An Experimental Analysis for Classifying Skin Diseases Using Soft Computing Techniques

Sanjeela Sagar* 1 and Seema Murkar2

1Assistant Professor, Department of IT, Vidyalankar School of Information Technology, Mumbai, India
2Assistant Professor, Department of IT, Vidyalankar School of Information Technology, Mumbai, India

1sanjeela.sagar@vsit.edu.in, 2seema.bhatkar@vsit.edu.in

Abstract: The skin is an important part of the human body that cosmetically protects the body from infection. Skin diseases is highly popping up in the society as reported by researchers and medical practitioners (Wiley, 2019). Therefore, there is need to address this issue by introducing efficient treatments with the help of technology. Detecting the skin disease in their early stage will help both the patient and practitioner. To address this issue, we selected some popular machine learning algorithms, this paper focuses on comparing the efficiency of four machine learning techniques such as Logistic Regression, Support Vector Machines (SVM) and Random Forest with Inception V3 model of Transfer Learning for classification of skin diseases into five classes Acne, psoriasis, scabies, rosacea, and fungal infection. We have trained and tested the model using the skin disease images from the Dermnet, Dermis and Dermnet NZ dataset. The results have shown that transfer learning has resulted maximum accuracy as compared to other machine learning algorithms mentioned above.

Keywords: Skin diseases, Machine Learning, detection, transfer learning

1. Introduction

According to WHO, about 90 crore people are affected by Skin diseases at any time (Moloo, 2019). A disease that may be common to everyone is skin diseases. Skin infections are a significant medical issue influencing a high extent of the populace in India (Abolfotouh MA, 2000). Skin diseases will place an important emotional and psychological burden on patients which will be worse than the physical impact (J Ayer, 2006). If skin diseases are not treated at earlier stage, then it may lead to complications in the body including spreading of the infection from one individual to the other. The skin diseases can be prevented by investigating the infected region at an early stage. In this paper, proposed system is used for the classification of five classes of diseases (Acne, psoriasis, scabies, rosacea, and fungal infection) using transfer learning with an inception V3 model and then its result has been compared with other machine learning techniques (Logistic Regression, Neural Network, SVM and Random Forest).
2. Literature Review

Dermatological disorder like Eczema, Psoriasis and Planus were studied in 2017 by using various image processing techniques like blurring, contrasting, gray scaling, etc. under the project granted by department of BioTechnology, GOI, and they extracted the features using CNN and classification was done using SVM with an accuracy of 95% (Soumya Sourav, 2017). In 2018, again a comparative study of machine learning algorithms for skin diseases like eczema, psoriasis and plaque were studied using color and texture features of images and found that Linear Discriminant has shown better performance on colors as feature base and SVM shown better results when texture features were focused (Parameshwar R. Hegde, 2018). Later a research which was partially supported by the National Library of Medicine and National Institute of Allergy and Infectious Diseases of the National Institutes of Health for basal cell carcinoma, seborrheic keratosis and psoriasis was done and by using Inception V3 algorithm, an accuracy of 85.65 was achieved (Xinyuan Zhang, 2017). Study on Skin lesions was done, and a mobile app was developed and by using pretrained CNN latency of the system was reduced (Xiangfeng Dai, 2019). Skin lesion segmentation was done using graph cut method which was then followed by Naïve Bayes classification and an accuracy of 94.3% was achieved to classify benign disease (V.R. Balaji, 2020). Dull razor, Gaussian filter, k-means color clustering, GLCM and ABCD were used as image pre-processing features and then multi class SVM was used to detect skin cancer on ISIC challenge dataset and an accuracy of 96.25% was achieved (M. Krishna Monika, 2020).

3. Proposed Methodology

i) Images were collected from Dermnet NZ, DermIS.net. Since most of the images were not of same size. Hence, first they were made equal size.

<table>
<thead>
<tr>
<th>Class Id</th>
<th>Class Label</th>
<th>Total Images</th>
<th>No. of Training Images</th>
<th>No. of Testing Images</th>
<th>No. of Validation Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acne</td>
<td>300</td>
<td>240</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>Psoriasis</td>
<td>79</td>
<td>63</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Scabies</td>
<td>78</td>
<td>62</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Fungal</td>
<td>76</td>
<td>60</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>Rosacea</td>
<td>51</td>
<td>41</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

ii) Watermark removal, image thresholding, segment based on color, background subtraction methods along with GLCM features were extracted.

iii) Extracted features were labeled according to their classes as Acne, psoriasis, scabies, rosacea and fungal infection. These machine learning algorithms Logistic Regression, Support Vector Machines (SVM) and Random Forest were chosen based on literature review and compared with Inception V3 model of Transfer Learning.
3.1. Logistic Regression

Logistic Regression is a classification model which uses a logistic function to frame binary output model. The output of the logistic regression will be a probability $0 \leq x \leq 1$, and can be used to predict the binary 0 or 1 as the output (if $x < 0.5$, output = 0, else output = 1). In this paper we have used Multinomial Logistic Regression as we have to classify more than two classes. The underline technique will be same as the logistic regression for binary classification until calculating the probabilities for each target. Once the probabilities were calculated. We need to transfer them into one hot encoding and uses the cross-entropy methods in the training process for calculating the properly optimized weights.

