International Journal of Advanced Science and Technology
Vol. 29, No. 7s, (2020), pp. 1085-1095

Skin Cancer Classification Using Machine Learning Techniques

Dr.S.Arivazhagan[1], V.Saranya Devi [2]
Centre for Image Processing and Pattern Recognition,
Department of Electronics and Communication Engineering,
Mepco Schlenk Engineering College, Tamil Nadu, India.
[1]sarivu@mepcoeng.ac.in, [2]sdsaranya67_ec@mepcoeng.ac.in

Abstract

Skin cancer is the most common form of malignancy that affects human populations. The Basal cell carcinoma, Squamous cell carcinoma and Melanoma are the most commonly occurring skin cancers on the human skin that occurs due to the abnormal growth of skin cells. All these three types of skin cancers are the most dangerous form of skin cancers whose treatment is possible only if it is detected in early stage. Therefore, the classification of these three cancers types are really challenging. In this paper, these three commonly occurring skin cancers on human skins are classified using machine learning techniques. The proposed technique consists of three stages, namely, preprocessing, feature extraction, and classification. The database consists of 150 images of the three cancer types which are taken from the Dermnet public database. Thus, each cancer types having 50 images are used in this paper. In the first stage, the images were preprocessed using two methods, the edge enhancement method and then the histogram equalization method. In the second stage, two different features vectors were used such as texture and statistical features. In the classification stage, the support vector machine classifier is used to classify the three cancer types as it uses a decision plane to separate a data set having different classes. The performance measures were analyzed for the texture and statistical feature vectors individually as well as combined. The overall classifier performance measures used are accuracy, sensitivity and specificity. A classificiation with a success of 97% accuracy has been obtained by using SVM classifier.

Keywords: *Basal cell carcinoma (BCC), Gray Level Co-occurrence Matrix (GLCM), Machine Learning (ML), Melanoma, Squamous cell carcinoma (SCC), Support Vector Machine (SVM)*

1. INTRODUCTION

Machine learning is finding its application in all fields. Developing Machine Learning based diagnostic tools are quite challenging to develop and it is important to choose the right decision making algorithm to achieve a better diagnostic accuracy. It is equally important to choose the right feature and Machine Learning algorithm to obtain the highest diagnostic accuracy. Many researches have been proposed to classify the skin cancers using machine learning techniques in the recent past.

An automated dermatological diagnostic system having two dependent steps were proposed [1]. At first this system detects the skin anomalies and then latter it identifies the diseased skin. To identify and diagnose the diseased skin the color image processing, K-means clustering algorithm and color gradient techniques were used. A set of global and local features are proposed for the detection of melanoma in [2]. These two different features were used to classify skin lesions. In [3], the image processing and data mining technologies are used for the detection and diagnosis for the diseased skin. In [4], the authors proposed an approach to detect the various kinds of diseases and used a dual stage approach method which is effectively combines Computer Vision and Machine Learning on clinically evaluated histopathological attributes to accurately identify the diseases. Image processing techniques are used to identify various skin diseases using ANN, edge detection and image filtering methods in [5]. In [6], deep learning algorithms are approached to diagnose the four common cutaneous diseases. The author proposed an image processing system to detect melanoma from nevus. For the removal of the noise from the skin lesion Gaussian filter is used and to segment out the lesion improved K-mean clustering algorithm is used. Finally, to classify the skin cancer into melanoma and nevus

ISSN: 2005-4238 IJAST

Support vector machine (SVM) is used. In [8], to recognize the skin disease from a given image, an automatic and effective visual analysis framework is implemented and a novel transfer learning model is used to explore the professional Web data in conjunction with the non professional data. In [9], a Bacterial Vaginosis is diagnosed by quantitative analysis of a three-step approach i.e., bacteria regions segmentation, overlapping bacteria splitting, and bacterial morphotypes classification. In [10], different types of image pre-processing methods are used to extract the features and for training and testing purposes Feed Forward Artificial Neural Network (FFANN) is used. The authors proposed a geometric features to differentiate between a benign lesion and a malignant one in [11]. The k-Nearest Neighbors (k-NN) Machine Learning (ML) algorithm is used to classify 15 lesions based on the ABD features. The results indicate that this technique is used to detect Melanoma skin cancer.

