International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 02, 2022
www.ijirse.com
ISSN: 2454-9665

SKIN DISEASE DETECTION AND CLASSIFICATION USING ARTIFICIAL NEURAL NETWORK

[1] S. Gopalakrishnan, [2] Dr. Ebenezer Abishek. B,

[3] Dr. A. Vijayalakshmi

[1] Research Scholar, [2,3] Associate Professor,

[1,2,3] Department of Electronics and Communication Engineering,

[1,3] Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai.

[2] Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai.

Author's Mail id: [1] sk.gopalakrishnan@outlook.com , [2] ebenezerabishek@gmail.com,

[3] Vijayalakshmi.se@velsuniv.ac.in.

ABSTRACT

Skin diseases are one of the common health problems that can be diagnosed by examining the physical structure of diseases and history. At present, diagnosis of skin disease is done by invasive techniques like clinical examination and histopathology. The tests are very efficient and helpful. But these methods require subject experts, more time, and have less data reproducibility. Since diagnosis needs an expert, making it reachable to all places is difficult. So this research came up with a solution using a clinical feature-based diagnosis tool to diagnose certain skin diseases. The tool was developed with the help of image processing and deep learning algorithms.

1. INTRODUCTION

Skin diseases such as contact dermatitis, lichen planus, and plaque psoriasis, are important health problems that impair the quality of life [1,2]. These diseases are classified based on symptoms like inflammation, lesions, redness, etc. The common indications of skin diseases are red or itchy skin, rashes, papules or plaques, irregular patterns of nails, etc. [3]. In addition, the lesions seen in dermatoses may vary with severity from, minor localized patches, to complete body coverage [4,5]. Based on these physical indications people search their skin problems on the internet and try to identify the disease [6]. But, the problem with this approach is that most of the time the information or images available on the internet are labeled incorrectly, which misleads people. The easy alternative for this problem is providing an authenticated online data source to prevent falsified information, and educate the public about skin diseases. A survey conducted by the researcher showed that currently, less than 20 working online databases are available worldwide [7,8], yet no authentic online databases are publicly available from India. These databases have limitations such as updated information, diseased image metadata, user registration, lack of disease statistics, diagnosis or prognosis tool, and patient-doctor interaction. Therefore, this research aimed to create a database for three common skin diseases along with their supportive information, and develop a deep learning tool to diagnose the same.

ISSN: 2454-9665

2. LITERATURE SURVEY

During the years, there are several studies has been done on skin disease classification using image processing and machine learning techniques. Before conducting this study a survey was carried to gather the available data on skin disease diagnosis using a machine learning approach. The data were summarized in three categories, i.e., multi-class, binary and other types of classification studies; to know the current scenario of the technology usage in skin disease diagnosis. The part comprises each literature in terms of data collection, data pre-processing, feature extraction and classification. The limitations of these studies were listed at the end along with the solutions to overcome the same.

[9] collected 813 images from five different skin diseases namely eczema, psoriasis, impetigo, melanoma, and scleroderma, with the aim of diagnosing it based on the color feature. Images were pre-processed using median filtering, image sharpening, and binary masking methods and then Red, Green, and Blue (RGB) color means were extracted from each image. Later, these color values were analyzed by Artificial Neural Network (ANN) classifier, and 90% classification accuracy was obtained.

[10] examined herpes, paedures dermatitis, and psoriasis diseases using 10 standard samples and 20 test samples. Gray Level Cooccurrence Matrix (GLCM) texture features were extracted after selecting the region of interest using median filtering and marker-controlled watershed algorithm along with clustering. The four GLCM features were tested using Support Vector Machine (SVM) classification and achieved 85%, 90%, and 95% accuracy respectively for the studied diseases.

[11] conducted a study to investigate the performance of a ANN tool for skin disease classification versus skin lesion characterization. In total, 75,665 images were collected from six online databases, i.e., AtlasDerm, Danderm, Derma, Dermanet, and DermQuest. These images were used to train a multi-class ANN for disease targeted and another multi-class ANN for lesion targeted classifications. The results showed 27.6% of top-1, and 57.9% of top-5 accuracy values, with an average precision rate of 0.42 using fine-tuning learning type.

[12] collected images of nevus and ringworms from Google images, with the aim of detecting the disease based on their shape. Sobel edge detection method was used to detect the shape and prior to that, images were enhanced by median and smoothing filters and image sharpening techniques. The decision was made using the ANN classifier and showed good accuracy.

