A Comparative Study on the Performance of Skin Disease Classification Based on Different Deep Residual Networks

Chengcheng Chen

*College of Arts & Sciences, American University, Washington, DC 20016, The United States*

*Email: cc4293a@student.American.edu*

Keywords: Deep learning; Dermatological image classification; Residual network; Convolutional neural networks.

Abstract: Billions of people are affected by skin illnesses, which are among the most prevalent diseases globally. The prognosis of patients depends on early diagnosis and treatment, yet conventional diagnostic techniques need human work. Convolutional neural networks (CNNs), have advanced medical image analysis significantly in recent years, especially in dermatological classification. Residual Network (ResNet), the residual learning-based CNN architecture, performs very well in image classification applications. In my study, classification experiments were conducted on a dermatological image dataset using three ResNet models with different depths: ResNet50, ResNet101, and ResNet152. I made a comparison of these models' performances. The experimental results show that ResNet50 performs the best in terms of classification accuracy (92%) and resource consumption for small and medium-sized datasets. ResNet101 and ResNet152 do not significantly outperform ResNet50 in terms of overall classification performance, although they perform well in some categories. Moreover, the training time for ResNet101 and ResNet152 is longer, and the resource consumption is higher, indicating that model depth is not always the sole determinant for improving classification performance. Future research can further improve classification performance by introducing larger datasets, utilizing more complex models (e.g., Vision Transformers), and incorporating advanced learning techniques.

# 1 Introduction

Skin disorders are one of the most common diseases in the world. The World Health Organization (WHO) states that the prevalence of dermatological disorders remains high globally and is diverse, including common types such as acne, skin cancer, and eczema (Dagne, 2018). The prognosis of patients with dermatological diseases greatly depends on prompt diagnosis and treatment, as these measures can stop the disease's progression, lower the risk of complications, and enhance the patient's general health (Thomsen, 2020). However, those in need may not be diagnosed in time, and recognizing skin conditions relies on the experience of the dermatologist and may be geographically limited. This reality highlights the need to identify skin diseases through innovative approaches (Hogarty, 2020). Deep learning approaches, particularly convolutional neural networks (CNNs), have achieved substantial advances in medical image processing, particularly in recognizing dermatological diseases (Li, 2021; Hogarty, 2020). CNNs are very suitable for the classification of skin disease images, among which Residual Networks (ResNet) based on residual blocks perform well in several image classification tasks. In 2015, ResNet indicated by He et al., solved the problem of gradient vanishing in the deep network by integrating residual connectivity, which significantly enhances deep network training (He, 2016). The ResNet model shows great potential in image classification tasks. The accuracy of the algorithm is comparable to that of a dermatologist, and the automated dermatology testing system can effectively help physicians improve diagnostic accuracy, which has the potential to be a skin cancer detection tool (Li, 2023).

Although deep learning algorithms have made tremendous advances in medical image categorization, the differences in performance and resource consumption of models with different depths still need to be further explored. Especially for small and medium-sized datasets, ResNet models with different depths (e.g., ResNet50, ResNet101, and ResNet152). The purpose of this work is to conduct comparative tests on skin disease classification tasks utilizing three ResNet models of varying depths (ResNet50, ResNet101, and ResNet152). This work will evaluate the performance of these models in terms of classification accuracy, training time, and resource consumption to explore the differences between the models. Compare the effect of model depth (number of layers) on classification accuracy and resource consumption and provide a recommendation for the best choice of model based on real-world results. These findings are beneficial for the advancement of medical identification tools, which can effectively assist people in recognizing skin diseases, ultimately improving their skin issues more accurately and conveniently, and providing targeted treatment.

# 2 Related WORKS

In image classification, CNNs are highly respected for their ability to learn features automatically. Compared to traditional machine learning algorithms, CNNs do not require manually designed feature extraction methods and are able to learn directly from raw image data. By introducing deep structure, CNNs are able to improve classification accuracy and effectively handle complex visual tasks. By using models already trained on big publically accessible datasets (e.g., ImageNet) and then adapting them for specific medical tasks, the performance of the models on small-scale medical datasets is greatly improved (Salehi, 2023).

CNNs' performance in medical image processing is mostly due to their multi-layered design, which is capable of capturing everything from low-level edge and texture information to high-level structural features (Yadav, 2019). This makes it ideal for detecting complex lesions, especially in areas such as dermatology, cancer and cardiovascular diseases. Meanwhile, the public distribution of a large number of labeled datasets has hastened the use of CNNs in medical image analysis (Sun, 2023).

# 3 METHODOLOGIES

**3.1 Dataset**

The dermatology dataset this work used was sourced from Kaggle and contains images of six categories of dermatology; Acne, Carcinoma, Eczema, Keratosis, Milia, and Rosacea, with each category containing 400 images each, for a total of 2400 images (Jay, 2023).