It is clear from the literature review that, only a few works have been concentrated on the classification of various skin cancer types. Few researchers have attempted to classify various skin diseases like psoriasis, eczema, acne etc., using different feature vectors with different classification algorithms. In this paper, commonly occurring skin cancers types of Basal cell carcinoma, Squamous cell carcinoma and Melanoma are classified using the Support Vector Machine (SVM) classifier method is attempted.

2. PROPOSED METHOD

The block diagram of the proposed method is shown in Figure 2.1. It involves the classification of the three commonly occurring skin cancer types namely, Basal cell carcinoma, Squamous cell carcinoma and Melanoma. The input images are taken from the Dermnet public database. The database consists of 150 images of the above three different skin cancer types and for each type of skin cancer, there are 50 dermatology images. First, the input images are pre-processed with some methods like edge enhancement and histogram equalization. Then the features are extracted from the pre-processed images. Finally, SVM classifier is used for the classification of the skin cancers as shown in Figure 2.1 and the performance of the classifier is analyzed.

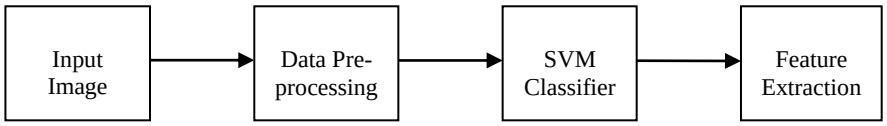
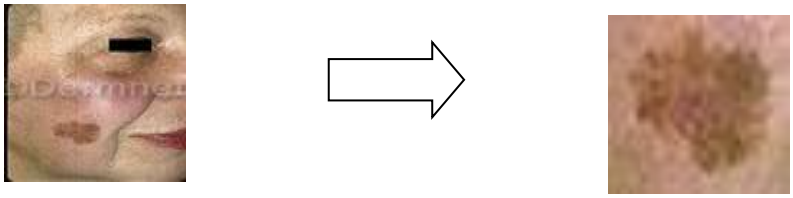


Figure 2.1: Proposed Method for skin cancer classification

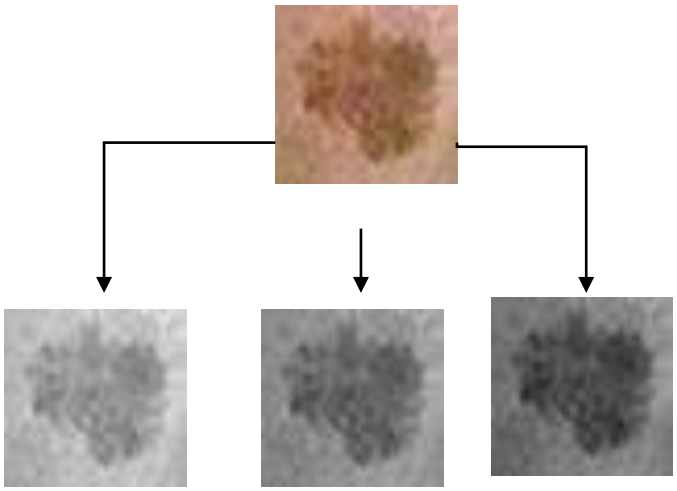
2.1 Pre-Processing

The Data Pre-processing stage which involves two methods namely, edge enhancement and histogram equalization. At first, the cancer affected regions are cropped in all images as shown in Figure 2.2 and separated as Red, Green and Blue channel (RGB separation) as shown in Figure 2.3.



(a) Original image **(b) Cropped image**

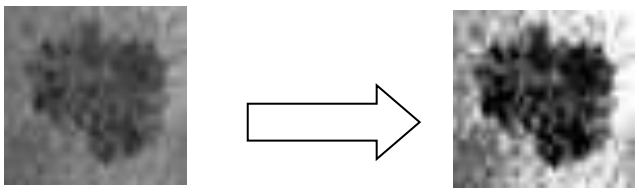
Figure 2.2: Skin sample image affected by skin cancer



