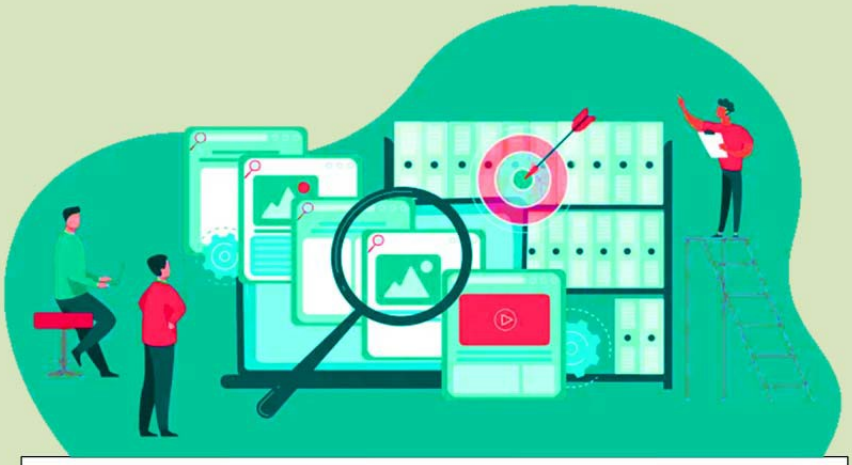


CHAPTER – 9

SKIN CANCER CLASSIFICATION IN IMAGE PROCESSING USING MACHINE LEARNING TECHNIQUES

Dr.K.Kasturi

ISBN : 978-93-91871-47-5

SKIN CANCER CLASSIFICATION IN IMAGE PROCESSING USING MACHINE LEARNING TECHNIQUES

Dr.K.Kasturi, Associate Professor, Department of Information Technology,School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai.

Abstract:

Skin is the outer most region of our body and it is likely to be exposed to the environment which may get in contact with dust, Pollution, micro-organisms and also to UV radiations. These may be the reasons for any kind of Skin diseases and also Skin related diseases are caused by instability in the genes this makes the skin diseases more complex. The human skin is composed of two major layers called epidermis and dermis. The top or the outer layer of the skin which is called the epidermis composed of three types of cells flat and scaly cells on the surface called SQUAMOUS cells, round cells called BASAL cells and MELANOCYTES, cells that provide skin its color and protect against skin damage. Machine Learning (ML) is a branch of

ISBN : 978-93-91871-47-5

Artificial Intelligence (AI) in which a computer algorithm analyses raw data and automatically learns discriminatory features needed for recognizing hidden patterns in them. Over the last decade, this field has witnessed striking advances in the ability of DL-based algorithms to analyses various types of data, especially images and natural language.

1. Introduction

Propensity of skin diseases to manifest in a variety of forms, lack and mal-distribution of qualified dermatologists, and exigency of timely and accurate diagnosis call for automated Computer-Aided Diagnosis (CAD). This study aims at extending previous works on CAD for dermatology by exploring the potential of Machine Learning to classify hundreds of skin diseases, improving classification performance, and utilizing disease taxonomy. We trained state-of-the-art Deep Neural Networks on two of the largest publicly available skin image datasets, namely Resnet50 and ISIC Archive, and also leveraged disease taxonomy, where available, to improve classification performance of these models.

2. Existing System

Frequent use of biopsies is also not encouraged by dermatologists. According to International Skin Imaging Collaboration, the number of unnecessary culture tests which are being performed

ISBN : 978-93-91871-47-5

vastly varies depending upon various parameters which include clinical setup, expertise of dermatologist, and the technology applied. Computer procedures and advancements in machine learning not only aid the dermatologists in early detection of melanoma but also avoid heavy expenses of melanoma detection and unnecessary biopsies. The literature review reveals that different practices have been used to develop computer-aided automatic diagnostics systems for the classification of skin cancer which take dermoscopic images as input and give classification results as benign or malignant at the output.

