MELANOMA DETECTION USING HYBRID CLASSIFIER

D.Selvaraj[1], D.Arul Kumar[2], D.Dhinakaran[3]

[1,2]Associate Professor, Department of ECE, Panimalar Engineering College, Tamilnadu, (India)

[3]Assistant Professor, Department of CSE, Peri Institute of Technology, Tamilnadu, (India)

ABSTRACT

In our proposed method an automatic melanoma classification system is developed. The input image is pre-processed, segmented and features are extracted. Based on the calculated features, the image is classified as cancerous or non-cancerous using KNN and SVM classifier. In the pre-processing stage noise is removed using Median filter and the image is enhanced using Adaptive Histogram Equalization. Later the image is segmented using Otsu thresholding, a novel method for acquiring accurate border of tumour of the selected images. Then features like entropy, mean, variance, skewness, kurtosis, correlation, energy, contrast, area and homogeneity are calculated from the segmented images. Later, SVM Classifier is trained with the extracted features. A total of 100 images have been used, out of which 45 are used for training and remaining images for testing. SVM classifier classifies the input image to be cancer affected or normal based on features extracted. If the image is cancer affected, then type of cancer is detected as malign tumor or benign tumor using KNN Classifier.

Keywords: ***KNN Classifier, Median Filter, Melanoma, Skin Cancer And SVM Classifier***

I. INTRODUCTION

Melanoma is named after the type of skin cell from which they arise. Basal cell cancer originates from the lowest layer of the epidermis, and is the most common but least dangerous skin cancer. Squamous cell cancer originates from the middle layer, and is less common but more likely to spread and, if untreated, become fatal. Melanoma, which originates in the pigment-producing cells (melanocytes), is the least common, but most aggressive, most likely to spread and become fatal if untreated. Most cases are caused by over-exposure to UV rays from the sun or sun beds. Treatment is generally via surgical removal. Melanoma has one of the higher survival rates among cancers. Most melanomas consist of various colors from shades of brown to black. A small amount of melanomas are pink, red or fleshy in color; these are called amelanotic melanomas which tend to be more aggressive. Warning signs of malignant melanoma include change in the size, shape, color or elevation of a mole. Other signs are the appearance of a new mole during adulthood or pain, itching, ulceration, redness around the site, or bleeding at the site.

Image segmentation plays a vital role in biomedical image processing to segment the medical input images in to meaningful regions. Even though several image segmentation and classification methods like otsu thresholding, region growing, region split and merge, active contour, Fuzzy C-means (FCM), Self Organizing Map (SOM), Support Vector Machine (SVM), Artificial Neural Networks are available, till image segmentation and classification remains a challenging problem.

In our proposed technique, initially the input melanoma image is preprocessed in order to remove the noise and make the image fit for the rest of the processes. Here we use median filter for removing the noise and CLAHE for image enhancement. Subsequently, the preprocessed image is segmented using the OTSU thresholding technique. After segmentation process, the feature such as mean, variance, Correlation, energy, Entropy, Skewness, Kurtosis, Contrast and Homogeneity are extracted from the regions and given to the SVM classifier for training. Later, the inage is classified as tumourous or normal with the help of trained SVM. Finally, the type of cancer is detected using KNN classifier.

The paper is organised as follows, A brief review of researches relevant to the melanoma detection and segmentation technique is presented in section 2. The proposed segmentation and classification technique is presented in section 3. The detailed experimental results and discussions are given in section 4. The conclusions are summed up in section 5.

II. REVIEW OF RELATED WORK

Many research works have been performed for the segmentation and classification of melanoma images. Some of the related works regarding the segmentation and classification of melanoma images are reviewed in the following section.

PaulWighton et al., [1] presented many sub problems in automated skin lesion diagnosis (ASLD) can be unified under a single generalization of assigning a label, from a predefined set, to each pixel in an image. The first model is based on independent pixel labeling using maximum a-posterior (MAP) estimation. The second model is based on conditional random fields (CRFs), where dependencies between pixels are defined using a graph structure. But the discrepancy between the objective and subjective performance of the CRF model implies that evaluation metric (pixel-wise sensitivity and specificity) is less than ideal.

Ho Tak Lau et. al.,[2] developed an automatic skin cancer classification system describing the relationship of skin cancer image across different type of neural network with different types of preprocessing. The collected images are fed into the system, and across different image processing procedures to enhance the image properties. Then the normal skin is removed from the skin affected area and the cancer cell is left in the image. The paper presented a study which can be concluded that there are some possible factors of low classification result. The image database is not feasible and too small; the variation between dermoscopy and digital image is large. The imaging processing methods are not unique and their variation is large.

