Skin Disease Detection using Machine Learning

Sayantan Nandy, Sindhu K

Abstract— Skin disease is a very common health issue all across the globe. It is one of the most common forms of infection occurring in people of all ages. Skin is one of the most delicate parts of the human body. Skin disease is one of the most difficult to predict due to irregular skin color, rough area, moles, burns, blisters, pimples etc. The above factors need to be suppressed in order to accurately detect the disease. In a developing country like India, it is very costly for many people to visit a doctor for their skin disease. So, the need for automatic skin disease detection is very important for patients as well as the dermatologist. It makes skin disease detection much easier, faster and cheaper. In this paper, we are developing a model that can detect 4 types of skin diseases, those are Eczema, Acne, Ichthyosis, Psoriasis (Bacterial disease). In the proposed model, we have 4 phases. In the first phase, we grayscale the image so that the system can recognize and control it, then in the second phase we do Image Segmentation to locate the ROI (Region of Interest i.e the main area or portion of the image from which the disease is to be detected) so that system can analyze the image more clearly. In the third phase, we do Image Pre-Processing using various techniques to remove the unnecessary distortions from the image and enhance the image quality for further processing. In the fourth and the final phase, after the image is processed, it is given as input to the Neural Network which is able to detect the skin disease with an accuracy of 95%.

Index Terms — Skin disease, Model, Grayscale, Image Segmentation, Region of Interest, Image Pre-Processing, Neural Network

1 INTRODUCTION

Skin is one of the most difficult parts of the human to diagnose due to its intricacy. In a lot of developing countries, it is very costly for people to seek advice from a dermatologist. The ever-present use of computers in developing countries has opened new areas for detection and diagnosis of skin diseases at low prices. Cameras present in every computer can be used and the image processing feature of the device can be utilized for disease detection. An algorithm has been designed that uses a four staged approach to address the problem. The first stage involves grey-scaling, second stage involves image segmentation, third stage involves image pre-processing and the fourth stage involves Machine Learning for a near effective solution.

Feature extraction is very crucial for predictive modeling applications. Skin disease is a major health problem in our country. So, it is the need of the hour to build a Machine Learning and Image Processing tool that will accurately determine whether a person is suffering from skin disease or not and what kind of skin disease it is. Several factors responsible for skin diseases are exposure to ultraviolet radiation, tanning, family history, environmental factors etc. They have a devastating impact on its well-being. Many suicidal attempts are seen in patients suffering from skin disease.

Almost 60% of the population in the United Kingdom suffer from skin diseases. In the developing countries, many people find it expensive to consult a dermatologist. Necessary tools that are required for early detection of skin disease are still not available in many countries. There is a need for intelligent expert systems for problems caused by skin disease. These experts can perform multi-class skin lesion classification to help people in early diagnosis. Various applications can be made to detect skin diseases. Nowadays people are suffering from various skin diseases. Many people are suffering from Psoriasis also known as skin cancer. Almost 10 million cases of psoriasis are encountered in India every year. Skin cancer has increased rapidly. If skin diseases are not treated properly at an early stage, then it may lead to several complications. The infection can spread from one person to another. The skin disease can be prevented if it is diagnosed at an early stage. The important factor in skin disease detection is skin tone and skin color. Processing of such images require quantitative discriminator to differentiate the disease.

An application can be developed to solve the problem in two stages. Image processing can be used for identification of the disease which is the first stage. Machine learning can be used for the verification of the disease which is the second stage. Some diseases can show features in the initial stage and they have their characteristic features in the upcoming stages. Biopsy here becomes necessary for diagnosis. Histopathological features can be also seen in these diseases. The issue can be solved by a machine learning model that can check the features which were obtained by skin sample analysis under a microscope. In the proposed work, a system has been built which detects four different kinds of skin diseases using machine learning and image processing techniques that will accurately determine whether a person is suffering from skin disease or not and what kind of skin disease it is.

