INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3456 - 3460
RECEIVED : 02.04.2018 PUBLISHED : 25.04.2018
April 25, 2018

DERMOTOLOGICAL DISEASE DETECTION USING IMAGE PROCESSING AND CONVOLUTION NEURAL NETWORK

[1]S.Niveditha, [2]Pullela Sneha, [2]Lipsa Sahu, [2]Semanti Roy
[1]Faculty, Department of Computer Science and Engineering
SRM Institute of Science and Technology
[2]UG Scholar, Department of Computer Science and Engineering
SRM Institute of Science and Technology
Corresponding Author: lipsa.sahu017@gmail.com

Abstract – Skin diseases constitute a major health problem around the world. Human skin diseases are one of the most unpredictable and are most difficult diseases to diagnose. In our paper we have proposed some methodologies that use computer vision-based techniques to detect various kinds of dermatological disorders. We have used different types of Deep learning algorithms such as Inception_v3, Mobilenet , Inception resnet_v2 for feature extraction and Machine learning algorithms such as Random forest , Logistic Regression and linear discriminative algorithm for training and testing purpose. This is followed by the ensemble features in order to analyse and identify the skin elements. We have considered area of interest as image extracted feature, average colour code of infected area, shape, area size and also the texture. The approach operates on two phases; pre-processing of the colour skin images is done at first to extract significant features and recognition of the skin elements is done later.
Index Terms - Skin Diseases, Mobilinet, Inceptioin resnet_v2, Machine Learning

1. INTRODUCTION

Dermatology remains unpredictable and complex terrain due to diagnosis complexity. Skin diseases are most prevalent and feared disease of this era, causing mental and physical damage to the patient. They must be treated at the initial stage to prevent havoc. They vary in the intensity of damage, and ease of treatment. Late detection could lead to permanent damage that would never be reversed. With increased industrial exposure and chemicals there's a greater risk of these diseases. Considering the similarity of symptoms makes the process of identification a more complex one. A less importance has been given to skin disease. Though we have a range of health apps, a negligible importance has been given to this field. With new emerging rise in technology and growth of neural networks has boosted to these fields. With a proper dataset, a better model could be trained to predict the disease. A wide range of Machine Learning and deep learning techniques are used to build a more accurate system. Dermatitis type may be different but some cases have common symptoms like swells, redness, itching. They have been ranked under top five occupational diseases. Bacterial infections are among major reason for skin infestation.

Still many don't know bacterial infection can cause akin disease, MYCIN was initially found for this purpose. Treatment or prediction is still a unaffordable for certain section of the population who might misunderstanding the symptoms lightly, or those far from development might never get them. Dermatologist with high expertise and knowledge becomes a grave question in those areas. Limelight must be given to prevent future crisis. Detecting different types of skin disease is a burdensome piece of work in computer vision. Figuring out

different features from c infected area and predicting the disease is our main purpose. We can use camera technology to exploit image processing to a considerable extend. This leads to inexpensive diagnosis. This might reduce the need for biopsy. But the difficulties remain high due to different skin textures and common symptoms that might confuse. Machine learning has gone beyond, stimulating like how a human might think. Different algorithms like LBP, HOG, are used in recent days based on need. Heraldic texture is used for those datasets with varying textures. Artificial Neural network reigns with its neurons and synapses. The structure is fixed within the hidden layers but the layers are not so, this help us to add any number of layers based on the actual need. But the actual problem that remains here is calculations of each stage count to huge making computation a tedious one. Connecting each neuron with each other in the next hidden layer and then trying to reduce it is a complete task and resource consuming one. These have been improved by Convolution Neural Network by eliminating certain weight calculation, thus reducing the number of calculations totally required at the end. We use a totally new technological hype Tensor Flow. Tensor Flow is an open source library for this purpose trained by Google. The ultimate task of computing the weights at each neuron, calculating the error during activation function, have been simplified by Tensor Flow. A wide range of classifiers is available like SVM, KNN, Logistic Regression, Linear Regression, Naive BaysClassifier, Random Forest, Decision Trees. Actually, they can be even well defined as classification type or regression type. If the prediction is to say a discrete value, say 'yes' or 'no' classifications used for continuous values, like a range between 0-100, we say regression. Many pre-defined models are openly available for future research. They were trained for general objects like identifying a panda for example. We use these pre-trained models and their weights for building our own systems of use. Many models like Inception, Resnet, MobileNet, Xception are currently under extensive use. Based on use different models can be adapted. For example any application on an android needs a light weight architecture. For those purpose a MobileNet would do good with light weight configuration. But largely due to poor dataset their classification remains little inaccurate. We overcame this problem by using a new technique called transfer flow technique. The layer before classification is fetched and connected to a better classifier. In this way we try maintaining the accuracy high.

