Performance Comparison of Machine Learning Based Classifiers for Melanoma Cancer Diagnosis

K.Vijila Rani
Research Scholar
Faculty of Information & Communication Engineering
Anna University, Chennai, Tamil Nadu, India.
Email: *vijilarani.k@gmail.com*

Abstract
In recent years one among the rising deadliest diseases is skin cancer. Skin cancer is one among the foremost difficult illness among numerous cancer kinds. This paper proposes an automated skin lesion analysis system for the first detection and classification of melanoma using image process techniques. First, dermoscopy image of skin is taken as image acquisition step, then pre-processing step for noise removal and post-processing step for image improvement. Identification of the diseased/abnormal portion of the skin is possible solely by correct delineation ways. Therefore here the processed image undergoes image segmentation. Second, options are extracted using feature extraction technique – ABCD parameter, GLCM, and FOS. A comparison of the performance of all feature sets is conferred during this paper so as to see what feature sets offer the most effective classification results. Numerous feature combinations are given because the input to the KNN, SVM & ANN classifiers. Performance is analyzed supported the accuracy of the learning classifier output.

Keywords -Skin Cancer, Otsu, Morphological, Feature extraction, KNN, SVM, ANN

INTRODUCTION

Skin cancer is a malignant tumor that unevenly grows in the skin. Melanoma cancer can spread out to all parts of the body through the lymphatic system and blood. The early detection of melanoma cancer can be helpful to save the human life. When melanomas occur, they typically arise from pigmented moles that are large, asymmetric, with irregular borders and coloration. Bleeding, itching, and a mass below the skin are additional signs of cancerous [1]. There are various types of dermoscopy equipment, but all of them use the same rule and allow registering skin images with x10 magnification and above. Due to light source integrated into dermatoscopy lens, there occurs to be a problem with skin reflections [6][7].

To overcome this problem, a liquid is used as a medium layer between the lens and the skin in the proposed method. Acquired image contains both healthy and unhealthy skin. Normally, taken image contains both tissue and unwanted particles.

There are many preprocessing steps are done in past papers to remove noise, hair& the enhancement of image. Gaussian [8], wiener [2] & gabor[3] filters are used to remove noise and thin hair particles from the acquired the image. There are some problems occur in this process.To avoid this, dull-laser software [9] [10] [11] [12] is used, which is not applicable for certain format (Bit map & png) of an image. Median filter [6] [8] is used in this proposed system, which is an efficient method to remove the noise but some blur particles occur. Top-hat transform is used to extract the wide hair particle, which is a morphological method [13].

We have to separate a healthy skin tissue from the affected area, for that many segmentation types are followed in this paper. Many segmentation processes are done in this paper[14].In paper [2], active contour model (ACM) is used, which is based on the principle of minimization of the energy defined on a closed curve comprising control points. Statistical region merging (SRM) process is start from pixel level to region level until convergence[11] [5]. Paper [6] deals with the region growing method, which begins with a pixel as the seed point and then calculate the average of the regions.

Otsu segmentation is used in this paper, which easily extracts the object from the background by setting an appropriate threshold value [4] [14] [12]. Erosion operation is used to erode the bright part present in the background of the image. Perfect segmented leads the best accuracy of further processing. The main aim of the proposed method is to detect &classify the skin tissue, it is a cancer skin or not by using SVM, ANN&KNN classifier. Improve the accuracy of skin classification & also find border error. Then performances are analysis based on the output of classifiers. Finally comparisons are made.

PROPOSED METHOD

In the proposed skin cancer detection and classification system, dermoscopic images are taken as input images.

