

Novel Method for Dermatological Disease Detection using Machine Learning

Geo George K

Dept. of Computer Science and Engineering,
Albertian Institute of Science and Technology
Ernakulam, India

Anjana R

Dept. of Computer Science and Engineering,
Albertian Institute of Science and Technology
Ernakulam, India.

Sourav Sujith

Dept. of Computer Science and Engineering,
Albertian Institute of Science and Technology
Ernakulam, India.

Albin Reji

Dept. of Computer Science and Engineering,
Albertian Institute of Science and Technology
Ernakulam, India

Chinnu Edwin A

Dept. of Computer Science and Engineering,
Albertian Institute of Science and Technology
Ernakulam, India.

Abstract - Dermatological diseases are prevalent in our world for many reasons such as atmospheric conditions, pollution, side effects on the skin, etc. These dermatological diseases can include diseases caused by fungal infections, allergies, viruses, bacteria, etc. Technological development in the field of medicine has made diagnosis [1] more accurate and less time-consuming. But this diagnostic method is expensive and ordinary people cannot afford it every time. The system takes an image of the infected skin part when fed the image, and the machine learning techniques and detected diseases are displayed as output. The proposed system is thus effective in areas where there are fewer dermatologists.

I. INTRODUCTION

Skin Diseases are common among all age groups, it can occur due to the lifestyle and change in the environmental conditions. It is seen that out of five people at least one is affected with skin disease of any class. Skin diseases are occurring in almost all groups of ages among people. The rate of skin disease has increased due to lifestyle and changing environments. It is observed that every one out of five people [2] are infected with any kind of skin disease. They are usually caused by factors like different organism's cells, a different diet, internal and external factors, such as the hierarchical genetic group of cells, hormones, and the immune system of conditions. These factors may act together or in a sequence of skin disease. There are chronic and incurable diseases, like eczema and psoriasis, and malignant diseases like malignant melanoma. Some of the medical researchers have found the availability of cures for these diseases if they are detected in the early stages. Normally seen human behavior [3] is ignoring these skin abnormalities by their own curing method and sometimes acting like it will automatically get cured as time passes by. Such practices will make the case worse. The available diagnosis procedure consists of large laboratory tests but this paper proposes a system which enables the users to predict the dermatological diseases using computer vision.

The system enables the user to upload the image of the infected portion of the skin area and get the best result of the highest closest skin disease class along with probability distribution of the possibilities of

the other classes of diseases which gives a clear idea about the probability of the detected disease. The details about the detected disease can be obtained by web scraping the result.

The model is implemented in python using TensorFlow framework and trained using Convolutional Neural Network. SoftMax classifier is used for the classification of the skin disease images which is a high precision classifier.

II. RELATED WORKS

A. Dermatological disease detection using image processing and machine learning

In this research [4], we propose an approach to detect these diseases. We use a two-step approach that effectively combines computer vision and machine learning with clinically assessed histopathological features to accurately identify disease. In the first step, various pre-processing techniques are applied to the skin disease image, followed by feature removal. In the second step, machine learning algorithms are used to identify diseases based on histopathological features observed in skin analysis. After training and testing six diseases, the system produced up to 95 percent accuracy.

B. Automatic Classification of Clinical Skin Disease Images with Additional High-Level Position Information

In this paper [5], we improve the SD-198 dataset to form the SD-198-P dataset, which contains 198 classified 6584 images with the location information of the imaged skin disease. In addition, we evaluate the performance of two coding methods, One-Hot coding and Multi-Category One-Hot coding. The results show that the accuracy of the classification network improves well when the hyperparameters are set to the correct value. The complexity of time is too great. In this paper, we improve the SD-198 dataset to form the SD-198-P dataset, which contains 198 classified 6584 images with the location information of the imaged skin disease. In addition, we evaluate the performance of two coding methods, One-Hot coding and Multi-Category One-Hot coding. The results show that the accuracy of the classification network improves well when the

hyperparameters are set to the correct value. The complexity of time is too great.

