International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

Skin Cancer Detection of Dermoscopic Images using K- Means Algorithm and PCA Algorithm

Jagdish Prasad Bandhaw, Ekata Patil,

M. Tech Scholar, Digital Electronics, Department of ETC, CCEM Raipur,CSVTU, Bhilai, C. G., India

Assistant Professor, Department of ETC, CCEM Raipur, CSVTU, Bhilai, C. G., India

ABSTRACT: Skin cancer is the most common type of cancer in the world and the incidents of skin cancer have been rising over the past decade. Even with a dermoscopic imaging system, which magnifies the lesion region, detecting and classifying skin lesions by visual examination is laborious due to the complex structures of the lesions. This necessitates the need for an automated skin lesion diagnosis system to enhance the diagnostic capability of dermatologists. In this study, the authors propose an automatic skin lesion segmentation method which can be used as a preliminary step for lesion classification. The proposed method comprises two major steps, namely preprocessing and segmentation. In the preprocessing step, noise such as illumination, hair and rulers are removed using filtering techniques and in the segmentation phase, skin lesions are segmented using the Otsu's segmentation algorithm. The k-means clustering algorithm is then used along with the colour features learnt from the training images to improve the boundaries of the segments.

I. INTRODUCTION

One of the biggest threats to human beings is cancer and it is the second leading cause of death. Of all the cancer types, skin cancer is the most common form which is rapidly increasing in the recent years. Skin cancer is the uncontrolled growth of abnormal skin cells. These cancer cells may spread from the skin into other tissues and organs if left unchecked. Dermoscopy is a non-invasive imaging method used to detect melanoma and other skin lesions. With the help of dermoscopy, subsurface structures of skin lesions can be analyzed and different types of lesions can be distinguished with better visualization. The most common types of skin cancer [1–3]are

- *Basal cell carcinoma (BCC):* BCC is the most common form of skin cancer. Generally, it begins to grow from the bottom of the outer skin layer and develops in the area that is exposed to sunlight for long term. BCC is slow in growth and hence can be treated very easily. BCC appears as small, smooth, shiny, pale or waxy lump which may be red or brown in colour with rough, dry or scalypatches.
- *Squamous cell carcinoma (SCC):* SCC is the second most common type of cancer. The development of SCC starts in the outer skin layer and infiltrates the underlying tissues if left unchecked. This cancer spreads to other parts of the body and at its early stage, SCC can be easily treated. SCC also appears as small shinny lumps with red or brown incolour.
- *Malignant melanoma (MM):* Melanoma is more dangerous but less common type of skin cancer. It occurs in the melanocytes (cells that produce pigment) and it is the leading cause of skin cancer mortality. Melanoma is asymmetric in shape with an irregular border and uneven incolour.

Among all the cancer types, melanoma is the most dangerous form of skin cancer which leads to the highest degree of the death rate. Melanoma has a very high tendency to grow and even after removing the lesions may sometimes grow back. It may intrude and damage nearby organs and tissues and may invade other parts of the body where it is very hard totreat.

Generally, skin cancer is screened by clinicians through visual examination. During the screening of cancer, the clinicians look for moles and other spots that are different in colour from the normal skin. Visual screening by dermatologists for skin cancer does not guarantee 100% detection and sometimes it may lead to potential harm. Potential harm includes unnecessary procedures such as skin biopsy for lesions that do not turn out to be cancer or sometimes the lesions might have been missed and not have gone for biopsy, resulting in death. From the literature, it has been shown that even for experienced dermatologist the diagnostic accuracy rate is 60% [4–6] with a visual

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

examination. As a result, there is a clear requirement for automatic detection systems for skin cancer which should be highly accurate and help in the overall analysis of a skin lesion.