2.1 Proposed System

Although the advancement in the dermatological equipment has increased the classification accuracy of melanoma, the technological developments and improvements in the area of machine learning and image processing have resulted in a medical breakthrough in diagnosis, detection, and classification of melanoma with much more accuracy and reliability. Automatic melanoma detection systems save a lot of time, money, and effort. Machine learning has proven to provide melanoma classification with improved and higher accuracies.

Problem definition

Here we can set precedence for classifying all 622 unique sub-classes in this dataset and achieved 67% accuracy and 98% AUC. On

ISBN : 978-93-91871-47-5

ISIC Archive we classified all 10 diseases with 93% average accuracy and 99% AUC. This study shows that Deep Learning has great potential to classify a vast array of skin diseases with near-human accuracy and far better reproducibility. It can have a promising role in practical real-time skin disease diagnosis by assisting physicians in large-scale screening using clinical or dermoscopic images.

2.1.1 Objectives of the Study

The precedence has been classifying all 622 unique sub-classes in this dataset and achieved 67% accuracy and 98% AUC. On ISIC Archive we classified all 10 diseases with 93% average accuracy and 99% AUC.

This study shows that Machine Learning has great potential to classify a vast array of skin diseases with near-human accuracy and far better reproducibility. It can have a promising role in practical real-time skin disease diagnosis by assisting physicians in large-scale screening using clinical or dermoscopic images.

3. Skin Cancer Classification

One of the most rapidly spreading cancers among various other types of cancers known to humans is skin cancer. Melanoma is the worst and the most dangerous type of skin cancer that appears usually on the skin surface and then extends deeper into the layers of skin.

ISBN : 978-93-91871-47-5

However, if diagnosed at an early stage; the survival rate of Melanoma patients is 96% with simple and econosamplemical treatments. The conventional method of diagnosing Melanoma involves expert dermatologists, equipment, and Biopsies. To avoid the expensive diagnosis, and to assist dermatologists, the field of machine learning has proven to provide state of the art solutions for skin cancer detection at an earlier stage with high accuracy. In this paper, a method for skin lesion classification and segmentation as benign or malignant is proposed using image processing and machine learning.

Melanoma is the worst and the most dangerous type of skin cancer that appears usually on the skin surface and then extends deeper into the layers of skin. However, if diagnosed at an early stage; the survival rate of Melanoma patients is 96% with simple and economical treatments. The conventional method of diagnosing Melanoma involves expert dermatologists, equipment, and Biopsies. To avoid the expensive diagnosis, and to assist dermatologists, the field of machine learning has proven to provide state of the art solutions for skin cancer detection at an earlier stage with high accuracy.

3.1 Pre-Processing of Image Database

The database is downloaded from publicly available images on the ISIC website. The database included the ISBI-2016 challenge which has RGB dermoscopic images along with their labels and segmentation

ISBN : 978-93-91871-47-5

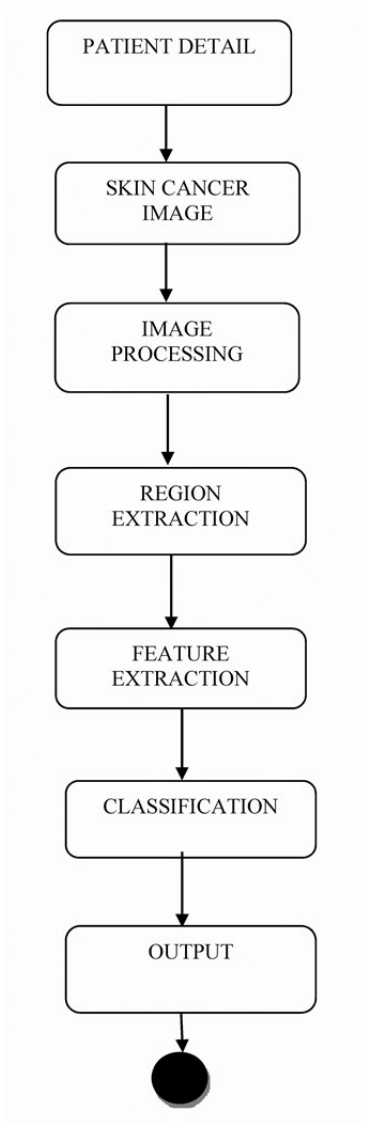
ground truths.As dataset images have different artifacts, so pre-processing is done on them to make them more meaningful. Pre-processing included image resizing, noise removal, contrast stretching, RGB to Gray conversion, and hair removal. For noise removal median filter of 3-by-3-by-3 is used. For contrast enhancement, a new method is proposed. Contrast enhancement significantly improves the results of segmentation in the next phase.