[13] segmented diseased lesions from the normal skin. A total of 45 plaque psoriasis images were captured in a 16-mega pixel camera and cropped manually before the segmentation process. CIE's L*a*b color space was used with Kmeans clustering for ROI segmentation. Post-processing was done using erosion followed by dilation. This segmentation process was compared with manually selected ground regions, and it achieved 93.83% accuracy.

This literature survey showed that it is possible to develop a skin disease diagnosis tool using image-driven features (79–82). But some limitations identified in the literature reviewed were:

1. Some studies created large training and testing sets from the initially collected images using image cropping. This may lead to lesion duplication, and hamper the efficiency of the classification process
2. Most of the studies did not give the justification for the disease or feature they had selected and few studies had collected the images from the patients visiting the hospitals
3. Most of the studies are not validated by dermatologists

To overcome these limitations, the present work was planned, and all the procedures supervised, by a dermatologist. Both training and testing set images were collected separately, and maximum possible lesions were cropped from them without any duplication. Three diseases that are differentially diagnosed with one another were selected for the study. Finally, clinical diagnosis procedure was studied and relevant features were extracted using the best suitable algorithms.

3. METHODOLOGY

3.1. MATERIALS

Two cameras and different lighting environments were used to acquire skin disease images so as to make the system more robust. The images were processed and results were tested by the programs written in MATLAB Integrated Development Environment (IDE) using other supporting packages. The details of the materials are listed in Table 1.

Table 1: List of materials used in the study

Sl. No.	Materials	Description
1	Cameras	1. Redmi 9Pro max Mobile camera
2	Lighting conditions	LED light, Flash light
3	Computer/ Laptop	HP 15q core i3 7th Gen (8GB RAM, 1TB HDD)
4	Platform	MATLAB 2019A
5	Supporting packages	Image Processing

3.2. DATA ACQUISITION

The data has been collected from the [14] ,[15] , hospitals and few images from Google. In this research totally 2500 images are collected from the sources. After finalize we used 1596 images for both testing and training. Sample image is shown in Fig 1.

3.3. PREPROCESSING

Hair removal was effective in all the tested images, but in dermoscopic images, it showed better performance and/ or results than the usual digital camera images. There were 58 different dermatoses images in the given dermoscopic image dataset. The tool performed better in many cases such as the presence of a minimal or moderate amount of noise, presence of noise using median filter in the focused region of the image, and lesions with red, white, and blue color (Fig. 2). Image sharpening features were incorporated in the tool to enhance the visual quality of the image.

International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 02, 2022
www.ijirse.com
ISSN: 2454-9665

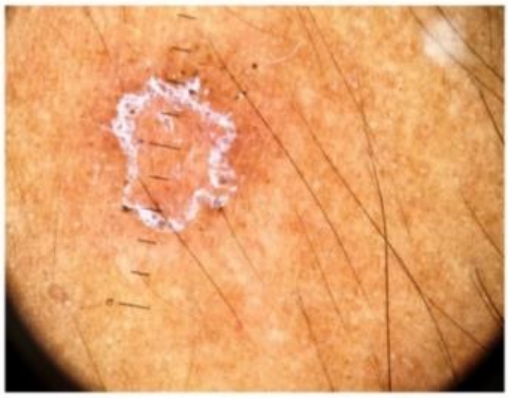

Fig 1: Input Image

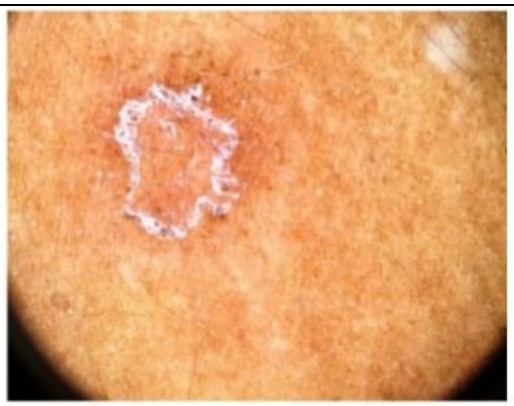

Fig 2 : Pre-processed Image

3.4. SEGMENTATION

Removing the unwanted regions from the image was helpful in terms of accurate results during the feature extraction, and reducing the processing time. Moreover, all input images contained more than one lesion in it. So a 256×256 pixel cropping window was used to get a large number of lesions from an image without duplicating it. Then MS-ROI segmentation algorithm was implemented to remove the unwanted region from the cropped image during the color feature extraction, Shown in Figure 3.

