Intelligent Classification of Melanoma Using SVM-Machine Learning

Manisha N L[1], Neethu V C[2], Swathy Venu K[3], Sheethal M S[4]
[1,2,3,4] *Sahrdaya college of engineering & technology*

***Abstract*- In recent days, skin cancer is seen as one of the most Hazardous form of the Cancers found in Humans. Skin cancer is found in various types such as Melanoma, Basal and Squamous Cell Carcinoma among which Melanoma is the most unpredictable. The detection of Melanoma cancer in early stage can be helpful to cure it. In this paper, an intelligent decision support system for melanoma classification is build as a health monitor robot to promote early diagnosis. This system is developed in order to categorize skin lesion as malignant and benign for these steps such as pre-processing, segmentation, feature extraction and finally classification is done. Dermoscopic images are used as the database which improves the performance during the prediction of the result. Multiple regions from skin lesion is taken into consideration which is the highlights of this project. A combination of support vector machine and random forest are used as the classifier which predicts the desired output.**

I. INTRODUCTION

The incidence of malignant melanoma continues to increase worldwide. This cancer can strike at any age; it is one of the leading causes of loss of life in young persons. Since this cancer is visible on the skin, it is potentially detectable at a very early stage when it is curable. New developments have converged to make fully automatic early melanoma detection a real possibility. The development of advanced technologies in the areas of image processing and machine learning have given us the ability to allow distinction of malignant melanoma from the many benign mimics that require no biopsy. These new technologies should allow not only earlier detection of melanoma, but also reduction of the large number of needless and costly biopsy procedures. Although some of the new systems reported for these technologies have shown promise in preliminary trials, widespread implementation must await further technical progress in accuracy and reproducibility. Dermoscopy is a non-invasive skin imaging technique which allows a magnified visualization of the skin surface and sub dermal structures. Dermoscopy images have played a significant role in increasing the survival rate of patients by assisting the early diagnosis of MM. Usually, a computerized dermoscopy image analysis model includes four aspects: pre-processing, segmentation, feature extraction and classification. Pigmented skin lesions are typically evaluated by dermatologists using the "ABCD" rule, which analyzes the Asymmetry, Border irregularity, Color variation and Different structures of a lesion. Based on the "ABCD" rule, many classification methods have been developed on diverse datasets of dermoscopy images. As the initial step, pre-processing is done where hair removal, resizing, gray-scale conversion, structuring element and also contours are found out. Here Otsu method is used for the segmentation. As a part of feature extraction we are collecting 12 features namely border irregularity, vertical asymmetry, horizontal asymmetry, compactness, perimeter, relative chromaticity of red, green and blue, ratio of red, green and blue, difference in lightness, chroma, color and hue and GLCM features such as entropy, IDM, homogeneity, mean1 and contrast in an arrary for each particular image. A combination of random forest and svm classifier is used for the classification section. Apart from other researches we have taken multiple skin lesion region that improves the accuracy.

II.PROPOSED SYSTEM

This research mainly deals with the intelligent classification of melanoma using SVM & Random forest algorithm which takes multiple regions from skin lesion. The major steps that helps the system to do the classification includes; pre-processing,

segmentation, feature extraction and classification. Fig.1 shows the system architecture of the proposed system.

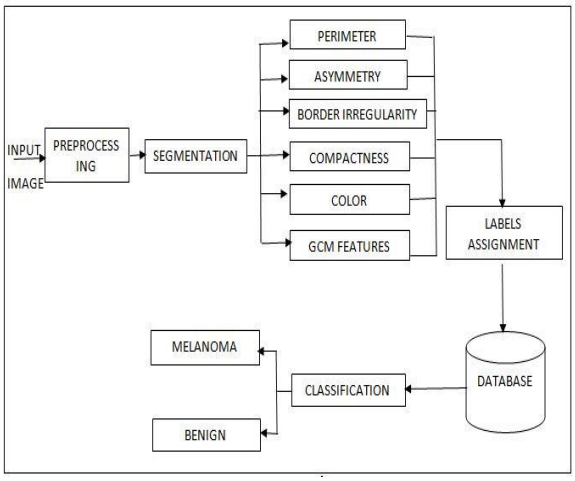


Fig.1 System architecture

Pre-processing

Pre-processing is a common name for operations with images at the lowest level of abstraction -- both input and output are intensity images. The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features important for further processing. The following steps are done in this research such as hair removal, contrast enhancement, gray-scale conversion and noise filtering.

