

Skin Disease Classification

Using CNN with Transfer Learning

Raj Kiran Chhatkuli¹, Biplove Pokhrel^{1*}, Roshan Subedi¹, Sudip Dahal¹, Deepak Bastola¹

¹Department of Electronics and Computer Engineering, Pashchimanchal Campus, Institute of Engineering, Tribhuvan University, Pokhara, Nepal

*biplove@wrc.edu.np

(Manuscript Received: 28th January, 2026; Revised: 10th February, 2026; Accepted: 17th February, 2026)

Abstract

Skin diseases are one of the most common public health issues and some of these conditions like melanoma are very dangerous to human lives unless detected at an early age. The latest development in deep-learning algorithms has shown a lot of potential in analyzing medical images, especially in the field of disease diagnosis based on visual information. This paper examines the process of automatic detection of dermatological disorders using dermoscopic visualization and convolutional neural networks (CNNs). This research is based on the HAM10000 dataset which consists of dermoscopic images of seven different categories of skin diseases. Both the standard CNN architecture and the pre-trained models of deep-learning are used: AlexNet is trained as a standard reference, and ResNet-18 and DenseNet are fine-tuned on the basis of the transfer learning. Strict data preprocessing guidelines and class-balancing strategies are followed to improve the performance of the model. Performance indicators such as accuracy, precision, recall, and F1-score are embraced to determine the effectiveness of the model. The results of the experiments prove that transfer learning significantly increases the performance of the classification, and the overall accuracy of DenseNet was the highest of the considered models. The results highlight the possibility of deep-learned diagnostic systems to be used successfully to help dermatologists diagnose skin diseases correctly.

Keywords: CNN, DenseNet, HAM10000, Skin disease classification, Transfer learning

1. Introduction

1.1 Background

Intelligent systems and automation are becoming more prevalent in a variety of industries, including healthcare. The use of machine learning and deep learning methods has become quite popular because of their capability of discovering complex patterns from large datasets (Krizhevsky et al., 2012). In the field of medicine, there is a massive amount of data created in everyday practice and research, which is why machine learning technologies are particularly advantageous in the diagnostic and decision-making process (Kassem et al., 2020).

Dermatological issues are one of the most prevalent health issues in the world. Many people often overlook the skin symptoms, or postpone medical evaluation, as they think them unimportant. Other skin conditions such as melanoma may become life-threatening if not diagnosed in a timely fashion (Esteva et al., 2017). The World Health Organization reported millions of new skin cancer cases reported worldwide each year. Dermoscopy has been found to aid skin lesion visual diagnosis but the diagnosis remains mainly subjective, depending on the dermatologist's expertise (Codella et al., 2018).

In this research, the goal is to devise an automatic dermatological classifier using Convolutional Neural Networks (CNNs). The research will focus on implementing novel deep learning and transfer learning methods to improve classification performance and build a tool to diagnose the disease early and reliably, by taking dermoscopic photographs of the lesions (He et al., 2016; Dosovitskiy et al., 2021). A set of 10,015 dermoscopic images are included in the HAM10000 dataset (Tschandl et al.,

2018), which is processed under standardized conditions and class balanced; three CNN architectures, namely AlexNet (Krizhevsky et al., 2012), ResNet-18 (He et al., 2016), and DenseNet (Gessert et al., 2020; Srinivasu et al., 2021), are evaluated.

1.2 Related Work

The deep learning models have been shown to be highly accurate in the diagnosis of dermatologic diseases and cutaneous malignancies in numerous studies. As there are a number of big image collections of dermoscopic images now, researchers have begun applying CNNs to assist dermatologists (Codella et al., 2018; Kassem et al., 2020).

Recently, Esteva et al. (2017) showed that a CNN trained on a large collection of skin lesion images can learn excellent features and achieve a high accuracy in the classification of skin cancer at dermatologists level. Similarly, Codella et al. (2018) shared the results of the International Skin Imaging Collaboration (ISIC) challenge, where CNN-based methods beat the dermatologists trained by the challenge.

Different CNN architectures like Inception and ResNet (He et al., 2016), DenseNet have been explored in order to further improve the classification. To address the lack of labelled medical images and improve the generalisation and convergence of pre-trained networks, transfer learning strategies have been employed in the fine-tuning of pre-trained networks to be applied to dermoscopic images (Kassem et al., 2020; Gessert et al., 2020; Srinivasu et al., 2021). Recently, transformer architectures have also been used for image recognition in medical images (Dosovitskiy et al., 2021), which opens new avenues for medical image analysis.