2 RELATED WORK

Archana Ajith et.al [1] proposed a skin disease identification technique with the help of image processing. This method is used inside a mobile that's why it is very accessible even in very remote areas where there are no proper facilities, and it is completely harm free to the person's skin. The person with skin disease provides a picture of the infected area of the skin as an input to the system. Image processing technique is performed on this image and the detected disease is displayed as the output. R. Bhavan et.al [2] proposed automatic skin disease identification for the patients and for skin doctors. They have used computer vision techniques to identify different types of dermatological skin diseases. Inception_v3, MobileNet, Resnet are the CNN algorithms that are being used

for feature extraction. This method fetches us efficiency of more than 90%. D.A. Gavrilov et.al [3] proposed a skin disease identification method to detect melanoma disease. Melanoma is a skin disease which is very dangerous. The visual accuracy to detect melanoma mainly depends on the doctor. If the doctor is experienced, then it can be detected very easily otherwise it's very difficult to detect melanoma in early stages. The method presented in this research paper is an algorithm in order to detect melanoma early using artificial deep CNN. M. Shamsul Arifin et.al [11] proposed a machine-based disease detection method in humans rather than doctor's diagnosis. When tested on a total of 2055 disease areas in 705 images of skin the system gave a disease detection accuracy of 95.99%. Soumya Sourav et.al [12] proposed that in cases of skin disease like Melanoma, detection of the disease in early stages is very important to increase the chances of the disease getting cured. The method used in this paper is capturing the infected area using a mobile. The patient with skin disease provides a picture of the infected area of the skin as an input to the system. The model is designed into three phases consisting of collection of data, designing, and developing the algorithm and finally testing. In this model algorithms used were CNN and SVM that gave an accuracy of 92.1%. Xiangfeng Da et.al [13] proposed a model which used a supervised Machine Learning algorithm, and it needed a huge number of images to train the model, storing and processing so much data poses a lot of system challenges for a mobile. Therefore, this study focuses on deploying the system to a cloud-based ML system which makes use of the Internet to send data intensive work onto cloud. However, this method has its own downsides related to latency and privacy. Rahul Nijhawan et.al [14] proposed an algorithm in order to detect nail diseases as it is still unexplored. This paper proposes a Neural Network (NN) framework to detect nail disease from pictures. Nail disease is a class of 11 diseases such as subungual hematoma, beau's lines, yellow nail syndrome, psoriasis, koilonychias, paronychia, pincer nails etc. This paper claims to have an accuracy of 91%. Rajaram Anantharaman et.al [15] proposed an algorithm that will be helpful for field workers to check for initial symptoms of orofacial diseases, using their mobile phone. For this tool, Oro Vision to detect mouth sores. The objective of this tool is to extend it to identify various types of mouth disease such as Thrush, Leukoplakia, Lichen Planus, etc. This paper claims to have an accuracy of 87%.

3 METHODOLOGY

The proposed model is designed in four phases which are as follows:

3.1 Phase 1

First phase of the model involves grey-scaling the input image. A greyscale image is a black and white image that contains information about intensity. It is used to convert a persistent tone image to a picture that a PC can recognize and control.

3.2 Phase 2

The second phase of the model involves Image Segmentation. It is used to locate disease boundaries and curves so that the computer can analyze the image more clearly.

3.3 Phase 3

Pre-processing is the usual name for operations with pictures at a very low level of abstraction. It is the most important step in any of the applications as it builds up the image for further analysis and prepares it so that any algorithm can be applied to it. The main aim of pre-processing is to improve the degree of image knowledge that includes removal of unwilling distortions and enhances some image properties vital for additional processing, through geometric transformation of pictures such as rotating, scaling and translating images.

3.4 Phase 4

Neural Network is a system or circuit of neurons, or from an advanced perspective, a counterfeit neural system, made out of fake neurons or nodes. Thus, a neural system is either a natural neural system, made up of genuine organic neurons, or a fake neural system, for tackling man-made brainpower (AI) issues.

