International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Melanoma Detection through K-Means Segmentation and Feature Extraction

Dr.R. Pon Periyasamy[1], V.Gayathiri[2]

[1] *Associate Professor, Department of Computer Science, Nehru Memorial College, Tamilnadu, India*
[2] *Assistant Professor, Department of Computer Application, GTN Arts College, Tamilnadu, India*

Abstract- *Skin cancer is the most common of all cancer and each year million cases of skin cancer are gave attention to. Damaging melanoma is responsible for 75% of the deaths caused needing payment to skin cancer every year. However, melanoma discovery can be possible through point extraction and order, which can lower the danger, if the melanoma is sensed at an early stage. Clustering algorithm is one of the most useful instruments used to note as being different features that can send in (writing) to melanoma. This make observations work uses the k-means clustering algorithm for putting into effect of color breaking down into parts. The offered careful way having among its parts of 2 Major sides (of a question) (1) preprocessing and (2) breaking down into parts using semi-supervised learning algorithm. In the preprocessing phase noise are taken away using coming through slowly way of doing and in the breaking down into parts phase skin wound are segmented based on clustering way of doing. K-means clustering algorithm is used to mass, group the preprocessed images and skin wound are made clean from these clusters based on the color point.*

Keywords: K- Means Clustering, Feature Extraction, ABCDE rule, Melanoma detection.

1.INTRODUCTION

Skin cancer is among the most commonly worked out cancers; of which damaging melanoma is by far its most war-like form. Although melanoma is less common than other types of skin cancer, its incidence has been quickly increasing over the last years [1]. Happily, when melanoma is worked out in an early stage, it can easily be gave attention to through a simple taking out with a cut of the wound. As an outcome of that, several diagnosis techniques have been had a look for to get well the early discovery of melanoma. Melanoma and Non-Melanoma are chief groups of skin cancers. Damaging melanoma is of several sub-types. Basal unit carcinoma and squamous unit carcinomas are 2 main types of non-melanoma skin cancers. Each sort of skin cancer is different from the other skin cancers in certain qualities.

Generally skin cancer is screened by clinicians through seeing observation. During the going-over of cancer, the clinicians look for moles and other spots that are different in color from the normal skin. The rule for cancer detection is called as ABCD rule [2] which is given by

- A : asymmetry (one half of the mole does not match the other half)
- B : border irregularity (edges of the mole are ragged, notched, or blurred)
- C : color (pigmentation of the mole is not uniform, with varying degrees of tan,brown, or black)
- D : diameter of more than inch (about the size of a pencil eraser)
- E : evolving (the mole is changing over time)

Seeing going-over of clinicians for skin cancer does not be responsible for 100% discovery and sometimes it may lead to possible & unused quality damage. possible & unused quality cause damage includes unnecessary procedures such as skin biopsy or taking out with a cut for wound that do not turn out to be cancer or sometimes the wound might have missed and not have gone for biopsy, coming out in death. As an outcome there is a clear thing needed for automatic discovery system for skin cancer which should be highly good at producing an effect and accurate. The offered careful way form of 2 Major steps (1) preprocessing which takes away the things like hair, ink marking and lighting-on bad, wrong points. (2) Breaking down into parts of wound using k-means clustering.

2. EXISTING METHOD

The having existence careful way is based on the and field, range growing. The careful way was not taken into account the spatial qualities. Normally spatial qualities are important for the serious, violent diseased growth (in body) discovery. In the based breaking down into parts the image is taken into account as having only 2 values either black or white. But the bit map image has in it 0 to 255 gray scale values. So sometimes it has nothing to do with the diseased growth (in body) units in addition. In Case of the field, range growing based breaking down into parts it needs more user effect on one another for the selection of the seed [3]. Seed is nothing but the inside middle of the diseased growth (in body) units; it may cause in number in being made up of parts of the same sort hard question. And also it will not give the pleasing outcome for all the images.

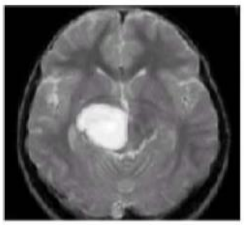

Fig-1: Input Image

Fig-2: Output Image

Fig.1 is the input image for thresholding. From the MR image itself we can see the tumor area but it is not enough for further treatment. For that it is given to the thresholding process. Fig2 is the output image for the thresholding. It consists of only two gray values .That is white as 1 and black as 0. The background value is assigned to binary value 0 and object gets the value 1. So we cannot extract the tumor from the image. This is the main drawback of the existing system. Due to that we go for the proposed method for tumor segmentation.