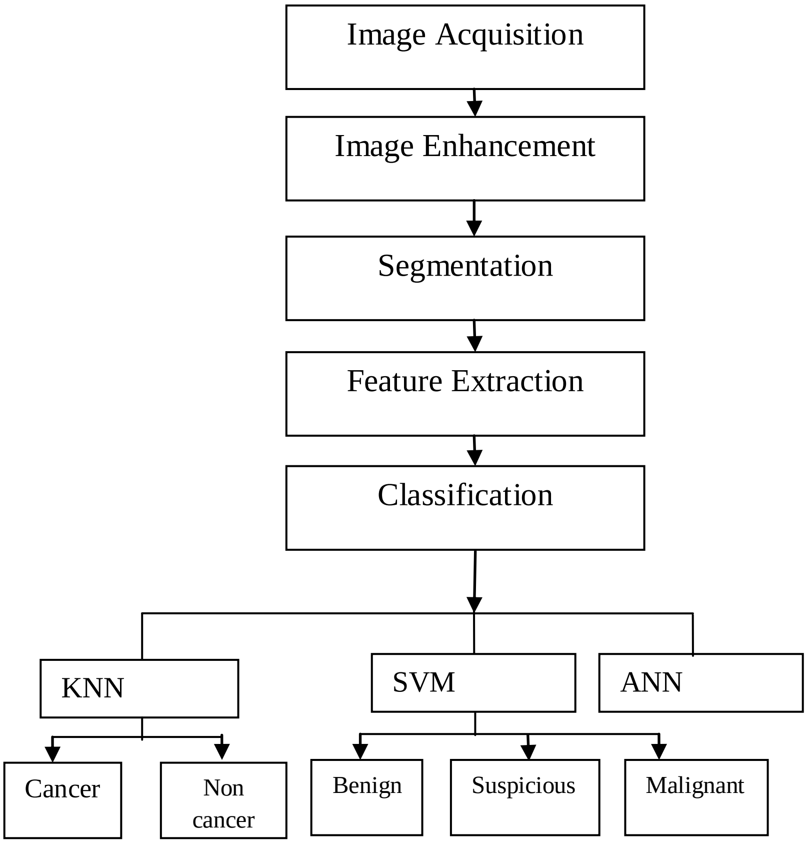

Fig.1 *Block Diagram of Skin cancer detection &classification*

The design of a system for melanoma detection and classification based on Image Processing techniques include the following stages:

Stage 1: Image Acquisition

Cancer affected skin images are captured using the dermoscopy. This type of device is called Dermatoscopy or Epiluminescence Light Microscopy (ELM)

(Fig.2). Acquired image is a high-resolution digital image which is also known as a Dermoscopic image (Fig.3). Oil is applied on the surface of affected skin area, and the lens of a microscope is placed directly on the surface of the skin. Pigmented structure of the epidermis is taken out with the help of this device by lighting at an angle of 460˚. Two advantages: examine the depth of infection & enhance the visibility of skin layers.

***Fig.2** Dermatoscopy & Dermatoscopic image*

Stage 2: Image Enhancement

The acquired dermoscopic image is always in digital format which is subjected to various Digital Image Processing Techniques. Image Processing consists of following procedures:

- First, the image is resized to 180x180 pixels.
- Second, Gamma correction can repeatedly enhance the image contrast.
- Then color (RGB) image is converted to gray scale image using mat lab command.

***Gamma correction**:* Gamma correction is a non-linear function, which is used to correcting brightness (lightness or darkness) of the image. The formula for calculating the resulting output is as follows:

$$I' = 255 \text{ X } \left(\frac{I}{255}\right)^{\gamma} \quad (1)$$

Where I' is the gamma corrected intensity value, I is the intensity value of original image &γ is the gamma value.

***Top-hat transform**:*In mathematical morphology and digital image processing to extracts small elements and details from given images, the top-hat transform is used. Top-hat transform can be divided into two types. They are: white top-hat transform&black top-hat transform.

***Dilation**:* it is the basic operators in mathematical morphological processing. Dilation on binary images is to make wider the areas of foreground pixels at their borders of the image. The areas of foreground pixels grown in size, but the background pixels are shrinks in their size.

***Median filter**:* For smoothing image from noise & thin hair, median filtering is used. Median filtering is used for denoising. In the median filter, the operation will be performed on non-linear function. Median filtering is used for reducing the occurrences of small structures like thin hairs and small air bubbles. In order to improve the result of future process & to preserve the edges of input image this type of filtering is done.