a) Hair removal

The Enhanced Dull Razor algorithm was used to remove hairs from images where morphological closing image processing was generalised to grey-level images, followed by identification of the narrow, elongated hair outline. Bilinear interpolation was implemented to substitute the identified pixels of the hairs. Matlab'rolfill' function was used to perform a fill operation using an interpolation method based on Laplace's equation. This step resulted in a smooth fill inward from the borders of the region of interest.

b) Contrast enhancement

Subsequently, the image clarity was enhanced by improving the shape and edges of the image. Image borders were sharpened using contrast enhancement. This process may also optimise subsequent segmentation accuracy.

c) Gray-scale conversion

RGB images of lesions, with 360×360 pixels in size, were transformed to greyscale by removing hue and saturation using a process which computes the weighted sum of the color components.

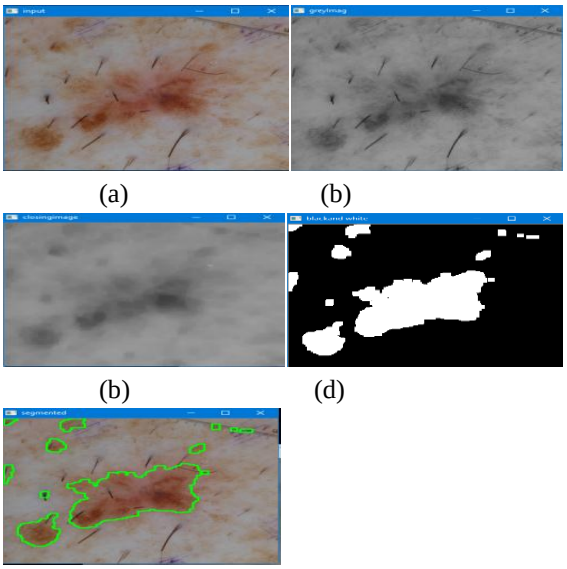
d) Noise filtering

Median filtering was used to effectively remove 'salt and pepper' noise, thus diminishing the effect of small objects such as thin hairs and outlying regions of pixels caused, for example, by air bubbles.

Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like Marching cubes.

Otsu's method, named after Nobuyuki Otsu, is used to automatically perform clustering-based image thresholding or the reduction of a gray-level image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalently (because the sum of pair wise squared distances is constant), so that their inter-class variance is maximal.



(a) (b)

(b) (d)

(e)

Fig.2. Segmentation results:(a)input image, (b)gray-scale image, (c)closing image, (d)black and white image, (e)segmented image.

Feature extraction

Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to over fit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.

1) *Shape*

Asymmetry: A melanocyte lesion may be diagnosed by a number of identifiers, of which one of the most significant is a lack of symmetrical morphology.In consideration of the symmetry feature, a number of factors are concurrently relevant, including colour, texture and morphology. A three-fold classification system can be derived from measuring symmetry, with three-class outputs representing total symmetry, a lack of symmetry along a single axis and a lack of symmetry along dual axe, respectively. The lesion asymmetry was evaluated by calculating the area with inner and outer of the lesion, using the formula shown as follows.

$$AI = \frac{\Delta AK}{AL} * 100$$

where AI represents asymmetry Index. ΔAK represents the area between the two halves of the lesion and AL denotes the lesion area.