3. PROPOSED METHODOLOGY

Offered framework take image as input, process it and then segmented the processed image using k-mean clustering algorithm. Then color and feeling of a material features get got from the parts.

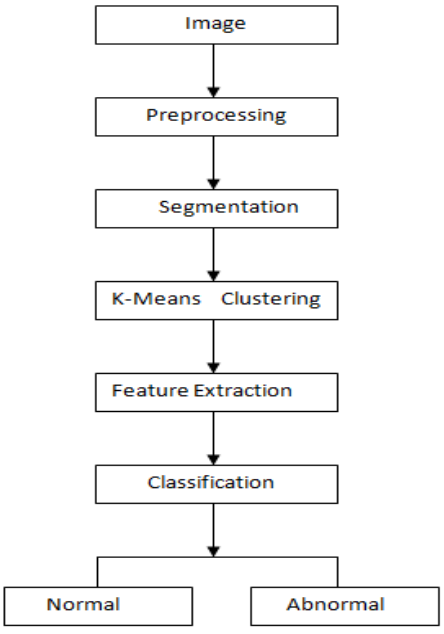

Fig-3: Block Diagram

A. Preprocessing

1) Image Scaling:

The size of the images in the training set varies from 1022_ 767 to 6708 _ 4439. To reduce the computational complexity, the images in the training set were resized down to 25% using bi-linear interpolation method [4], if the height or width of the original image is above 1500. If the width or height of the image is less than 1500 the images size is retained as such.

2) Hair removal: Some of the images in the training put have within hair which makes the breaking down into parts process hard. The existence of hair also makes the point extraction process for further image observations requesting. So we have removed the hairs using Frangi vesselness algorithm [5].

B. Segmentation

Breaking down into parts is the process of isolating the diseased part from normal skin based on the being made up of parts of the same sort of the bit of picture. Being made up of parts of the same sort of the bit of picture may be strong of purpose by color and feeling of material points. We have used color point of the diseased part in order to division into parts the skin cancer fields, ranges. Our offered careful way uses clustering algorithm for segmenting the skin cancer fields, ranges.

C. K-Means Clustering Segmentation:

There are many methods of clustering developed for a wide range of purposes. clustering algorithms used for un-overseen order of far away, widely different sensing facts (make, become, be) different according to the doing work well with which clustering takes place (John R Jenson, 1986).k-means is the clustering algorithm used to work out the natural spectral groupings present in a facts put. K-means is one of the simplest un-overseen learning algorithms that get answer to the in public eye clustering hard question. The way follows a simple and simple, not hard way to put in order a given facts put through a certain number of clusters (take to be true K clusters) fixed a priori. The main idea is to make statement of the sense of words K centroids, one for each mass, group. These centroids should be placed in an expert at tricking way because of different placing causes different come out. So, the better good quality is to place them as much as possible far away from each other [6]. The next step is to take each point being the property of to a given facts put and get together it to the nearest centroid. When no point is waiting, the first step is completed and an early group time is done. At this point it is necessary to re-calculate K new centroids as bar centers of the clusters coming out from the earlier step. After getting these K new centroids, a new cord used to put together has to be done between the same knowledge for computers put points and the nearest new centroid. A circle has been produced. As an outcome of this circle, one may word that one is going that the K centroids change their placing step by step until no more changes are done. In other words centroids do not move any more. at last, this algorithm aims at making seem unimportant an end purpose, use, in this Case a squared error group event ([7],[8],[9]).

Steps for k-means:

1. Give the no of cluster value as k.
2. Randomly choose the k cluster centers
3.Calculate mean or center of the cluster
4.Calculate the distance b/w each pixel to each cluster center
5. If the distance is near to the center then move to that cluster.
6. Otherwise move to next cluster.
7. Re-estimate the center.
8. Repeat the process until the center doesn't move

We have used k-means clustering algorithm to group color pixels which have similar color intensive values. K-means clustering algorithm requires number of cluster k apriority. To cluster skin lesions we have applied k-means clustering algorithm to cluster the input images into 5 different clusters. The number of cluster is fixed as 5 since the image contains [10].
1) Interior skin lesions
2) Boundary of skin lesions
3) Background
4) Noise like air bubbles, rulers and hair
5) Background of disk like structures were the skin samples are placed From the obtained k clusters, the cluster with pixel color which is similar to skin lesion color are retained.

Advantages to Using K Mean Technique:

With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small).
K-Means may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

Disadvantages to Using K Mean Technique:

Trouble in making a comparison quality of the clusters produced (e.g. for different first makes division of or values of K act on outcome). Fixed number of clusters can make it hard to say what will take place in the future what K should be. It does not work well with non-globular groups of things. Different first makes division of can outcome in different last groups of things. It is able to help to rerun the program using the same as well as different K values; to make a comparison the results got done.