3.2 Segmentation

After the Pre-Processing, image segmentation is performed to extract the region of interest i.e. skin lesion. We used the OTSU thresholding algorithm for the segmentation of gray scale Pre-Processed images. Background triangles that exist in a few images on four sides are then removed by creating a mask and adding it to the binary image.

ISBN : 978-93-91871-47-5

3.3 Features Extraction and Reduction



ISBN : 978-93-91871-47-5

After the segmentation, features of texture, shape, and color are extracted from segmented skin lesions. To extract the texture features, segmented lesions mapped on original images are converted from RGB to grayscale first. Low pass approximation coefficients matrix (CA) of level 3 wavelet decomposition result in new images of reduced size. Afterward, texture features are obtained via gray level co-occurrence matrix of approximation coefficients matrix.

3.4 Smote Sampling

Data of ISIC- ISBI has a class imbalance problem. Almost 80% of data is "benign" and 20% of data is "malignant". As a result, it would be poor training, poor training accuracy, and poor classification accuracy on test data. Hence data needs to be balanced first. For thispurpose Synthetic Minority Over Sampling (SMOTE) technique is implemented on the final feature vector . SMOTE technique resulted in data being 50% benign and 50% malignant.

3.5 Features Standardization

Features vector is standardized using the method of "Linear scaling to unit variance" . Average and standard deviation of each column of the feature vector is computed. Then the average of each column is subtracted from the original feature vector's respective column to get zero mean feature vector. Afterward zero mean feature

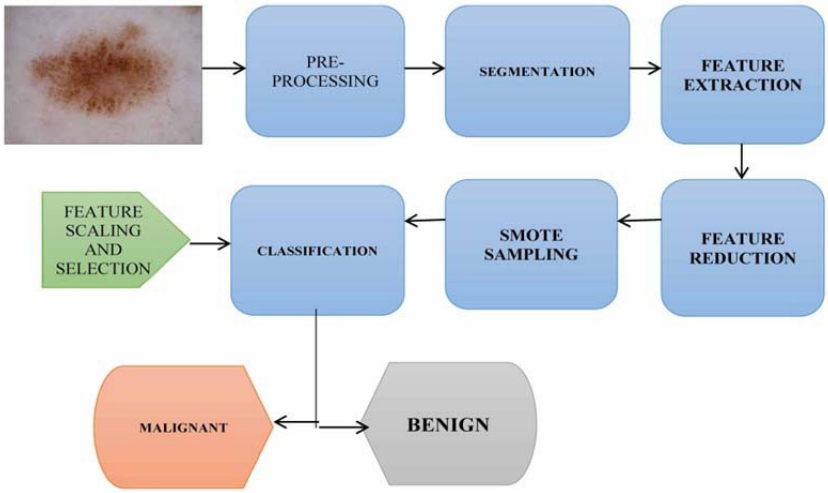
ISBN : 978-93-91871-47-5

vector's each column is divided by the standard deviation of that column to get the feature vector of unit variance. Features are scaled in such a way that in the feature vector each feature is normalized in a range from 0 to 7 by using the Min-Max Mapping Algorithm