4 COMPONENTS OF METHODOLOGY

The first process is grayscaling the image so that segmentation can be done easily and accurately. Gray scaling and segmentation are the most important part in image pre-processing and they must be done properly so that the region of interest of the image can be detected accurately. The more accurate is the region of interest, the better are the results. In grayscaling each and every color is grayscaled according to its intensity. A more intense color is given a darker shade of gray and a less intense color is given a dull shade of gray. So, in a similar manner all colors in the image are grayscaled and the resultant image is obtained.

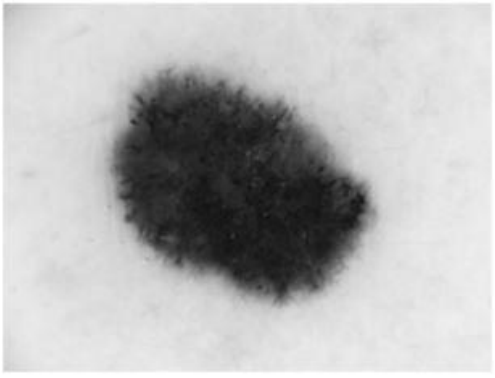

Fig. 1. Grey scaled Image of a picture

In the proposed model, the K-means algorithm has been used for image segmentation. In the K-Means segmentation algorithm, first the image has been reshaped and a matrix was obtained from it. The matrix was then converted to 32 decimal float form for the calculation. After that a criterion for the K-means segmentation algorithm was decided. The criteria is the termination crite-

ria for the iteration. So, when the criteria is met, the algorithm would stop. So, two conditions are possible i.e either the specified accuracy(epsilon) is reached or stop after a specified number of iterations. A K for the K-means algorithm is defined i.e the number of clusters. K=5 or K=7 can be taken but we take K=5 because if we take K=7, it will take more time to process the image. So, a mean RGB value is taken after taking the RGB of all the points in the image and then segmented into 5 different clusters based on the variation of the RGB points from the mean RGB value. Points with little difference in intensity of color are assigned to the same cluster. The center of each cluster is obtained and then we get a center of all the cluster centers and finally obtain a clustered center image. The pixels of the clustered center image is converted to 8-bit integers. The final segmented image is obtained by flattening the clustered center image.

The output of the clustered and segmented image is:

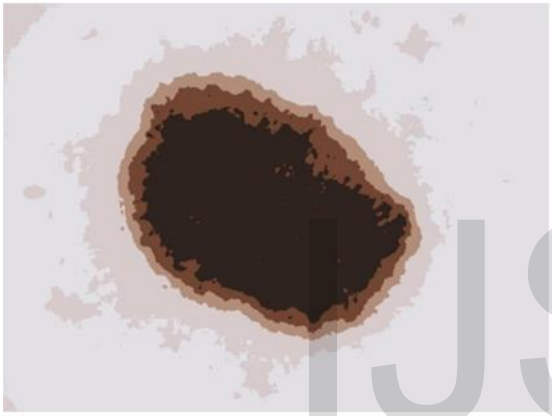

Fig. 2. Image Segmentation

In the image pre-processing technique, median blur is done first. Median blur was considered than Gaussian blur because Gaussian blur uses a gaussian kernel which consists of many filter box and other things and also it involves many comparisons and calculations and takes much more time than median blur. The median blur processes the middle of the considerable number of pixels under the particular window and the focal pixel is supplanted with this middle value. This is profoundly compelling in evacuating salt-and-pepper commotion. This lessens the commotion viably. The piece size must be a positive odd whole number. After the median blur is done, thresholding of the image is performed.

For each pixel, a similar limit boundary of pixel value is applied. On a very less chance, if the pixel boundary value is smaller than the edge, it is set to 0, else it is set to the most extreme value. The function cv.threshold is utilized to apply the thresholding. The main thing of importance is the source picture, which ought to be a grayscale picture. The subsequent thing of importance is the estimated edge pixel value which is utilized to order the pixel values. The third thing of importance is the most extreme value which is calculated near to pixel estimate surpassing the limit. OpenCV gives various kinds of thresholding which is given by the fourth parameter of the method. Essential thresholding as portrayed is done by utilizing the method cv.THRESH_BINARY.

