International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Survey on Skin Disease Classification Models

Namitha S J[1], Nikhilesha N[2], Siddarth S Bellur[3], Soumil S Sinha[4], Ojus M S[5]

[1]*Assistant Professor, Dept. of Computer Science & Engineering, B.N.M. Institute of Technology, Karanataka, India*
[2-5] *B.E. Student, Dept. of Computer Science & Engineering, B.N.M. Institute of Technology, Karanataka, India*

---***---

Abstract - *Currently skin disease identification is performed by medical professionals. This process of manual recognition is slow and possesses a degree of subjectivity which is hard to quantify. Therefore, there is a scope to develop a technology-assisted model for skin disease detection and its classification. In an effort in developing such a model, a survey on the research works on the technology-assisted skin disease classification models was conducted. The information obtained from the survey has been presented in the proposed paper.*

Key Words*: *Dermatology, Skin Disease Classification, Support Vector Machine (SVM), Gray-Level Co-occurence Matrix (GLCM), Convolutional Neural Networks (CNN)

1. INTRODUCTION

Dermatology is the branch of medicine concerned with the diagnosis and treatment of diseases of skin, hair and nails. Skin disease is an abnormal condition of the skin. The skin helps in protecting the body from harmful bacterial, fungal and parasitic infections. Hence, the diagnosis of skin disease is crucial. Some common skin diseases are: Acne, Dermatitis, Eczema, Melanoma etc.

Individuals suffering from skin diseases may be subjected to pain, dysfunction, distress, social problems and anxiety. Skin diseases can affect people not only physically, but also mentally, as contracting and living with a disease can alter the affected person's perspective on life. Many skin diseases can lead to complications if not treated at an initial stage. Human mindset tends to presume that most skin conditions are not as fatal as described thereby applying their own curing methods. However, if these remedies are not apt for that selective skin problem then it does nothing to cure it.

The World Health Organization has emphasized the severity of skin diseases in India which affects 10 to 12 percent of India's population. But there are around 6000 dermatologists to cater to a population of 134 crore. Most of the dermatologists are concentrated in the cities. Due to this, rural areas are lacking dermatologists to provide a consultation and remedy. The current practices used by dermatologists include biopsy, scrapings, patch testing and prick test which are invasive methods of detection. In patch testing and prick test, the allergen is directly applied to the infected area. The skin might take a long time, even several days, to react to the allergen. The individuals living in rural areas simply cannot afford the luxury of time to get a diagnosis of what they are suffering from. So the need arises to develop a system that is efficient and accurate and can be deployed in rural areas where people facing difficulties to schedule a consultation with a dermatologist can get the diagnosis done and be better informed.

Hence a survey on the various skin disease classification models was carried out and the findings have been described in the paper.

2. METHODS FOR SKIN DISEASE CLASSIFICATION

There are several research-works on the various skin disease classification models. Brief appraisals on the available literatures are illustrated in this section.

2.1 Feature vector generation using Discrete Cosine Transform and Discrete Wavelet Transform

The Discrete Cosine Transform (DCT) transforms an image of the infected portion of the skin from spatial domain to frequency domain. In the DCT transformed image, most of the visually significant information is concentrated on the upper left corner of the 2-D matrix. Hence the feature vector is generated by scanning the 2-D matrix in a zigzag manner.

The Discrete Wavelet Transform (DWT) decomposes the signal into mutually orthogonal set of wavelets. It treats the image of the infected portion of the skin as 2-D signals which after decomposition is divided into four sub-bands: A1, H1, V1 and D1. Each of the sub-bands (feature vectors) contains coefficients that represent certain important features of the original image.

The feature vectors of the test images and that of the training images are compared in order to classify the skin disease [1].

The advantage is that it is a better method for skin disease classification compared to the other image processing techniques like median filtering and sharpening filtering with efficiency up to 80%. The limitation is that images of different skin tones and types are not considered.

2.2 Feature extraction using GLCM, dimensionality reduction using PCA and classification using SVM

The feature set used for this study, that uses Principal Component Analysis (PCA) based CADx system for psoriasis risk stratification and image classification, are (i) 11 Higher order spectra (HOS) features, (ii) 60 texture features, and (iii) 86 color feature sets and their seven combinations. Aggregate 540 image samples (270 healthy and 270 diseased) from 30 psoriasis patients of Indian ethnic origin are used in their database. Machine learning using PCA is used for dominant feature selection which is then fed to Support Vector Machine (SVM) classifier to obtain optimized performance.