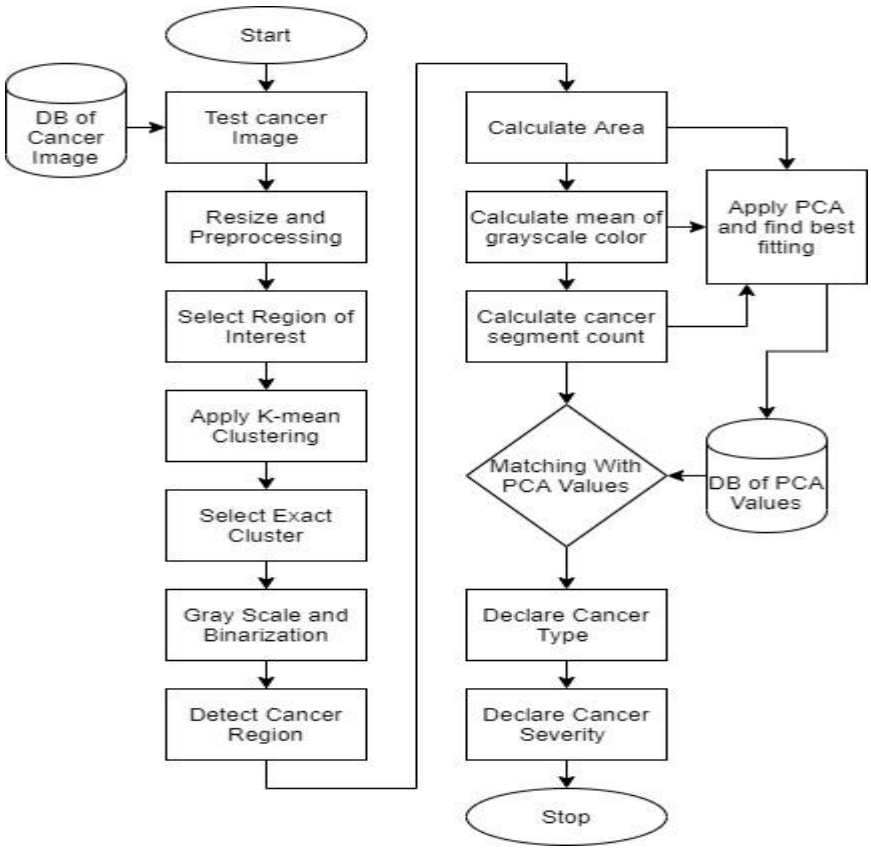
The basic approach in developing an automatic diagnostic tool for skin lesion analysis includes segmentation, feature extraction and classification. Among these three tasks, segmentation is important since the accuracy of segmentation directly affects successive tasks. However, segmentation of a lesion is very hard because lesion shape, size, boundaries and colour vary greatly with the difference in the type of skin and texture. Apart from these variations, the lesion boundaries are also irregular and the transition from the lesion to skin is very smooth making the discrimination of normal skin and lesion very complex. Moreover, dermoscopic images include various artifacts such as air bubbles, hair, ink markings and so on, making the segmentation process challenging.

Recently, many methods for automatic segmentation and classification of skin lesions from dermoscopic images have been proposed to address these issues. The methods used for segmentation of the lesion can be widely classified as region- based, edge-based and threshold-based segmentation algorithms [7,8].

A saliency-based lesion segmentation framework for segmenting lesion regions from dermoscopic images was proposed in [9]. The proposed approach separates the background by analyzing the image region, boundaries and colour characteristics of dermoscopic images. Lesions are then localized by constructing a Bayesian framework using reconstruction error estimated by comparingthesimilaritiesofpixelsinthelesionregionandbackground. Segmentation of lesions in dermoscopic images using a fixed-grid wavelet network (FGWN) was proposed by Sadri *et al.* [10]. The proposed FGWN takes R, G, B values of the dermoscopic image as input and use orthogonal least squares algorithm to estimate network weights and optimise the network structure. The output of the network describes the boundary of the skin lesion region. A new data model for the graph-cut method was proposed in [11] which takes the colour feature, texture and shape information for segmentation. The method automatically derives a seed pixel to train the classifier from a coarse initial segmentation of the pigmented skin lesions obtained by clustering and morphological operators.

II. PROPOSED METHODOLOGY

We propose a semi-supervised learning technique to automatically segment skin lesion in a given image. The method includes two major steps namely preprocessing and segmentation. The preprocessing step is applied to filter out the artifacts present in an image and this filtered image is further used for segmenting the skin lesion. The segmentation is done with Otsu's and k-means clustering algorithms. The workflow of the proposed method is depicted in Fig. 1.



| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

Fig1. The workflow of the proposed method

[1] We have arranged various kind of cancer image from multiple sources like hospital, datacenter, case studies and reference patient. These images have taken in appropriate image capture condition and ambient light. We have maintain basic of image parameters, dpi and frame size. We are passing one by one image for training and testing.

[2] J = imresize(I,scale) returns image J that is scale times the size of I. The input image I can be a grayscale, RGB, binary, or categorical image.We optionally can resize images using a GPU (requires Parallel Computing Toolbox™).