Stage 3: Segmentation

Segmentation plays a major role in the detection of skin cancer through digital image processing, which maximizes the accuracy of further steps. Before segmenting a cancer affected area, we must consider the size, lesion shapes, color, and textures of the input image. It refers to the partitioning an image into two or more put out of joint regions that are unique on some property. In order to separate an object (region of interest) from background, an optimal thresholding is used which is called Otsu's method. This method partitions the image into two classes based on differentiating analysis**.** In this paper, gray scale image is taken for segmentation which leads to an effective separation of the image into two classes. The output of segmented image contains two intensity values (0, 1). Normally, background pixel value is 0 but sometimes background area pixel value is 1 because some properties of objects are similar to the background. To avoid that problem, morphological process is done. The main aim of segmentation stage is to extract the lesion area from the healthy skin images. For example in a given gray scale image, represented in L gray levels {0,1,2,………L},this method divide the gray scale image pixel values into two classes S0={0,1,2………t} & S1={t+1,t+2……………..L-1}. Consider the number of pixels in the i_{th} gray level image is n_i , where n is the total number pixels in an image. The probability of occurrence of gray scale image formula is given by [14]:

p_i=n_i/n (2)

Stage 4: Feature extraction

To diagnose skin lesion automatically, we will follow feature extraction process. There are three feature extraction processes to be considered. They are ABCD-rule of dermatoscopy, FOS, GLCM.

ABCD Parameter

ABCD is defined as asymmetry, border structure, color variation, and diameter of the lesion. If the two half patterns of the skin lesion are equal, it is called the symmetric pattern. The degree of symmetry is determined using the value of asymmetry feature i.e. Asymmetry Index (AI).

Asymmetry Index: Asymmetry Index is computed with the following equation:

AI = (A1+A2)/2Ar (3)

Border Irregularity: To calculate the amount of border irregularity, four different measures are found: compactness index, fractal index, edge abruptness.

1) Compactness Index: Compact Index can be determined by using the following equation:

CI= (P2L) = (4AL) (4)

Where, PL = Perimeter of the Lesion.

AL = Area of the Lesion.

2) Fractal Index: Box-counting method is used to calculate the fractal set. It returns two variables whose differential log ratio gives the fractal dimension as the mean value along 4-8.

3) Edge abruptness: Edge abruptness is calculated using the following equation

EI= ((Max Min) % 6+2)/100 (5)

Where Max= length of the major axis.Min=length of the minor axis.Axis lengths are calculated using *region props* function. In order to calculate the edge variations, mean & variance must be known. It is the difference between the variance and means.

4) Pigment transition: Gradient magnitude and direction of an image are measured. When the gradient operator is applied at the borders of the image, values outside the image bounds are set to nearest image border value. Three types of gradient operators are available, but here Prewitt mask is used as a gradient operator. Finally, mean and variance values are calculated.

Color Index: Color image is converted to HSV image for checking how many colors are present in the input image.

Diameter: To calculate Diameter the *region props* function is used to get the minor axis length of the lesion region. The resultant value is converted into mm value, and the value is assigned to diameter.

Stage 5: Classification methods

KNN Classifier

An instance-based learning method called the K-Nearest Neighbor or K-NN algorithm has been used in many applications in areas such as statistical pattern recognition, image processing. In order to perform the classification process, take 80 melanoma images which contain

40 cancerous and 40 non-cancerous images. KNN classifier output is displayed in a confusion matrix of 3x3. Various combinations of features are combined and given to the classifier. From the classifier result, we find the accuracy of the image.

SVM Classifier

Support Vector Machine (SVM) is typically used for learning classification. Two main properties of SVM are: it achieves high generalization by maximizing the margin, and it supports an efficient learning of nonlinear functions. To perform the classification process, take 54 melanoma images which contain 18 benign images, 18 suspicious images and 18 malignant images. SVM classifier output is displayed in a confusion matrix of 3x3. Various combinations of features are combined and given to the KNN classifier. From the classifier result, we find the accuracy of the image. When compared to other learning classifier SVMs have several advantages.

ANN Classifier

A classifier classifies the given datasets into cancerous and non-cancerous. Here a computer basedclassifier implemented in MATLAB software is used for classification purpose. Figure 7 shows that, there are 13 features, the classifier network consists of 13 inputs. Number of hidden neurons taken is 4 and one output neuron.Then the feature selection method is used to minimize the number of input features. Finally the input feature taken for this classifier is 9. The activation function used is tan sigmoid function. The output of the network is 0 or 1. Zero indicates anon-cancerous or benign condition and one indicates Cancerous condition or malignant melanomacondition. ANN is trained using Back propagationalgorithm, by giving known values of features and desired output. Weights are initialized randomly.

During each epoch, the weights are updated so that error between desired output and actual output is minimum. Sixty image datasets were taken for classification.From the ANN classifier, we plots 5 various characteristics. They are Performance, train state, error histogram, confusion, receiver operating characteristics.

