Comparative Analysis of Deep Learning and Machine Learning Models for Skin Disease Classification

Nure Jannat[1*] and Prity Rani Das[2*]

[1*]Institute of Information Technology, NSTU, Sonapur, Noakhali, 3802, Chittagong, Bangladesh.
[1*]Institute of Information Technology, NSTU, Sonapur, Noakhali, 3802, Chittagong, Bangladesh.

*Corresponding author(s). E-mail(s): jannat2515@student.nstu.edu.bd; prity2515@student.nstu.edu.bd;

Abstract

This paper compares different deep learning and machine learning models for classifying skin diseases. It looks at how well each model performs and discusses their strengths and weaknesses to help the automated skin disease diagnosis in the future.

Keywords: CNN Architectures, Transfer Learning (TL), Machine Learning, Skin Disease

1 Introduction

Skin diseases are becoming more common around the world. Some common diseases are acne, vitiligo, hyperpigmentation, nail psoriasis, and SJS-TEN are hard to diagnose sometimes because they look similar. If we can identify these conditions early, patients have a better chance of getting the right treatment and recovering faster. Sometimes doctors struggle to tell these diseases apart just by looking, especially when symptoms overlap. In this situation, technology can be useful. With the rise of machine learning and deep learning, we now have the chance to research and develop an intelligent system that can help to analyze the skin images and identify the disease that might be present. These systems can work fast and accurately, helping doctors make quicker decisions and improving patient care. In this research, we explore how machine learning and CNNs can be used to identify different skin diseases and classify them. We use a custom dataset of skin disease images to train and test different models, aiming to support healthcare professionals in diagnosing skin conditions more effectively.

1.1 Research Problem

Detecting skin diseases from dermatoscopic images is a challenging task, especially when dealing with rare conditions, overlapping symptoms, and low-quality or unprocessed image data. Sometimes skin problems can be signs of other health issues. Therefore, it is critical to identify and treat skin conditions as soon as possible to prevent them from getting worse or spreading. Our study focuses on developing a reliable classification model using various machine learning algorithms (such as SVM, Random Forest, and KNN), applying transfer learning, and using different Convolutional Neural Networks (CNN) architectures. Our aim is to accurately classify five types of skin conditions - acne, vitiligo, hyperpigmentation, nail psoriasis, and SJS-TEN to improve early diagnosis and medical decision-making.

2 Literature Review

Many research activities have been done previously for accurately identifying skin diseases. All researchers proposed different approaches using different techniques and different learning algorithms to classify the skin diseases. In this section, some of the related work will be analysed to extract their ideas, techniques and research gaps.

Sadik, Rifat[1] proposed a solution to recognize 5 classes of skin disease using CNN architectures, MobileNet and Xception. They used a transfer learning method which uses the pre-trained imagenet dataset. They also compare their proposed model with other CNN architectures which are ResNet50, InceptionV3, Inception-ResNet, and DenseNet. Accuracy, precision, recall, and F1-score are used to evaluate the performance of the model where MobileNet achieved 96.00% accuracy, and the Xception model reached 97.00% accuracy. But this method takes a huge computation time.

Agarwal, Raghav[2] used 11 different kinds of convolutional neural networks to identify the skin diseases and comparing the results they found out that ResNet152 beat other CNN models. This research used eight most common skin disorders and applied transfer learning. Before applying the transfer learning and CNN model, they applied some preprocessing(resizing, Standardization). This technique made an input layer, a convolutional layer, a batch normalization layer and an activation layer. For the activation layer, the softmax activation function was used for global average pooling. Adam optimizer was used to adjust the weights. To calculate the loss, Categorical cross-entropy is also used. The performance was evaluated using accuracy, precision and recall.

Skin diseases have dangerous effects on skin and may keep spreading over time. To identify these skin problems Pugazhenthi, V., Sagar K. Naik[3] suggested a solution to detect the presence of skin problems using the Decision Tree algorithm. In their proposed solution, images were pre-processed using Contrast Enhancement, and Grayscale Conversion. They also used Global Thresholding for image segmentation and GLCM and IQA for extracting features from the image. Finally, the decision tree was applied to identify the four kinds of skin diseases.