[3] A region of interest (ROI) is a portion of an image that you want to filter or operate on in some way. The toolbox supports a set of ROI objects that we can use to create ROIs of many shapes, such circles, ellipses, polygons, rectangles, and hand-drawn shapes. After creation, you can use ROI object properties to customize their appearance and functioning. In addition, the ROI objects support object functions and events that we can use to implement interactive behavior. For example, using events, your application can execute custom code whenever the ROI changes position. As a convenience, the toolbox includes a parallel set of convenience functions for ROI creation. For example, to create a rectangular ROI, we can use images.roi.Rectangle or the corresponding convenience function drawrectangle.

[4] K-means clustering is a partitioning method. The function k-means partitions data into k mutually exclusive clusters and returns the index of the cluster to which it assigns each observation. K-means treats each observation in our data as an object that has a location in space. The function finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. We can choose a distance metric to use with k-means based on attributes of our data. Like many clustering methods, k-means clustering requires you to specify the number of clusters k before clustering.

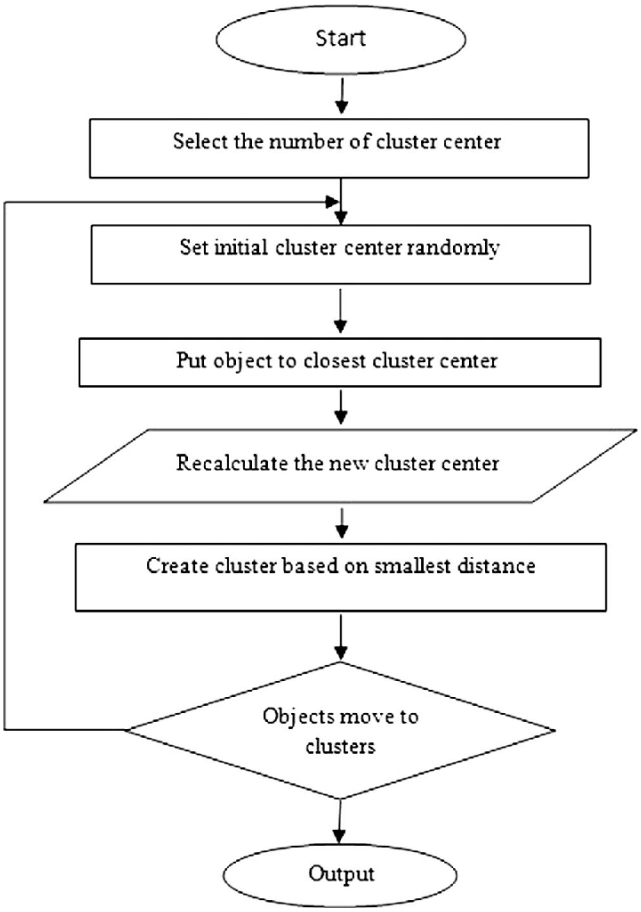


Fig2. The workflow of the K-mean Clustering Algorithm.

Unlike hierarchical clustering, k-means clustering operates on actual observations rather than the dissimilarity between every pair of observations in the data. Also, k-means clustering creates a single level of clusters, rather than a multilevel hierarchy of clusters. Therefore, k-means clustering is often more suitable than hierarchical clustering for large amounts of data.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

Each cluster in a k-means partition consists of member objects and a centroid (or center). In each cluster, k-means minimizes the sum of the distances between the centroid and all member objects of the cluster. K-means computes centroid clusters differently for the supported distance metrics.

We can control the details of the minimization using name-value pair arguments available to k-means; for example, we can specify the initial values of the cluster centroids and the maximum number of iterations for the algorithm. By default, k-means uses the k-means++ algorithm to initialize cluster centroids, and the squared Euclidean distance metric to determine distances.

K-means clustering has followingsteps:

a) Compare k-means clustering solutions for different values of k to determine an optimal number of clusters for our data.

b) Evaluate clustering solutions by examining silhouette plots and silhouette values. We can also use the evalclusters function to evaluate clustering solutions based on criteria such as gap values, silhouette values, Davies-Bouldin index values, and Calinski-Harabasz index values.

c) Replicate clustering from different randomly selected centroids and return the solution with the lowest total sum of distances among all the replicates.

[5] I = rgb2gray(RGB) converts the truecolor image RGB to the grayscale image I. The rgb2gray function converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance.