The results conclude that CNN-based techniques for dermatologic disease classification are effective based on the evidence amassed. However, issues of class imbalance and dataset size are still present. The current study is based on the HAM10000 dataset (Tschandl et al., 2018) to assess the transfer learning with different CNN models.

2. Methodology

2.1 Dataset Description

The dataset used in this research is HAM10000, obtained from the International Skin Imaging Collaboration (ISIC). The dataset contains 10,015 dermoscopic images categorized into seven different skin disease classes. This dataset is widely used for research purposes and provides reliable labeled data for training deep learning models.

2.2 Data Preprocessing and Splitting

All images are resized and normalized before being used for model training. Since the dataset is highly imbalanced, oversampling is applied to the training data to ensure that each class has a comparable number of samples. The dataset is divided into training, validation, and testing sets using a 64:16:20 ratio. This split helps the models generalize better and provides a reliable evaluation of performance.

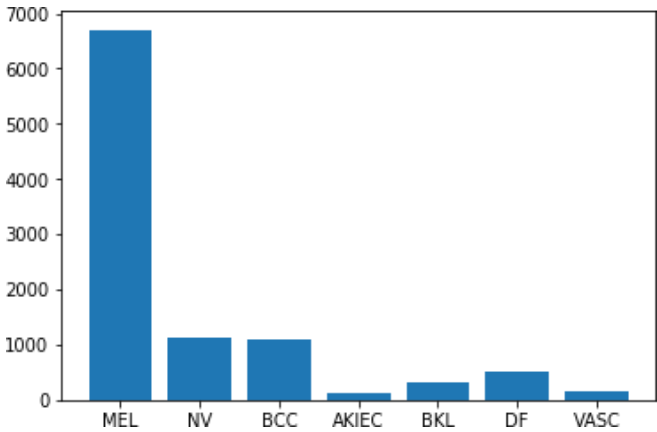


Fig. 1: Distribution of data with respect to class

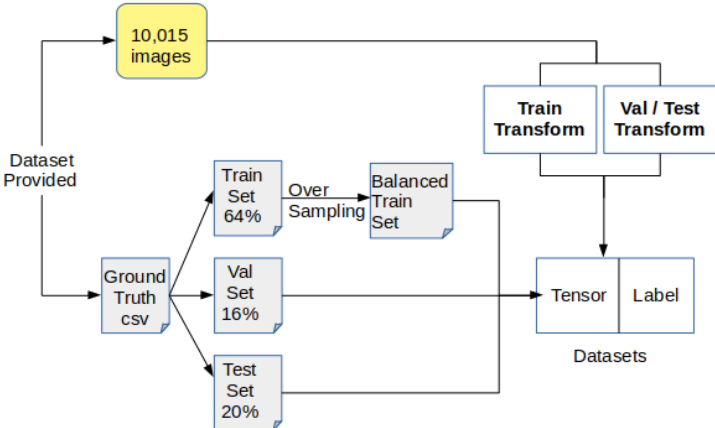


Fig. 2: Block diagram for preprocessing

2.3 System Architecture