C. Skin Lesion Classification by Ensembles of Deep Convolutional Networks and Regularly Spaced Shifting

Skin damage is caused by several factors [6] such as allergies, infections, exposure to the sun, etc. These skin diseases have become a challenge for medical diagnosis due to their visual similarity, where image classification is an important task to obtain an adequate diagnosis of the lesions. In this work, we propose an improved convolutional neural network with a test time periodic transfer technique for classification of skin lesions. The permutation technique constructs multiple versions of the test input image, which are permuted using permutation vectors located on a regular grid in the plane of possible permutations. These changed versions of the test image are then fed to each classifier in the ensemble. Finally, all the results of the classifiers are combined to get the final result

D. Skin Diseases Detection Based on Different Segmentation Technique

In this paper [7], image processing techniques such as adaptive thresholding, edge detection k-means clustering and morphology-based image segmentation were used to detect skin disease from given images. The disadvantage of this is that the influence of some classes reduces the classification system. The obtained image set was pre-processed with blurring, noise reduction and post-processed. Depending on the pattern present in the processed image (related to a specific disease), the disease is detected in the output of the corresponding input image.

III. DATASET

For this project we have used two types of datasets. The first dataset is used for identifying the skin disease types is the HAM 10000 dataset. In the HAM10000 Dataset a large version of dermatoscopic images of common pigmented skin lesions are found. It consists of 10000 images for detecting pigmented skin lesions. The authors collected these images from different populations acquired and stored by different modalities.

The dataset contains the seven classes of the skin

disease images which are the collection of important diagnostic categories which are listed below:

- Actinic keratoses and intraepithelial carcinoma (akiec).
- Basal cell carcinoma(bcc).
- Benign keratosis-like lesion (blk).
- Dermatofibroma (df).
- Melanoma (mel).
- Melanocytic nevi (nv).
- Vascular lesion (vase).



Fig.1. Sample Image Disease images in HAM 10000 Dataset.

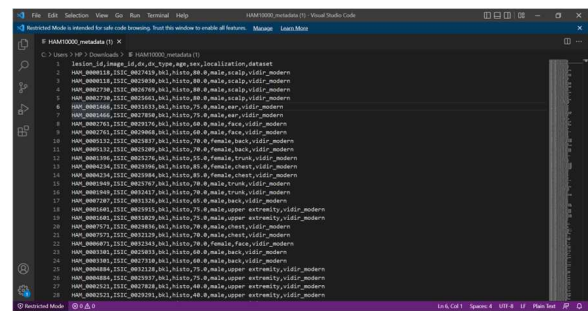


Fig.2. Sample Image of Metadata File in HAM 10000 Dataset.

The second dataset used is a lively collected healthy skin images of different age groups and different gender which are used for identifying healthy skin or no disease inputs images.

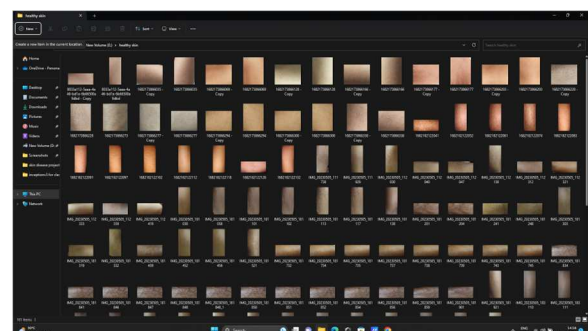


Fig.3. Sample Image Disease images of Healthy skin image dataset.

IV. METHODOLOGY

Common methods of diagnosing skin diseases include patient history and symptom analysis, skin scraping, visual inspection, dermoscopic examination, and skin biopsy. However, these diagnostic methods are tedious, time-consuming and prone to subjective diagnosis. Most of them require experience and excellent visual perception of the dermatologist. Advanced and powerful medical imaging techniques can also be used to diagnose skin diseases. However, these technologies are complex, expensive, and limited to centralized healthcare facilities, leaving under-resourced populations with limited access to dermatology services. The proposed works showed promising results in the diagnosis of various skin diseases based on clinical images. However, most of the work relied on the availability of public online databases, focusing on cancer and tumors and designed to diagnose specific body parts. In addition, the materials that can be collected and used consist mainly of white leather. In addition, diagnostic performance, including reported accuracy, is not satisfactory.

1. Data Preprocessing

The images resized to 300*300*3 and then converted to a numpy array. The numpy arrays are then stacked to create a numpy stack. The labels are converted to a numpy array which are later one hot encoded. The numpy stack of images undergoes normalization by division and the Image data generator is used for shearing, rotation, width shift, height shift, horizontal and vertical flip by image augmentation.

2. Image Classification

The proposed method uses 2 different models for image classification. First model is used for binary classification of healthy and unhealthy skin images. The second model is activated only if the image is predicted as unhealthy skin. It is a multiclass image classifier for predicting skin images into 7 different classes of diseases.