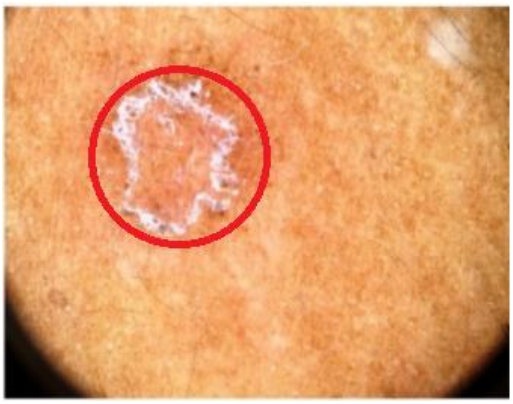

Fig 3: segmented Image

3.5. CLASSIFICATION

ANN model is used to classify the segmented images with GLCM matrix's. this system will provide the type of the diseases and other info. The main purpose of the Convolution layer is to extract the features from the input image. There may be presence of more than one convolution layer in the network. The first convolution layer is responsible for acquiring the low-level features such as edges, color, sharpen, gradient orientation, etc. Further layers acquire the high-level features of the image of the dataset. It works as a mathematical operation that has an image matrix and filter as its two inputs.

3.5.1. CONFUSION MATRIX AND PERFORMANCE MEASURES

The confusion matrix is also known as error matrix. It is a matrix to describe the performance of a segmentation model on a set of test data for which the true values are known. The number of cases correctly and incorrectly predicted by classifier is summarized in this matrix and helps to predict the various performance measure, like accuracy, sensitivity etc.

The performance of a segmentation system is the ability of the system to correctly predict the test data to its actual class. Accuracy is the performance measure that is used for assessment of the goodness of a Classification system. Accuracy for the four-class Classification system is defined in terms of class accuracy and overall accuracy. The confusion matrix formula and table has been shown in Figure 4.

- **Class accuracy:** It is the ratio of correctly Segmented cases of class to the total number of cases in that class with respect to ground truth.
- **Overall accuracy:** It is overall correctness of the Classification system. It is the ratio of correctly Segmented cases (diagonal elements) from each class to the total number of cases.

		Condition Phase (Worst Case)		
		Condition Positive/ Shaded	Condition Negative/ Unshaded	
Testing Phase (Best Case)	Test Positive/ Shaded	True positive shaded T_p (*Correct*)	False positive shaded F_p (*Incorrect*)	Precision/Positive Predictive Value (PPV) $\frac{T_p}{T_p+F_p}$ x 100%
	Test Negative/ Unshaded	False negative unshaded F_n (*Incorrect*)	True negative unshaded T_n (*Correct*)	Negative Predictive Value (NPV) $\frac{T_n}{T_n+F_n}$ x 100
		Sensitivity/Recall Rate (RR) $\frac{T_p}{T_p+F_n}$ x 100%	Specificity Rate (SR) $\frac{T_n}{T_n+F_p}$ x 100%	

Fig 4: basic confusion matrix with formula

The confusion matrix for two-class Classification system is shown in table 2. Here the entries are True Negative (TN), False Positive (FP), True Positive (TP) and False Negative (FN). Where TP is the number of skin disease correctly Segmented as the disease in question; FP is the number of skin disease without disease wrongly Segmented as disease in question; FN is the number of skin disease wrongly Segmented as without the disease; TN is the number of skin disease correctly Segmented as without disease.

International Journal of Innovative Research in Science and Engineering
Volume 08, Issue 02, 2022
www.ijirse.com
ISSN: 2454-9665

Table 2: Confusion matrix for a SS-ROI Classification system

	Actual - Skin diseases	***Actual - Not a Skin diseases***	***Total***
Predicted - Skin diseases	TP=893	FP=3	896
Predicted - Not a skin diseases	FN=5	TN=695	700
Total	898	698	1596

Accuracy is expressed by the overall rate of correctly and wrongly classified classes and can be defined as:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} * 100\% \quad (1)$$

$$Accuracy = \frac{893 + 695}{893 + 5 + 3 + 695} * 100\% = \frac{1588}{1596} * 100\% = 0.994987 * 100\% = 99.498\%$$

However, for unbalanced set, accuracy may not be a good criterion for evaluating a Classification system. The other measures of diagnostic test are specificity, sensitivity, positive predictive value and negative predictive value. These are defined as:

Specificity is also called true negative rate. It measures the proportion of actual negatives cases that are correctly identified.