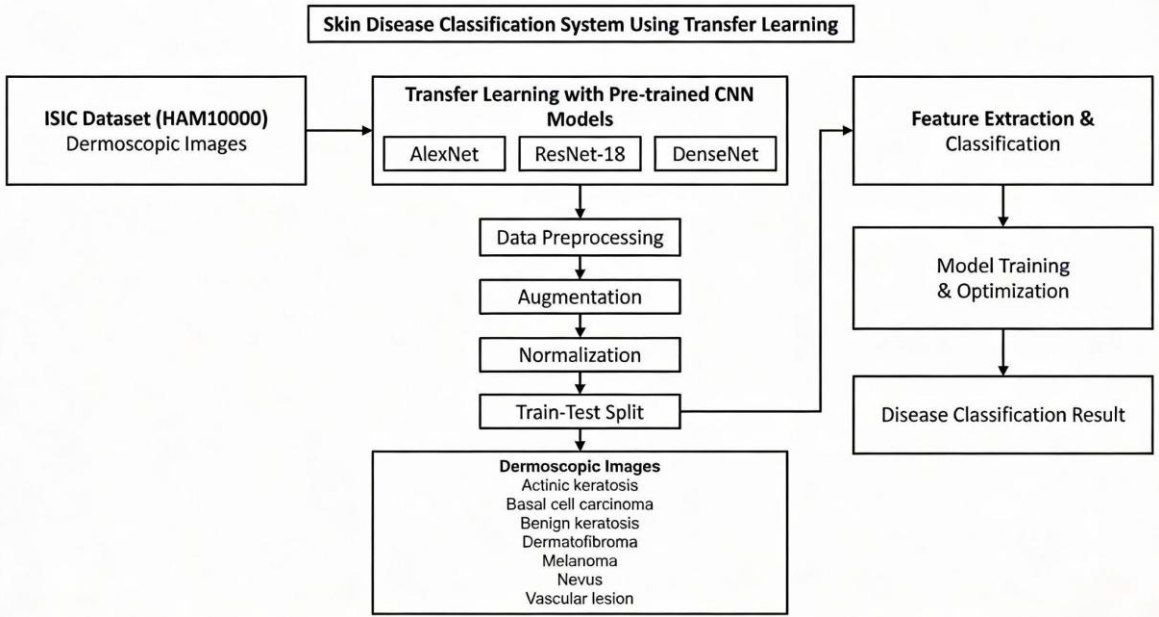


Fig. 3: System Architecture

2.4 Model Selection and Training

Three convolutional neural network models are used in this research:

- i) **AlexNet:** Used as a baseline CNN model and trained from scratch.
- ii) **ResNet-18:** A pre-trained deep network using residual connections, fine-tuned on the HAM10000 dataset.
- iii) **DenseNet:** A pre-trained model with dense connectivity, fine-tuned for skin disease classification.

Transfer learning is applied to ResNet-18 and DenseNet by modifying the final fully connected layers to match the number of output classes. All models are trained using the Adam optimizer and cross-entropy loss function for 50 epochs.

2.5 Hyperparameters Used

Batch size: 32

Loss criteria: Cross entropy

Learning Parameter: $1e-4$

Optimizer: Adam

Epochs: 50

3. Results and Discussion

3.1 Testing and Evaluation

All models were evaluated and the results were recorded. The training loss and validation loss for each model were visualized using Seaborn. The plots of epochs versus loss are shown in the figures below for AlexNet, ResNet-18, and DenseNet respectively.

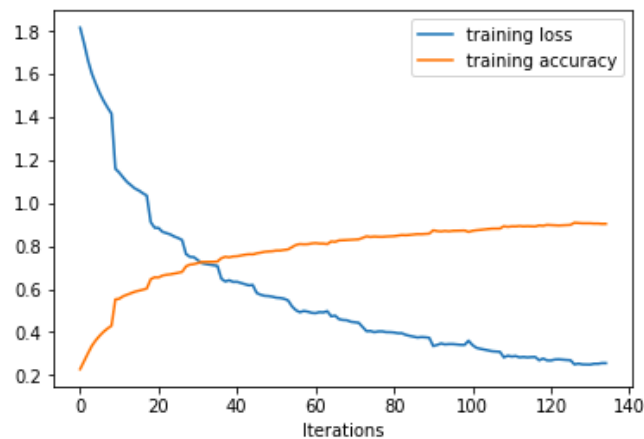


Fig. 4: Train loss- Train accuracy for Alexnet

As shown in Fig. 4, the training loss for AlexNet decreased with increasing iterations, while the training accuracy improved with an increasing number of epochs.

