Skin Disease Classification Using CNN Algorithms

Raghav Agarwal¹ and Deepthi Godavarthi², *

¹School of Computer Science & Engineering (SCOPE), VIT-AP University, Amravati, Andhra Pradesh, India
²School of Computer Science & Engineering (SCOPE), VIT-AP University, Amravati, Andhra Pradesh, India

Abstract

INTRODUCTION: Dermatological disorders, particularly human skin diseases, have become more common in recent decades. Environmental factors, socioeconomic problems, a lack of a balanced diet, and other variables have all contributed to an increase in skin diseases in recent years. Skin diseases can cause psychological suffering in addition to physical injury, especially in people with scarred or disfigured faces.

OBJECTIVES: The use of artificial intelligence or computer-based technologies in the detection of face skin disorders has advanced dramatically over time. Even for highly experienced doctors and dermatologists, identifying skin disorders can be tricky since many skin diseases have a visual affinity with the surrounding skin and lesions.

METHODS: Today, the majority of skincare specialists rely on time-consuming, traditional methods to identify disorders. Even though several research have demonstrated promising results on the picture classification job, few studies compare well-known deep learning models with various metrics for categorizing human skin disorders.

RESULTS: This study examines and contrasts various skin illnesses in terms of cosmetics and common skin concerns. Our dataset includes over 25000 of the eight most common skin disorders. Convolutional neural networks have shown imaging performance that is comparable to or greater than that of humans. We used 11 different network algorithms to identify the illnesses in the sample and compared the results.

CONCLUSION: To adjust the format of incoming photographs, we do certain image pre-processing and image scaling for each model. ResNet152 beat other deep learning methods in terms of recall, accuracy, and precision on a test dataset of 1930 images.

Keywords: Skin disease classification, Transfer Learning, Deep Learning, Medical Images, Clinical Decision Support System

Received on 19 June 2023, accepted on 07 September 2023, published on 02 October 2023

Copyright © 2023 R. Agarwal et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetph.9.4039

¹Corresponding author. Email: deepthi.g@vitap.ac.in

1. Introduction

Human skin disease is a chronic ailment that affects people of all ages all over the world. Skin disorders that are severe enough to necessitate medical treatment or that receive insufficient attention might result in bodily damage or even death. Clinicians must still detect most skin problems by hand and examine the patient's symptoms. However, the following factors make it difficult to diagnose skin conditions. Skin disease classification necessitates a considerable amount of information; thus, a diagnosis given by a non-expert doctor may result in inadequate therapy or other complications. Second, reaching hospitals or basic treatment in rural places might be difficult, especially during the COVID-19 epidemic. Furthermore, few people go to the doctor unless they have a serious illness. As a result, addressing skin diseases too late may worsen the situation and exacerbate the symptoms.

According to the 2018 English Skin Establishment Report, 5.4 million new cases of skin sickness are recorded in the United States each year, and one in every five persons is believed to have a lifelong cutaneous risk. An estimated 60% of English people suffer from a skin ailment. Skin problems can be fatal and severely impair a person's daily activities, interpersonal interactions, and internal organs. This illness may also be characterized by dysfunctional behaviors that may lead to suicide, social isolation, or even despair.

Skin disorders affect 10% to 12% of Indians. The skin functions as a sensor of its environment in addition to providing physical protection. It is the biggest organ in the
human body, with seven layers of ectodermic tissue protecting bones, muscles, and internal organs. Poor personal hygiene, rising pollution, climate change, and hazardous UV radiation can all aggravate skin disorders. Cancer incidence may increase by two to three percent for every percentage point decrease in ozone. In India, photosensitive and infectious skin disorders are extremely widespread. It is critical to address skin disorders as soon as they emerge to prevent complications that impact more than just the skin. People's quest for efficient solutions is motivated by the following reasons: Given India's explosive population growth, it is critical to begin providing high-quality care to each individual as soon as possible. Skin problems mirror diseases such as AIDS, tuberculosis, and others that can be fatal if not treated promptly. Even mild skin issues are expensive, limiting treatment possibilities. As a result, it is critical to create economical and effective approaches for identifying skin disorders. Manual traditional equipment, including that employed in the medical sector and other disciplines, has been largely supplanted by automated technology in the modern world.