2. RELATED WORK

Detection of different types of skin diseases from colour images is a very gruelling chore in computer vision. Figuring out different facets from the colour skin images of the infected area of different skin diseases and recognising them with a high accuracy rate is the primary purpose of this research.

Researchers are working on a variety of algorithms that can be used to recognise different types of skin diseases. Kabari et al crated a artificial neural network system that prescribes diagnosis and routine treatment for skin diseases and the accuracy rate of the proposed paper is 90%. M. SHAMSUL et al proposed an automated dermatological diagnostic system. They have deployed different pre-processing algorithms like ours and used feed forward back propagation artificial neural network for training and testing purpose. Shuzlina et al embellished a prototype with back propagation neural network to behold the different elements.Nidhal et al proposed a prototype using feed forward artificial neural networks that identifies psoriasis . Hashim et al used different matlab tools for image processing and Wahab et al proposed a texture classification system. Gerald et al have introduced an

approach that produces accurate overlays of thermal and visual medical images. Chang et al have put forward a prototype that detects the skin defects with an accuracy rate of 98.0%.

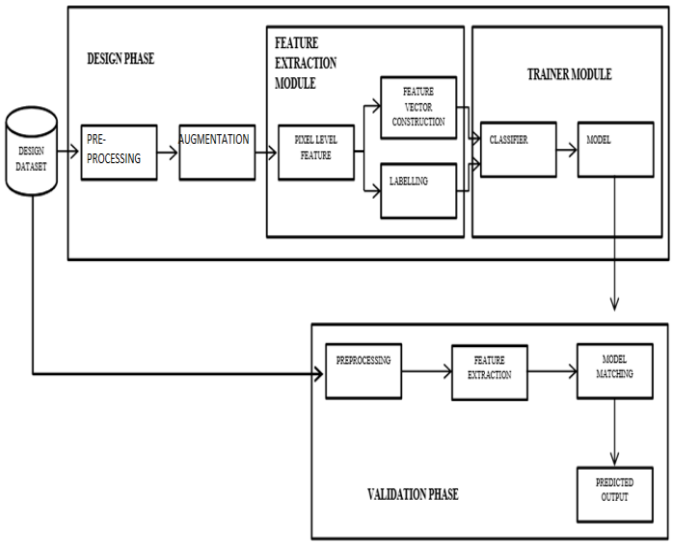
In our paper we have used a similar approach in order to identify different types of skin diseases from colour images. The main difference between their system and our system is, they have considered area of interest as image extracted feature . we have also considered average colour code of the infected area, shape and texture.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

The approach operates on two phases; pre-processing of the colour skin images is done at first to extract weighty features and recognition of different skin elements is done later.

At first, we are applying different image processing algorithms on the colour skin images to find some visual pattern and suggestive attributes like average colour code of infected area, infected area size in terms of pixels and shape or edge detection of an infected area. And then we train the system with the extracted features.

The system has two main phases, namely the DESIGN Phase and VALIDATION Phase. Both phases have common sub-modules except for the final model matching step. Number of images always makes a difference in the accuracy. It's always better to have maximum number of images per class. We bring up this count by making several operations like inter-zooming operation. A shift of one pixel may not be a different one for human eye but definitely it adds to new score in machine learning. We uniformly batch process all images to one common format, here we have taken .jpg format as global one. We crop down the area of concern. This way we help up bring the accuracy at later stage.



3.1 FEATURE EXTRACTION

Two different pre-trained models are used for making an analysis on accuracy. Both have similar architecture but differ only in their hidden layers and the datasets used to model them. MobileNet , InceptionV3 and inception_resnet_v2 are used for feature extraction. MobileNet is considered to have light weight architecture and fast model, more preferred for

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3456 - 3460
RECEIVED : 02.04.2018 PUBLISHED : 25.04.2018
April 25, 2018

mobiles and embedded application. With small size (17MB), they are based on streamlined architecture that uses deep-wise separate convolutions. Though theses process same as inception these have light weights. Inception v3 consists of two parts, a convolutional neural network, fully connected layer. The model pulls out general attributes from input images in the first phase and categorises them based on those features in the second phase. To build a new model to categorise our original dataset, we reuse the feature extraction part and re-train the systemisation part with our dataset,in transfer learning.