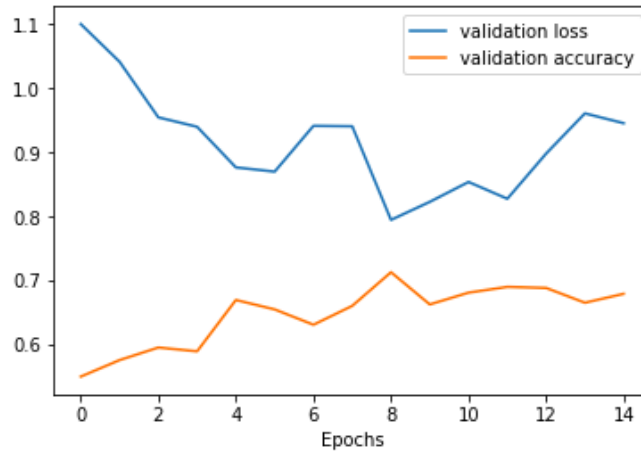


Fig. 5: Validation loss- Validation accuracy for Alexnet

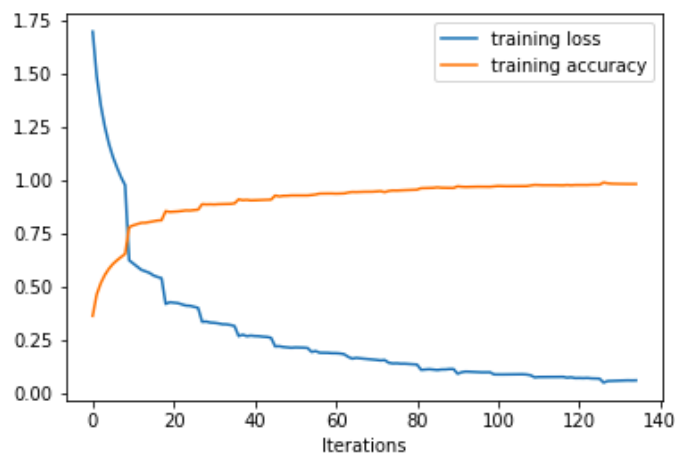


Fig. 6: Train loss- Train accuracy for Resnet18

As shown in Fig. 6, the training loss of ResNet-18 expressed a consistent downward trend as the number of iterations increased and the training accuracy increased as the number of epochs increased.

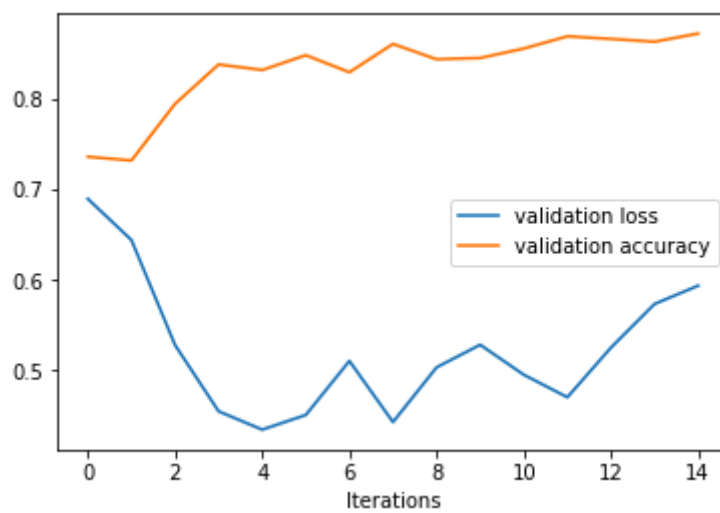


Fig. 7: Validation loss- Validation accuracy for Resnet18

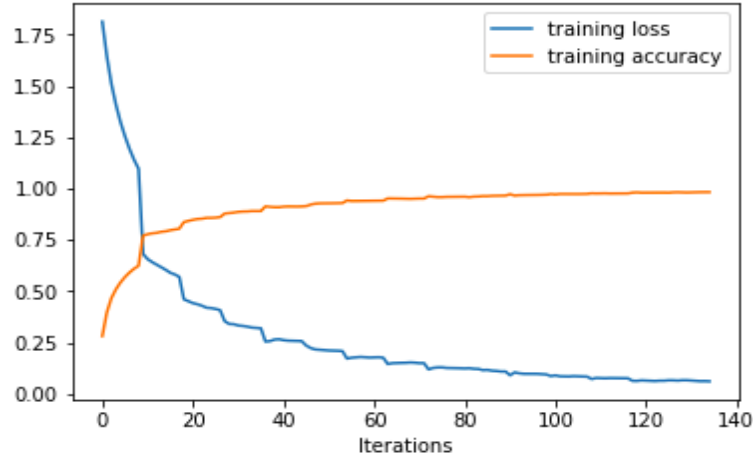


Fig. 8: Train loss- Train accuracy for DenseNet

As Fig. 8 shows, training loss of DenseNet dropped gradually as the number of iterations increased, whereas the training accuracy rose with the number of epochs.

Table 1: Comparison of Precision, Recall and F1score for different architectures.

Model	Precision	Recall	F1 Score
DenseNet	0.81260	0.74911	0.77802
AlexNet	0.54051	0.59446	0.55996
ResNet	0.73681	0.73044	0.72799