The first model is implemented by transfer learning and fine tuning the Inception V3 model. The Inception v3 is loaded with the ImageNet training weights. The custom dataset for binary classification

of images: healthy and unhealthy skin is loaded and preprocessed. The data is then split as training data and testing data as 80% and 20% of the entire data. The validation data is 20% of the training data. The Inception model is set to non-trainable and layers are frozen. The global maximum pooling layer is appended onto the base Inception v3 model whose layer is mixed10. Full connected layers of 512 nodes and dropout layers are added to the existing model. The final dense layer of 2 nodes with SoftMax activation function is appended to the existing model. The model is trained on training data for 100 epochs on the P100 GPU with 16GB. The weights of the model after training are saved as h5 and hdf5 files. The layers of the Inception v3 model are unfroze and then the entire model is loaded with the weights that were previously saved. The model continues to learn the features of the skin images by training on the dataset.

The inception v3 pretrained model is 48 layers deep and then 4 additional layers are appended to the model to obtain the probability distribution for 2 classes. During the second training, callbacks are implemented for the model. The callbacks are based on the early stopping and the learning rate reduction. The early stopping and reduction of learning rate on plateau is based on the validation loss during the training. The proposed method uses Adam optimizer for reduction of learning rate of the model with initial LR as 0.001. Categorical cross entropy is used as the loss function for the model with accuracy metric.

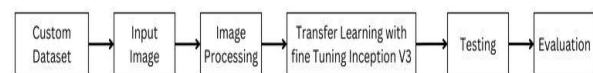


Fig.1. Working flowchart for prediction of healthy or unhealthy skin image.

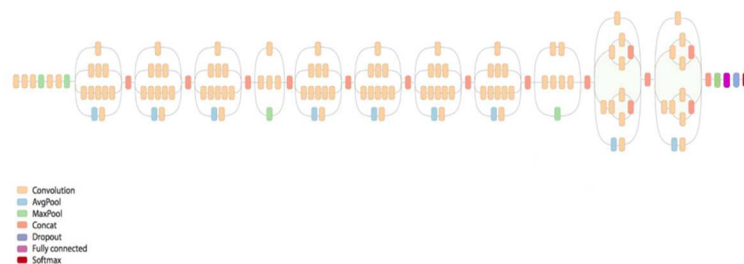


Fig.2. Architecture diagram of inception v3.

The second part of the proposed method is used for image classification on 7 different classes of skin diseases. The second model is Convolutional Neural Network implemented in a sequential form. The CNN consists of 2 blocks of a convolutional layer, followed by batch normalization and maximum pooling layers. 3 blocks of convolutional layers and batch normalization layers are appended to the model. This is followed by maximum pooling layer. The Flatten layer and two fully connected layers with dropout, and a final dense layer is added to the existing layers of the model.

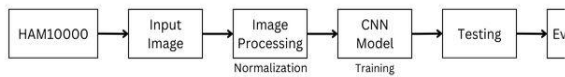


Fig.3. Working flowchart for the prediction of the type of skin disease.

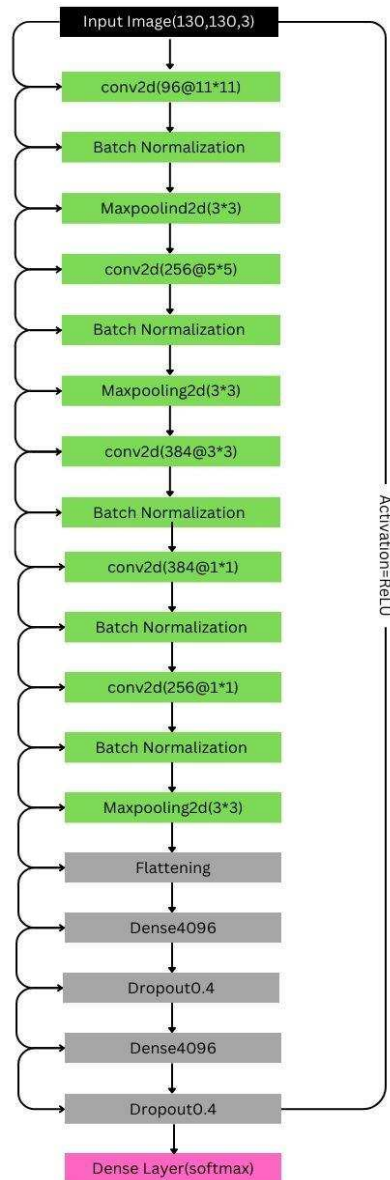


Fig.4. CNN Architecture Diagram.

