Automated System for Prediction of Skin Disease using Image Processing and Machine Learning

Lakshay Bajaj
Department of Biotechnology,
Delhi Technological University,
Shahbad Daulatpur, Main
Bawana Road, Delhi-110042,
India

Himanshu Kumar
Department of Biotechnology,
Delhi Technological University,
Shahbad Daulatpur, Main
Bawana Road, Delhi-110042,
India

Yasha Hasija
Department of Biotechnology,
Delhi Technological University,
Shahbad Daulatpur, Main
Bawana Road, Delhi-110042,
India

ABSTRACT

Dermatology is one of the major field of prescription which is worried about the analysis and treatment of skin disorders. Skin diseases are among the most widely recognized medical issues around the world. Regardless of being common, their determination is quite troublesome and requires broad knowledge and expertise in the area. Skin disease might cause severe health and monetary consequences for patients if not detected and controlled early. Early recognition can forestall the condition from worsening. This research paper presents the development of an automated skin disease diagnosis system which takes images of a skin disease as an input by the user and predicts the type of skin disease. The system uses a dual stage approach for detection and prediction process which effectively amalgamates image processing and machine learning. In the 1st stage, the image of the skin condition is subject to numerous types of pre-processing techniques followed by feature extraction. The extracted features for each image are then converted to a feature vector. In the second stage, the feature vectors are fed to a machine learning algorithm (artificial neural networks) to identify disease and predict accordingly. On training and testing for 5 diseases (eczema, psoriasis, impetigo, melanoma, and scleroderma) system produces an overall prediction accuracy of 90%.

General Terms

Dermatology, image processing, machine learning

Keywords

Skin diseases, pre-processing techniques, artificial neural networks

1. INTRODUCTION

Medical Imaging is now extensively used in medical science so that internal structures of the human body can be determined and diagnosed. As it helps to look inside in the body without actually having to open the body [1]. Hence, medical imaging undoubtedly has become an important tool for the detection and evaluation of the treatment of a disease. Skin detection is one of the elementary subjects in image processing. In many instances, example human detection and tracking, visual recognition and face identification, a skin detection system are desirable/required. The concept of skin detection lays foundation for skin diseases detection. Towards this direction, several techniques and approaches have been introduced as of now. In a large proportion of the proposed approaches, researchers have attempted to classify and isolate a feature vector for each and every pixel of image, wherein most of the techniques make use of color spaces to extract required features [2]. In most developing countries, prices are too steep for majority of the population to consult a dermatologist. For example, in a developing country like India, most of the people are reluctant to go to a dermatologist for their skin disease problem as they are not able to afford the prices. Every year significantly great number of people in the developing nations suffer due to diverse range of skin diseases. Therefore, the widespread of skin disease has made it is an indispensable need for both the patients and dermatologists to have an automated skin disease detection system especially in developing countries [3].Our application involves a layered approach, with two stages: first, we make use of Image Processing for identification as well as extraction of region of interest and the second stage utilizes Machine Learning for a fairly accurate prediction of the disease. This system is capable of detecting 5 main skin disorders namely psoriasis, melanoma, scleroderma, eczema, impetigo.

2. RECENT WORKS

In the past two decades, strong impulse has been given to developing automated systems capable of assisting physicians in medical imaging task [4].The paper proposed by Muhammad Zubair Asghar et al [5] presents an online expert system to detect certain skin diseases that make use of forward-chaining with depth-first search method. However, using a system based on set rules and symptoms is not feasible due to the various manifestations of a single skin disease. The system created by us would be superior in performance in such a manner as the issue we are addressing is probabilistic in itself and thus we require a system which has the ability to adapt and learn from the underlying pattern that exists in the skin disease which can be inferred by the image. Another system that was proposed by A.A.L.C. Amarathunga et al [6] introduces a data mining unit to the system of skin detection. However, it lacks the choice of attributes essential for detection. Neither the data source nor the attributes used for learning/testing have been mentioned. M. SHAMSUL et al [7] proposed a system that uses an image processing system with pre-processing algorithms and a feed forward neural network**.** Florence et al [8] categorize the image as a bacterial or viralskin infection using image processing techniques. Damilola et al [9] developed a system that assembles pigmented skin lesions image results, analysis, comparing observation and conclusions by medical experts using prototyping methodology. Our system makes use of image processing with pre-processing algorithms and feed forward back propagation method in artificial neural networks that are discussed in the following section.