Ahammed, Mostafiz, Md Al Mamun,Mohammad Shorif Uddin[4] proposed a complete methodology for skin disease classification using handcrafted features and traditional machine learning algorithms. The proposed approach included digital hair removal techniques via black hat morphological operations, noise reduction through Gaussian filteringand lesion segmentation using GrabCut algorithm. Feature extraction relied on GLCM and statistical measures followed by classification with SVM, KNN, and decision trees. This approach achieved high accuracy on the ISIC 2019 and HAM10000 datasets (up to 97.% with SVM), especially after applying random oversampling to address data imbalance. This research also highlighted segmentation limitations and advocated for the future use of deep learning and ensemble techniques to enhance lesion detection and classification.

Many research has proposed models for the classification of skin diseases, focusing on various image processing and machine learning techniques. Similarly, Shaden Abdulaziz AlDera, Mohamed Tahar Ber Othman[5] utilize a model following the preprocessing steps such as resizing, median filtering, and grayscale conversion to enhance image quality. Feature extraction techniques include Gabor, entropy, and Sobel filtering. Machine learning algorithms like Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (K-NN) are commonly used for skin disease classification. Among these, SVM has been found to perform exceptionally well, achieving an accuracy of 90.7%. Comparative studies show that models with large datasets and diverse features perform well. Additionally, the lack of labeled skin disease images remains a challenge that indicates further improvements in classification can be achieved with larger and more diverse datasets.

Melanoma, a potentially deadly skin cancer, has a rising incidence, particularly in Australia and New Zealand. Early detection is crucial but challenging. Viswanatha Reddy Allugunt[6] proposed a two-stage Convolutional Neural Network (CNN) for detecting melanoma classification, achieving 88.83% accuracy. This proposed model successfully differentiates various melanoma types like malignant, superficial spreading, and nodular melanoma. Compared to traditional machine learning methods like decision trees and random forests, the proposed CNN achieved higher accuracy and greater computational efficiency. However, this model was tested on a single dataset from Dermatology Resource, suggesting the need for future research on more diverse datasets to enhance its generalizability.

Dip et al. (2024)[7] proposed a transfer learning approach for equitable skin disease diagnosis across diverse skin tones. They benchmarked models like MedViT and RETFound using the Diverse Dermatology Images (DDI) dataset. MedViT-base showed the highest performance, especially after domain adaptation with HAM10000. Their method significantly improved model fairness and accuracy. This study highlights the value of inclusive datasets and cross-domain pre-training in dermatological AI.

3 Dataset Description

This study uses a dataset of 9,548 dermatoscopic images showing different skin conditions. The images are divided into five categories: acne (1,148 images), vitiligo (2,016), hyperpigmentation (700), nail psoriasis (2,520), and SJS-TEN (3,164). These images were collected from hospitals and online sources in different countries. The dataset was published on Mendeley Data on July 30, 2024, by Sharun Akter Khushbu.

The images in this dataset have not been pre-processed, which makes it more challenging to train machine learning models. Since the dataset focuses on rare skin diseases, it is especially useful for testing how well transfer learning techniques can classify skin conditions using medical images.

4 Results

This section represents the results of applying the methodologies from the seven selected research papers to classify 5 skin diseases.

The Decision Tree model[3] achieved an accuracy of 49%, with moderate performance across the skin disease categories. It performed best on Nail Psoriasis (precision of 0.54, recall of 0.56). But struggled with conditions like Hyperpigmentation and Acne showing lower precision and recall. The results suggest that the model can identify certain skin diseases but it needs improvement.