EXPERIMENTAL RESULTS AND DISCUSSIONS

Table.1*Extracted features of Skin Images*

Image no	ABCD Feature					GLCM Feature				FOS Feature				
	Asymmetry	Border irregularity	Color	Diameter	TDS	Contrast	Correlation	Energy	Homogeneity	Mean	Standard deviation	Skewness	Kurtosis	Entropy
1	0.157	4.1376	3	4	4.11835	0.008566	0.97740	0.6123	0.995717	0.006	0.01080	2.01778	6.76021	2.3850
2	0.518	3.2295	3	4	4.49716	0.005587	0.968634	0.8163	0.997207	0.013	0.018964	1.384442	3.37708	2.98339
3	0.162	5.0918	4	4	4.72106	0.005338	0.978297	0.7487	0.997331	0.009	0.012767	1.655632	4.73059	2.77234
4	0.167	3.7503	3	4	4.09217	0.005214	0.976312	0.7746	0.997393	0.011	0.013236	1.244131	3.33805	2.91287
5	0.114	3.6505	3	4	4.01382	0.006642	0.973369	0.7439	0.996679	0.011	0.019105	2.274177	7.10785	2.68715
6	0.107	3.4877	3	4	3.98798	0.005742	0.983522	0.6463	0.997129	0.006	0.010301	1.605176	4.28244	2.05357
7	0.139	6.9425	3	4	4.37624	0.008752	0.956693	0.7892	0.995624	0.008	0.010716	1.578003	4.13867	2.57759
8	0.362	3.1396	3	4	4.28458	0.004593	0.969224	0.8461	0.997703	0.010	0.01203	0.826578	2.47775	2.77483
9	0.451	4.9752	3	4	4.58421	0.009621	0.976871	0.5744	0.995189	0.011	0.01429	0.945757	2.65592	2.23901
10	0.201	4.4543	3	5	4.70752	0.00748	0.982728	0.5598	0.99626	0.010	0.008143	0.419944	2.28365	2.81441