Three different protocols are implemented using three kinds of feature sets. Reliability index of the CADx is computed. The algorithm Grey Level Co-occurrence Matrix (GLCM) is used for feature extraction from the image, Principal Component Analysis (PCA) for dimensionality reduction of extracted features and Support Vector Machine (SVM) for classifying whether the image represents an affected patient or a healthy person.

The paper gave rise to outstanding results by providing an efficient methodology for psoriasis disease detection. It presented a comparative study of computer aided diagnostic systems for disease identification by using different feature sets. This allows for disease identification and classification without the need of any physical examinations of the affected area of skin such as biopsy, scraping or patch and prick testing [2].

The advantage of the paper was that 100% accuracy was achieved on testing the system. The paper does have limitations such as the small size of the dataset (540 images with 270 images being of healthy skin samples and the remaining 270 images being of patients affected by psoriasis) from 30 patients, all of whom originate from a single geographical area, with limited environmental variations (India).

2.3 Automated segmentation using C-means and Watershed algorithms with feature extraction by GLCM and IQA

In this method, 45 digital images were collected from MIT BMI database consisting of warts, benign skin cancer and malignant skin cancer and normal skin images. These images were subjected to various pre-processing techniques such as resizing, conversion and contrast enhancement. Then these images were segmented using C-means and Watershed algorithms individually. Feature extraction was performed using Grey Level Co-occurrence Matrix (GLCM) and Image Quality Assessment (IQA) methods for examining texture which gave the statistical parameters of each algorithm respectively. Here, different types of skin diseases commonly classified are Benign Skin Cancer, Malignant Skin Cancer and Warts using multi-class Support Vector Machine (SVM). C-means algorithm produced better segmentation results with an accuracy of 98% compared to watershed algorithm (92% accuracy) in segmenting the skin cancer images. A computer aided diagnostic tool was implemented to diagnose the different diseased state of skin cancer [3].

The advantage is that C-means, GLCM and SVM give an accuracy of 96% - 98%. The limitation is that the dataset consists of only 45 images.

2.4 Deep learning ensembles for skin disease recognition in dermoscopic images

As expertise is in limited supply, automated systems capable of identifying disease could save lives, reduce unnecessary biopsies, and reduce costs. Toward this goal, use of a system that combines recent developments in deep learning with established machine learning approaches, creating ensembles of methods that are capable of segmenting skin lesions, as well as analyzing the detected area and surrounding tissue for melanoma detection. The system is evaluated using the largest publicly available benchmark dataset of dermoscopic images, containing 900 training and 379 testing images. New state-of-the-art performance levels are demonstrated, leading to an improvement in the area under receiver operating characteristic curve of 7.5% (0.843 vs. 0.783), in average precision of 4% (0.649 vs. 0.624), and in specificity measured at the clinically relevant 95% sensitivity operating point 2.9 times higher than the previous state-of-the-art (36.8% specificity compared to 12.5%).

Compared to the average of 8 expert dermatologists on a subset of 100 test images, the system produces a higher accuracy (76% vs. 70.5%), and specificity (62% vs. 59%) evaluated at an equivalent sensitivity (82%). Hand coded feature extractors, sparse coding methods, fully convolutional neural network and deep residual network are used [4]. The advantage is that a higher melanoma diagnostic accuracy (75-84%) is achieved compared to existing system (72%). It is tough to calibrate the system for both dermoscopic images as well as clinical images, which is the limitation.

2.5 Using a dataset consisting of a large collection of multi-source dermoscopic images of common pigmented skin diseases

The training of neural networks for automated diagnosis of pigmented skin lesions is hampered by the small size and lack of diversity of available datasets of dermatoscopic images. This problem is tackled by releasing the HAM10000 ("Human Against Machine with 10000 training images") dataset. Dermatoscopic images from different populations acquired and stored by different modalities were collected. Given this diversity, it was necessary to apply different acquisition and cleaning methods and developed semi-automatic workflows utilizing specifically trained neural networks. The final dataset consists of 10015 dermatoscopic images which are released as a training set for academic machine learning purposes and are publicly available through the ISIC archived.