SookpotharomSupot et. Al.,[3] proposed a new method of segmentation to locate the skin lesion is proposed by. The method consists of two stages; image pre-processing and image segmentation. As the first step of image analysis, pre-processing techniques are implemented to remove noise and undesired structures for the images using median filtering. In the second step, the fuzzy c means (FCM) thresholding technique is used to segment and localize the lesion. The border detection results are visually examined by an expert dermatologist and are found to be highly accurate. The proposed method gives more reliability and visually precise boundary tracing for a range of images. It is robust from noise and unwanted objects that gave the challenged problem in the most of the segmentation methods.

Thresholding is a simple but effective method to separate objects from the background. A commonly used method, the Otsu method improves the image segmentation effect obviously. It's simpler and easier to implement. Its half belongs to object and half belongs to background. Then a new weight to the Otsu method is applied by Kritika Sharma, Chandrashekhar Kamargaonkar, Monisha Sharma [4]. This weight can make sure that the result threshold value will always reside at the valley of the two peaks or at the bottom rim of a single peak. Moreover, it ensures that both the variance of the object and the variance of the background keep away from the variance of the whole image. The components of the histogram cover a broad range of the gray scale and it is reasonable to conclude that an image, whose pixels tend to occupy the entire range of possible gray levels and, in addition, tend to be distributed uniformly, will have an appearance of high contrast and will exhibit a large variety of gray tones.

A new segmentation method that combines the advantages of fuzzy C mean algorithm, thresholding and level set method is presented by AmmaraMasood, Adel Ali Al-Jumaily[5]. 3-class Fuzzy C mean thresholding is applied to initialize level set automatically and also for estimating controlling parameters for level set evolution. The proposed method showed reasonably good accuracy for segmentation of skin lesion images with an average true detection rate of 92.6% and quite reduced false positive and false negative error i.e. 4.66% and 7.34% respectively. Comparative analysis proved that it works well even in the presence of different artifacts present in skin images.

According to Howard Lee .Yi-Ping Phoebe Chen [6], a new approach to segment different types of skin cancers using fuzzy logic approach is developed. An optimum threshold segmentation algorithm based on type-2 fuzzy sets algorithms to delineate the cancerous area from the skin images is proposed. By using the 3D color constancy algorithm, the effect of color changes and shadows due to skin tone variation in the image can be significantly reduced in the preprocessing stage. The proposed method showed more tolerance at the border. Proposed skin cancer segmentation using the fuzzy algorithm approach demonstrated a good segmentation result while utilizing global features such as RGB color features. However, biological properties of different skin cancer types have not yet been utilized to fine tune the segmentation result.

One application of image processing is to reconstruct the original scene from the low quality images. Considering the idea of histogram, many histogram analyzing based methods have been studied recently. However, some methods require users to set some parameters or condition, and cannot get the optimal results automatically. To overcome those short come, Qieshi Zhang, Hiroshi Inaba andSei-ichiroKamata [7] presented an Adaptive Histogram Separation and Mapping (AHSM) method for Backlight image enhancement. First, the histogram by binary tree structure with the proposed Adaptive Histogram Separation Unit (AHSU) is separated. And then mapping the LowDynamic Range (LDR) histogram partition into High Dynamic Range (HDR). By doing this, the excessive or scarcity enhancement can be avoided. The experimental results show that the proposed method can gives better enhancement results, also compared with some histogram analyzing based methods and get better results.

The value of ‘k’ in KNN classifiers based on a n-sized training set is evaluated by C.Alippi, M.Fuhrman and M.Roveri [8]. The paper relates the generalization error of a k-nn classifier with k and n through a functional that allows identifying the optimal k. It depends on the application itself and cannot be solved in a closed form. However the functional can be numerically solved and an approximation of the optimal k derived. This method provides optimal results and it is seen that when the number of training samples increases the generalization error becomes more robust with respect to k.

III. PROPOSED METHOD

Our proposed method consists of 4 phases namely preprocessing, segmentation, feature extraction and classification. In preprocessing phase, the noise is removed using median filter and the image is enhanced using CLACHE. We have used OTSU thresholding technique for segmentation. After segmentation process, the features such as mean, variance, Correlation, energy, Entropy, Skewness, Kurtosis, Contrast and Homogeneity are extracted from the regions and given to the SVM classifier for training. Later, the inage is classified as tumourous or normal with the help of trained SVM. Finally, the type of cancer is detected using KNN classifier. The block diagram of the proposed technique is shown in Fig. 1.