In the proposed model, the adaptive threshold algorithm has been used. For thresholding, generally a global value is preferred, but it won't be a good idea if the image has different lighting conditions in different areas. For the simple thresholding, the global value v = 127 has been used. Here, the calculation of the algorithm decides the limit for a pixel dependent on a little district around it. So, various limits for various districts of a similar picture are obtained which gives better outcome for pictures with changing brightness.

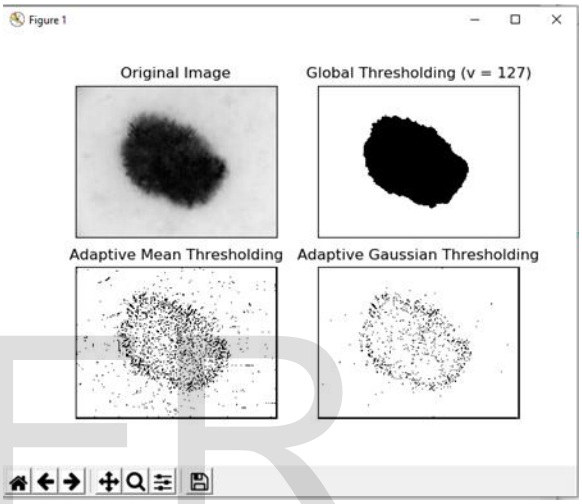

Fig. 3. Preprocessed Image of a skin

So, there are two types of adaptive thresholding that is the adaptive gaussian thresholding and the adaptive mean thresholding algorithm. Both have been used, which gives us the ROI more accurately so that proper analysis can be done further. In the adaptive mean thresholding algorithm, the mean of the surrounding area is taken as the threshold value minus the constant C. In the adaptive gaussian thresholding algorithm, the gaussian-weighted sum of the surrounding areas is taken as the threshold value minus the constant C.

5 EXPERIMENTAL SETUP

We have used the AWS platform for the neural network training so that the neural network is trained faster. We have used the IaaS (Infrastructure as a Service) service provided by AWS in which we have used the Amazon EC2 instance on which 64-bit Red Hat Linux Server was running. The Linux server was connected to a ssh terminal and then the neural network model and the dataset was migrated on the server.

The configuration of our linux server is as follows:

- Type: HTTP
- Protocol: TCP
- Port Range: 80

The dataset used was obtained from [16]. The model was trained on the training dataset consisting of 4000 images and the neuro python file was obtained, a copy of which was made and put into the local system for testing the neural network model.

The tools used are:

- Python
- Numpy
- Tensorflow
- Keras
- OpenCV

6 RESULT

As training on the normal computer would have taken much time so the training of 4000 images was done on Amazon Web Service using Convolutional Neural Network algorithm. After training, a train_data.npy was obtained which is a neuro python file that maps the result to which kind of disease the skin has.

After the training was done, the train_data.npy file was taken from the AWS console that we got as output after training the neural network model and then a runtime model was made that can detect whether a skin is diseased or not by giving the image of the skin using the train_data.npy file. Also, we have taken the IMG_SIZE variable which is the image size that we will give as the input parameter to the convolutional neural network and we have taken the MODEL_NAME variable which is the name of the model trained. The images have been resized to length: breadth ratio as 2:1 so that there is not much variation in the accuracy of the model. After resizing, we have done the grayscaling of the image which gave us a black and white image based on the color intensity of the regions. After grayscaling, we did the image segmentation using K-means to detect the boundaries of the infected area. After K-means we did image pre-processing so as to remove the noise, distortions in the segmented image and make the ROI clearer for analysis. After the image pre-processing is done, a neural network classifier is used to detect the disease. Finally, the detected disease is shown as the output.

The efficiency of disease detection by this method is 95% accurate. Advancement in image processing and machine learning will definitely improve the accuracy of disease detection in the future. The use of these types of technologies is a must to ensure safety and security for the future generation. The final confusion matrix that is obtained after testing is done is shown in Table I.