No age group is more vulnerable to skin disorders. Skin disorders affect people of all ages, including babies and the elderly. Skin disorders can be difficult to identify, especially when many conditions display the same or similar symptoms. Because of the intricacy of the skin's structure and the illnesses' seeming proximity, determining the actual kind of skin problem can be difficult. Diagnosing skin disorders can be difficult due to a variety of variables, including the difficulty in segmenting and examining the skin due to the presence of hair, sweat, and other unattractive aspects. Because many skin conditions resemble one another, different dermatological conditions can be difficult to distinguish from colored photographs. This is a significant problem for computer vision. One of the biggest disadvantages of large-scale biomedical image processing features is the difficulty in structuring and extracting critical information from data. Digital photographs taken with a camera, which may have noisy images, poor quality, or inconsistent lighting, aggravate skin problems. Because of lesion irregularities such as skin and hair coloring, it may be difficult to detect skin disorders. [1–5] Skin issues [6] are usually unexpected and difficult to diagnose. Some skin disorders may be difficult to identify and categorize. Many people in other nations cannot afford to see a dermatologist. [7]. As a result of mobile phone use, skin disease detection has become more economical in underdeveloped nations. Image processing in devices such as cameras and cell phones is used to evaluate skin disorders. We can address an issue in two steps by developing a purpose. In the first stage, image processing is used to identify skin problems, and then a deep learning system is used. Skin illnesses are difficult to diagnose in both the early and late stages due to changes in the distinguishing properties of the skin, such as color and texture [8]. Deep learning technologies developed for analyzing skin disease sample data might be applied to solve this problem.

The use of computer-assisted technology in this industry facilitates diagnostics and helps to reduce mistakes. These automated procedures usually make use of machine learning technology based on artificial intelligence. This paper presents an automated deep learning-based technique for diagnosing skin diseases. Diseases are predicted using a deep convolutional neural network (CNN).

The following is the paper's structure: In this section, beginning with section 2, we have provided a quick synopsis of the literature review. Section 3 then outlines the article's strategy. Section 4 presents the results and opinions. Section 5 closes the papers with a brief comment on future research.

![Figure 1. Sample representation of the dataset.](image)

2. Related Work

A lot of previous articles have discussed dermatological categorization. [9] carried conducted the study by improving a deep convolutional neural network (CNN) that has previously been taught to spot skin cancer. [10] used the DermNet dataset to test the efficiency of traditional machine-learning approaches.

[11] used transfer learning with AlexNet to identify a melanoma, typical nevi, and atypical nevi. [12] used transfer learning with VGGNet to classify skin lesions as benign or malignant. [13] used transfer learning on Google Net to classify the eight basic kinds of skin illnesses (basal cell carcinoma, melanoma, melanocytic nevus, and so on).
[14] used Google Net as a transfer learning strategy to classify different skin diseases. In [15], many deep learning models were utilized, including VGG16, MobileNet, VGG19, and InceptionV3.

[16] proposed developing an Android app that could diagnose seven ailments using transfer learning and the MobileNet methodology. They employed an uneven dataset using data augmentation strategies such as under-sampling, oversampling, and default preprocessing. They use a simple technique and have encouraging results; however, they don't provide the datasets they used or the specific categories of skin illnesses that were useful. The approach has a 94.4% accuracy rate.

[17] categorized seven disorders using two alternative methods. The first way is the Alone CNN model, while the second is the One-versus-All and CNN strategy. The preprocessing step is skipped in this manner. Combining the one-versus-all-and-alone CNN themes is smart and wonderful. It was feasible to achieve accuracy rates of up to 92.9%.

[18] described a strategy for categorizing the seven skin disorders. The steps of this CNN-based approach are preprocessing, a deep learning algorithm, model training, model validation, and classification. They didn't say anything about the conditions or the dataset they utilized. Neither CNN nor data pre-processing was mentioned. Approximately 93% of the categories were correct.

[19] used the fine-tuning direct technique and pre-trained Google's EfficientNet-b4 with seven auxiliary classifiers applied to each intermediate layer group to identify fourteen diseases with 94.8% accuracy. To expose the hidden image properties that were explored, the proposed approach employs t-distributed Stochastic Neighbour Embedding. The dataset and sickness kind are not supplied.