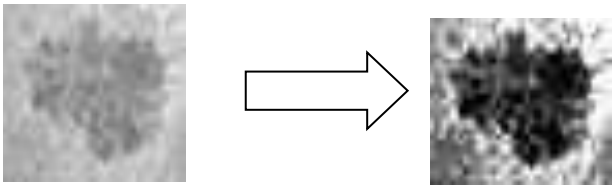
Red channel Green channel Blue channel

Figure 2.3: RGB Channel separation

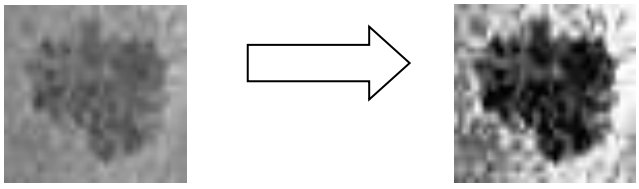
Then the edge sharpening method is used to highlight the edges of the skin affected areas in the image and then the histogram equalization method is used to improve the contrast of an image as shown in Figure 2.4.



(a) Sharpened Red channel (b) Histogram Equalized Red channel



(c) Sharpened Green channel (d) Histogram Equalized Green channel



(e) Sharpened Blue channel (f) Histogram Equalized Blue channel

Figure 2.4: Output of pre-processing skin cancer image

2.2 Feature Extraction

2.2.1 Gray Level Co-occurrence Matrix (GLCM) Feature

The Gray Level Co-occurrence Matrix (GLCM) is the two dimensional matrix of joint probabilities *p(i,j)* between pairs of pixels separated by a distance *'d'* in a given direction *'r'*. The second order image histogram referred to as the Grey Level Co-occurrence Matrix (GLCM) of an image offers greater information about the inter-pixel relationship, periodicity and spatial grey level dependencies. This matrix is a source of fourteen texture descriptors. Haralick defines fourteen GLCM based textural features to characterize texture statistics of an image. In 1973, Haralick et al, [15] described the textural features based on gray-tone spatial dependencies and the illustrated that the textural features were applicable to the image classification applications.

2.2.2 Statistical Feature

Statistical methods can be further classified into first-order (one pixel), second order (two pixel) and higher-order (three or more pixels) statistics. The basic difference is that first-order statistics estimate the properties (e.g. average and variance) of individual pixel values, ignoring the spatial interaction between image pixels, whereas, second and higher order statistics estimate the properties of two or more pixel values like variance and do not consider pixel neighborhood relationships. Images can be represented with higher order statistical parameters computed from co-occurrence or run-length matrices or from the frequential approaches. The common statistical features are given in eqn (1) to eqn (5).
Where,
I(i,j) be the intensity value of an image.
N be the number of discrete intensity levels in the image
σ be the standard deviation

1. Mean (μ)

$$= \frac{1}{N^2} \sum_{i,j=1}^{N} I(i,j) \tag{1}$$

2. Variance

$$= \sum_{i,j=1}^{N} (i-\mu)^2 I(i,j) \tag{2}$$

3. Standard Deviation(σ)

$$= \sqrt{\frac{\sum_{i,j=1}^{N} [I(i,j)-\mu]^2}{N^2}} \tag{3}$$

Skewness

$$= \frac{1}{\sigma^4} \sum_{i,j=1}^{N} (i-j)^3 I(i,j) \tag{4}$$

4. Kurtosis

$$= \frac{1}{\sigma^4} \sum_{i,j=1}^{N} (i-j)^4 I(i,j) \tag{5}$$

3. CLASSIFICATION

In this work, dermatology images of basal cell carcinoma, squamous cell carcinoma and melanoma are classified using the SVM classifier. The features used are the texture and statistical feature vectors. A multiclass SVM is used because of three cancer types are involved. The kernel function of Gaussian Radial Basis function is used. Finally the performance evaluation is done using accuracy, sensitivity and specificity. The experiments are conducted using Matlab2018a.

4. RESULTS AND DISCUSSION

In this paper, 150 images of three skin cancer types from Dermnet database are classified using SVM classifier. The classification is done by using (i) texture features derived from red, green and blue planes, (ii) statistical features derived from original RGB image and (iii) combined texture and statistical features, derived from the original images. The results obtained are shown in Table 4.1.