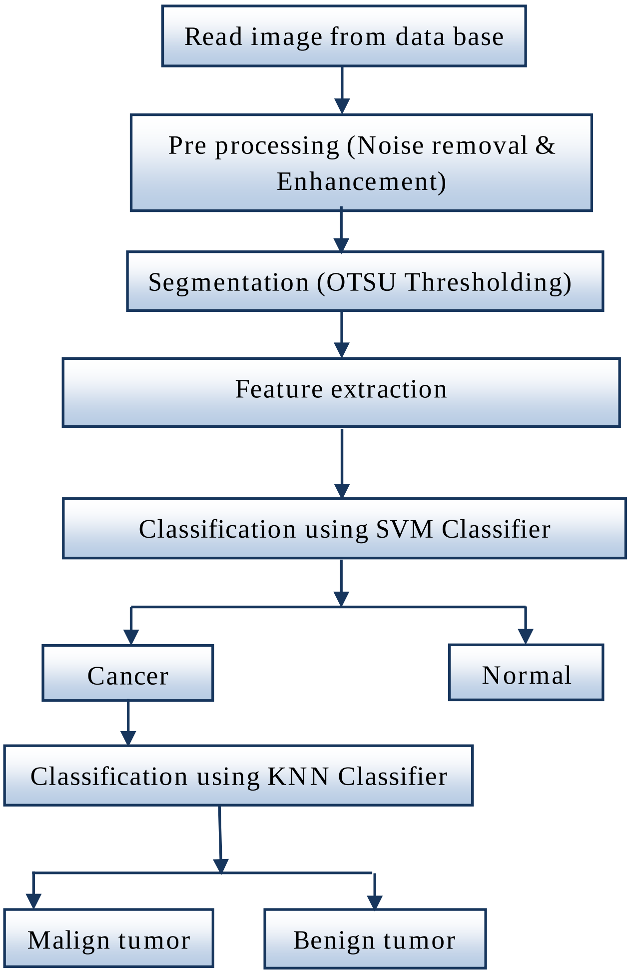


Fig. 1 Block diagram of proposed approach

3.1 Preprocessing

The first step in image processing is to read an image from data base. Select an image which is to be detected as cancerous or not from database available. Steps which are done prior to processing of an image are called pre processing. It includes processes like noise removal, image enhancement, resizing of image and equalization of image. These are done in order to make the image suitable than original image for specific applications. If the image selected is a color image, it is converted to gray scale image using 'rgb to gray conversion' command. Then the intensity variation of gray level image is shown in graph as in Figure 3. Its value varies from 0 to 255.

Resizing of image is done for accurate processing of image. Image can be resized to any size of our interest. There are many factors which cause noise in an image. It may be due to environmental conditions, quality of image sensing elements or interference in transmission channel. In order to avoid this, "filtration" is done. Here, median filter is used for removing noise. The median filter replaces each pixel in the input image by the median of gray levels in neighborhood. This filter is a type of smoothing spatial filter.

Median filter is useful for removing isolated lines or pixels while preserving spatial resolutions. It provides less blurring than linear filters. The filtering procedure consists of three steps:

- Arrange the pixel values in the window in increasing or decreasing order.
- If window size is odd, the middle value is the median. If the window size, K is even, the average of two values in the middle is the median.

Once noise is removed, we go for image enhancement. We use CLAHE (Contrast Limited Adaptive Histogram Enhancement) for this purpose. Histogram equalization also known as histogram linearization is a process of automatically determining a transformation function which produces an output image with uniform histogram. The transformation function modifies the pixels based the gray level content of an entire image. These techniques are used to enhance details over small areas in an image.

3.2 Segmentation

Otsu thresholding technique is used for segmenting the enhanced melanoma image. Otsu method is the commonly used thresholding technique. Otsu method is simple and effective to implement. Otsu's thresholding technique is based on a discriminant analysis which partitions the image into two classes C_1 and C_2 at gray levels 'k' such that $C_1 = \{ 0,1,2,3,\ldots, k\}$ and $C_2 = \{ k+1, k+2,\ldots, L-1\}$where, 'L' is the total number of gray levels of the image. Let 'n' be the total number of pixels in the given image and 'n_i' be the number of pixels at the i^{th} gray level.