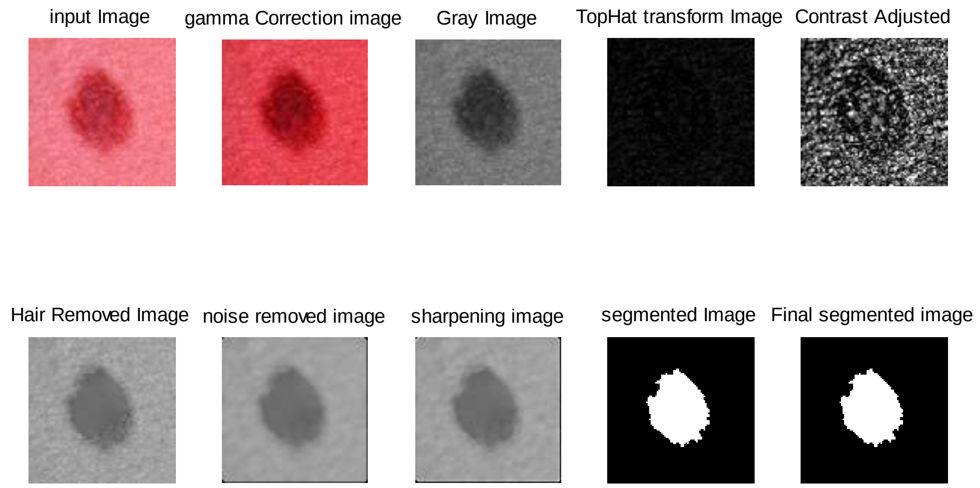

Fig.3 *Benign skin image-segmentation*

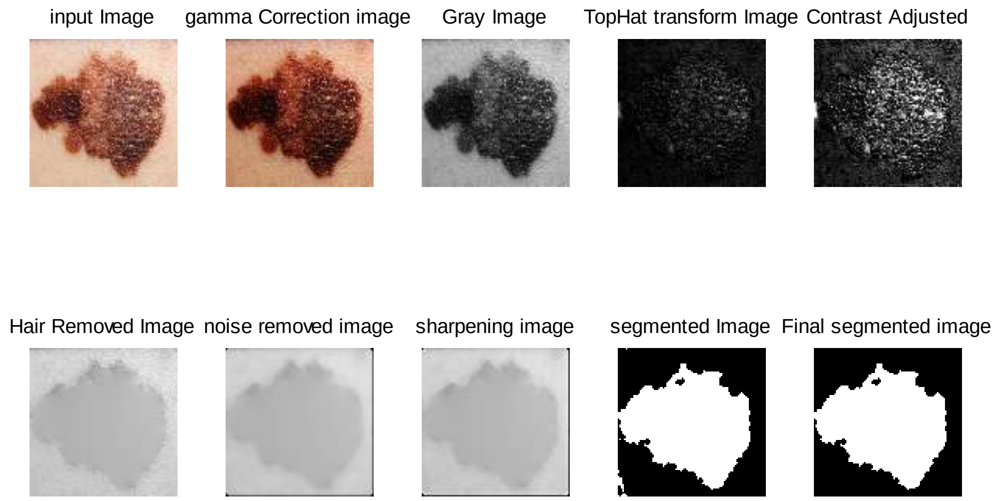

Fig.4 *Suspicious skin image-segmentation*

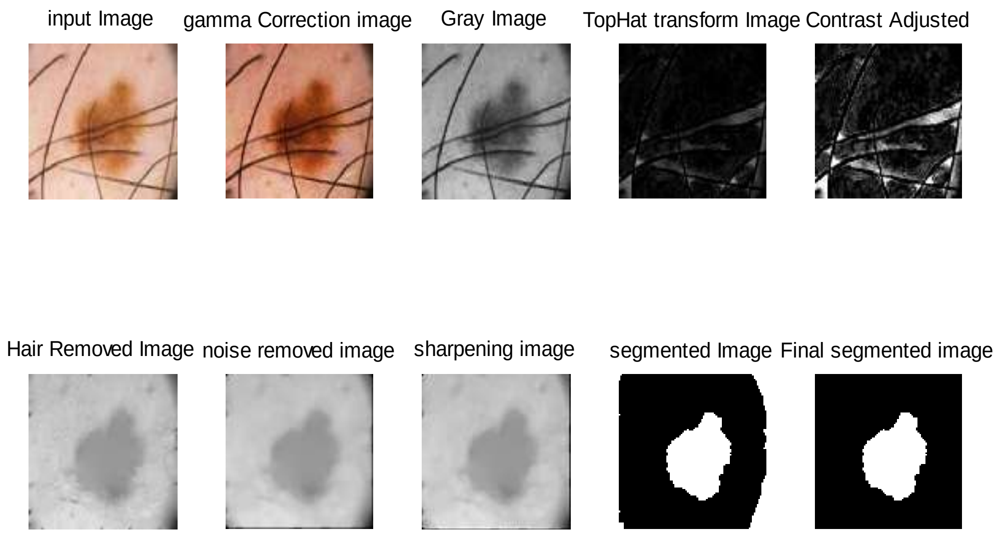

Fig.5 *Malignant skin image-segmentation*

Table.2 *KNN Classifier standard metrics values of skin based on various feature combination*

Features	Sensitivity	Specificity	accuracy	precision	similarity	border error
ABCD	93.3	100	96.7	100	96.6	3.3
GLCM	86.7	86.7	86.7	86.7	86.7	13.3
FOS	73.3	66.7	70	68.8	71	30
ABCD+GLCM	93.3	100	96.7	100	96.5	3.3
ABCD+FOS	93.3	80	86.7	82.4	87.5	13.3
GLCM+FOS	80	66.7	73.3	70.6	75	26.7

Table.2 shows that the standard metrics values of various features, these are output from the KNN classifiers. From the table, we can understand which feature combinations gives the higher accuracy of the classifier. Border error denotes how many skin images are represented wrongly. ABCD and ABCD + GLCM give the higher accuracy of 96.7%. GLCM and ABCD + FOS give the accuracy of 86.7%. FOS and GLCM + FOS give the worst accuracy.