ANN, CNN, KNN, and GAN are included in [20] for classifying graphic images. Each calculation has its advantages and merits. Deep CNN performs comparative analysis with the highest accuracy of 98.79%. It gives the best results compared to other algorithms.

For the two-class classification of [21], the accuracy achieved by RESNet50 is 50.50%, VGG is 82.49%, SVM is 83.48%, Vit 84.31%, and CNN 97.61%.

Regarding three-class classification, [22] studied many machine learning and deep learning algorithms. The support vector machine achieved the highest accuracy among all machine learning approaches (92.54%), and the pre-trained Xception model achieved the highest accuracy (99.33%) among all deep learning models.

In [23], the CNN model Inception-ResNetV2 is mainly used to solve the 23-class skin classification problem. Approximately 93% of the categories were correct.

### 3. Methodology

#### 3.1. Data Collection

The dataset serves as the foundation for all algorithms, models, and systems. Approximately 25000 color images of skin disorders were imported for this project using several Kaggle [24] datasets, each of which focuses on a different type of skin disease. Fig.1 shows the sample representation of the dataset.

Eight common skin conditions are covered by around 25000 colored photographs in the collection: dermatofibroma, melanocytic nevus, melanoma, squamous cell carcinoma, actinic keratosis, basal cell carcinoma, benign keratosis, and vascular lesion. The photographs in the collection depict various body parts. Following the import of the dataset, 80% of the images are used for training, 10% for validation, and 10% for testing.

#### 3.2. Pre-processing

Image pre-processing is the process of improving image quality and producing the desired outcomes. Images may contain disturbing components such as air bubbles, hair, and other noises. Image Pre-processing includes:

**Image resizing**

Because the gathered photos come in a range of sizes, the input image into the recommended CNN is shrunk to 224 by 224 dimensions. The image's data was still intact at this size.

**Image Standardization**

As part of data normalization, the input image is turned into a set of pixels in the [0,1] range. The pixel values of each picture, which vary from 0 to 255, are normalized to a range of pixels from 0 to 1 by dividing each pixel by 255.

#### 3.3. Methods

Transfer learning is a strategy for enhancing model performance on smaller datasets by reusing models trained on larger datasets. In this part, we'll look at ways to improve classification accuracy by combining transfer learning with previously trained CNN models and feature extraction from MRI images.

**VGG16**

The well-known CNN model architecture VGG16 was introduced at ILSVRC-2014. Figure 2 depicts the model's construction. It outperforms AlexNet by sequentially replacing many 3x3 kernel-sized filters with large kernel-sized filters. To build the VGG16 model, we first set the input image size to 224x224. The top of the model is removed, the global average layer is added, and a prediction layer with softmax activation is added before importing the pre-trained VGG16 model from Keras using the Imagenet
dataset weights. We use the Adam optimizer, cross-entropy loss, and 50 iterations to train the model.

**VGG19**
To adapt the VGG19 model to our objective, the input layer will be modified to take images as input. The model's output layer will be replaced with a new softmax layer with eight nodes to represent the various forms of skin disease. To focus on our aim, we will alter the weights of the remaining layers while freezing the weights of the model's initial few layers, which are in charge of feature extraction. Because it has more layers than the VGG16 model, the VGG19 model can capture images with more complex attributes.

**ResNet**
In this study, three alternative ResNet models—ResNet50, ResNet101, and ResNet152—were employed. Deep convolutional neural network models known as ResNet models have succeeded in image classification applications. The skip connections in the ResNet models allow the models to create and learn residual functions. ResNet50 has 50 layers and is somewhat shallower than the other ResNet models. ResNet101 is a deeper model than ResNet50 since it includes 101 layers. ResNet152 is the most complex ResNet model, with 152 layers. The ResNet models were pre-trained using the ImageNet dataset, which offers a solid basis for employing transfer learning to fulfill our aim of recognizing skin illness.