3. ARCHITECTURE AND METHODOLOGY

The system amalgamates two approaches that are image processing and machine learning algorithms represented in figure 1.

The first stage makes use of different image processing unit to isolate and extract the region of interest that is the diseased region from the image.

Initially those pre-processing algorithms were applied on colour skin images and further image processing algorithms were applied. The Second stage makes use of machine learning unit, the input images are fed to an artificial neural network to train network and predict an accurate result. For all the process to take place we needed data in the form of skin disease image, particularly of eczema, psoriasis, impetigo, melanoma, and scleroderma. The training data set was obtained from [10].
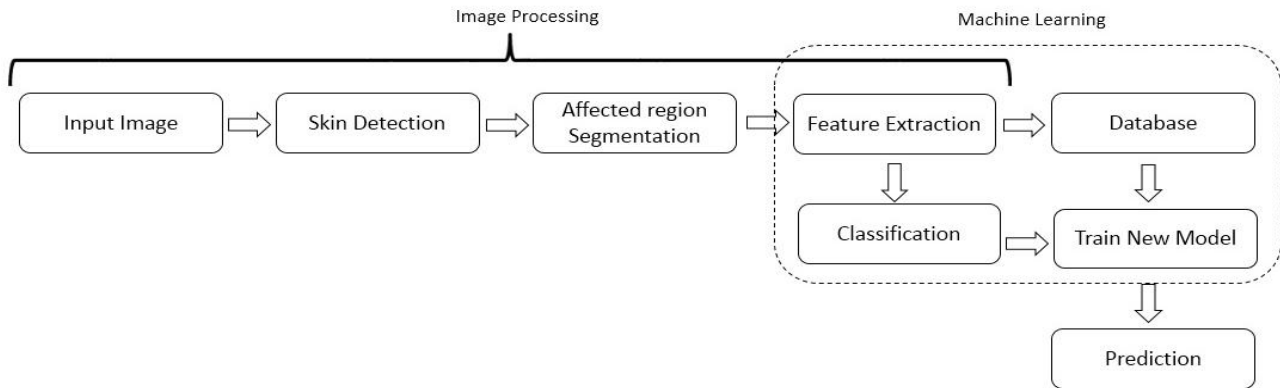

Fig 1: Brief Flowchart of the methodology used

3.1 Image Processing Unit

The colour images taken from the database were subjected to pre-processing algorithms before feature extraction. However, there are various image processing methods available to extract the required region such as thresholding, compression based methods, clustering, region growing, watershed transformation, etc. From previous works, we found that the edge-based segmentation method based on active contour, can be used to segment the affected region more accurately [11]. The disease region was segmented using the active contour and pre-processing algorithm. Both the methods were compared and gave equal results but it was found that active contour had a flaw, the user needed to mark the region around the infected area. Therefore, Pre-Processing Algorithm was given preference over active contour because of its autonomous nature and reducing the involvement of user.