![Logistic Model](image1.png)

**Figure 1: Logistic Model** (Polamuri, 2020)

3.2 Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. It can easily handle multiple continuous and categorical variables. SVM constructs a hyperplane in multidimensional space to separate different classes. SVM generates optimal hyperplane in an iterative manner, which is used to minimize an error. The core idea of SVM is to find a maximum marginal hyperplane (MMH) that best divides the dataset into classes.

![Support Vector machines](image2.png)

**Figure 3: Support Vector machines** (Navlari, 2019)
3.3. Random Forest

Random forest is a supervised learning algorithm. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

![Random Forest Diagram](image)

*Figure 4: Random Forest (Donges, 2019)*

3.4. Transfer Learning

Conventional machine learning and deep learning algorithms, so far, have been traditionally designed to work in isolation. These algorithms are trained to solve specific tasks. The models have to be rebuilt from scratch once the feature-space distribution changes. Transfer learning is the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired for one task to solve related ones. So, we have used Transfer Learning Inception V3 architecture to build our model. Inception v3 architecture model trained on ImageNet images, and train a new top layer that can recognize other classes of images. The top layer receives as input a 2048-dimensional vector for each image. We train a softmax layer on top of this representation. The softmax layer contains 5 labels of 5 classes, this corresponds to learning 5 + 2048*5 model parameters corresponding to the learned biases and weights.

4. Implementation Training and Testing

4.1. Logistic Regression

In this technique 75% of data was used for training and remaining 25% was used for testing. Multinomial Logistic Regression was used with 10- fold cross validation and stratified sampling method was chosen.
4.2. Support Vector Machine (SVM)

We built SVM model with Radial Basic Function (RBF) Kernel, cost (C): 1.00, Regression loss epsilon: 0.10 and have trained this model for 100 steps.

4.3. Random Forest

Random forest model was built with 10 trees.
4.4 Transfer Learning

Inception V3 model was used with ‘relu’ as activation function and ‘adam’ for model optimization.

```
<table>
<thead>
<tr>
<th>Actual</th>
<th>Acne</th>
<th>Fungal</th>
<th>Psoriasis</th>
<th>Rosacea</th>
<th>Scabies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acne</td>
<td>559</td>
<td>11</td>
<td>8</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Fungal</td>
<td>24</td>
<td>70</td>
<td>11</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>12</td>
<td>6</td>
<td>127</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Rosacea</td>
<td>31</td>
<td>6</td>
<td>6</td>
<td>52</td>
<td>5</td>
</tr>
<tr>
<td>Scabies</td>
<td>18</td>
<td>33</td>
<td>12</td>
<td>7</td>
<td>90</td>
</tr>
<tr>
<td>Σ</td>
<td>644</td>
<td>126</td>
<td>164</td>
<td>83</td>
<td>153</td>
</tr>
</tbody>
</table>
```

**Figure 8. Confusion Matrix for Transfer Learning:**

5. Results and Discussions

**Table 2: Result of the Experiment**

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>CA</th>
<th>F1</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.924</td>
<td>0.733</td>
<td>0.693</td>
<td>0.734</td>
<td>0.733</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.847</td>
<td>0.655</td>
<td>0.618</td>
<td>0.621</td>
<td>0.655</td>
</tr>
<tr>
<td>Neural Network</td>
<td>0.924</td>
<td>0.768</td>
<td>0.760</td>
<td>0.756</td>
<td>0.768</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.930</td>
<td>0.776</td>
<td>0.769</td>
<td>0.765</td>
<td>0.776</td>
</tr>
</tbody>
</table>

It is observed from the experiment that random Forest and SVM are not the good classifier models in this dataset and Logistic regression and Inception V3 model performed well. Inception v3 model has the capacity to extract the features for training and then it could be used for classification.

6. Conclusion

Considering the benefits offered by the Transfer Learning model to recognize images we try to use them for skin disease classification. We have implemented an inception V3 model by retraining final layer of ImageNet for the task of classifying skin diseases and have achieved high accuracy as compared to Logistic Regression, Neural Network, SVM and Random Forest. We will try to increase the accuracy of inception V3 model by adding more dataset. We will train and test this model on other types of skin diseases and with healthy skin.
7. References