After preprocessing the data, the model can be trained. Initially, the photos were resized to a predetermined size of 224x224 pixels for utilization in the deep learning model. Additionally, to enhance data diversity, techniques such as horizontal flipping, random rotation, and random scaling were employed. These strategies effectively prevent model overfitting while also improving its generalization capacity.

**3.2 Models**

The study utilizes three ResNet models, specifically ResNet50, ResNet101, and ResNet152. These models are based on residual learning technique that tackles the problem of gradient vanishing in deep neural networks by incorporating residual connections. This approach enhances training efficiency without compromising accuracy (He, 2016).

Train ResNet50, a compact model with 50 layers, on small datasets. This model is suitable for limited CPU resources or when the data's complexity does not require a deeper network. Evaluate its performance in various picture classification tasks, including dermatological image analysis.

Train an AI model using ResNet101 architecture, which consists of 101 layers. This architecture strikes a balance between ResNet50 and ResNet152, offering improved capability to capture intricate features in images. It is particularly beneficial for tasks demanding advanced comprehension of visual input. Employ this approach for complex image classification tasks where accurate classification heavily relies on visual details. The additional layers facilitate enhanced feature extraction, potentially resulting in superior performance on challenging datasets.

ResNet152, the deepest of the three models with 152 layers, is designed to tackle the most complex tasks and large datasets. Its extensive depth allows it to learn a vast array of features, which can be beneficial for discerning even the most minute differences in images. However, this increased depth also comes with a trade-off: higher training time and resource consumption. The model's complexity necessitates more computational power and time to train, which can be a limiting factor in certain practical applications. Despite the aforementioned challenges, it has been demonstrated that ResNet152 is a viable and efficient solution for intricate image classification tasks demanding utmost precision.

The assessment will consider various parameters including classification accuracy, training efficiency, and resource utilization. By doing so, it is expected to gain insights into the optimal balance between model complexity and performance, which can inform future developments in dermatological image classification and beyond.

# 4 Experiment and Results

**4.1 Training Details**

During the training process of the model, employ transfer learning technique by initializing the model with pre-trained weights from ImageNet. Freeze the base model layer and only fine-tune the top layer. Divide the dataset into three sets: training, validation, and testing, with proportions of 80%, 10%, and 10% respectively. Optimize using Adam optimizer with a learning rate of 1e-4. Utilize sparse categorical crossentropy as the loss function and evaluate classification accuracy using appropriate evaluation metrics. Evaluate model performance on both training and validation sets for 20 training cycles. Finally, assess performance on the test set by measuring classification accuracy, precision, recall, F1-score along with other relevant metrics.

**4.2 Performance Comparison**

In this study, three different ResNet models (ResNet50, ResNet101, and ResNet152) were used to classify dermatological images, and their classification accuracies, precision, recall, and F1 scores were compared, as demonstrated in Table 1.

ResNet50 has the highest classification accuracy of 0.9200, while its precision (0.9237), recall (0.9200), and F1 score (0.9202) indicate its stable performance in the skin disease classification task.

ResNet101 performs slightly less well overall, with an accuracy of 0.8800, and although it performs well on specific categories (e.g., Keratosis and Milia), its overall Rec (0.8800) and F1 (0.8803) are slightly lower than that of ResNet50.

Table 1: Performance comparison of different models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Acc | Pre | Rec | F1 |
| ResNet 50 | .9200 | .9237 | .9200 | .9202 |
| ResNet 101 | .8800 | .8883 | .9000 | .8803 |
| ResNet 152 | .9160 | .9240 | .9160 | .9162 |

ResNet152 performs similarly to ResNet50, with an accuracy of 0.9160, precision of 0.9240, recall of 0.9160, and F1 score of 0.9162. Although ResNet152 performs very well on some specific categories (e.g., Carcinoma and Keratosis) with very high accuracy and recall, the performance decreases when dealing with categories like Eczema and Milia.

Firstly, ResNet50 shows consistently high precision and recall in all categories, especially in the categories of Keratosis and Carcinoma where it achieves almost perfect classification accuracy. Secondly ResNet101, although slightly inferior in overall performance, performs well in the Acne and Rosacea categories with a high recall of 0.90, suggesting that the model is superior in handling these categories. However, the slightly lower recall when dealing with categories such as Carcinoma suggests that the model may be deficient in some complex categories. Finally, ResNet152 showed very good results on some categories, especially the Keratosis and Carcinoma categories, with precision and recall close to 100%. However, the low recall on the Milia category suggests that there may be some misclassification in this category.