3.2.1 Algorithm

Step1: Compute probabilities of each intensity level

Step2: Compute various thresholds T = 1, 2, ….. , max intensity and i) upgrade w_i and m ii) Compute

$s_B^2(k)$

Step 3: Select threshold value having max $s_B^2(k)$

3.3 Feature Extraction

Feature is a parameter of interest to describe an image. Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. There are many features we can extract from an image such as energy, entropy, mean, standard deviation, variance, color, texture etc.; we have considered extracting features like entropy, mean, contrast, correlation, energy, homogeneity, skewness, kurtosis and variance.

3.4 Classification

Two classifications are done here using two different efficient classifiers and hence the name "hybrid classifier". First classifier, SVM (Support Vector Machine) classifies the image to be cancerous or not depending on the features extracted. If the outcome of SVM classifier says that the input image is diseased, then second classifier KNN classifier is used. It classifies severity stage of skin cancer as malign or benign tumor.

3.4.1 SVM Classifier

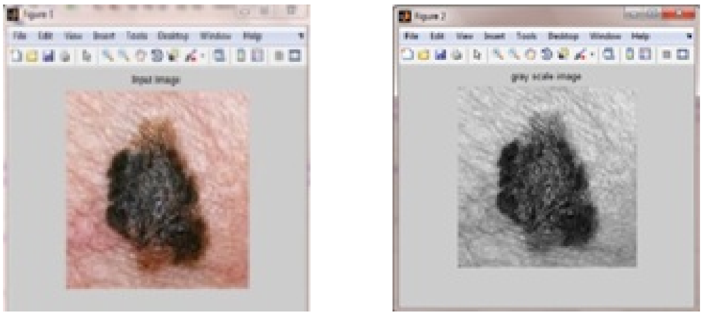
In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

3.4.2 k-NN Classifier

In pattern recognition, the *k*-Nearest Neighbors algorithm (*k*-NN) is a non parametric method used for classification. Here the input consists of the *k* closest training examples in the feature space. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor. *KNN* is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The *k*-NN algorithm is among the simplest of all machine learning algorithms. Based on the output of KNN classifier, the severity stage of skin cancer is detected.

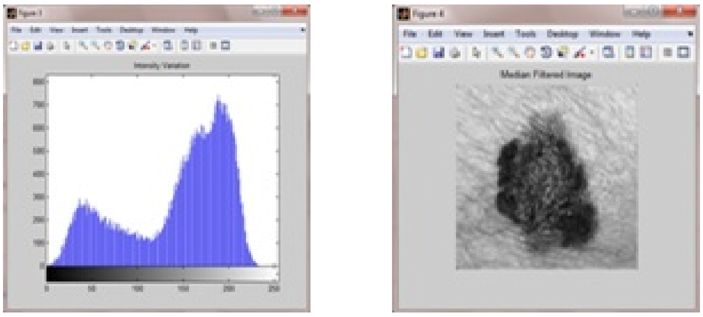
IV. EXPERIMENTAL RESULTS AND DISCUSSION

Here, totally 100 images have been used out of which 45 are taken for training and remaining have been used for testing. Calculated feature values of various input skin images are tabulated. Results show that 27 images are detected to be normal, 13 are benign tumor and 15 images are malign tumor. Comparison of various features for normal, benign and malign tumor images are shown in figure.



(A) (B)

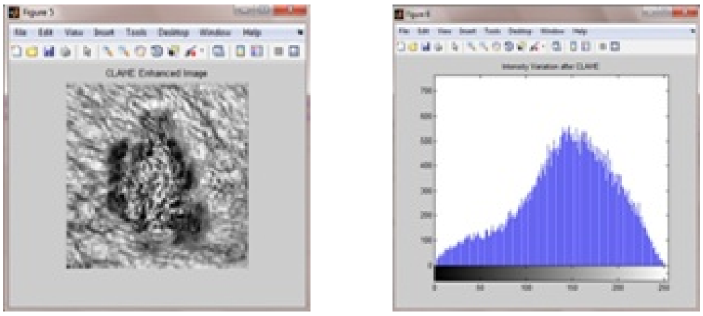
Fig. 2: (a) Input skin cancer image, (b) Converted gray scale image of input melanoma image



(A) (B)

Fig. 3: (a) Intensity variation before filtering, (b) Input melanoma image after median filtering

The input image is first preprocessed by converting color image to gray scale image, resizing it to required size and later filtering it using median filter as shown in Fig. 2 and Fig. 3. The enhanced image and segmented image is shown in figure 4 and 5. The intensity variation of the melanoma image before and after median filtering is shown in Fig. 3 (A) and Fig. 4(B).