Table 4.1: Skin Cancer Classification using Texture, Statistical and combined Texture and Statistical Features

S.No.	Attributes	Accuracy (%)	Sensitivity (%)	Selectivity (%)
Texture based SVM				
1	RGB	93.33	96.10	90.23
2	R	94.33	95	93.45
3	G	94.12	95	90.60
4	B	93.55	94	89.67
Statistical based SVM				
5	Statistical Features	94.33	96.67	92.33
Texture and Statistical based SVM				
6	Texture and Statistical Features	94.67	100	92.13

From, the Table 4.1, it is observed that the combined Textures and Statistical features results in slightly improved performance over individual Texture or Statistical features.

In the skin cancer classification obtained using, combined texture and statistical features presented in Table 4.1, 50% of the images in each skin cancer category are used for training and the remaining 50% the images are used for testing.

Table 4.2: SVM Results – Texture Features

S.No.	Training and Testing Ratio	Accuracy (%)	Sensitivity (%)	Selectivity (%)
1	25:75	89.43	90	87.77
2	30:70	90.24	92.30	88.67
3	40:60	92.13	95.67	89.50
4	50:50	93.33	96.10	90.23
5	60:40	95.10	96.34	91

6	70:30	95.23	96.44	92.46
7	75:25	96.67	98	94

Table 4.2 shows the SVM Results for the texture features which is extracted from the combined red, green and blue channel and the results are obtained for the performance parameters at various training and testing ratios.

Table 4.3: SVM Results – Red Channel Features

S.No.	Training and Testing Ratio	Accuracy (%)	Sensitivity (%)	Selectivity (%)
1	25:75	90	92	88.67
2	30:70	91.56	94.33	90.13
3	40:60	93.33	94.67	92.22
4	50:50	94.33	95	93.45
5	60:40	94.67	95.77	92
6	70:30	95	96.67	92.33
7	75:25	95.45	97	93

Table 4.3 shows the SVM results, obtained for the red channel features.

The SVM results obtained for the green channel features are shown in Table 4.4.

Table 4.4: SVM Results – Green Channel Features

S.No.	Training and Testing Ratio	Accuracy (%)	Sensitivity (%)	Selectivity (%)
1	25:75	90	91.63	87.50
2	30:70	91.13	94	88.77
3	40:60	93.33	94.23	90
4	50:50	94.12	95	90.60
5	60:40	94.33	95.77	91.23
6	70:30	95	96	91.68
7	75:25	95.23	97	92

The SVM results obtained for the blue channel features are shown in Table 4.5.

Table 4.5: SVM Results – Blue Channel Features

S.No.	Training and Testing Ratio	Accuracy (%)	Sensitivity (%)	Selectivity (%)

1	25:75	91.23	93	88
2	30:70	91.56	93.33	88.33
3	40:60	92.13	94	89
4	50:50	93.55	94	89.67
5	60:40	94.23	95	90
6	70:30	94.67	95	91.33
7	75:25	95	96	92

The SVM results obtained for the statistical features are shown in Table 4.6.

Table 4.6: SVM Results – Statistical Features

S.No.	Training and Testing Ratio	Accuracy (%)	Sensitivity (%)	Selectivity (%)
1	25:75	92.47	94.67	90
2	30:70	93.33	95.33	90.45
3	40:60	94.13	96.45	92
4	50:50	94.33	96.67	92.33
5	60:40	94.67	98	94
6	70:30	95	98.12	94.23
7	75:25	96.23	98.34	94.45

Table 4.7: SVM Results – Combined Texture and Statistical Features

S.No.	Training and Testing Ratio	Accuracy (%)	Sensitivity (%)	Selectivity (%)
1	25:75	80.44	85.47	75.42
2	30:70	82.33	86.67	78.33
3	40:60	85.00	90.12	82.50
4	50:50	94.67	100	92.13
5	60:40	95.00	100	93.17
6	70:30	95.97	100	93.57
7	75:25	97.22	100	95.83

By comparing the performance analysis in Table 4.7 for varying Training and Testing ratios, it is observed that the combined texture and statistical features provides the best accuracy of 97.22% for Training and Testing ratio of 75:25.

However, the whole skin cancer classification is repeated by varying training and testing ratio from 25:75 to 75:25 and the results obtained are presented from the Table 4.2-4.7.