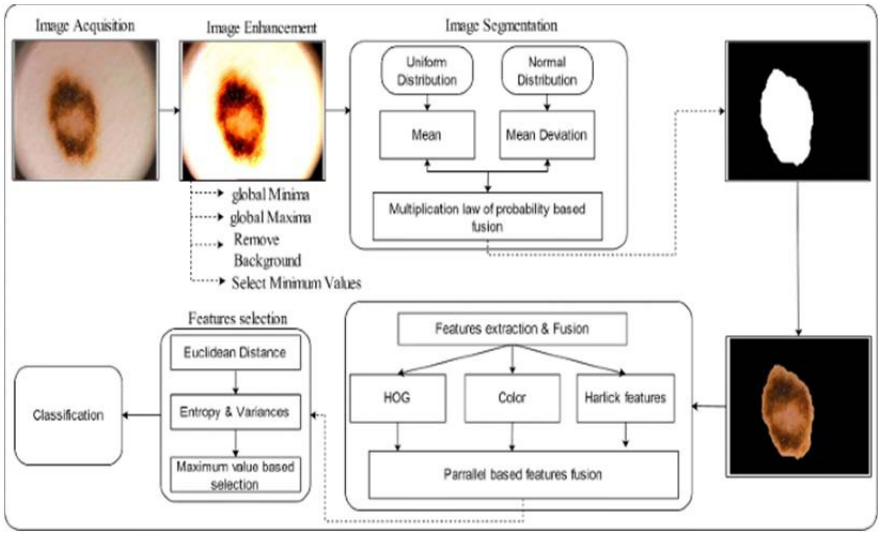
4. Classification

In the final step, we classified data as benign or malignant. Numbers of classifiers including Quadratic Discriminant, SVM (Medium Gaussian), and Random Forest are then used for training and their performance is evaluated on test data. Classifier models are later evaluated on test data. SVM divides the data into distinguishable classes to determine hyper plane with maximum margin. Proposed features selection based on wrapper methods gives remarkable and promising results when Random Forest classification is performed. The Random Forest technique generates decision trees on samples of data and predictions from each tree undergo the process of voting for classification. More votes from different trees mean more likely it belongs to that particular class.

ISBN : 978-93-91871-47-5



Architecture Diagram Fig-1



Architecture Diagram Fig-2

ISBN : 978-93-91871-47-5

5. Conclusion

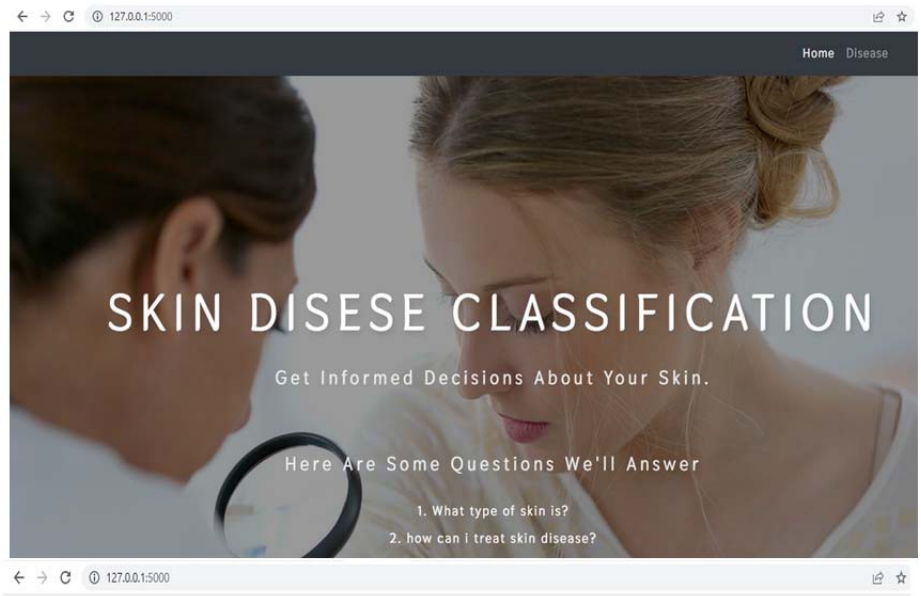
In this paper, a method of skin cancer classification using machine learning and image processing is implemented. In the first step, a novel method of contrast stretching based on the mean and standard deviation of pixels for dermoscopic images enhancement is proposed. Then OTSU thresholding is performed for segmentation. Shape, color, and texture features are extracted in the second step, and shape features are then reduced using the PCA. The class imbalance problem of the ISIC dataset is overcome using the SMOTE sampling technique. In the third step, features are standardized and scaled and then a novel approach of features selection based on wrapper methods for selecting the most optimum features is proposed. The recommended system is tested on an openly available dataset i.e. ISIC-ISBI 2016 and it is concluded that the proposed wrapper method for feature selection in combination with the Random Forest classifier gives promising results as compared with other classifiers.