3.2 Feature Extraction

In order to improve the accuracy of feature extraction pre-processing algorithms were used. The algorithms used were grey scale, sharpening filter, median filter, smooth filter, binary filter, and RGB extraction. RGB values of the image were stored beforehand converting them to grey scale image. Sharpening filter was then applied to enhance the subtle details of the grey scale image. Median filter was then used to remove any noise from the image using a 5*5 matrix. The next filter used was the mean filter or smoothening filter which replaced each and every pixel with average (mean) value of its neighbor including itself. The resulting image was then converted into binary image which provides the basis of extraction of diseased region. The result of the applied filter is represented in figure 2. The Sobel operator was then used to detect edges which facilitated us in detecting the shape of the affected region. The affected/diseased region was extracted and compared to RGB color stored before converting to grey scale which resulted in extraction of only the colored diseased region of the image.
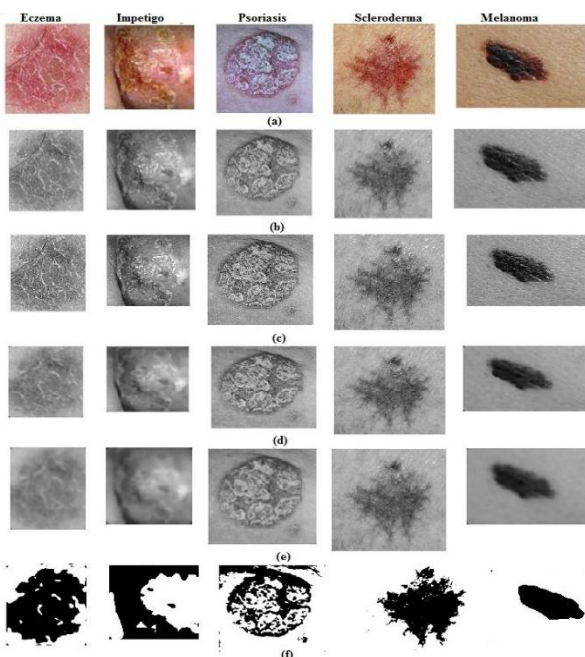

Fig 2: Result of Pre-processing algorithms. (a) Original Color skin image, (b) Grey Scale image, (c) sharpening filter, (d) Median filter, (e) Smooth Filter, (f) Binary Mask
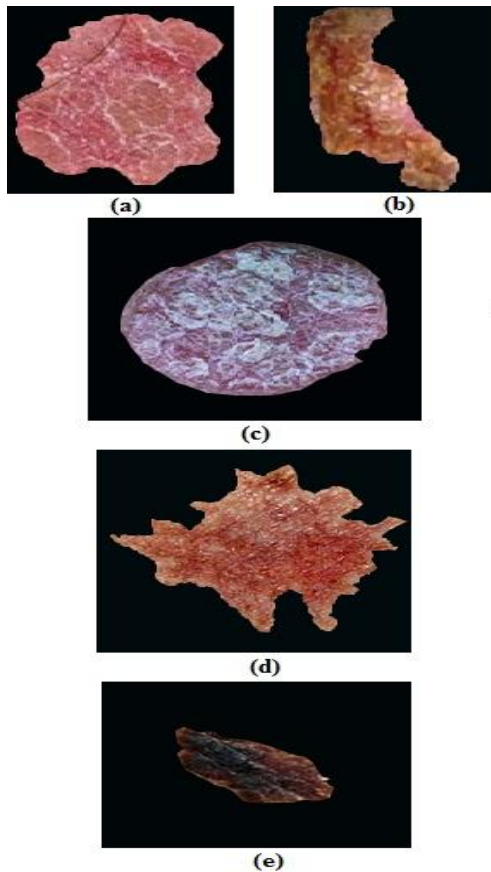
Fig 3: Segmented Disease region (a) Eczema (b) Impetigo (c) Psoriasis (d) Scleroderma (e) Melanoma

This was done by super-imposing the binary image and the original color image pixels and then elimination of those pixels which were not present in the binary affected region. Figure 3 represents the resulting images.

The resulting diseased region was then converted into a feature vector. Similarly, each image in the training set was converted into feature vector in MATLAB and classified together so that it can be then fed to the machine learning unit for training and testing. Figure 4 represents the feature vector obtained from each image. Each feature vector contains 25 feature values from each image in column-wise orientation.

3.3 Machine Learning Unit

We used feed forward back propagation artificial neural network (ANN) for this step. The ANN comprised of one input layer, two hidden layers and one output layer. Since, we have used a supervised machine learning algorithm, the feature table obtained after image processing was fed to the input layer and their corresponding results to the output layer for training. A deep learning technique using the neural network tool in MATLAB which trains the model using an identity function was used to initialize the parameters of the model, the weight matrices W.