(A) (B)

Fig. 4: (A) Enhanced melanoma image after filtering, (B) Intensity variation after median filtering

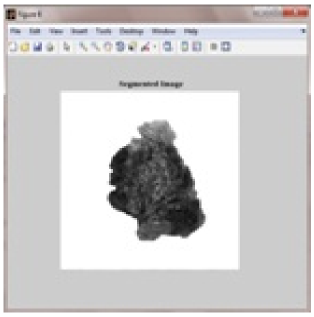


Fig. 5: Segmented melanoma image

Fig. 6: First order gradient features of melanoma image

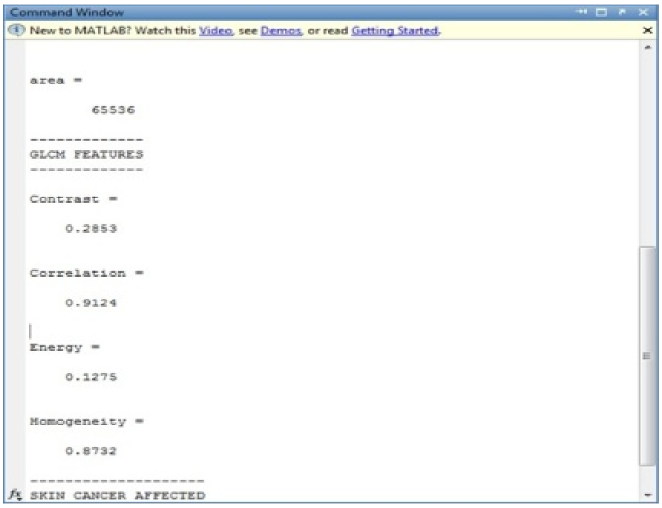


Fig. 7: GLCM features of melanoma image

The calculated first order gradient and GLCM feature values are shown in Fig. 6 and Fig. 7. Based on the calculated features values, the SVM classifer classified the image as a cancer image. so, the input image is fed to the KNN classifer for further classification. This classfier classified the image as Benign. The screensot of the final result is shown in Fig. 8.

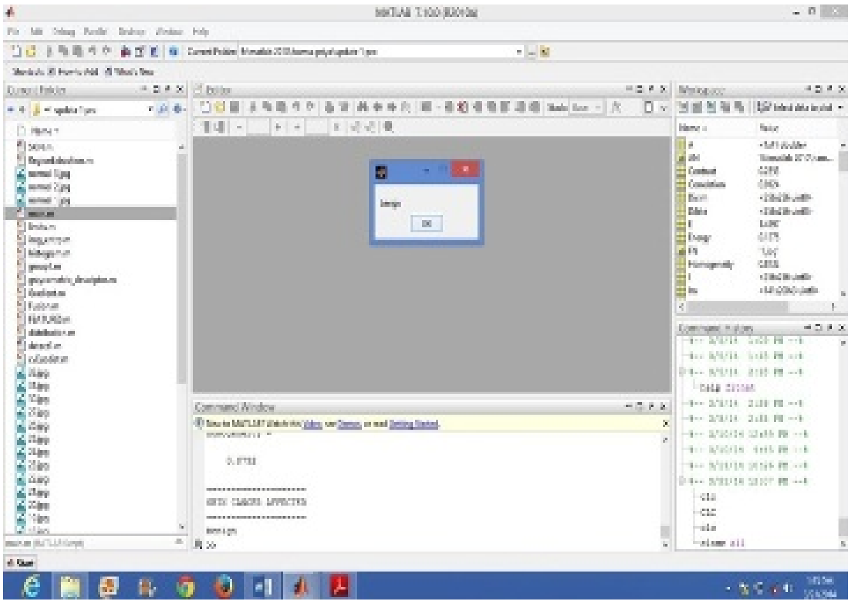
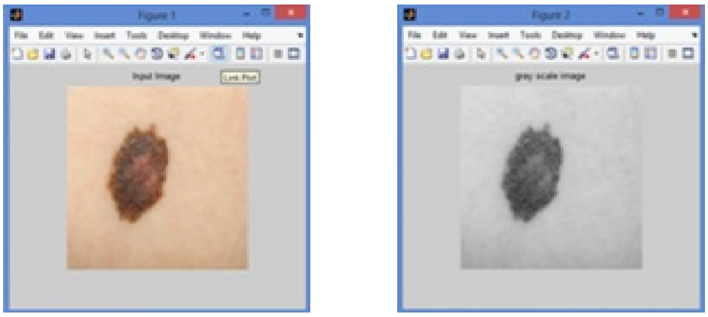
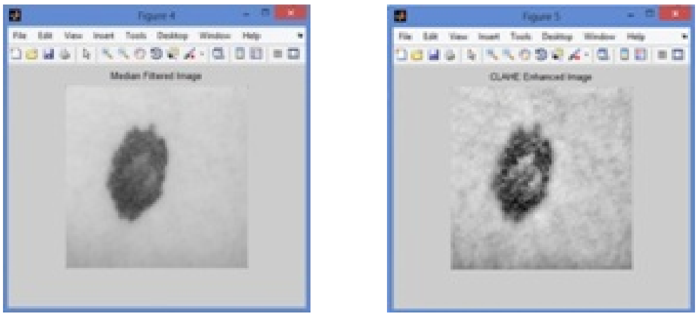