Border Irregularity: Irregularities occurring in the edge of a malignant lesion offer useful information concerning that lesion's nature. Typically, the edge of a malignant lesion usually exhibits four factors of interest, i.e. density, fractal dimension, radial variability and the extent to which its contour exhibits small irregularities. To identify the lesion border irregularity, we use the following four border irregularity indexes as shown below:

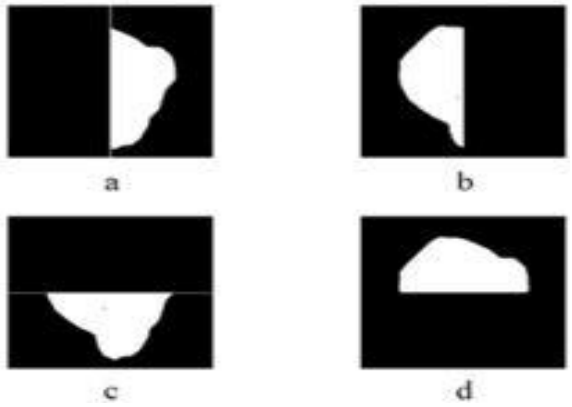


Fig.3. Calculating the border irregularity index, a:border right, b:border left, c:border bottom, d:border upper.

$$I = \frac{ab}{2\pi(a^2 + b^2)} \frac{P^2}{\Delta A}$$

where I represent irregularity with a and b representing the lengths of major and minor axes of the lesions. P represents the perimeter of the lesion and ΔA indicates the area of corresponding.

Compactness: Another relevant feature is the degree to which the lesion can be described as compact. In order to determine this aspect, a comparative analysis is performed between the lesion's boundary and a circle with a circumference of the same length. It is the former of these two numerical values that presents a challenge in its assessment. One solution to this issue is to use the proportions of the most easily measured values of maximum and equivalent lesion diameter is defined as:

$$C = \frac{4\pi P^2}{\Delta A}$$

where C represents the compactness.

2) Color

The range of colour types utilised in diagnosing a melanocyte lesion can be broadly categorised into the following types: black, grey-blue, brown (dark), brown (light), red and white, which are indicators for a malignant skin lesion. The dermatological analysis allows for the determination of whether a colour category exists in a particular image and if so, where it exists. This positional information is noted via a binary mask application, with image segmentation performed by the dermatological professional. In this paper, three type of colour space including HSV, RGB and LAB are used.

Relative chromaticity of red, green and blue: The chromaticity of red, green and blue in relation to each other is also significant. The chromaticity of red in relation to green and blue is expressed as follows:

$$Rel_{Chroma,Red} = \frac{r_L}{r_L + g_L + b_L} - \frac{r_S}{r_S + g_S + b_S}$$

In the above expression, rL, gL and bL represent the mean red, green and blue appearing within the lesion, respectively. The chromaticity of green and blue in relation to red/blue and red/green respectively are expressed in a similar way.

Ratio of red, green and blue: In the case of red, the ratio represents the average of the red constituent present in a lesion divided by the mean colour of the surrounding non-lesion skin. The ratio for red is expressed as follows:

$$Ratio_{Red} = \frac{r_L}{r_S}$$

Difference in lightness, Chroma, colour, and hue: Finally, the factors exhibited regarding the lesion's tone, Chroma, colour and hue are examined in relation to the nonlesion skin are considered. The lesion's tone, Chroma and hue are represented by $LL*$, $CC*$ and $HH*$ respectively in the following equation.

$$(\Delta L^*) = L^*_{Lesion} - L^*_{Skin}$$

$$(\Delta C^*) = C^*_{Lesion} - C^*_{Skin}$$

$$(\Delta Clr^*) = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

$$(\Delta H^*) = \sqrt{(\Delta Clr^*)^2 + (\Delta L^*)^2 + (\Delta C^*)^2}$$

where

$$\Delta a^* = a^*_{Lesion} - a^*_{Skin}$$

$$\Delta b^* = b^*_{Lesion} - b^*_{Skin}$$

3) Texture

The Gray Level Co-occurrence Matrix[1] (GLCM) and associated texture feature calculations are image analysis techniques. Given an image composed of pixels each with an intensity (a specific gray level), the GLCM is a tabulation of how often different combinations of gray levels co-occur in an image or image section. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity (a.k.a. image texture) at the pixel of interest. The features used are entropy, energy,

1. Energy feature

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2$$

2. Entropy feature

$$Entropy = \sum_{i,j=0}^{N-1} -\ln\left(P_{ij}\right) P_{ij}$$

3. Contrast feature

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}\left(i-j\right)^2$$

4. Homogeneity feature

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+\left(i-j\right)^2}$$

where:

P_{ij} = Element i,j of the normalized symmetrical GLCM

N = Number of gray levels in the image as specified by Number of levels in under Quantization on the GLCM texture page of the Variable Properties dialog box.