• Input Layer: Input from the user (25 neurons)

• Hidden Layer: Composed of sigmoid function (40 neurons)

• Output Layer: Softmax layer

Once the input and output data were fed the network was trained. Further, the system was validated and tested by using the tenfold cross validation process because of its unique quality which makes sure that there is no overlapping of the training data and the test data, thereby, making the test results fairly accurate and dependable. The system was trained using 25 features as input and made use of 40 neurons in the hidden layer. Figure 5 represents how our system is trained by using 25 features from each image. Output layer represents 5 neurons because we used a softmax layer [12] approach as it gives probabilistic results for all 5 diseases.

FEATURE TABLE

	IMG 1	IMG 2	IMG 3	IMG 4	IMG 5	IMG 6	IMG 7	IMG 8
1	222	190	211	150	210	175	128	174
2	216	187	209	156	203	181	135	168
3	227	188	211	145	203	185	140	168
.								
24	208	207	145	203	171	211	204	173
25	195	193	173	209	169	215	208	171

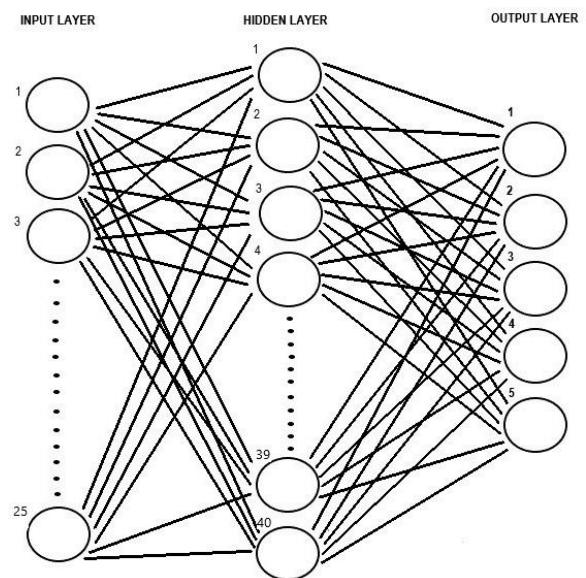**Fig 4: Feature Table**

Fig 5: Feed Forward Back Propagation Neural Network

In **forward propagation**, a set of weights are applied to the input data for the calculation of an output. Weights are selected randomly for the first forward propagation.

In **back propagation**, margin of error is measured for output and accordingly, weights are adjusted to decrease the error.

4. RESULTS AND DISCUSSION

The approach of using 2 stage process for prediction of skin diseases, where the disease region is converted into a feature vector and then used for training of network is novel to best of our knowledge with an overall accuracy of 90%.Correlation of our work with related works in this area has uncovered stark contrasts in the execution and performance. None of the current solutions for detection and prediction of skin diseases deals with the five skin diseases that we have proposed in approach. . Out of the 813 images of 5 diseases, 70% of the images were used for training, 15% for validation and

remaining 15% for testing. Figure 5 represents regression graph and their respective accuracy rate.
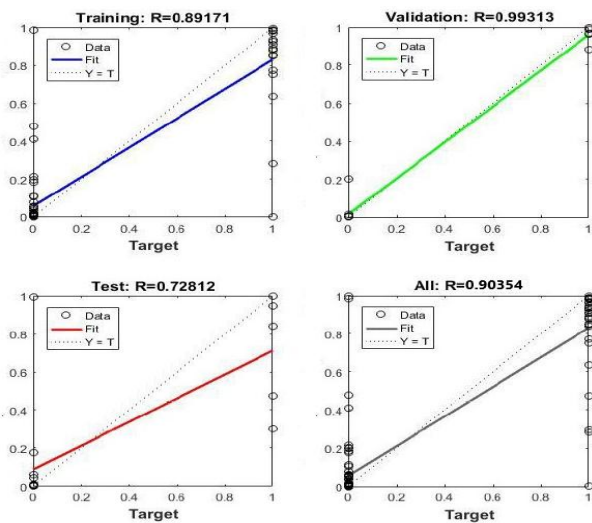

Fig 6: Regression graph representing training, validation, test and overall accuracy rate