This model contains above 85 million learnable parameters and 2752 non-learnable parameters. The HAM 10000 dataset is used for the training of the model. The proposed method uses this custom CNN along with HAM 10000 whose classes contain an increased number of images by up sampling them due the bias present in the dataset. The processed data is split into training and testing data in 8:2 ratio with validation data having 20% of the training data.

The input of the CNN is a 3-dimensional tensor onto which the kernels are applied. There are 96, 256, 384, 384, 256 numbers of filters applied in the first to the fifth convolutional blocks respectively. The batch normalization helps in mitigating the effects of covariate shift and helps the model to converge to a global optima if possible. The normalization helps in the input to be normalized so that it has zero mean and unit variance. Maximum pooling helps in down sampling the tensors in the intermediate levels so as to reduce the complexity and training time of the model. All pooling layers are implemented with a pool size of 3*3. The model is trained with callbacks as was implemented in the first model along with sparse categorical cross entropy as the loss function. The Adam optimizer uses adaptive learning rates and depends upon the gradient of loss function to update the parameters of the convolutional layers so as the move in the direction of the negative gradient of the loss function. The CNN uses the ReLU as the activation function in the convolutional layers. This introduces non-linearity into the model so the model can learn complex features that can be represented as feature maps in the intermediate layers of the CNN. The model is trained on a P100 GPU with 16GB with 100 epochs. The data is split into batches of 128 tensors for training of the model.

V. OBSERVATION

The training of the first inception model resulted in an accuracy of 98% as shown in fig1. Training the CNN model using clinical images resulted in an accuracy of 93% and the testing of the image resulted in an accuracy of 73% as shown in fig 3 and fig 4. The web app predicts results of real time skin images also that falls under the seven skin diseases categories. If the healthy skin image or is tested for prediction the results will be displayed as no diseases /unknown diseases. The results of the skin illness are displayed on a graphical user interface. A short description about the diseases is also made available to users using web scribing.

```

Test Accuracy: 98.810%
8/8 [=====] - 1s 67ms/step
      precision    recall  f1-score   support

   h1         0.99      0.98      0.99         130
   uh1        0.98      0.99      0.99         122

 accuracy                   0.99         252
 macro avg         0.99      0.99      0.99         252
 weighted avg      0.99      0.99      0.99         252
    
```

Fig.1. Test accuracy of Inception Model.

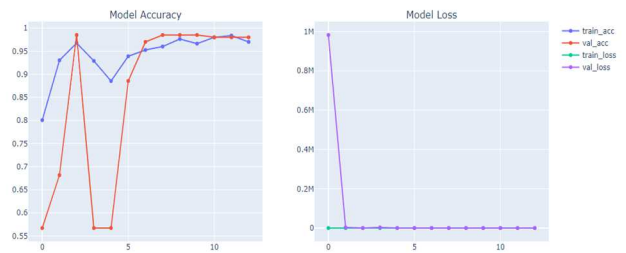


Fig.2. Model accuracy and Model Loss of Inception Model

```

Test Accuracy: 93.254%
      precision    recall  f1-score   support

   nv         0.98      0.99      0.98         431
   mel        0.94      0.93      0.93         432
   bkl        0.93      0.89      0.91         420
   bcc        0.89      0.92      0.90         433
   akiec      0.90      0.92      0.91         431
   vasc      0.97      0.97      0.97         428
   df         0.92      0.92      0.92         434

 accuracy                   0.93        3009
 macro avg         0.93      0.93      0.93        3009
 weighted avg      0.93      0.93      0.93        3009
    
```

Fig.3. Test accuracy of CNN Model.

```

Test Accuracy: 73.875%
22/22 [=====] - 0s 12ms/step
      precision    recall  f1-score   support

   nv         0.90      0.97      0.93         79
   mel        0.74      0.66      0.70        100
   bkl        0.66      0.65      0.65        114
   bcc        0.63      0.64      0.64         98
   akiec      0.65      0.62      0.63         97
   vasc      0.94      0.90      0.92        110
   df         0.67      0.77      0.72         91

 accuracy                   0.74        689
 macro avg         0.74      0.74      0.74        689
 weighted avg      0.74      0.74      0.74        689
    
```

Fig.4. Test accuracy of CNN Model.