Classification

Random forest and SVM are used as the classifier Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes(classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set.

III.EXPERIMENTAL RESULTS

Since most of them, however, consist of 720 pixels in the longest dimension, other images that did not meet this criterion were resized to 720 pixels in the longest dimension, with the aspect ratio being maintained. In each case, the lesion appears in the centre of the image with non -lesion skin visible in the corners of each image. This required a mean run-time between 5 and 10ms, although this will vary depending upon which base SVM model is utilised. By using the formula,

$$Accuracy = \frac{(Total\ of\ detected\ lesions)}{(Total\ dataset\ images)} * 100$$

$$TP$$

We have got 69% for svm and 96% for random forest classifiers as a result out of 297 test images applied. Hence, Random forest is considered to be the best classifier that pedicts more accurately than svm classifier from the table shown below:

CLASSIFICATION METHODS	TEST IMAGES	ACCURATE RATE	FALIURE RATE	ACCURACY
SVM	297	205	92	69%
RANDOM FOREST	297	285	12	96%

IV.CONCLUSION

This research proposes an intelligent decision support system for the identification of benign and malignant skin lesions from dermoscopic images, which offers the prospect of achieving improved and more accurate classification of lesions from images.

The proposed system employs pre-processing such as dull razors and median filters to remove hair and other noise. Then, the images were segmented using a pixel limitation technique to separate lesions from image background. Feature extraction is subsequently conducted. The features extracted by our system reflect the well-known asymmetry, border irregularity, colour variegation and diameter (ABCD) of dermatology and the epiluminescence microscopy (ELM) criteria. They focus primarily on the size, shape, colour and local parameters of lesions with some additional consideration of the lesion edges. From the experiments we have concluded that Random forest algorithm is the best with 90% and 94% accuracy for malignant and benign.

REFERENCES

[1] Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N. and Halpern, A., 2016. Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint arXiv:1605.01397.

[2] Flores, E. and Scharcanski, J., 2016. Segmentation of melanocytic skin lesions using feature learning and dictionaries. Expert Systems with Applications, 56, pp.300-309.

[3] Maier, T., Kulichova, D., Schotten, K., Astrid, R., Ruzicka, T., Berking, C. & Udrea, A. 2015. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. J Eur Acad Dermatol Venereol, 29, 663-7.

[4] Moan, J., Grigalavicius, M., Baturaite, Z., Dahlback, A. and Juzeniene, A., 2015. The relationship between UV exposure and incidence of skin cancer. Photodermatology, photoimmunology & photomedicine, 31(1), pp.26-35.

[5] Abuzaghleh, O., Faezipour, M. and Barkana, B.D., 2016. A Portable Real-Time Noninvasice Skin Lesion Analysis System to Assist in Melanoma Early Detection and Prevention.

[6] Sboner, A., Eccher, C., Blanzieri, E., Bauer, P., Cristofolini, M., Zumiani, G. & Forti, S. 2003. A multiple classifier system for early melanoma diagnosis. Artificial Intelligence in Medicine, 27, 29-44.

[7] Schmid, P. 1999. Segmentation of digitized dermatoscopic images by two-dimensional color clustering. Medical Imaging, IEEE Transactions on, 18, 164-171.

[8] Gilmore, S., Hofmann-Wellenhof, R. & Soyer, H. P. 2010. A support vector machine for decision support in melanoma recognition. Experimental dermatology, 19, 830-835.

[9] Ramlakhan, K. & Shang, Y. A Mobile Automated Skin Lesion Classification System. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, 7-9 Nov. 2011 2011. 138-141.

[10] Barata, C., Ruela, M., Francisco, M., Mendonça, T. & Marques, J. S. 2014. Two systems for the detection of melanomas in dermoscopy images using texture and color features. Systems Journal, IEEE, 8, 965-979.