Out of 5 skin diseases used, the detection rate of Eczema and Psoriasis was 92.5% and 91.6% respectively. One of the proposed methods presented in [3] obtained only 88% detection accuracy for Eczema. Similarly, detection rate of 88% and 89.3% for Eczema and Psoriasis respectively could be obtained in [9] which is less than the detection rate obtained by our system. Although, detection rate of Scleroderma was not up to our expectations, however, to best of our knowledge, no system as of yet has been proposed to successfully detect Scleroderma. Figure 6 shows individual detection rate of diseases.
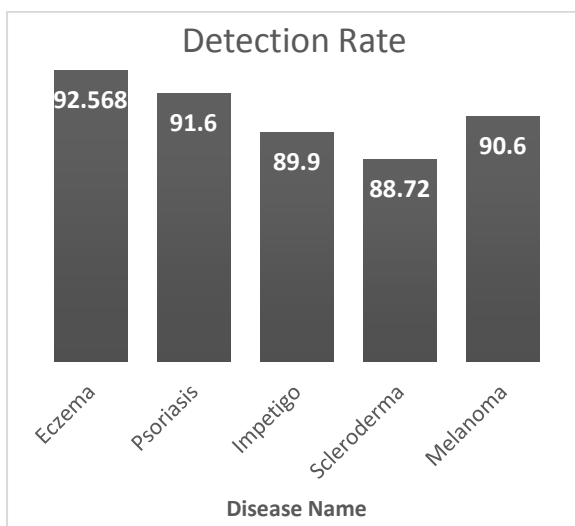

Fig 7: Detection Rate of Each Disease

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Aarthipoornima E, Jeyaseelan T, Medical imaging modalities: A survey, in: 2016 Int. Conf. Emerg. Trends Eng. Technol. Sci., IEEE, 2016: pp. 1–4. doi:10.1109/ICETETS.2016.7603066.

[2] Shervan F. E, Mohammad S, Farshad T, An Innovative Skin Detection Approach Using Color Based Image Retrieval Technique, Int. J. Multimed. Its Appl. 4 (2012) 9. doi:10.5121/ijma.2012.4305.

[3] Rahat Y, Md. Ashiqur R, Nova A, Dermatological disease detection using image processing and artificial neural network, in: 8th Int. Conf. Electr. Comput. Eng., 2014: pp. 687–690. doi:10.1109/ICECE.2014.7026918.

[4] Radu D, Stefan M, Dan P, Medical images classification for skin cancer diagnosis based on combined texture and fractal analysis. WSEAS Transactions on Biology and Biomedicine, July 2010.

[5] Muhammad Z. A, Asghar Mj, Sheikh S, Shakeel A, Diagnosis of Skin Diseases using Online Expert System, International Journal of Computer Science and Information Security, June 2011

[6] A.A.L.C. Amarathunga, E.P.W.C. Ellawala, G.N. Abeysekara, C.R.J. Amalraj, Expert System For Diagnosis Of Skin Diseases, Int. J. Sci. Technol. Res. 4 (2015) 174–178.

[7] M. Shamsul Arifin, M. Golam Kibria, Adnan. F, M. Ashraful Amini, Hong Y, Dermatological disease diagnosis using color-skin images, Published in Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

[8] Florence T,Ernest M, Fred N. K, An image-based diagnosis of virus and bacterial skin infections, International Conference on Computing and ICT Research,2011

[9] Damilola A. O, Olidayo O. O, Soloman A. O, Automating skin disease diagnosis using image classification, Published in Proceedings of the World Congress on Engineering and Computer Science 2013 Vol II WCECS 2013, 23-25 October, 2013, San Francisco, USA

[10] Diepgen TL, Yihune G et al. Dermatology Online Atlas. Published online at: http://www.dermis.net/doia/

[11] Teck T. T, Li Z, Ming J, , An intelligent decision support system for skin cancer detection from dermoscopic images in 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)

[12] Er. Shrinidhi G, Ansari N, Ansari Z Shaikh R,An Innovative Approach for Skin Disease Detection Using Image Processing and Data Mining. In International Journal of Innovative Research in Computer and Communication Engineering.