[6] BW = im2bw(I,level) converts the grayscale image I to binary image BW, by replacing all pixels in the input image with luminance greater than level with the value 1 (white) and replacing all other pixels with the value 0 (black).

This range is relative to the signal levels possible for the image's class. Therefore, a level value of 0.5 corresponds to an intensity value halfway between the minimum and maximum value of the class.

[7] Edge detection is an image processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation and data extraction in areas such as image processing, computer vision, and machine vision.

[8] bwarea estimates the area of all of the on pixels in an image by summing the areas of each pixel in the image. The area of an individual pixel is determined by looking at its 2-by-2 neighborhood. There are six different patterns, each representing a different area:

a. Patterns with zero on pixels (area = 0)

b. Patterns with one on pixel (area = 1/4)

c. Patterns with two adjacent on pixels (area = 1/2)

d. Patterns with two diagonal on pixels (area = 3/4)

e. Patterns with three on pixels (area = 7/8)

f. Patterns with all four on pixels (area = 1)

g. Each pixel is part of four different 2-by-2 neighborhoods. This means, for example, that a single on pixel surrounded by off pixels has a total area of 1.

[9] M = mean(A) returns the mean of the elements of A along the first array dimension whose size does not equal 1.If A is a vector, then mean(A) returns the mean of the elements.If A is a matrix, then mean(A) returns a row vector containing the mean of each column.If A is a multidimensional array, then mean(A) operates along the first array dimension whose size does not equal 1, treating the elements as vectors. This dimension becomes 1 while the sizes of all other dimensions remain the same.

[10] In bwlabel, Run-length encodes the input image.Scan the runs, assigning preliminary labels and recording label equivalences in a local equivalence table.Resolve the equivalence classes.Relabel the runs based on the resolved equivalence classes.

[11] Principal Component Analysis Algorithm Steps

1. Find the mean vector.

```
m = sum(parameters)/12;
```

2. Assemble all the data samples in a mean adjusted matrix.

```
z = zeros(12, 3);
for i = 1:3
z(:,i) = m(i);
end
```

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

c = parameters - z;

3. Create the covariance matrix.

```
cm = zeros(3, 3);
for i = 1:3
for j = 1:3
cm(i,j) = sum(c(:,i).*c(:,j))/11;
end
end
```