The graphical representation of the skin cancer classification results for the different feature vectors of the ratio 75:25 is shown in Figure 4.1. The overall accuracy comparision for the different feature vectors are shown in Figure 4,2.

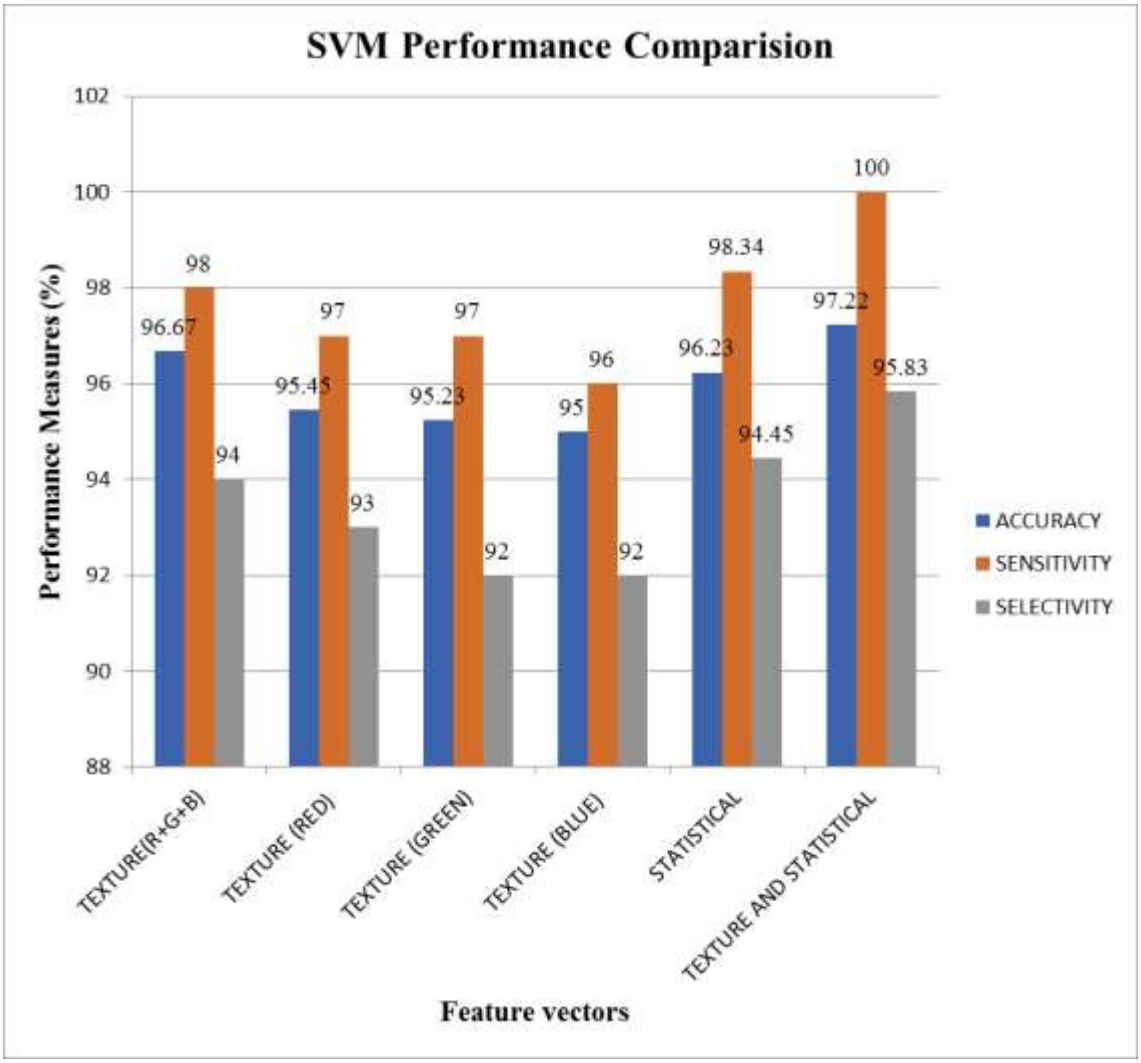


Figure 4.1: Graphical representation of SVM Performance Comparison

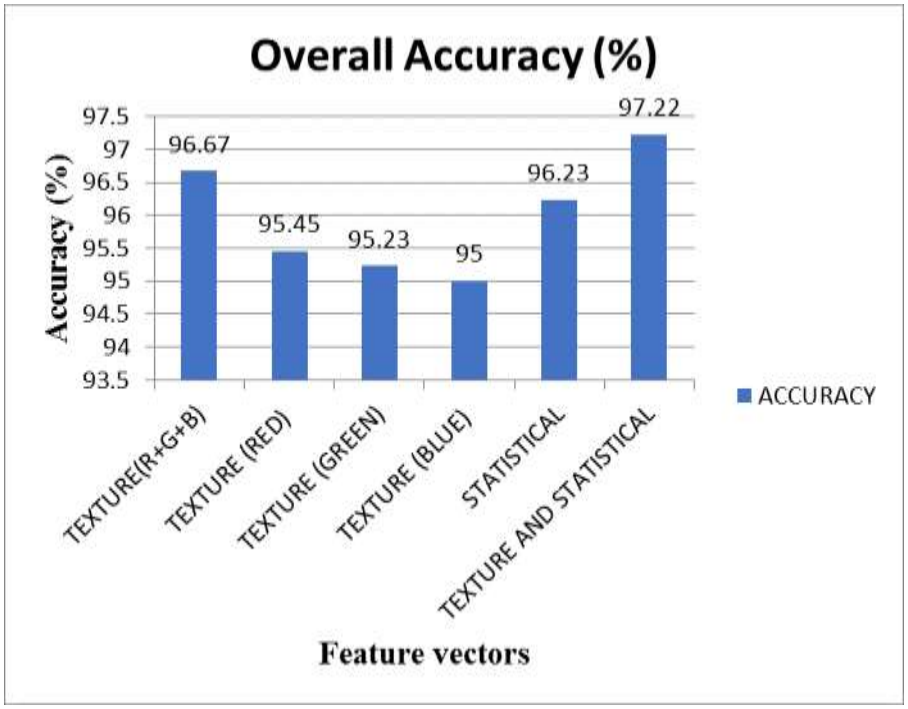


Figure 4.2 Graphical representation of overall accuracy

5. CONCLUSION

The classification of skin cancer types of basal cell carcinoma, squamous cell carcinoma and melanoma are performed in this work. Initially, preprocessing method namely, edge enhancement and histogram equalization is performed to extract the features. Two different feature vectors such as Gray Level Co-occurrence Matrix features and statistical features are extracted from the preprocessed images. The classification of the three cancer types are performed using support vector machine (SVM) classifier. The performance evaluation is done for the support vector machine classifier (SVM) using accuracy, sensitivity and selectivity parameters.

From the observation of this work, the selectivity is comparatively lower than the accuracy and sensitivity. From the experiments, it is concluded that the proposed method of combining texture and statistical features provides the best accuracy of 97.22%.

REFERENCES

1. VM. S. Arifin, M. G. Kibria, A. Firoze, M. A. Amini, and H. Yan, "Dermatological disease diagnosis using color-skin images" International Conference on in Machine Learning and Cybernetics (ICMLC), vol.5, pp. 1675–1680, July 2012.
2. Catarina Barata, Margarida Ruela, Mariana Francisco, Teresa Mendonça, and Jorge S. Marques, "Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features" IEEE Systems Journal, vol.8, no. 3, September 2014.
3. Amarathunga, E. Ellawala, G. Abeysekara, and C. Amalraj, "Expert system for diagnosis of skin diseases," International Journal of Scientific & Technology Research, vol.4, no. 01, pp. 174–178, January 2015.
4. V.B.Kumar, S.S.Kumar, and V.Saboo, "Dermatological disease detection using image processing and machine learning," International Conference on Artificial Intelligence and Pattern Recognition (AIPR), October 2016.