Fig. 8: The screenshot of the classified result



(A) **(B)**

Fig. 9: (a) input skin cancer image, (b) converted gray scale image of input melanoma image



(A) **(B)**

Fig. 10: (a) input image after median filtering, (b) input enhanced image

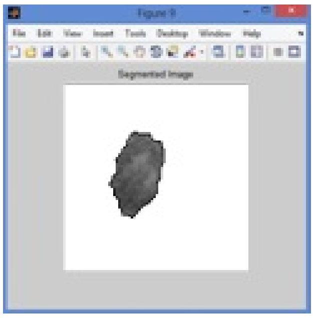


Fig. 11: Segmented image

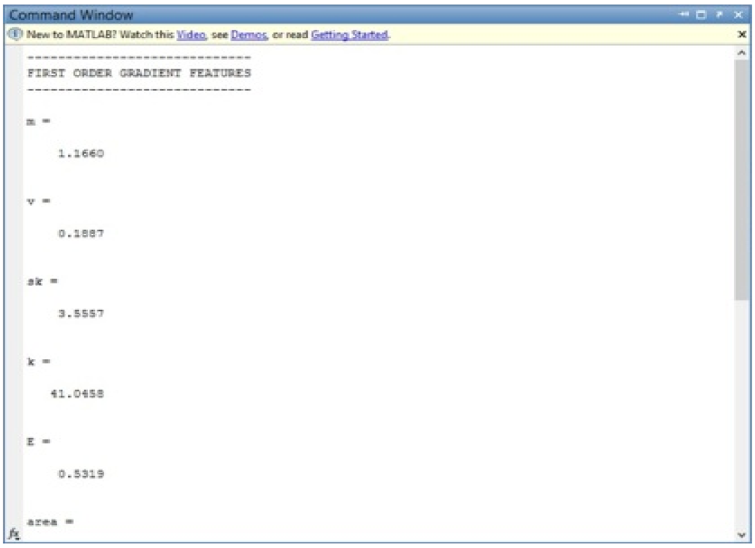


Fig. 12: First order gradient features of melanoma image

Fig. 13:GLCM features of melanoma image

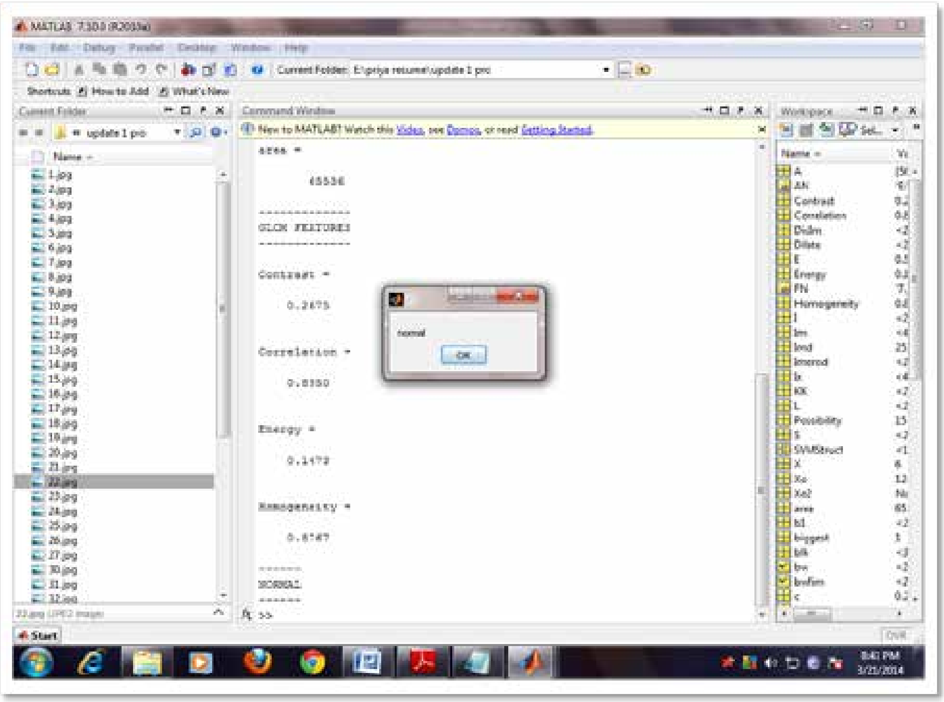


Fig. 14: The screenshot of the classified result

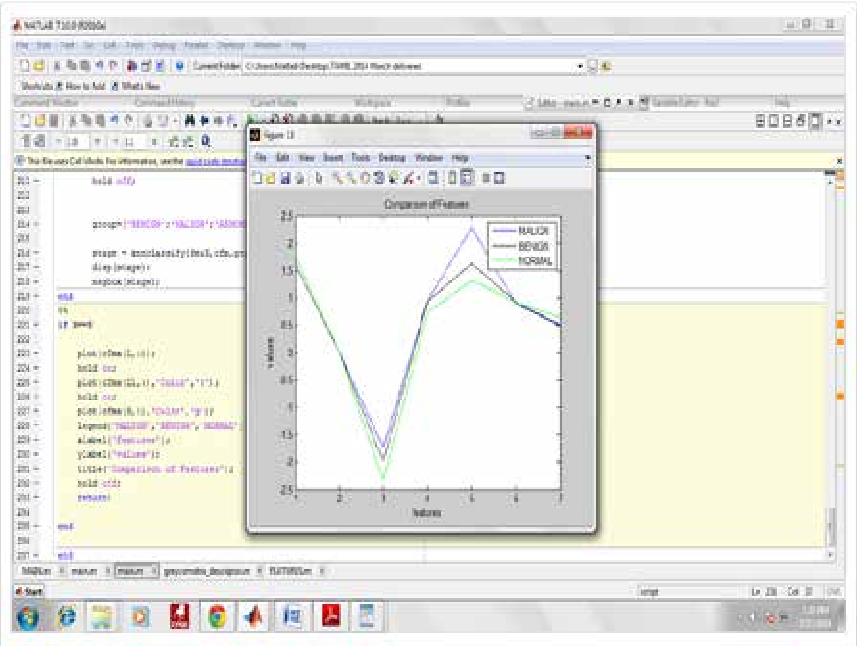
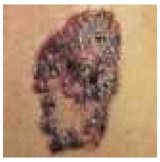
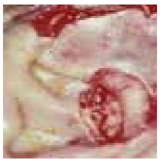
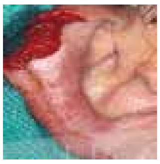
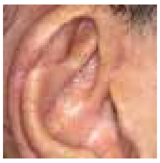


Fig. 15: Comparison between feature values of normal, benign and malignant images

Table 1 calculated feature values of different melanoma and non- melanoma images

IMAGES	M	V	E	ENE	S	CNT	COR	K	H	RESULT
	1.5976	0.5531	0.5197	0.7275	5.5098	0.2853	0.9124	109.9113	0.8732	Benign
	1.6624	0.6188	0.8557	0.5279	4.2582	0.0946	0.9617	43.5705	0.9531	Benign
	2.2224	1.1146	1.7452	1.1284	0.9545	0.2961	0.9092	12.1516	0.8731	Malign
	1.7791	0.7159	0.6141	0.5085	1.9156	0.1952	0.9064	21.3572	0.9080	Benign
	1.0002	0.0578	0.1259	0.2837	0.5599	0.0725	0.9705	9.6587	0.9649	Normal
	1.1660	0.1887	0.2319	0.1473	3.5557	0.2675	0.8950	41.0458	0.8767	Normal
	1.1115	0.1527	0.0147	0.2484	0.2662	0.0607	0.9755	5.2423	0.9697	Normal
	1.2462	0.2571	0.2150	0.2119	2.3509	0.2453	0.8364	36.7244	0.8874	Normal