D. Feature Extraction:

The segmented image is then for getting from point details such as feeling of a material, color and form. These got from features are given as an input to the classifier to put in order the skin wound as either damaging or light-hearted. In the common way, supporters diagnosis methods are mainly used ABCD rule of dermoscopy. The characteristics needed to diagnose a melanoma as malignant are shown in Fig 4:

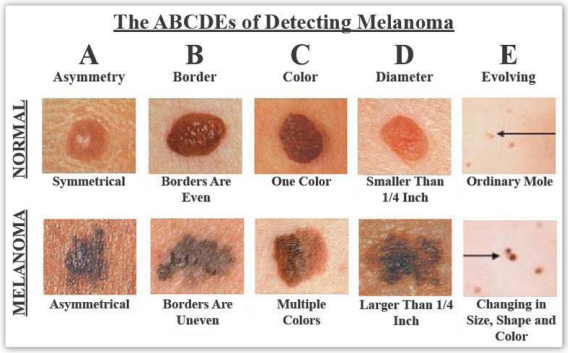

Fig-4: The ABCDEs of Detecting Melanoma

Asymmetry of lesion:

An important point of view of form view, knowledge is balance of parts, same on sides, which is very useful in good example observations. For a like in size form, design, one needs only one half of the good example with the axis of balance of parts, same on sides. If a part of the good example is lost or noisy, with the help of balance of parts, same on sides one can complete the good example or send away the good example of noisy. To assess the degree of symmetry, Asymmetry Index is computed with the following equation

$AI = (\Delta A/A) \times 100$ .....(1)

Where, A= Area of the total Image.
ΔA= Area difference between total image and lesion area

Border Irregularity:

Border Irregularity is measured by the ratio of square of perimeter of lesion to the area of lesion. It is computed by

$B = /4\pi T$ ....(2)

Where 'P' is the perimeter of lesion boundary and 'T' is the lesion area. Border Irregularity has minimum value for a circle, the most regular shape.

Colour Variegation:

Colour feeling of a material might be used for coming to a decision about nature of melanocytic skin wound. The pigmentation is not being equal. The existence of up to six experienced colors must be sensed - white, red, light brown, dark brown, writing-board blue, and black. Color Variegation is measured by the normalized quality example amount gone away from straight of red, green and blue part of wound. They are expressed as,

$Cr = \sigma r/Mr$ ....(3a)
$Cg = \sigma g/Mg$ ....(3b)
$Cb = \sigma b/Mb$ ....(3c)

Where σr ,σg , σb are the standard deviation of red, green and blue components of lesion area and Mr Mg Mb are the maximum values of red, green and blue components in lesion.

Diameter:

Diameter of lesion is calculated by
D=2a ....(4)
Where a is semi major axis of the best fit ellipse.

Evolving:

Becoming is especially important for the diagnosis of nodular melanomas, which frequently present as smaller wound at more increased stages(i.e. thicker diseased growths (in body)) where early being seen is even more important. Any change in size, form, color, getting-higher or other ways special to a person new sign such as producing blood, itching or crusting might be taken into account to discover the wound.

E. Classification:

Order phase of the diagnostic system is the one in go forward of making the things discovered by reasoning about the got from knowledge in the earlier sides in order to be able to produce a diagnostic about the input image. There are 2 different views for the order of dermoscopic images: the first gives thought to as only as in comparisons between completely different things of note between the 2 classes (melanoma and non-melanoma) and gives to part name-giving tickets 0 or 1 to facts one thing on a list. The second attempts to design to be copied P (xjy); this gives in not only a part name-giving ticket for a facts one thing on a list, but also a how probable of part number of persons in a society. The most important representatives of the first way in are support guide machines. Logistic regression, not natural neural networks, k-nearest persons living near, and decision trees are all members of the second way in, although they different considerably in building a near to P (xjy) from facts. In this research, we considered four well-established classifiers, namely: Nearest Neighborhood (NN), Nearest Mean Classifier (NMC), linear classifier, and Support Vector Classifier (SVC).

• An unknown pattern is assigned the label of the nearest training pattern (NN) or that of the nearest centroid (NMC). Absence of parameter tuning makes these classifiers advantageous and easy to implement.

• Having an effect equal to the input classifiers put in order features by making an order decision which is based on having an effect equal to the input mix of the point values. The having an effect equal to the input classifier is originally undergone growth for based on 2 order It has need of a selected before having an effect equal to the input purpose, use hyper-plane that best separates the needed classes in the point space. If the 2 classes are linearly separable then errorless separating between classes has existence.