This benchmark dataset can be used for machine learning and for comparisons with human experts. Cases include a representative collection of all-important diagnostic categories in the realm of pigmented lesions. More than 50% of lesions have been confirmed by pathology, while the ground truth for the rest of the cases was either follow-up, expert consensus or confirmation by in-vivo confocal microscopy. The algorithm utilized for the study is an automated method similar to Han *et al.* and fine-tuned Inception V3 architecture with Stochastic Gradient Descent[5]. The advantage is that it provides a huge data set to train a system for melanoma. Although, training the system takes lot of time and memory resources, which is a limitation.

2.6 Bagging, AdaBoost and Gradient Boosting classifier technique for skin disease prediction

Here, a new method which applies six different data mining classification techniques, and then develop an ensemble approach using Bagging, AdaBoost and Gradient Boosting classifier techniques is used to predict classes of skin disease. Furthermore, a feature importance method is utilized to select the most salient 15 features which will play a major role in prediction. A subset of the original dataset is obtained after selecting the 15 features, to compare the results of six machine learning techniques, and an ensemble approach is applied to the entire dataset. The ensemble method is compared with the subset obtained from the feature selection method. The outcome shows that the dermatological prediction accuracy of the test dataset is increased as compared to the use of an individual classifier, and improved accuracy is obtained as compared with the feature selection subset method. The ensemble method and feature selection applied to dermatology datasets yields a better performance as compared to individual classifier algorithms. The ensemble method provides a more accurate and effective skin disease prediction [6].

The algorithms used here are Bagging, AdaBoost and Gradient Boosting classifier. The efficiency of 96.71% has been achieved. The advantage is that it could be of efficient use for erythamato-squamus disease with improvement in speed and accuracy. The limitation is that large amount of numerical data can be a strain on the system.

3. CONCLUSIONS

As the method of generating feature vectors using the Discrete Cosine transform and Discrete Wavelet Transform for skin disease classification results in an accuracy up to 80%, the machine learning and the deep learning approaches are preferred. The machine learning algorithms such as Support Vector Machine and feature extraction algorithm such as Grey Level Co-occurrence Matrix are very efficient and can be used. Combining the above feature extraction algorithm with Image Quality Assessment, classification algorithms such as C-means and Watershed algorithms can be used on a smaller dataset for accurate classification. Another approach advised is an ensemble approach of machine learning and deep learning algorithms for classification. Such an implementation is used to produce state-of-the-art performance levels in terms of accuracy and specificity.

A good dataset for skin disease classification is the HAM10000 dataset that contains 10015 dermatoscopic images. The images are also very diverse in nature. Such diversity is essential in a dataset to train a Convolutional Neural Network model to be highly efficient. Lastly, Bagging and Boosting techniques, such as AdaBoost and Gradient Boosting are another solution for image classification. However, other algorithms performing a multitude of tasks such as data mining and feature selection need to be implemented in tandem. An ensemble approach such as the above yields efficient results.

REFERENCES

[1] A. Ajith, V. Goel, P. Vazirani 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.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] Shrivastava VK, Londhe ND, Sonawane RS, Suri JS. Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: A first comparative study of its kind, Comput Methods Programs Biomed, 2016 Apr; 126:98-109.

[3] M. S. Manerkar, U. Snekhalatha, S. Harsh, J. Saxena, S. P. Sarma and M. Anburajan, "Automated skin disease segmentation and classification using multi-class SVM classifier," 3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS 2016), Tadepalligudem, 2016, pp. 1-6.

[4] N. C. F. Codella et al., "Deep learning ensembles for melanoma recognition in dermoscopy images," in IBM Journal of Research and Development, vol. 61, no. 4/5, pp. 5:1-5:15, 1 July-Sept. 2017.

[5] Tschandl, P. et al. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5:180161 doi: 10.1038/sdata.2018.161 (2018).

[6] Verma, A., Pal, S., Kumar, S. (2019). 'Classification of Skin Disease using Ensemble Data Mining Techniques', Asian Pacific Journal of Cancer Prevention, 20(6),pp.1887-1894.