Table I
CONFUSION MATRIX

N = 4000	Predicted: NO	Predicted: YES
Actual NO	1611	85
Actual YES	109	2195

TP – Actual: YES and Predicted: YES
TN – Actual: NO and Predicted: NO
FP – Actual: NO and Predicted: YES
FN – Actual: YES and Predicted: NO
N(Total elements) = TP + TN + FP + FN

The formula for calculating the accuracy is:

$$((TP+TN))/N$$

So, from the formula the accuracy is (1611+2195)/4000 i.e 0.95 hence 95%.

7 CONCLUSION

Disease detection in a skin is a very crucial part of saving human lives. Skin disease is one of the most difficult to predict due to irregular skin color, rough area, moles, burns etc. These factors must be suppressed to accurately detect the disease. In many countries, it is very costly for many people to visit a doctor for their skin disease. So, the need for automatic skin disease detection is very important for patients as well as the dermatologist.

In this research, we trained the Convolutional Neural Network from scratch by training it on a dataset of 4000 images. We were also able to do live disease detection of the skin by giving live images as input to the model. Machine learning methods for detecting disease can help us save valuable time and money. It is very easy to use, very reliable.

REFERENCES

[1] A. Ajith, V. Goel, P. Vaziran and M. M. Roja, "Digital dermatology: Skin disease detection model using image processing," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 168-173.

[2] U. Niyaz, A. S. Sambyal and Devanand, "Advances in Deep Learning Techniques for Medical Image Analysis," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan Himachal Pradesh, India, 2018.

[3] Gavrilov, D.A., Melerzanov, A.V., Shchelkunov, N.N. et al. Use of Neural Network-Based Deep Learning Techniques for the Diagnostics of Skin Diseases.

[4] S. Kolkur and D. R. Kalbande, "Survey of texture-based feature extraction for skin disease detection," 2016 International Conference on ICT in Business Industry \& Government (ICTBIG), Indore, 2016.
[5] V. B. Kumar, S. S. Kumar and V. Saboo, "Dermatological disease detection using image processing and machine learning," 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, 2016.
[6] P. B. Manoorkar, D. K. Kamat and P. M. Patil, "Analysis and classification of human skin diseases," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 1067-1071.
[7] J. Rathod, V. Waghmode, A. Sodha and P. Bhavathankar, "Diagnosis of skin diseases using Convolutional Neural Networks," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore.
[8] Y. Hasija, N. Garg and S. Sourav, "Automated detection of dermatological disorders through image-processing and machine learning," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017.
[9] N. Hameed, A. M. Shabut and M. A. Hossain, "Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine," 2018 12th International Conference on Software, Knowledge, Information Management \& Applications (SKIMA), Phnom Penh, Cambodia, 2018, pp. 1-7.
[10] V. B. Kumar, S. S. Kumar and V. Saboo, "Dermatological disease detection using image processing and machine learning," 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, 2016.
[11] M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraful Amin and Hong Yan, "Dermatological disease diagnosis using color-skin images," 2012 International Conference on Machine Learning and Cybernetics, Xian, 2012.
[12] Y. Hasija, N. Garg and S. Sourav, "Automated detection of dermatological disorders through image-processing and machine learning," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017.
[13] X. Dai, I. Spasić, B. Meyer, S. Chapman and F. Andres, "Machine Learning on Mobile: An On-device Inference App for Skin Cancer Detection," 2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC), Rome, Italy, 2019.
[14] R. Nijhawan, R. Verma, Ayushi, S. Bhushan, R. Dua and A. Mittal, "An Integrated Deep Learning Framework Approach for Nail Disease Identification," 2017 13th International Conference on Signal-Image Technology \& Internet-Based Systems (SITIS), Jaipur, 2017.
[15] R. Anantharaman, V. Anantharaman and Y. Lee, "Oro Vision: Deep Learning for Classifying Orofacial Diseases," 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, 2017.
[16] https://isic-archive.com/