4. Compute the Eigen vectors and Eigen values.

```
L = eigs(cm);
[V,D] = eigs(cm);
```

5. Compute the basis vectors.

```
scores = c*V;
```

6. Represent each sample as a linear combination of basis vectors.

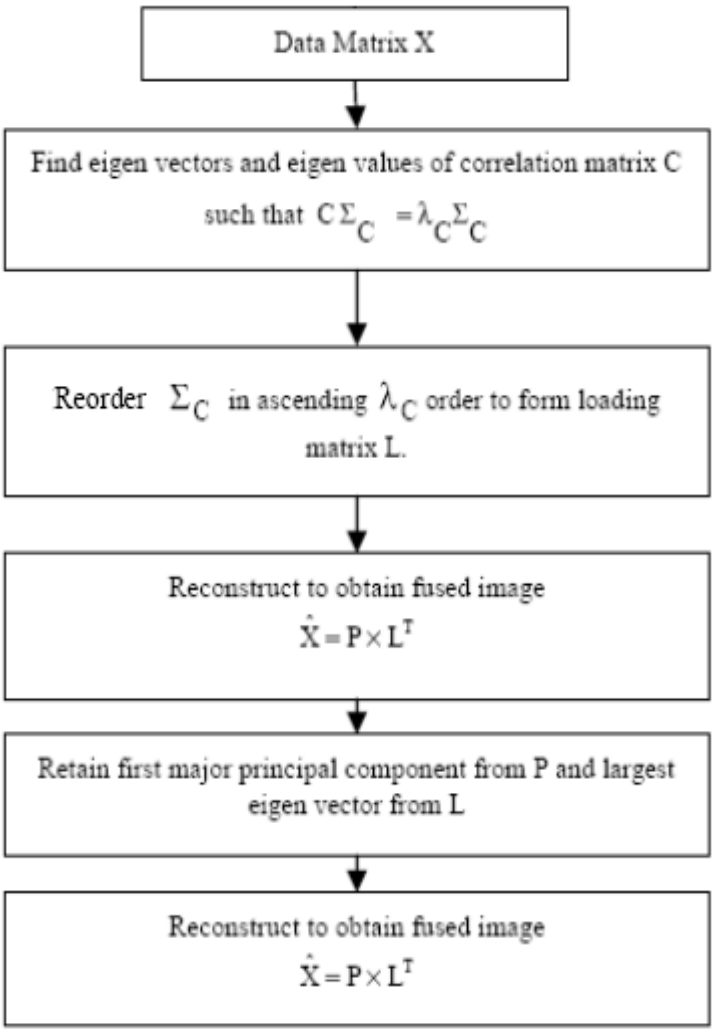


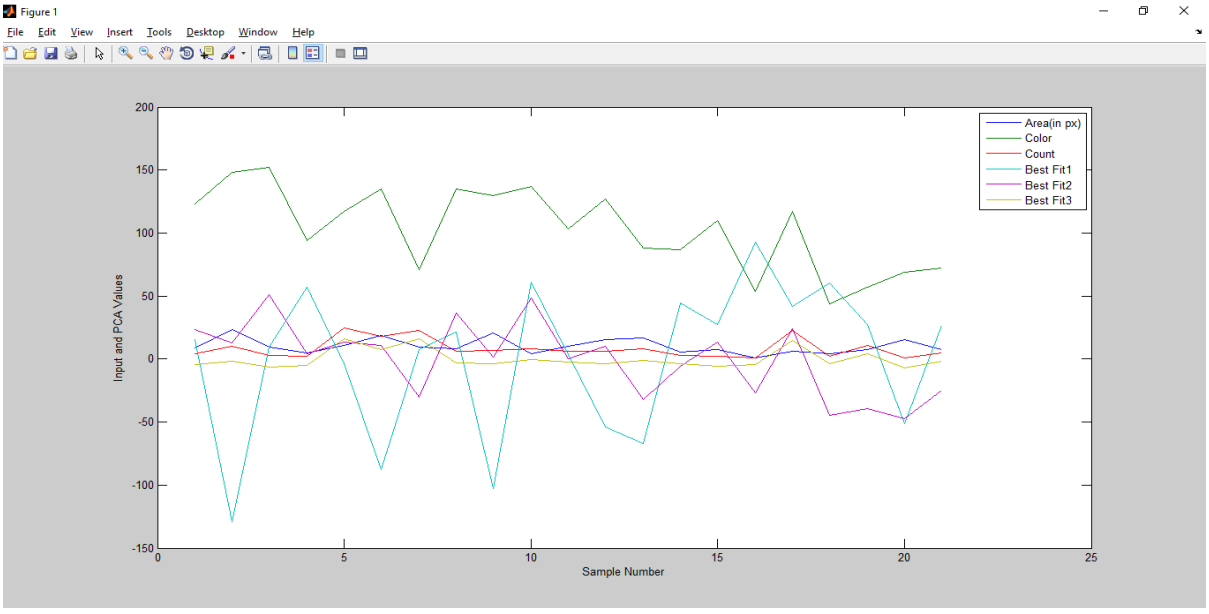
Fig3. The workflow of the PCA Algorithm.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

Input Parameters			PCA Fitting		
Area(in px)	Color	Count	Best Fit1	Best Fit2	Best Fit3
875	123	4	151.7215505	23.26633636	-4.564151595
2319	148	10	-1292.468333	12.99363914	-1.908492778
934	152	3	92.03226387	50.77025665	-6.493771634
457	94	2	570.309746	4.484678424	-4.97873851
1067	117	25	-40.12639237	13.15474619	16.24744544
1901	135	18	-874.2968518	10.48382242	7.219117772
950	71	23	77.96890247	-29.99927678	15.76615507
813	135	6	213.4040995	36.8355464	-2.792234596
2052	130	7	-1025.101103	1.471540044	-3.90996357
416	137	8	610.2299773	48.63017692	-0.125894309
990	103	6	37.24073785	0.515744062	-2.207439742
1569	127	6	-542.1713434	10.29455738	-3.943802688
1700	88	8	-672.1823645	-31.83522118	-1.077420122
583	87	3	444.5166569	-5.573202451	-4.010362485
751	110	2	276.0071832	13.26088138	-5.968565938
102	54	1	926.1836707	-26.80650284	-4.196637681
609	117	23	417.7400383	24.33403745	15.08334589
425	44	2	603.5236129	-44.69518951	-3.502380226
757	57	11	271.2833028	-39.5943546	4.5199508
1538	69	1	-509.7488166	-47.04576371	-7.240366483
764	72	5	263.9334628	-24.94645174	-1.915792615