The result from 11 CNN architecture[1][2] are in the following table with training and validation data:

Model	Accuracy	Precision	Recall	Loss
VGG16	96.02%	96.43%	95.53%	0.1079
VGG19	96.31%	96.71%	95.95%	0.1066
ResNet50	79.05%	83.87%	74.48%	0.5616
ResNet101	75.81%	81.70%	69.28%	0.6350
ResNet152	87.82%	90.40%	85.34%	0.3404
Xception	99.78%	99.79%	99.78%	0.0099
DenseNet121	99.68%	99.77%	99.68%	0.0087
DenseNet169	99.67%	99.76%	99.67%	0.0075
DenseNet201	99.56%	99.62%	99.54%	0.0126
MobileNet	99.13%	99.14%	99.12%	0.0273
InceptionV3	98.82%	98.94%	98.79%	0.0269

Table 1 Training Data Performance Comparison of Models for Skin Disease Classification

By comparing the result of 11 CNN architecture, we found that DenseNet121 got 96.50% , DenseNet169 got 96.74%, MobileNet got 96.01% validation accuracy. Among all of them DenseNet201 got the highest validation accuracy with 97.13%.

DenseNet201 achieves the highest validation accuracy due to its deep architecture with 201 layers which allows it to capture more complex features and improve performance. The dense connectivity between layers enables efficient feature reuse, better gradient flow, and enhanced generalization which makes it more effective for

Model	Accuracy	Precision	Recall	Loss
VGG16	92.37%	92.83%	92.02%	0.9075
VGG19	91.39%	92.14%	91.11%	0.9033
ResNet50	68.03%	72.97%	63.34%	0.9347
ResNet101	67.09%	73.42%	61.90%	0.9133
ResNet152	76.54%	79.37%	73.67%	0.9148
Xception	94.85%	95.53%	94.82%	0.3128
DenseNet121	96.50%	96.50%	96.46%	0.2808
DenseNet169	96.74%	96.81%	96.64%	0.2757
DenseNet201	96.74%	97.13%	97.09%	0.2646
MobileNet	94.54%	96.07%	95.90%	0.1656
InceptionV3	94.54%	94.70%	94.47%	0.2382

Table 2 Validation Data Performance Comparison of Models for Skin Disease Classification

skin disease classification. The increased number of depth helps DenseNet201 to learn hierarchical features and contributes to achieve highest accuracy compared to other models like DenseNet121, DenseNet169, and MobileNet.

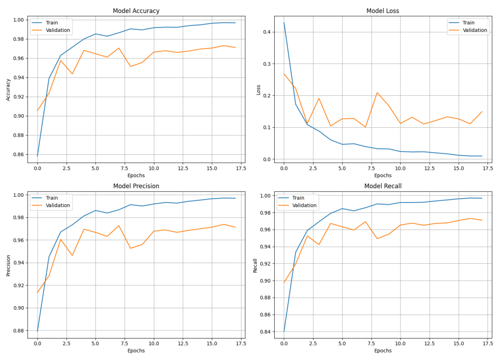

Fig. 1 Accuracy, Precision, Recall and Loss for CNN Model DenseNet201 with Training and Validation Data

For the performance of different traditional machine learning classifiers for skin disease classification, we tested three models: Support Vector Machine (SVM) with RBF kernel, k-Nearest Neighbors, and Decision Tree[4]. The classification results in terms of precision, recall, and F1-score for each class are summarized in following Table . Additionally, the overall accuracy, macro-average, and weighted average scores are provided for comprehensive comparison.