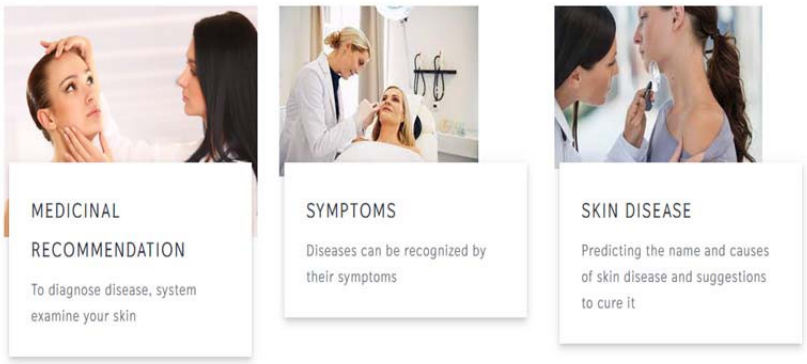
Future Scope

This paper will provide directions for the prevention and early detection of skin cancer in POC based on a comprehensive review of available data. Due to limited research, there are few resources providing insights for evaluating darkly pigmented lesions in POC. Diagnostic features for different types of skin cancers were recorded and various possible risk factors were considered.

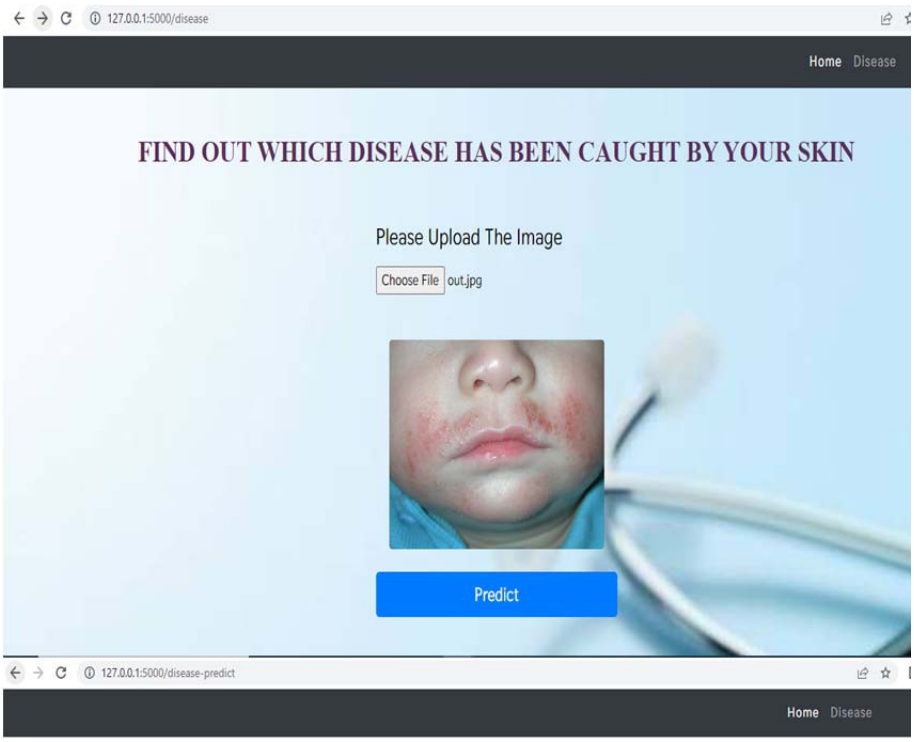
ISBN : 978-93-91871-47-5

Sample Screen Shots

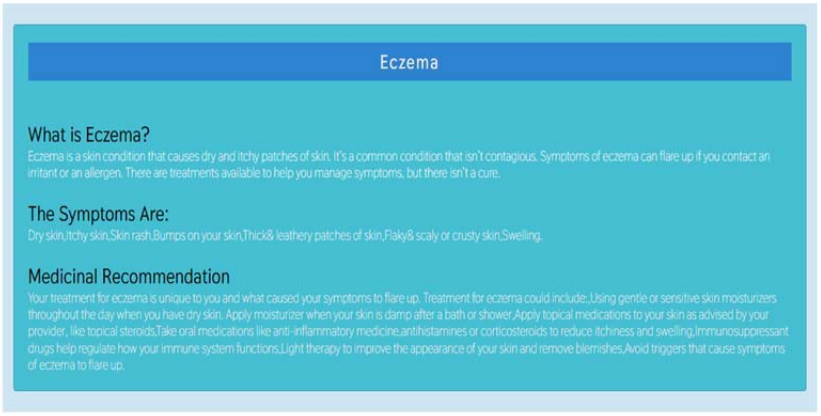




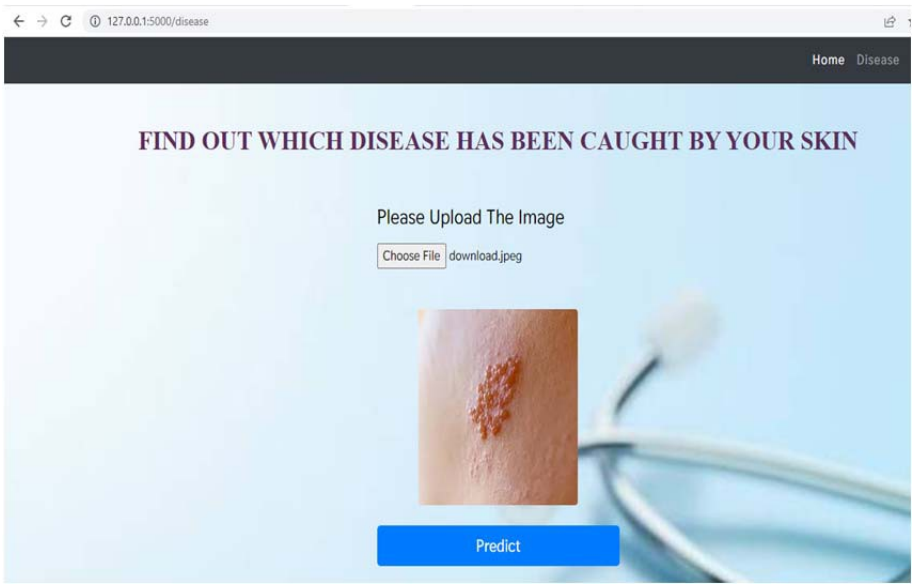
ISBN : 978-93-91871-47-5

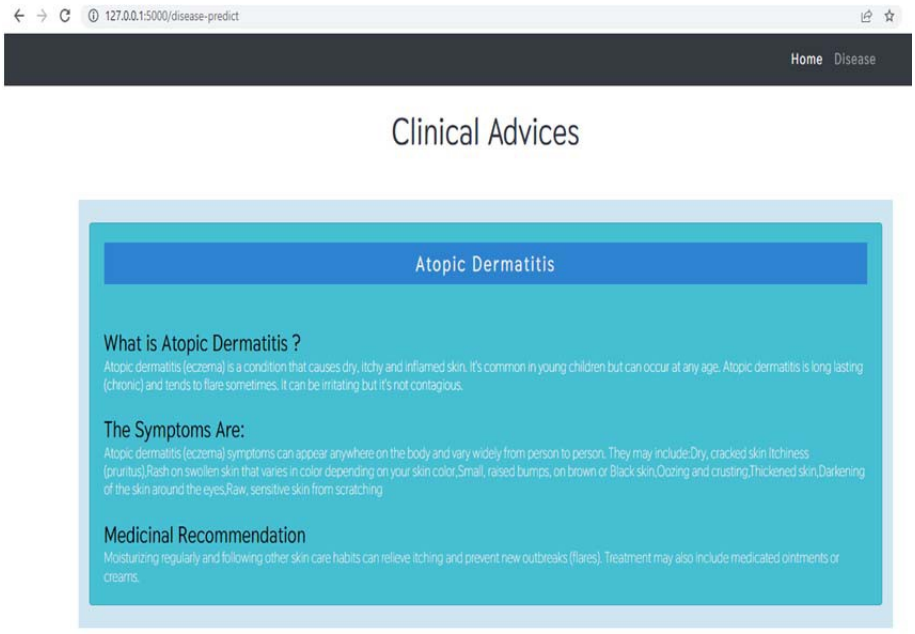


Clinical Advices



ISBN : 978-93-91871-47-5





ISBN : 978-93-91871-47-5

REFERENCES

[1] D. Schadendorf et al., “Melanoma,” Lancet, vol. 392, no. 10151, pp. 971–984, 2018.