**InceptionV3**
The InceptionV3 model's convolutional neural network design was initially revealed at the ImageNet Large-Scale Visual Recognition Challenge. To extract complex characteristics from input data, the InceptionV3 model includes convolutional layers with variable filter sizes and average pooling layers. The ImageNet dataset's more than one million annotated pictures from 1,000 categories were used as pre-training data for the InceptionV3 model's 48 convolutional layers.

**DenseNet**
In our analysis, we employed three distinct DenseNet models: DenseNet121, DenseNet169, and DenseNet201. DenseNet models, a type of convolutional neural network, primarily addressed the vanishing gradient issue in deep networks. DenseNet models incorporate direct skip connections between each layer, which improves gradient flow and network information flow during backpropagation. DenseNet121 has 121 layers, DenseNet169 has 169 layers, and DenseNet201 has 201 layers. As the number of layers grows, models get more complicated and contain more layers. Because all DenseNet models were previously trained on the ImageNet dataset, transfer learning provides a solid foundation for classifying skin disorders.

**MobileNet**
In our research, we employed the MobileNet model. MobileNet, a lightweight convolutional neural network model, was created for embedded and mobile vision applications. The MobileNet model employs depthwise separable convolutions to minimize processing costs and parameter counts while retaining excellent accuracy. When compared to other deep learning models, the MobileNet model is relatively simple, with only 28 layers. The ImageNet dataset was used to pre-train the MobileNet model, which provides an appropriate initialization for transfer learning to our aim of recognizing skin conditions.

**Xception**
The Xception model has excelled in a variety of image classification applications, including medical image analysis. By utilizing depth-wise separable convolutions, the model's capacity to generalize to new data is improved while the number of parameters is reduced. We classified MRI scans for skin cancer into groups with distinct subgroups in this investigation. The model's hyperparameters will be changed, and its effectiveness will be measured using accuracy, recall, and precision. We will train, validate, and test the model using the Kaggle dataset, which includes images of skin cancers. Table 1 lists the hyperparameters reported in each model.

<table>
<thead>
<tr>
<th>Hyper-parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image size</td>
<td>224 x 224</td>
</tr>
<tr>
<td>Weight</td>
<td>ImageNet</td>
</tr>
<tr>
<td>Epochs</td>
<td>50</td>
</tr>
<tr>
<td>Batch size</td>
<td>32</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Loss</td>
<td>Categorical cross-entropy</td>
</tr>
</tbody>
</table>

### 3.4. Fine-tuning
Transfer learning models' pre-trained weights were downloaded. The models are made up of an input layer, a convolutional layer, a batch normalization layer, and an activation layer. There were no obvious higher strata. The Transfer learning model's layers were completely frozen, but the weights survived. The Transfer learning model has been pre-trained except for the final layer. The output layer is made up of the global average pooling layer and a dense layer with a softmax activation function. The Adam optimizer and the categorical cross-entropy loss function were used to generate the model [25]. The model was trained using 50 epochs and 32 batches. The model's
efficacy was assessed using the test set. Figure 2 depicts the proposed fine-tuned design.

![Proposed Fine-tuned Architecture](image)

**Figure 2. Proposed Fine-tuned Architecture**

### 4. Results & Discussion

#### 4.1. Experimental Setup

The training plan used for this continuing work meets the following criteria: AMD Ryzen 7 5800H with a 3.20 GHz Radeon graphics processor. Our computer was built with a 512 GB SSD, a 64-bit operating system, and 32 GB of RAM. The experiment was carried out using an NVIDIA RTX 3050 GPU.

#### 4.2. Performance Evaluation

**Accuracy**

This formula determines the proportion of occurrences properly categorized in the test dataset. When there is a class imbalance or when misclassifying one class would cost more than misclassifying the other, accuracy might be deceptive.

\[
\text{Accuracy} = \frac{tN + tP}{tN + tP + fN + fP} \tag{1}
\]

**Precision and Recall**

Precision is the ratio of observed positives to expected positives, as opposed to recall, which is the percentage of projected positives to actual positives. When there is a class imbalance, these metrics are useful because they give a more thorough knowledge of how each class is doing using the model.

\[
\text{Precision} = \frac{tP}{tP + fP} \tag{2}
\]
\[
\text{Recall} = \frac{tP}{tP + fN} \tag{3}
\]

True positives (tP), true negatives (tN), false positives (fP), and false negatives (fN) are the units of measurement for the aforementioned metrics [26]. Fig. 3 provides a proper comparison of all deep learning models.

