	98.50 ± 0.10
Densenet-161	63.51 ± 0.68	84.46 ± 0.46	98.49 ± 0.06
SE_ResNeXt-101	64.03 ± 0.77	84.26 ± 0.66	98.48 ± 0.08
NASNet	60.69 ± 0.72	81.09 ± 0.61	97.90 ± 0.03
Ensemble	66.74 ± 0.64	86.26 ± 0.54	98.77 ± 0.07

Table. *Detailed results of classification for individual classifiers and their aggregators on HAM10000.*

2. Our work and comparison

The SE-ResNeXt-101 model has a runtime of 4 hours and 30 minutes, which is equivalent to 270 minutes. Average runtime of 5.4 minutes per epoch. This indicates that this model has a relatively long average runtime per epoch.

The ResNet-152 and VGG16 models have a runtime of 2 hours and 30 minutes, which is equivalent to 150 minutes. An average runtime of 3 minutes per epoch. Compared to SE-ResNeXt-101, these two models have a shorter average runtime per epoch.

Model	Top-1 Accuracy (%)	Top-2 Accuracy (%)
Resnet-152	73.07 ± 0.02	71.69 ± 0.10
SE-ResNeXt-101	84.24 ± 0.18	82.50 ± 0.14
Vgg16	68.20 ± 0.22	68.70 ± 0.25

Table *. Detailed results of 7-ary classification for individual classifiers on HAM10000.*

1. ResNet-152:

- Top-1 Accuracy: 73.07%
- Top-2 Accuracy: 71.69%
- Validation accuracy: 73.07%
- Test accuracy: 71.69%
- Validation loss: 0.814018
- Test loss: 0.798522

2. SE-ResNeXt-101:

- Top-1 Accuracy: 84.24%
- Top-2 Accuracy: 82.50%
- Validation accuracy: 78.93%
- Test accuracy: 82.50%
- Validation loss: 0.6020
- Test loss: 0.4753

3. VGG16:

- Top-1 Accuracy: 68.20%
- Top-2 Accuracy: 68.70%
- Validation accuracy: 68.70%
- Test accuracy: 68.20%
- Validation loss: 0.942204
- Test loss: 0.961099

Based on these metrics, we can make the following observations:

- SE-ResNeXt-101 has the highest Top-1 and Top-2 accuracies and the lowest losses on both the validation and test sets. It has the best overall performance among the three models compared.
- ResNet-152 shows relatively stable accuracy and losses on both the validation and test sets, but its performance is lower compared to SE-ResNeXt-101.

VGG16 has the lowest accuracy and highest loss among the three models, indicating lower prediction capabilities compared to the other two models

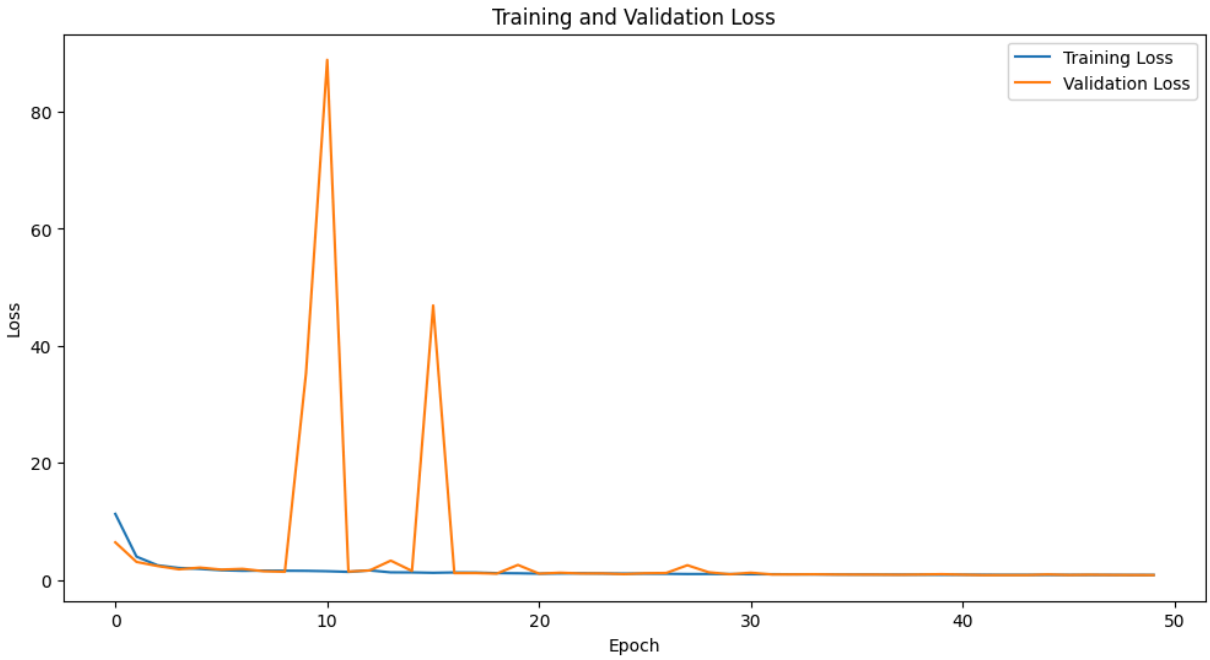

Fig. *The validation loss shows irregular fluctuations initially but gradually decreases towards the end, while the training loss continuously decreases throughout the training process. This pattern suggests the following observations:*