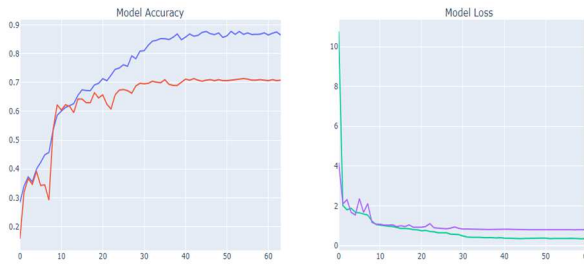


Fig.5. Model accuracy and Model Loss of CNN Model.



Fig.8. Image of a Healthy Skin.

On the implementation of the models on a web application for a dermatological disease detection using an uploaded image of the skin portion on fig 6 and fig 8, the results from fig 7 and fig 9 are observed respectively.



Fig.6. Image of a skin disease.

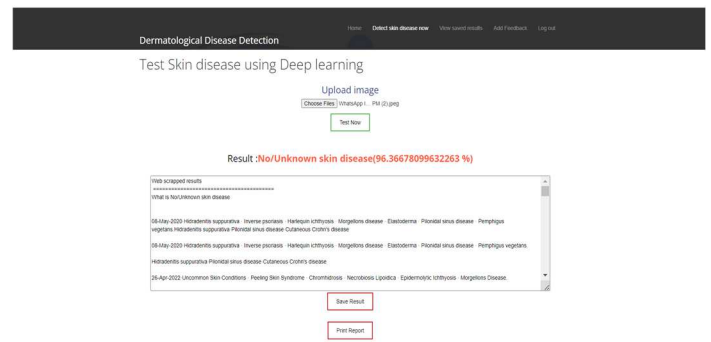


Fig.9. Tested Result on web application.

VI. CONCLUSION

This study proposes a web-based automatic diagnosis for seven common skin diseases based on deep learning techniques using clinical image and patient clinical data, and achieved average accuracy, precision, recall and F1 score, respectively. The results show that the developed system provides an excellent diagnosis for seven categories of skin diseases. The developed diagnostic system can be used as a decision support for dermatologists, family doctors, rural health doctors and patients in the diagnosis of skin diseases.

VII. REFERENCES

[1] Arifin, S., Kibria, G., Firoze, A., Amini, A., & Yan, H. (2012) "Dermatological Disease Diagnosis Using Color-Skin Images." Xian: International Conference on Machine Learning and Cybernetics. W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

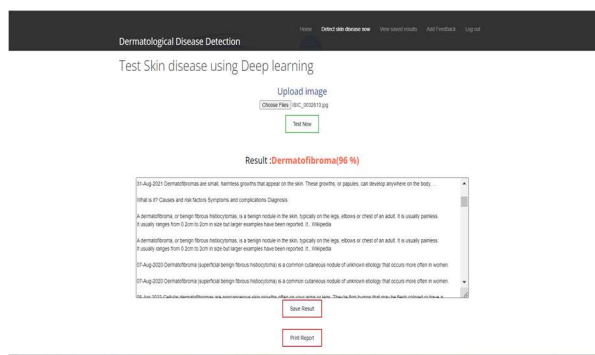


Fig.7. Tested result on the web application.

[2]. Nawal Soliman ALKolifi ALEnezi “A method of skin disease detection using image processing and machine learning” at 16TH International learning & Technology conference 2019

[3] Pravin S. Ambed, A S Shirsat “A image analysis system to detect skin diseases” at IOSR Journal of VLSI and signal processing Volume 6 Issue 5 Ver 1 e-ISSN 2013-4200

[4] Li-Sheng Wei, Quan Gan and Tao ji “Skin Disease Recognition Method Based on Image Color and texture features”Y. Computational and Mathematical Methods in Medicine Volume 2018, Article ID 8145713, 10 pages <https://doi.org/10.1155/2018/8145713>

[5] R. Sumithra, M. Suhilb, and D. S. Guruc, “Segmentation and classification of skin lesions for disease diagnosis,” *Procedia Computer Science*, vol. 45, pp. 76–85, 2015.

[6] Rahat Yasir, Md Ashiqur rehman and Nova Ahmed “Dermatological Disease Detection using image processing and Artificial Neural Network” 8th International Conference on Electrical and Computer Engineering 20-22 December, 2014, Dhaka, Bangladesh

[7] R. Yasir, M. S. I. Nibir, and N. Ahmed, “A skin disease detection system for financially unstable people in developing countries,” *Global Science and Technology Journal*, vol. 3, no. 1, pp. 77–93, 2015.