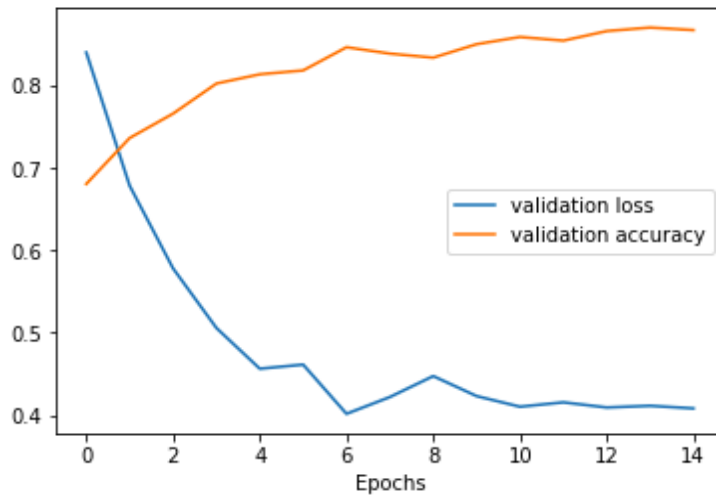


Fig. 9: Validation loss- Validation accuracy for DenseNet

Table 2: Comparison of Accuracy for different architecture

Model	Accuracy
DenseNet	86.86969
Alexnet	72.14128
Resnet18	84.12381

The performance metrics (F1-score, precision, recall, accuracy) are evaluated for each model and are shown in the Figure above. The models are compared on the basis of the performance metrics. The pre-trained models are much more effective than AlexNet, probably because it is not able to learn complex features enough, which leads to an oscillating and non-convergent validation loss. Comparatively, DenseNet performs better as it has an accuracy of 86.87%.

The confusionMatrix of AlexNet, ResNet 18, and DenseNet was obtained and the graphs are shown in the figures below respectively.



Fig. 10: Confusion Matrix for Alexnet

The actual and predicted values were used to plot the confusion matrix in Alexnet architecture as shown above.

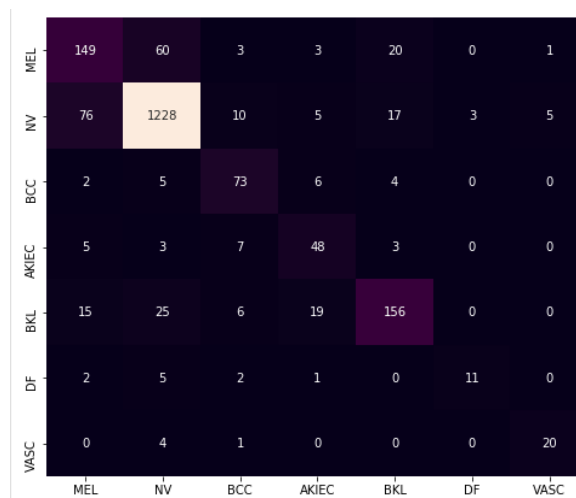


Fig. 11: Confusion matrix for Resnet18

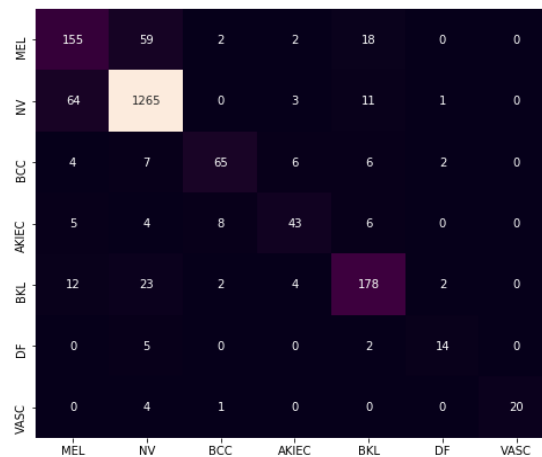


Fig. 12: Confusion matrix for DenseNet

The performance of all models is evaluated using accuracy, precision, recall, and F1-score. Training and validation results show that the pre-trained models converge faster and perform more stably compared to the AlexNet model.

AlexNet shows lower accuracy and unstable validation loss, indicating limited learning capability for complex skin lesion patterns. ResNet-18 performs significantly better due to its deeper architecture and residual connections. DenseNet achieves the highest overall performance, with an accuracy of approximately 86.9%.

The results clearly demonstrate that transfer learning improves classification performance, especially when the dataset size is limited. DenseNet shows better feature reuse and generalization across different skin disease classes.

4. Conclusions

This study constituted a deep learning-based model to classify skin disease through dermoscopic images. HAM10000 data was used to implement and test classical CNN models and models based on transfer learning. DenseNet was one of the best tested models both in terms of accuracy and general stability.

The findings validate the notion that deep learning models and more specifically, transfer learning models can be useful to support dermatologists to diagnose skin diseases early and accurately. The next phase of work can involve larger datasets, model ensembling, and deployment in real-time with the help of mobile or web-based applications.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.
- Codella, N. C. F., Gutman, D., Celebi, M. E., et al. (2017). Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging. *arXiv preprint arXiv:1710.05006*.

- 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*.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*.
- Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. <https://doi.org/10.1038/sdata.2018.161>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. arXiv:2010.11929.
- Kassem, M. A., Hosny, K. M., & Fouad, M. M. (2020). Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. *IEEE Access*, 8, 114822–114832.
- Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (2020). Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX*, 7, 100864.
- Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, 21(8), 2852.