IMAGES	M	V	E	ENE	S	CNT	COR	K	H	RESULT
	1.5702	0.5420	0.7486	0.7311	2.1349	0.1810	0.9578	8.2059	0.9133	Benign
	1.0975	0.2918	0.1341	0.2309	3.1586	0.3044	0.7607	42.1937	0.8640	Normal
	2.3432	1.1153	1.5673	1.1279	2.3867	0.1442	0.9754	59.1696	0.9312	Malign
	2.4261	1.2338	1.4644	1.2105	2.2530	0.1615	0.9153	15.2236	0.9206	Malign
	1.8322	0.6131	0.8422	0.6390	1.0508	0.1039	0.9513	11.2167	0.9486	Benign
	2.1439	1.2239	1.6632	1.1249	0.5898	0.2216	0.9479	16.4620	0.9029	Malign
	1.7584	0.5970	0.5873	0.5889	-0.9255	0.1358	0.9442	2.2801	0.9329	Benign
	2.4144	1.2315	1.3929	1.1289	0.4272	0.2231	0.9544	7.1833	0.8996	Malign

IMAGES	M	V	E	ENE	S	CNT	COR	K	H	RESULT
	2.4830	1.1100	1.2766	1.1676	1.3286	0.1265	0.9662	11.3186	0.9419	Malign
	1.7076	0.7374	0.6137	0.7135	5.3478	0.2609	0.9481	41.8162	0.8884	Benign
	2.5435	1.0784	1.7586	1.2365	3.2266	0.1411	0.9260	54.3418	0.9386	Malign
	1.2578	0.0528	0.1082	0.2464	1.5375	0.1472	0.9122	7.9463	0.9282	Normal
	1.7584	0.5970	0.5873	0.5889	-0.9255	0.1358	0.9442	2.2801	0.9329	Benign
	1.2853	0.0287	0.2852	0.1899	0.7658	0.1628	0.9272	5.3689	0.9204	Normal
	2.2224	1.1146	1.7452	1.1284	0.9545	0.2961	0.9092	12.1516	0.8731	Malign
	1.1085	0.1649	0.1931	0.0266	0.0502	0.0570	0.9471	4.0766	0.9741	Normal

IMAGES	M	V	E	ENE	S	CNT	COR	K	H	RESULT
	1.7923	0.6210	0.8803	0.7710	5.1837	0.3173	0.8687	77.1542	0.8627	Benign
	1.8667	0.8213	0.7190	0.6934	1.0656	0.2148	0.9247	0.1934	0.8990	Benign
	1.2233	0.0086	0.1577	0.2202	-0.5770	0.0985	0.9868	0.2202	0.9668	Normal

The feature values for different cancer and non-cancer images are calculated and tabulated in table 1. The same process as image 1 is repeated for the second image. Now, the second image is considered as normal image. The output of the image 2 at various stages are shown in figure 9 to 14. Similarly, various images can be processed.

V. CONCLUSION

This paper presents a method to detect type of cancer using hybrid classifier. Input image is subjected to pre processing stages like noise removal and enhancement. Here, totally 100 images have been used out of which 45 have been used as training images and 55 images are used as testing images. After enhancement, image is segmented using OTSU Thresholding. Features like mean, variance, entropy, energy, etc., are calculated. Based on the features extracted, image is classified to be cancer affected or not using SVM (Support Vector Machine) classifier. If the image is detected to be cancerous, then the severity of cancer is classified using KNN classifier. The severity of cancer is classified to be malignant or benign. From the results obtained, we can see our proposed hybrid method received a better quantity rate for all input images.

REFERENCES

1. Alippi.C, Fuhrman.M and Roveri.M, (2008), ‘k-NN classifiers: Investigating the k=k (n) relationship’hoi, Mihwa. “Contesting Imaginaires in Death Rituals during the Northern Song Dynasty.” PhD thesis., University of Chicago, 2008.
2. AmmaraMasood and Adel Ali Al-Jumaily, (2013), ‘Fuzzy C Mean Thresholding based Level Set for Automated Segmentation of Skin Lesions’.
3. Ho Tak Lau and Adel Al-Jumaily, (2009) ‘Automatically Early Detection of Skin Cancer: Study Based on Neural Network Classification’.
4. Howard Lee · Yi-Ping Phoebe Che, (2013), ‘Skin cancer extraction with optimum fuzzy thresholding technique’.

5. Kritika Sharma,ChandrashekharKamargaonkar and Monisha Sharma, (2012), 'An Improved Image Segmentation Algorithm Based on Otsu Method'.
6. PaulWighton, Tim K. Lee, GregMori, Harvey Lui, David I. McLean and Stella Atkins, (2011) 'Conditional Random Fields and Supervised Learning in Automated Skin Lesion Diagnosis'.
7. Qieshi Zhang, Hiroshi Inaba and Sei-ichiroKamata, (2010), 'Adaptive Histogram Analysis for Image Enhancement'.
8. SookpotharomSupot, (2009), 'Border Detection of Skin Lesion Images Based on Fuzzy C-Means Thresholding'