1. Early overfitting: The initial fluctuations in the validation loss may indicate early signs of overfitting. During the early stages of training, the model may learn specific characteristics of the training data, resulting in a continuous decrease in the training loss. However, when applied to the validation data, the model's performance may not generalize well, leading to an increase in the validation loss. This can happen when the model is too complex or the training data does not sufficiently represent real-world data.
2. Convergence: The gradual decrease in the validation loss suggests model convergence. This indicates that the model has learned and improved its predictive ability on the validation data. Despite the initial fluctuations, as the model continues to learn and adjust, it approaches a minimum loss value.
3. Performance evaluation: Despite the initial fluctuations, if the training loss consistently decreases and the validation loss gradually decreases towards the end, it indicates the positive performance of the model. Your model demonstrates the ability to learn from the training data and generalize well when applied to the validation data.

In conclusion, the observed pattern in the training and validation losses suggests that your model initially experiences overfitting but gradually converges and performs well on the validation data.

3. Discussion

Although this study used three different models, including ResNet-152, SE-ResNeXt-101, and VGG16, to classify skin inflammation images from the HAM10000 dataset, there are still some limitations that need to be considered to ensure the objectivity and accurate evaluation of our results.

- Sample size: The HAM10000 dataset is a large dataset consisting of 10,000 skin inflammation images from seven different disease types. However, using a fixed sample may lead to some limitations. This dataset may not fully reflect the diversity and variations of skin inflammation in reality. Therefore, the application of the results of this study needs to be carefully considered, especially for rare cases or cases that are not within the scope of the dataset.
- Reliability of diagnosis: In this study, the diagnosis of skin inflammation is determined based on the image classification process using three models: ResNet-152, SE-ResNeXt-101, and VGG16. Although all three models achieved relatively high accuracy, the diagnosis of skin inflammation can still have limitations. There may be cases where the classification process cannot accurately differentiate between similar disease types or there may be errors in classifying specific cases. These limitations can affect the reliability of the diagnosis results and the practical application of the method.
- Modeling and algorithm limitations: Although this study used three different models, all three have their own limitations. These models include ResNet-152, SE-ResNeXt-101, and VGG16, and each model has its own characteristics and limitations. The selection and tuning of hyperparameters, performance evaluation, and data processing can also affect the accuracy and stability of the results. The limitations in modeling can affect the accuracy and stability of the results and should be considered when applying this method in practice.

Potential development directions:

Although this study has achieved some notable results in classifying skin inflammation images using three models, ResNet-152, SE-ResNeXt-101, and VGG16, there are still many potential development directions to further research and improve our results. Below are some important development directions that we consider:

- Expand the dataset: Although the HAM10000 dataset provided a good platform for research, expanding the dataset by adding images from multiple sources and different hospitals can help increase the representativeness and diversity of skin inflammation. This will help improve the diagnostic ability and practical application of the method.
- Model optimization: Although the three models have shown relatively good results, fine-tuning and optimizing the hyperparameters and model structures can improve their performance. Further research can focus on exploring more powerful optimization methods to achieve higher accuracy and better generalization ability.
- Integration with other methods: The research can explore the possibility of combining deep learning models with other methods such as image segmentation or attribute classification to improve the diagnosis of skin inflammation. Combining information from multiple sources and methods can increase accuracy and provide additional information about the pathology and attributes of the skin.
- Building real-world applications: Finally, an important potential development direction is to build real-world applications based on the results of this study. Developing mobile applications or online tools can provide diagnostic support for doctors and end-users while collecting real-world data to improve the model and evaluate its performance in a real-world environment.

VI. OUR PROJECT'S DRAWBACK

Unfortunately, due to resource limitations, our project had to be conducted on Google Colab. This constrained our ability to delve deeper into exploring additional matrices to evaluate the model's performance. Regrettably, we could not push further in our endeavor to gain a more comprehensive understanding of the model's capabilities.

The restricted computational resources provided by Google Colab hindered our ability to extend our analysis beyond the initial scope. Consequently, we were unable to explore advanced evaluation methods and delve into intricate details that could have provided valuable insights into the model's strengths and weaknesses.

The confinement to Google Colab, while offering convenience and accessibility, presented significant limitations in terms of computational power and memory. As a result, we were unable to conduct extensive experiments and thoroughly investigate a wider range of evaluation matrices, which could have enriched our understanding of the model's behavior and performance across various dimensions.