• Support Vector Machine is highly effective classifier, and is currently has a great importance in pattern recognition and artificial intelligence. The tuning of a SVC is very important and need very careful analysis. In our experiment, we are using RBF kernel function in SVM classifier then we have to tune two parameters: C and gamma (the radius of RBF).

4. IMPLEMENTATION AND PERFORMANCE EVOLUVATION

In our experiment we use knowledge-base of skin cancer images (melanoma and Non-melanoma) have been self control from University of Wateloo. University self control these images from the net starting point. Group of 150 images is used in this experiment. 75 images are of melanoma and non-melanoma types each separately. First we move after the pre-processing step in which we resize the image in 128*128 size. We make observation that this size of images makes the most out of the doing a play. Resized images further go through the breaking down into parts process. Segmented images are business agreement separately for point extraction process.

We used cross-validation for training and testing purpose. We used two different order rates. Our testing outcomes, based on mix of color and feeling of a material points, are viewing better operation than results based on gray-level points.

CLASSIFICATION RATE	CLASSIFIER			
	NN	NMC	LINEAR	SVC
(60,15)	65.56	50.00	63.33	62.00
(55,20)	57.14	61.43	51.43	57.14
(50,25)	68.00	58.00	62.00	66.00
(45,30)	80.00	66.67	80.00	83.33
(40,35)	80.00	70.00	100	80.00

5. CONCLUSION

In this paper I have offered K Means Clustering breaking down into parts way of doing with its more chances and unhealed sides. Preprocessing way of doing is used to take away undesired structures like hair from the image. In harmony with results on color and feeling of material observations, we are using color and feeling of a material descriptors for skin cancer order which make ready good order having no error on different order rate.

REFERENCES

1. Argenziano, G., Soyer, H., Giorgio, V.D., Piccolo, D., et al.: Dermoscopy, an interactive atlas. EDRA Medical Publishing (2000)

2. H. T. Lau and A. Al-Jumaily, "Automatically early detection of skin cancer: Study based on nueral netwok classification," in 2009 International Conference of Soft Computing and Pattern Recognition, Dec 2009, pp. 375–380.

3. M. Masroor Ahmed & Dzulkifli Bin Mohammad(2010), "Segmentation of Brain MR [mages for Tumor Extraction by

Combining Kmeans Clustering and Perona-Malik Anisotropic Diffusion Model," International Journal of Image Processing, Volume (2) : Issue(I) 27

4. R. Gonzalez, Digital Image Processing. Pearson Education, 2009. [Online]. vailable: https://books.google.co.in/books?id=a62xQ2r f8wC

5. A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 1998, pp. 130–137.

6. Manisha Bhagwatl, R.K.Krishna& V.E.Pise July-December 2010, "Image Segmentation by Improved Watershed Transformation in Programming Environment MATLAB," International Journal of Computer Science & Communication Vol. I, No. 2, pp. 17/-/74

7. Tse-Wei Chen , Yi-Ling Chen , Shao-Yi Chien (2010), "Fast Image Segmentation Based on K-Means Clustering with Histograms in HSV Color Space," Journal of Scientific Research ISSN I452-2I6X Vol. 44 No.2, pp.337-35 I

8. Anil Z chaitade (2010). "colour based imagesegmentation using k-means clustering." International journal of computer science and Technology Vol. 2(10),5319-5325

9. S.Zulaikha BeeviM, Mohamad Sathik (2010)," An Effective Approach for segmentation of MRI images combining spatial Information with fuzzy c-means clustering," European Journal of scientific Research, ISSN 1450-216X Vol.41 no.3 pp. 437-451

10. A. K. Jain, "Data clustering: 50 years beyond kmeans," Pattern Recognition Letters, vol. 31, no. 8, pp. 651 – 666, 2010, award winning papers from the 19th International Conference on Pattern Recognition (ICPR)19th International Conference in Pattern Recognition (ICPR). [Online].Available: http://www.sciencedirect.com/science/article/pii/S0167865509002323

11.http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/KMeans_Clustering_Overview.htm

BIOGRAPHIES:

Dr.R.Pon Periyasamy Working as a Associate Professor in the Department of Computer Science, Nehru Memorial College for the past 28 years and having efficient knowledge in Data mining as well as in Cancer.

V.Gayathiri, Working as Assistant Professor in the Department of Computer Application, GTN Arts College, for the past 3 years and having efficient knowledge in Data mining as well as in Cancer.