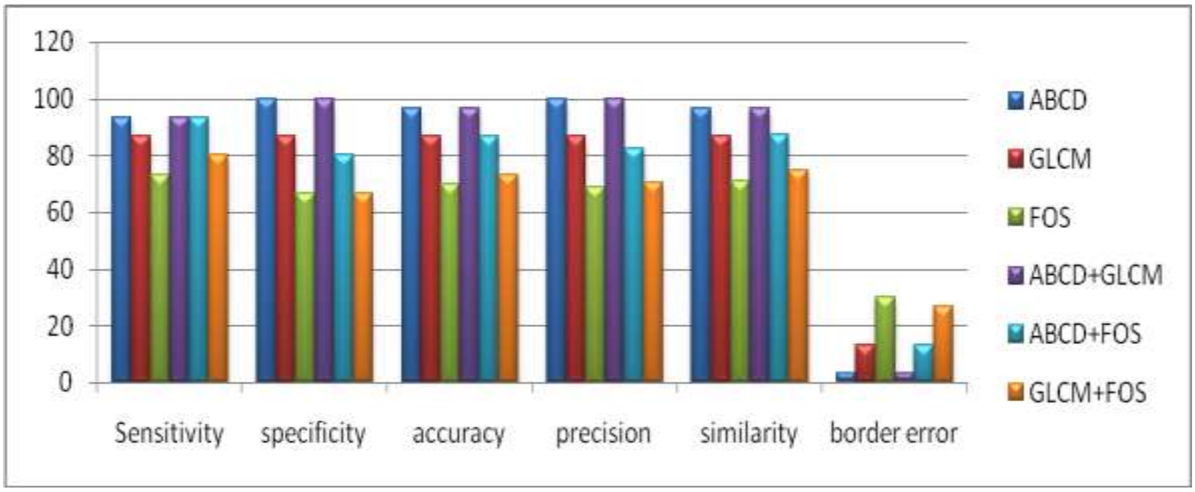

Fig.6 *KNN Classifier standard metrics values of skin based on various feature combination*