Despite these constraints, we endeavored to make the most of the available resources and conducted a comprehensive analysis within the given limitations. However, it is important to acknowledge that our project's findings should be interpreted within the context of the constrained environment in which it was executed.

Moving forward, it would be beneficial to explore avenues that provide more extensive computational resources, allowing for a more thorough exploration of evaluation matrices and enabling a deeper understanding of the model's performance. This would provide a more robust and comprehensive assessment, paving the way for future advancements in our research endeavors.

VII. BIBLIOGRAPHY

[1]	PE and LeBoit*, World Health Organi-zation Classification of Tumours: Pathology and Gienetics of Skin Tumours,* IARCPress, 2006.
[2]	P. Tschandl, C. Rosendahl and a. H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data,* vol. 5, 2018.
[3]	K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using efficientnets – a first step towards preventing skin cancer," *Neuroscience Informatics,* vol. 2, 2022.
[4]	R. Garg, S. Maheshwari and A. Shukla, Decision Support System for Detection and Classification of Skin Cancer Using CNN, Singapore: Springer Singapore, 2021.
[5]	L. Faes, S. K. Wagner, D. J. Fu, X. Liu, E. Korot, J. R. Ledsam, T. Back, R. Chopra, N. Pontikos, C. Kern, G. Moraes, M. K. Schmid, D. Sim and K. Balaska, "Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study," *The Lancet Digital Health,* pp. e232-e242, 2019.
[6]	N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC),," 2019.
[7]	M. Combalia, N. C. F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, and J. Malvehy, "Bcn20000: Dermoscopic lesions in the wild," 2019.
[8]	N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," 2017.
[9]	M. A. Kassem, K. M. Hosny, and M. M. M. Fouad, "Skin lesions classification into eight classes for isic 2019 using deep convolutional neural network and transfer learning,," *IEEE Access,* vol. 8, 2020.
[10]	B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and a. M. H. Yap, "Analysis of the isic image datasets: Usage, benchmarks and recommendations,," *Medical Image Analysis,* vol. 75, 2022.
[11]	X. Y. a. W. L. A. Gong, "Dermoscopy image classification based on stylegans and decision fusion," *IEEE Access,* vol. 8, 2020.
[12]	T. Shanthi, R. Sabeenian, and R. Anand, "Automatic diagnosis of skin diseases using convolution neural network," *Microprocessors and Microsystems,* vol. 76, 2020.
[13]	M. N. Bajwa, K. Muta, M. I. Malik, S. A. Siddiqui, S. A. Braun, B. Homey, A. Dengel, and S. Ahmed, "Computer-aided diagnosis of skin diseases using deep neural networks," *Applied Sciences,* vol. 10, 2020.
[14]	T. A. Rimi, N. Sultana and M. F. A. Foysal, "Derm-nn: Skin diseases detection using convolutional neural network," *4th International Conference on Intelligent Computing and Control Systems (ICICCS),* 2020.

[15]	P. Tschand, C. Rosendahl and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data,* vol. 5, 2018.

[16]	Yu, H.; Miao, X.; Wang, H. Bearing Fault Reconstruction Diagnosis Method Based on ResNet-152 with Multi-Scale Stacked Receptive Field. Sensors 2022
[17]	Ramadan, M.K.; Youssif,A.A.A.; El-Behaidy, W.H. Detection and Classification of Human-Carrying Baggage Using DenseNet-161 and Fit One Cycle. Big Data Cogn. Comput. 2022
[18]	Jianxiang Dong. 2020. Focal Loss improves the Model Performance on Multi-Label Image Classification with Imbalanced Data. In proceedings of the 2nd International Conference on Industrial Control Network And System Engineering Research (ICNSER2020). Association for Computing Machinery, New York, NY, USA, 18–21.
[19]	L. G. Falconí, M. Pérez and W. G. Aguilar, "Transfer learning in Breast mammogram Abnormalities Classification with Mobilenet and Nasnet," *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Osijek, Croatia, 2019, pp. 109-114
[20]	Brinker, T.J.; Hekler, A.; Utikal, J.S.; Grabe, N.; Schadendorf, D.; Klode, J.; Berking, C.; Steeb, T.; Enk, A.H.; von Kalle, C. Skin cancer classification using convolutional neural networks: Systematic review. J. Med
[21]	Han, S.S.; Lim, W.; Kim, M.S.; Park, I.; Park, G.H.; Chang, S.E. Interpretation of the Outputs of a Deep Learning Model Trained with a Skin Cancer Dataset. J. Invest. Dermatol. 2018, 138, 2275
[22]	Navarrete-Dechent, C.; Dusza, S.W.; Liopyris, K.; Marghoob, A.A.; Halpern, A.C.; Marchetti, M.A. Automated dermatological diagnosis: Hype or reality? J. Invest. Dermatol. 2018, 138, 2277–2279.
[30]	A novel method for peanut variety identification and classification by Improved VGG16 HaoyanYang1 , Jiangong Ni2 , JiyueGao2 , Zhongzhi Han2* & Tao Luan1*