5. N.Yadav, N.Yadav, and V.K.Narang "Skin diseases detection models using image processing: A survey," International Journal of Computer Applications (IJCA), (0975–8887) vol. 137, no.12, March 2016.
6. X.Zhang, S. Wang, J. Liu, and C. Tao ''Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge,'' Med. Inform. Decis. Making, vol. 18, no. 2, November 2017.
7. Muhammad Qasim Khan , Ayyaz Hussain, Saeed Ur Rehman , Umair Khan, Muazzam Maqsood , Kashif Mehmood, And Muazzam A. Khan "Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer" IEEE, vol:7, pp. 90132-90144, July 2019.
8. Yingjie Xia, Luming Zhang , Lei Meng, Yan Yan, Liqiang Nie, Member and Xuelong Li, "Exploring Web Images to Enhance Skin Disease Analysis Under A Computer Vision Framework " IEEE Transactions On Cybernetics, vol. 48, no. 11, November 2018.
9. Youyi Song, Liang He, Feng Zhou, Siping Chen, Dong Ni, Baiying Lei and Tianfu Wang, "Segmentation, Splitting, And Classification Of Overlapping Bacteria In Microscope Images For Automatic Bacterial Vaginosis Diagnosis", Journal of Biomedical And Health Informatics, Vol. 21, no. 4, July 2017.
10. Nalini Bodasingi, Narayanam Balaji, (2017), "Classification of multiple diseases based on wavelet features", The Journal of Engineering, vol. 2017, Iss. 4, pp. 110–118, January 11, 2017.
11. Rebecca Moussa, Firas Gerges, Christian Salem, Romario Akiki, Omar Falou, and Danielle Azar (2016), "Computer-aided Detection of Melanoma Using Geometric Features", 3rd Middle East Conference on Biomedical Engineering (MECBME). ISSN:16467837, 2016.
12. M. Emre Celebi, Azaria Zornberg, (2014), "Automated Quantification of Clinically Significant Colors in Dermoscopy Images and Its Application to Skin Lesion Classification", IEEE Systems Journal, VOL. 8, no. 3, pp. 980-984, September 2014.
13. Ioannis Valavanis, Ilias Maglogiannis and Aristotelis A. Chatziioannou, "Exploring Robust Diagnostic Signatures for Cutaneous Melanoma Utilizing Genetic and Imaging Data", IEEE Journal Of Biomedical And Health Informatics, Vol. 19, no. 1, January 2015.
14. Juan Lu, Ed Kazmierczak, Jonathan H. Manton, and Rodney Sinclair, (2013), "Automatic Segmentation of Scaling in 2-D Psoriasis Skin Images", IEEE Transactions On Medical Imaging, Vol. 32, no. 4, April 2013.
15. Robert M. Haralick, K. Shanmugam, and It'shak Dinstein, "Textural Features for Image Classification", IEEE Transactions On Systems, Man And Cybernetics, Vol.SMC- 3, no. 6, November 1973.
16. Jeremy Kawahara , Sara Daneshvar , Giuseppe Argenziano, and Ghassan Hamarneh, "Seven-Point Checklist And Skin Lesion Classification Using Multitask Multimodal Neural Nets", IEEE Journal Of Biomedical And Health Informatics, Vol. 23, No. 2, March 2019.
17. R. Yasir, M. A. Rahman, and N. Ahmed, (2014), "Dermatological disease detection using image processing and artificial neural network," in 8th International Conference on Electrical and Computer Engineering: Advancing Technology for a Better Tomorrow, ICECE. Pan Pacific Sonarga on Dhaka, pp. 687–690, 2014.
18. Zhe Wu, Shuang Zhao, Yonghong Peng, Xiaoyu He,Xinyu Zhao,Kai Huang, Xian Wu, Wei Fan, Fangfang Li, Mingliang Chen, Jie Li, Weihong Huang, Xiang Chen And Yi Li, "Studies On Different CNN Algorithms For Face Skin Disease Classification Based On Clinical Images", Data-Enabled intelligence for digital health, Vol.7, pp. 66505-66511, May 22 2019.
19. V. K. Shrivastava, N. D. Londhe, R. S. Sonawane, and J. S. Suri, (2015), "Reliable and accurate psoriasis disease classification in dermatology images using

comprehensive feature space in machine learning paradigm", Expert Systems with Applications, vol. 42, no. 15-16, pp. 6184–6195, 2015.
20. M. A. Sheha, M. S. Mabrouk, and A. Sharawy, (2012), "Automatic detection of melanoma skin cancer using texture analysis," International Journal of Computer Applications, vol. 42, no. 20, pp. 22–26, 2012.

ISSN: 2005-4238 IJAST