![Precision, Recall, Accuracy, and Loss Plot](image)

**Fig. 3.** Precision, Recall, Accuracy, and Loss Plot: (a) VGG16 (b) VGG19 (c) ResNet50 (d) ResNet101 (e) ResNet152 (f) Xception (g) DenseNet121 (h)
DenseNet169 (i) DenseNet201 (j) MobileNet (k) InceptionV3.

Table 2. Metric measurement of several transfer learning models

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy(%)</th>
<th>Precision(%)</th>
<th>Recall(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Valid</td>
<td>Test</td>
</tr>
<tr>
<td>VGG16</td>
<td>96.29</td>
<td>68.69</td>
<td>67.51</td>
</tr>
<tr>
<td>VGG19</td>
<td>95.50</td>
<td>68.01</td>
<td>68.91</td>
</tr>
<tr>
<td>ResNet50</td>
<td>96.15</td>
<td>71.99</td>
<td>71.45</td>
</tr>
<tr>
<td>ResNet101</td>
<td>95.86</td>
<td>72.15</td>
<td>72.75</td>
</tr>
<tr>
<td>ResNet152</td>
<td>95.78</td>
<td>74.24</td>
<td>73.01</td>
</tr>
<tr>
<td>Xception</td>
<td>63.97</td>
<td>59.42</td>
<td>60.62</td>
</tr>
<tr>
<td>DenseNet121</td>
<td>73.65</td>
<td>63.51</td>
<td>64.66</td>
</tr>
<tr>
<td>DenseNet169</td>
<td>72.77</td>
<td>64.14</td>
<td>64.30</td>
</tr>
<tr>
<td>DenseNet201</td>
<td>80.87</td>
<td>64.76</td>
<td>65.54</td>
</tr>
<tr>
<td>MobileNet</td>
<td>74.85</td>
<td>69.89</td>
<td>59.74</td>
</tr>
<tr>
<td>Inception</td>
<td>59.91</td>
<td>56.54</td>
<td>59.02</td>
</tr>
</tbody>
</table>

4.3. Results

Deep learning techniques that were recently released were used to analyze and automatically learn the distinguishing qualities required to find hidden patterns in raw data. We used a large, eight-class dataset of dermatological disorders that was freely available to the public for our study. We provide standard measures for the model prediction module, such as accuracy, precision, and recall, to evaluate the competency of each model. According to the outcomes of the investigations, all deep learning-based models outperform traditional machine learning models. Furthermore, we can classify skin diseases using the Restnet152 model at a rate of nearly 75% across all metrics.

RestNet152 looks to beat all other deep learning models because the more layers it has, the better it is at detecting hidden properties in difficult input data. Table 2 examines the precision, recall, and accuracy of each skin disease categorization model.

5. Conclusion and Future Remarks

Skin problems are a common and unpleasant ailment. As a consequence of recent advances in deep learning in medical imaging, we are encouraged to employ digital images to diagnose skin issues. Transfer learning is one of the most popular and advanced approaches for assessing medical pictures [27]. Component extraction substantially aids in the classification of skin disorders. We do extensive testing using seven deep-learning approaches.

ResNet152 outperformed all others with a great accuracy of 74.24% in validation and 73.01% in the test dataset, precision of 75.45% in validation and 75.30% in the test dataset, recall of 73.04% in validation and 71.71% in the test dataset.

The bulk of writers concentrate on identifying and categorizing a specific set of skin problems. The proposed strategy efficiently overcomes several difficulties, including unequal distribution of photos for each kind of skin disease in the dataset under consideration, differences in image lighting, similarities between some of the skin conditions in the dataset, and others. Furthermore, because the method is used for multiple body parts, training the network is more difficult.

Because there hasn't been much research in this field, there's a lot of research fields for new categorization or automatic machine analysis methods to be developed. Future performance enhancements will include increasing the size of our database and introducing new deep neural network models, both of which will increase the model's accuracy.
References