# 5 Discussions

During the training process, it can be observed through the performance on the validation set that ResNet50 has the strongest generalization ability and avoids overfitting. The deeper ResNet101 and ResNet152, on the other hand, despite their excellent performance on some categories, do not bring significant improvement in the overall performance, but instead may incur the risk of overfitting on small datasets. Therefore, model selection should consider the size of the dataset, computational resources, and the complexity of the task.

By comparing these three ResNet models, it can be seen that the depth of the model is not always the only determinant of classification accuracy. Although ResNet101 and ResNet152 are deeper than ResNet50, their improvement in classification accuracy is not significant, and they even underperform ResNet50 in some cases. ResNet50, with its relatively shallow network structure and stable performance, is the best choice, especially when computational resources are limited or the dataset is small.

Future research efforts should focus on enhancing the dermatologic categorization capabilities of diagnostic systems. This can be achieved by improving the size and comprehensiveness of databases used for training these systems. The current dataset may not provide enough diversity to enable the model to generalize findings across a wider range of skin diseases. To address this, future research should aim to incorporate extensive and diverse medical image datasets into the training process. These datasets should include a wide collection of photographs depicting various dermatological disorders, capturing different severity levels, and representing images from different demographic groups. By doing so, the author can improve the model's ability to learn and generalize effectively while reducing the risk of overfitting.

Secondly, the exploration of more sophisticated deep learning models is a promising avenue for advancement. The introduction of cutting-edge architectures, such as Vision Transformers (ViT), could revolutionize the way models perceive and classify images. ViT models have demonstrated remarkable potential in image classification tasks, particularly in their ability to capture holistic global features within images. In future, Models like ViT can greatly aid the development of medical image recognition.

Finally, the application of ensemble learning techniques provides a great contribution to further improving the classification system. By integrating multiple deep learning models that possess unique advantages and perspectives, ensemble learning can create a synergistic effect, thereby enhancing performance.For example, if I create a combination of the ResNet model, DenseNet model, or EfficientNet model, this integration, if applied properly, can enhance the comprehensive understanding and accuracy of skin disease images. This integrated approach has improved the efficiency of the skin disease recognition system.

With these anticipated improvements, the future of dermatological classification systems is bright, with the potential to achieve significant leaps in both diagnostic precision and operational efficiency. These advancements are expected to contribute to a transformative impact on the field of dermatology, offering more reliable diagnostic support and paving the way for personalized and precise treatment plans for patients with skin conditions.

# 6 conclusions

ResNet50 performs optimally in classification tasks with 92% accuracy and near-perfect performance on several categories (e.g., Keratosis and Carcinoma). Due to its shallower model, ResNet50 is able to process data more efficiently, avoiding the risk of overfitting, while balancing performance and computational resource requirements. This makes ResNet50 ideal for processing small to medium-sized medical image datasets. Despite being deeper, ResNet101 and ResNet152 failed to significantly outperform ResNet50 in terms of overall classification accuracy. While it performed well on some categories (e.g., ResNet152's precision and recall were close to 100% on Carcinoma), when dealing with more complex categories (e.g., Milia and Eczema) its performance is not as good as expected. This suggests that on smaller datasets, too deep a network may lead to excessive model complexity, which in turn affects its generalization ability.

**References**

Dagne, D. D. A. 2018. Recognizing neglected skin diseases: WHO publishes pictorial training guide. *World Health Organization Global, Geneva, Switzerland*. URL: https://www.who.int/news/item/08-06-2018-recognizing-neglected-skin-diseases-who-publishes-pictorial-training-guide. Last Accessed: 2024/09/09

He, K., Zhang, X., Ren, S., & Sun, J. 2016, Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

Hogarty, D. T., Su, J. C., Phan, K., Attia, M., Hossny, M., Nahavandi, S., ... & Yazdabadi, A. 2020. Artificial intelligence in dermatology—where we are and the way to the future: a review. *American journal of clinical dermatology*, *21*, 41-47.

Jay, D. 2023. Dermatology dataset. URL: https://www.kaggle.com/code/jayrdixit/dermatology-dataset. Last Accessed: 2024/09/09.

Li, M., Jiang, Y., Zhang, Y., & Zhu, H. 2023. Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, *11*, 1273253.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. 2021. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. 2023. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, *15*(7), 5930.

Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., & Yang, X. 2023. Machine learning methods in skin disease recognition: a systematic review. *Processes*, *11*(4), 1003.

Thomsen, K., Iversen, L., Titlestad, T. L., & Winther, O. 2020. Systematic review of machine learning for diagnosis and prognosis in dermatology. Journal of Dermatological Treatment, 31(5), 496-510.

Yadav, S. S., & Jadhav, S. M. 2019. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big data*, *6*(1), 1-18.