[2] S. Gupta and H. Tsao, “Epidemiology of melanoma,” Pathol. Epidemiol. Cancer, pp. 591–611, 2016.

[3] F. Nachbar et al., “The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions,” J. Am. Acad. Dermatol., vol. 30, no. 4, pp. 551–559, 1994.

[4] “The International Skin Imaging Collaboration (ISIC).” https://www.isicarchive.com/#!/topWithHeader/tightContentTop/about/isicArchive (accessed Oct. 04, 2020).

[5] H. Alquran et al., “The melanoma skin cancer detection and classification using support vector machine,” 2017 IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. AEECT 2017, vol. 2018- Janua, pp. 1–5, 2017.

[6] U. B. Ansari and M. E. Student, “Skin Cancer Detection Using Image Processing Tanuja Sarode 2,” Int. Res. J. Eng. Technol., vol. 4, no. 4, pp. 2395–56, 2017, [Online]. Available: https://www.irjet.net/archives/V4/i4/IRJET-V4I4702.pdf.

[7] Rao, K. S., Yelkar, P. S., Pise, O. N., & Borde, S. (2021). Skin Disease Detection using Machine Learning. International Journal of Engineering Research & Technology (IJERT), 9(3), 2278-0181.

[8] Bhadula, S., Sharma, S., Juyal, P., & Kulshrestha, C. (2019). Machine learning algorithms based skin disease detection. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(2), 4044-4049.

[9] Kumar, V. B., Kumar, S. S., & Saboo, V. (2016, September). Dermatological disease detection using image processing and machine learning. In 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR) (pp. 1-6). IEEE.

[10] Kolkur, M. S., Kalbande, D. R., & Kharkar, V. (2018). Machine learning approaches to multi-class human skin disease detection. International Journal of Computational Intelligence Research, 14(1), 1-12.