3.2 TRAINING MODULE

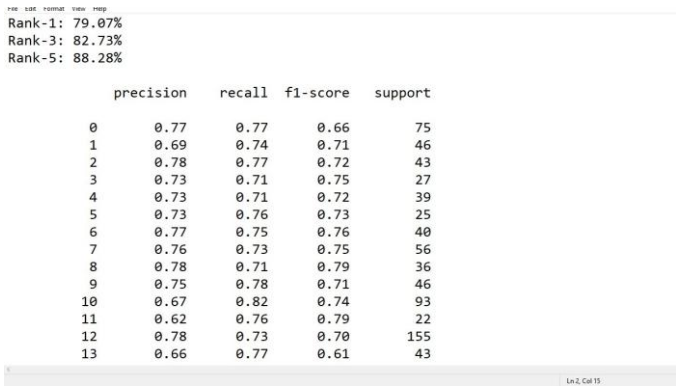
The features extracted above module will be trained. A classifier is used for predicting the output it compares the score of an input image against each class and picks the closest match as the predicted class. We have taken Logistic Regression, Linear discriminative analysis and Random forest for classification purpose. This type of classifier works best on continuous inputs rather than a discrete vale like ‘yes’ or ‘no’.

3.3 ENSEMBLE METHODS

The term ensemble methods generally refer to training a large number of models where the exact value of large depends on the classification task and then combining their output predictions via voting or averaging to yield an increase in classification accuracy. Techniques such as AdaBoost and Random Forests are the quintessential examples of ensemble methods. In Random Forests, we train multiple Decision Trees and use our forest to make predictions. Random Forest consists of multiple decision trees aggregated together. Each decision tree votes on what it thinks the final classification should be. These votes are tabulated by the meta-classifier, and the category with the most votes is chosen as the final classification. The same concept can be applied to deep learning and Convolutional Neural Networks. Here we train multiple networks and then ask each network to return the probabilities for each class label given an input data point. These probabilities are averaged together, and the final classification is obtained.

4.EXPERIMENTS AND RESULTS

90 % of the data collected is used for training and rest 10 % is used for testing the models.The matrices considered for evaluation are precision, recall ,F1 score and support value.



```
Rank-1: 79.07%
Rank-3: 82.73%
Rank-5: 88.28%

              precision    recall  f1-score   support

           0       0.77      0.77      0.66        75
           1       0.69      0.74      0.71        46
           2       0.78      0.77      0.72        43
           3       0.73      0.71      0.75        27
           4       0.73      0.71      0.72        39
           5       0.73      0.76      0.73        25
           6       0.77      0.75      0.76        40
           7       0.76      0.73      0.75        56
           8       0.78      0.71      0.79        36
           9       0.75      0.78      0.71        46
          10       0.67      0.82      0.74        93
          11       0.62      0.76      0.79        22
          12       0.78      0.73      0.70       155
          13       0.66      0.77      0.61        43
```

5. TESTING

The testing was done in two phases. Phase one in testing included the exhaustive testing of the system, which included the white box and the black box testing. In the white box testing, the system's code was tested on hypothetical datasets. The black box testing was performed on the system by running it on the developer system, on the real-time recorded dataset which was obtained from the mentioned source. The result of the phase one obtained was below the satisfactory level. In comparison between white box and black box, the result of black box testing was more acceptable. The testing in this phase was slow paced and exhaustive.
The next phase of testing was done on the real-run system where the real-time dataset was fed into the system and black box testing was performed on the system. The system outperformed the testing system when compared to its run on the developer system. The achievements were achieved after the alterations and adjustments which were realized from the phase one testing. The phase two testing was quick and exhaustive. The result of the testing was acceptable when compared to phase one testing.

6. CONCLUSION AND FUTURE WORK

The novel method of using a dual stage system has given very promising results in identification of skin diseases with accuracies upto 80 %. Comparison of our work with related works in this domain has revealed stark differences in the implementation and performance. None of the existing solutions handle these many no of diseases that we have proposed in this paper and and have implemented ensemble features .A fast and correct diagnosis is essential for the patient’s life. For this reason a computational-based diagnosis tool is a must.
As part of our future work, we would like to make the system develop immune to the varying skin colours.Our focus in this system has been on the twentyof the most common dermatological diseases. We intend to continue working with the doctors to come up with better feature functions in order to broaden the number of diseases that the system can detect with a larger dataset. And we also shall concentrate on reducing the latency of the system.

REFERENCES

1. Catarina Barata, M.Emre Celebi,Jorge S.Marques, ”Improving Dermoscopy Image Classification Using Color Constancy”, IEEE Journal of biomedical and health imformatics, volume19,no.3,may2015
2. Vinayshekhar Bannihatti Kumar, Sujay S Kumar, Varun Saboo, "Dermatological Disease Detection using Image Processing and Machine Learning", Artificial Intelligence and Pattern Recognition (AIPR) International Conference on, pp. 1-6, 2016.
3. D.S. Zingade, Manali Joshi, Viraj Sapre, Rohan Giri, ”Skin Disease Detection using Artificial Neural Network”, International Journal of Advance and Research Development, Volume 4, Special Issue 5, Dec.-Engineering2017.
4. Vinayshekhar Bannihatti Kumar, Sujay S Kumar, Varun Saboo. "Dermatological disease detection using image processing and machine learning" , 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), 2016