$$Specificity = \frac{TN}{TN + FP} * 100\% \quad (2)$$

$$= \frac{695}{695 + 3} * 100\%$$

$$= 0.995702 * 100\%$$

$$= 99.57\%$$

Sensitivity is also called true positive rate or recall. It measures the proportion of actualpositives cases that are correctly identified.

$$Sensitivity = \frac{TP}{TP + FN} * 100\% \quad (3) = \frac{893}{893 + 5} * 100\% = 0.994432 * 100\% = 99.44\%$$

Positive predictive value (PPV) is also called precision. It measures the proportion of correctly Segmented positive cases among all cases which are predicted positive by the Classification system in test set.

$$Positive\ predictive\ value = \frac{TP}{TP + FP} * 100\% \quad (4) = \frac{893}{893 + 3} * 100\% = 0.9966 * 100\%$$

$$= 99.66\%$$

Negative predictive value (NPV) is the proportion of correctly Segmented negative cases which are predicted negative by the Classification system in the test set.

$$Negative\ predictive\ value = \frac{TN}{TN + FN} * 100\% \quad (5)$$

$$= \frac{695}{695+5} * 100\%$$
$$= 0.9928 * 100\%$$
$$= 99.28\%$$

4. CONCLUSION

Difficulties in the diagnosing skin diseases arise because of the spreading of the skin diseases all over the world, which make it a challenge to the dermatologist to recognize the different skin diseases easily, a computer aided system is proposed to resolve these difficulties, so a Deep learning model based on bag of features algorithm is designed which use ANN as a classifier and MS-ROI for feature extraction, the core model is developed using MATLAB. The developed system performs the required work with accuracy 99.49% within the dataset and 99.10% with the external data.

5. REFERENCES

[1] Gopalakrishnan.S, Dr.Ebenezer Abishek.B, Dr. A. Vijayalakshmi, Dr. V. Rajendran, " Analysis And Diagnosis Using Deep-Learning Algorithm On Erythemato-Squamous Disease", Volume 69 Issue 3, 52-57, March 2021, doi:10.14445/22315381/IJETT-V69I3P210.

[2] Gopalakrishnan.S, Dr.Ebenezer Abishek.B, Dr. A. Vijayalakshmi, Dr. V. Rajendran, " An MS-ROI based Detection and Segmentation of Erythemato-Squamous Disease ", Volume 69 Issue 8, 260-267, August, 2021. doi:10.14445/22315381/IJETT-V69I8P231.

[3] Lim HW, Collins SAB, Resneck JS, Bolognia JL, Hodge JA, Rohrer TA, et al. The burden of skin disease in the United States. J Am Acad Dermatol. 2017; 76(5):958-972.e2.

[4] Dawson AL, Dellavalle RP, Elston DM. Infectious skin diseases: a review and needs assessment. Dermatol Clin. 2012; 30(1):141–51.

[5] Sehgal V, Khurana A, Pandhi D. Biologics in dermatology: an integrated review. Indian J Dermatol. 2014; 59(5):425.

[6] New Zealand Dermatological Society. Dermnet [Internet]. Internet. 1998.

[7] University of Iowa Carver college of Medicine. United States: Dermatology. Internet. 2013.

[8] Elmasri R, Navathe SB. Introduction to databases. In: Fundamentals of Database Systems. 6th ed. Pearson Education International; 2011. p. 3–6.

[9] Bajaj L, Kumar H, Hasija Y. Automated system for prediction of skin disease using image processing and machine learning. Int J Comput Appl. 2018; 180(19):9–12.

[10] Wei L, Gan Q, Ji T. Skin disease recognition method based on image color and texture features. Comput Math Methods Med. 2018; 2018:1–10

[11] Liao M, Shi B, Bai X, Wang X, Liu W. TextBoxes: A fast text detector with a single deep neural network. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17). 2016. p. 4161–7

[12] Kalaiarasi S, Kumar H, Patra S. Dermatological disease detection using image processing and neural networks. Int J Comput Sci Mob Appl. 2018; 6:109–18.

ISSN: 2454-9665

[13] Munia TTK, Rizwan I, Aymond A, Mackinnon N, Daniel L, Al-hashim M, et al. Automatic clustering-based segmentation and plaque localization in psoriasis digital images. In: Proceedings of 2017 IEEE Healthcare Innovations and Point of Care Technologies. Bethesda; 2017. p. 113–6.

[14] http://archive.ics.uci.edu/ml/datasets/dermatology

[15] http://biogps.org/dataset/tag/skin%20disease/