Table1: Input parameters for all DB images and PCA best fitting values.



| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

Fig4. Input and PCA value vs Sample plot

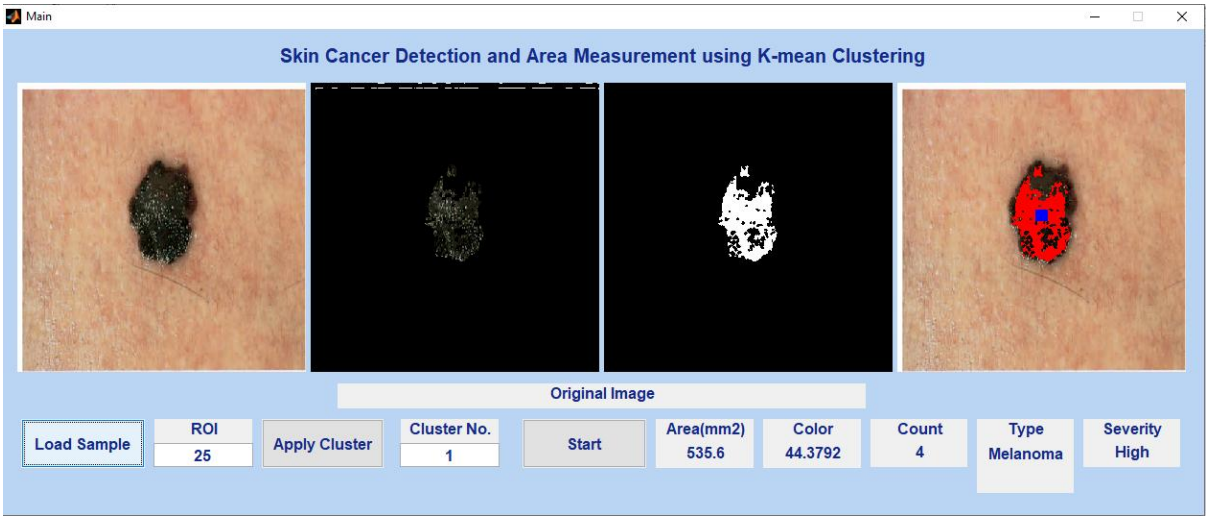


Fig5. GUI of the proposed method

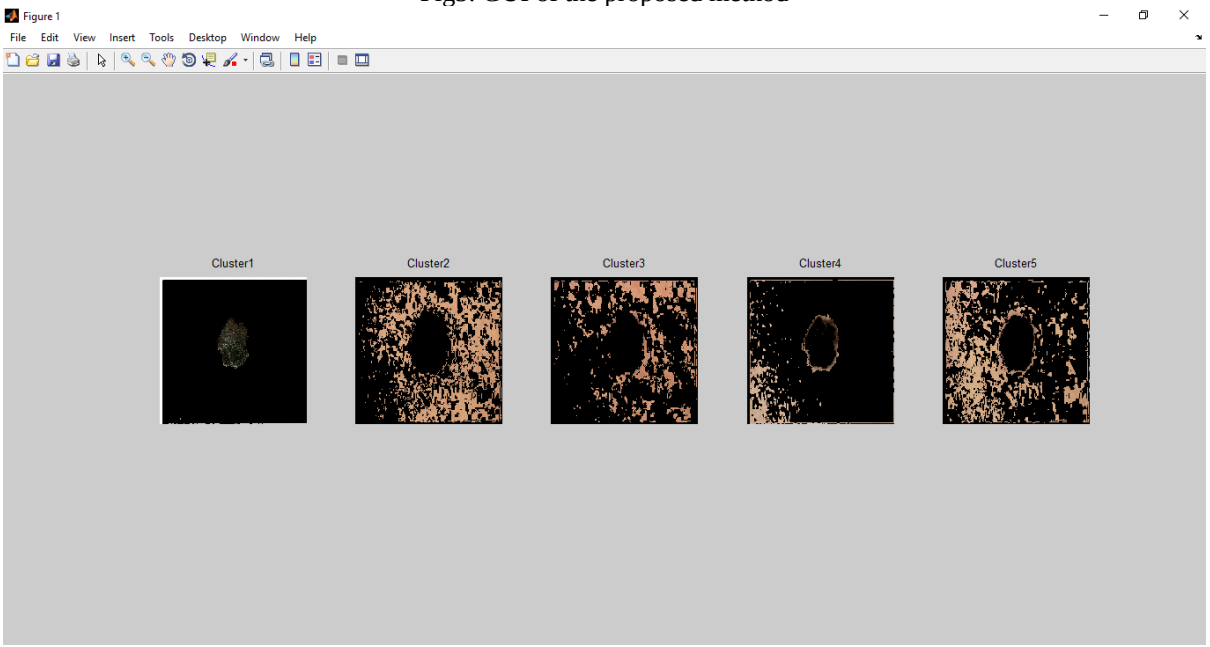


Fig6. Different types of cluster for test image

III. CONCLUSION AND FUTURE WORK

In this paper, we have presented an automated method to segment skin lesion regions using a semi-supervised learning technique. This system helps dermatologists to automatically locate lesion regions in dermoscopic images for further diagnosis. Our method comprises two major steps, namely preprocessing and segmentation. During preprocessing, artifacts such as air bubbles, hair and ink marks are filtered out. To correct the cancer image, we have appliedregion of interest.

From the filtered image, lesion regions are segmented using Ostu's and k-mean clustering algorithms. The Ostu'salgorithm segments the foreground image (lesion region) and k-means clustering and PCA algorithms are applied to obtain the lesion region with improved boundaries. To ascertain the performance of the proposed algorithm we have tested our method with different type of cancer image and got accuracy till almost 95%. Deep learning techniques may be tried in the future to further improve accuracy and Dice coefficientvalues.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 7, July 2020||

REFERENCES

1. Chuang, T.-Y., Popescu, A., Su, W.D., *et al.*: 'Basal cell carcinoma: a population-based incidence study in Rochester, Minnesota', *J. Am.Acad.*
2. *Dermatol.*, 1990, **22**, (3), pp. 413–417
3. Sigurdsson, S., Philipsen, P.A., Hansen, L.K., *et al.*: 'Detection of skin cancer by classification of Raman spectra', *IEEE Trans. Biomed. Eng.*, 2004, **51**, (10), pp. 1784–1793