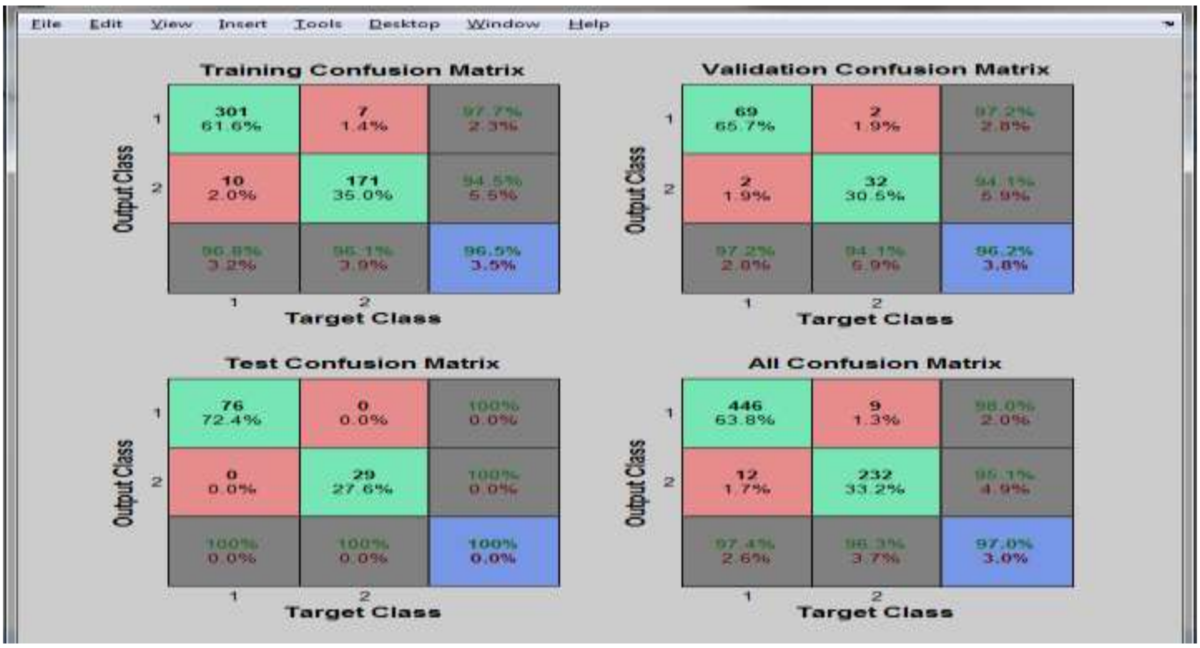

Fig.7. *Confusion Matrix*

Table.3 *Comparisons of classifier accuracy and error values*

Classifier	KNN	SVM	ANN
Accuracy (%)	96.7	92.7	97.3
Error (%)	3.3	7.3	2.7

CONCLUSION

In this paper machine learning techniques used, it proves that this is the best method for diagnosis of skin cancer. The image of carcinoma is taken, and it's subjected to numerous pre-processing techniques for noise, hair removal, and image improvement. Cancerous part gets separated from the skin by the strategy of Otsu segmentation. The options of the segmental region are extracted and supported the options pictures are classified as Cancerous or Non-cancerous victimization KNN classifier and Benign, Suspicious, Malignant using SVM classifier. The proposed system definesan effective way to detect the skin lesion more accurately and faster by segmenting the lesions images.Moreover, it has got good accuracy and higher levels of quality images. All the above processes are done in Matlab software. From the experimental result, the KNN classifier, SVM classifier& ANN classifier accuracy will be 96.7%, 92.7%& 97.3% based on the features. Bag of visual words classifier is used to classify the dataset based on image categories and the accuracy will be 97%.

REFERENCES

1. Abdul Jaleel.J, SibiSalim and Aswin.R.B, “Artificial Neural Network Based Detection of Skin Cancer,”International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 1, Issue 3, ISSN 2278 – 8875, September 2012.
2. Jeyaramya.V, Navarajan.J, Prathipa.R and Ashok kumar.I, “Detection of Melanoma Skin Cancer using Digital Camera Images,” Journal of Engineering and Applied Sciences Vol. 10, No. 7, ISSN 1819-6608, April 2015.
3. JishaGopinath, Maya John, “Feature Extraction by Gabor Filter and Classification of Skin Lesion using Support Vector Machine,”International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue10, ISO 3297:2007, October 2015.
4. Jose Fernandez Alcon, AlinaCiuhu. C, “Automatic Imaging System With Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis,” IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, Vol. 3, No. 1, February 2009
5. Alexander Wong, Paul Fieguth,“Automatic Skin Lesion Segmentation via Iterative Stochastic Region Merging,”IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, Vol. 15, No. 6, November 2011
6. Nadia Smaoui, SouhirBessassi,” A developed system for melanoma diagnosis,” 2013, International Journal of Computer Vision and Signal Processing, 3(1), 10-17(2013)
7. Mariam A.Sheha, Mai S.Mabrouk, AmrSharawy, “Automatic Detection of Melanoma Skin Cancer using Texture Analysis,” International Journal of Computer Applications (0975 – 8887) Vol 42– No.20, March 2012
8. A.A.L.C. Amarathunga, E.P.W.C. Ellawala, “Expert System For Diagnosis Of Skin Diseases,”International Journal of Scientific & Technology Research, Vol. 4, Issue 01, ISSN 2277-8616, January 2015.
9. Aswin.R.B1, Abdul Jaleel .J, Sibi

Salim, "Implementation of ANN Classifier using MATLAB for Skin Cancer Detection," International Journal of Computer Science and Mobile Computing, pg.87 – 94, ISSN 2320–088X, December 2013,

10. Messadi.M, Ammar.M, Cherifi.Hand Bessaid.A, "Interpretable Aide Diagnosis System for Melanoma Recognition,"Bioengineer &Biomedical Sci 2014, http://dx.doi.org/10.4172/2155-9538.1000132.
11. Sarika Choudhari, SeemaBiday, "Artificial Neural Network for Skin Cancer Detection,"International Journal of Emerging Trends & Technology in Computer Science (IJETTCS).
12. Palak Mehta, Bhumika Shah, "Review on Techniques and Steps of Computer Aided Skin CancerDiagnosis," International Conference on Computational Modeling and Security (CMS 2016), Doi: 10.1016/j.procs.2016.05.238.
13. Dr. Abbas Hanon. Alasadi, BaidaaM.ALsafy, "Early Detection and Classification of Melanoma Skin Cancer,"I.J.Information Technology and Computer Science, DOI: 10.5815/ijitcs.2015.12.08, November 2015.
14. Nilkamal S. Ramteke and Shweta V. Jain,"Analysis of Skin Cancer Using Fuzzy and Wavelet Technique – Review & Proposed New Algorithm,"International Journal of Engineering Trends and Technology, Vol.4, Issue.6,ISSN: 2231-538, June 2013.
15. Indira, Jyotsnasupriya.P, "Detection & Analysis of Skin Cancer using Wavelet Techniques,"International Journal of Computer Science and Information Technologies, Vol. 2 (5), ISSN:0975-9646,1927-1932, 2011.