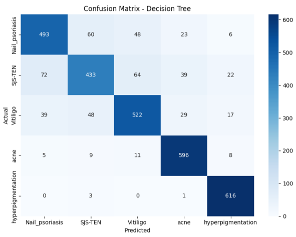

Fig. 2 Confusion Matrix using Decision Tree

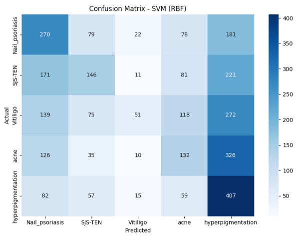

Fig. 3 Confusion Matrix using SVM

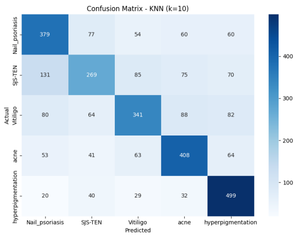

Fig. 4 Confusion matrix using KNN

Classifier	Accuracy	Precision	Recall	F1-Score
SVM	0.32	0.35	0.32	0.29
KNN	0.60	0.59	0.60	0.59
Decision Tree	0.84	0.84	0.84	0.84

Table 3 Performance of Classifiers

We evaluated the performance of three machine learning models—Support Vector Machine (SVM) with RBF kernel, k-Nearest Neighbors (KNN), and Random Forest—on a five-class skin disease dataset[5]. The models were assessed using accuracy, macro-averaged precision, recall, and F1-score metrics. The summarized results are provided in Table -

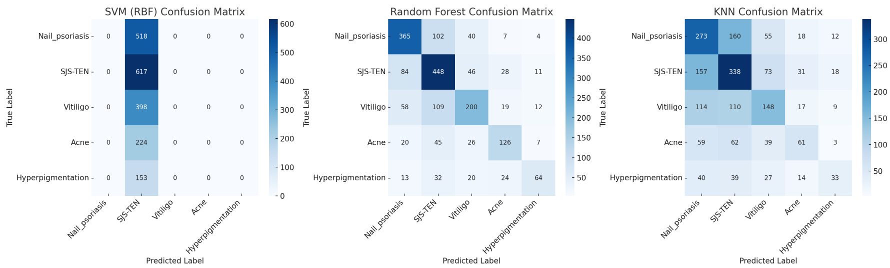

Fig. 5 Confusion metrics of the classifiers [5]

Classifier	Accuracy	Precision	Recall	F1-Score
SVM	0.32	0.06	0.20	0.10
KNN	0.45	0.44	0.39	0.40
Random Forest	0.63	0.63	0.58	0.60

Table 4 Average performance metrics of classifiers

To evaluate the performance of the proposed CNN model and compare it against classical machine learning methods, we trained and tested three different classifiers: Convolutional Neural Network (CNN), Decision Tree (DT), and Random Forest (RF). Each model was evaluated using four key metrics: accuracy, macro-averaged precision, macro-averaged recall, and macro-averaged F1-score. The CNN was trained for 20 epochs, and performance was evaluated on the validation set[6].

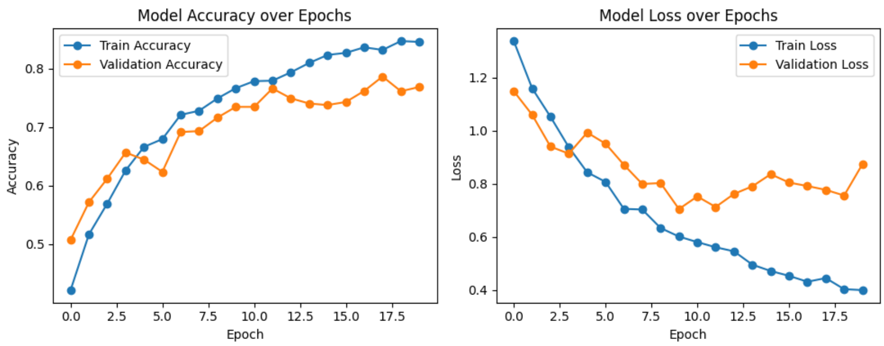

Fig. 6 Model Accuracy and loss over Epochs (CNN) [6]