4. Camacho, F., Gildea, J., Goldenberg, G., *et al.*: '*Managing skin cancer*' (Springer, Berlin, Heidelberg, 2010), pp.17–35
5. Fleming, M.G., Steger, C., Zhang, J., *et al.*: 'Techniques for a structural analysis of dermatoscopic imagery', *Comput. Med. Imaging Graph.*, 1998, **22**, (5), pp. 375–389
6. Grin, C.M., Kopf, A.W., Welkovich, B., *et al.*: 'Accuracy in the clinical diagnosis of malignant melanoma', *Arch. Dermatol.*, 1990, **126**, (6), pp. 763– 766
7. Morton, C.A., Mackie, R.M.: 'Clinical accuracy of the diagnosis of cutaneous malignant melanoma', *Br. J. Dermatol.*, 1998, **138**, (2), pp.283–287
8. Celebi, M., Iyatomi, H., Schaefer, G., *et al.*: 'Lesion border detection in dermoscopy images', *Comput. Med. Imaging Graph.*, 2009, **33**, (2), pp. 148– 153
9. Silveira, M., Nascimento, J.C., Marques, J.S., *et al.*: 'Comparison of segmentation methods for melanoma diagnosis in dermoscopy images',*IEEE*
10. *J. Sel. Top. Signal Process.*, 2009, **3**, (1), pp. 35–45
11. Ahn, E., Kim, J., Bi, L., *et al.*: 'Saliency-based lesion segmentation via background detection in dermoscopic images', *IEEE J. Biomed. Health Inf.*, 2017, **21**, (6), pp. 1685–1693
12. Sadri, A.R., Zekri, M., Sadri, S., *et al.*: 'Segmentation of dermoscopy images using wavelet networks', *IEEE Trans. Biomed. Eng.*, 2013, **60**, (4), pp. 1134– 1141
13. Kehichian, R., Gong, H., Revenu, M., *et al.*: 'New data model for graph-cut segmentation: application to automatic melanoma delineation'. 2014 IEEE Int. Conf. on Image Processing (ICIP), Paris, France, October 2014, pp. 892– 896
14. Abbas, Q., Fondon, I., Rashid, M.: 'Unsupervised skin lesions border detection via two-dimensional image analysis', *Comput. Methods Programs Biomed.*, 2011, **104**, (3), pp.e1–e15