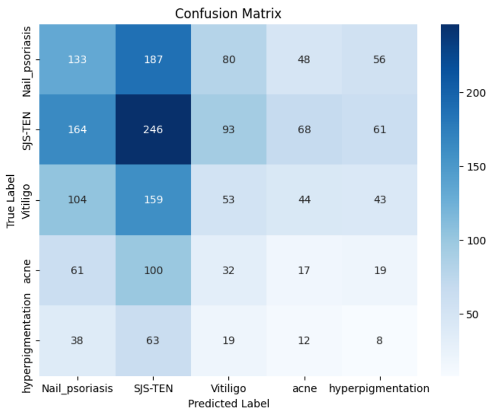

Fig. 7 Confusion matrix of CNN [6]

Model	Accuracy	Precision	Recall	F1-Score
CNN	0.7689	0.1832	0.1832	0.1812
Decision Tree	0.5947	0.59	0.57	0.56
Random Forest	0.7827	0.85	0.73	0.77

Table 5 Performance comparison of CNN, Decision Tree and Random Forest models

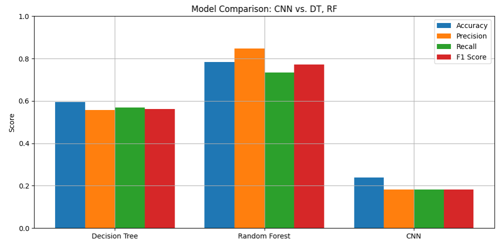

Fig. 8 Performance comparison of CNN, Decision Tree and random forest models [6]

The model initially showed strong improvement from Epoch 1 to Epoch 2, with loss decreasing significantly and accuracy peaking at 87.18%[7]. However, starting from Epoch 3, the performance began to decline. Both loss and accuracy deteriorated over the next epochs, suggesting potential overfitting, learning rate issues, or instability during training. Further tuning or early stopping might help stabilize the training process.

Epoch	Average	Train Accuracy
1	0.9115	74.60
2	0.4866	87.18
3	0.6852	76.74
4	1.1144	54.95
5	1.4051	37.11

Table 6 Epoch-wise Average and Train Accuracy

5 Conclusion

While traditional machine learning models like SVM, KNN, and Random Forest show moderate success in classifying skin diseases, the results indicate limitations in handling complex and visually similar conditions. The current accuracy levels suggest that stronger, deep learning-based classifiers—particularly CNNs or transformer-based architectures—may provide better performance. Additionally, expanding the dataset with more diverse and higher-quality images is essential to improve generalization, especially for underrepresented or rare conditions. Future work will focus on leveraging advanced deep learning techniques and larger, more varied datasets to build more robust and reliable diagnostic tools.

References

[1] Sadik, R., Majumder, A., Biswas, A.A., Ahammad, B., Rahman, M.M.: An in-depth analysis of convolutional neural network architectures with transfer learning for skin disease diagnosis. Healthcare Analytics **3**, 100143 (2023)

[2] Agarwal, R., Godavarthi, D.: Skin disease classification using cnn algorithms. EAI Endorsed Transactions on Pervasive Health and Technology **9**(1) (2023)

[3] Pugazhenthi, V., Naik, S.K., Joshi, A.D., Manerkar, S.S., Nagvekar, V.U., Naik, K.P., Palekar, C.G., Sagar, K.: Skin disease detection and classification. International Journal of Advanced Engineering Research and Science (IJAERS) **6**(5), 396–400 (2019)

[4] Ahammed, M., Mamun, M.A., Uddin, M.S.: A machine learning approach for skin disease detection and classification using image segmentation. Healthcare Analytics **2**, 100122 (2022)

[5] AlDera, S.A., Othman, M.T.B.: A model for classification and diagnosis of skin disease using machine learning and image processing techniques. International Journal of Advanced Computer Science and Applications (IJACSA) **13**(5) (2022)

[6] Allugunti, V.R.: A machine learning model for skin disease classification using convolution neural network. International Journal of Computing, Programming and Database Management **3**(1), 141–147 (2022)

[7] Dip, S.A., Arif, K.H.I., Shuvo, U.A., Khan, I.A., Meng, N.: Equitable skin disease prediction using transfer learning and domain adaptation. In: AAAI Fall Symposium Series (FSS-24) (2024)