International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-05, Issue-04, July 2019

# Classification of Melanoma Skin Cancer using Machine Learning Technique

[*]Amruta S. Shitole, [#]Sunita kulkarni

[*]MTech Student School of Electronics and Communication, [#]Professor, Dr. Vishwanath Karad, MIT World peace University, Pune, India.

[*]amu9shitole@gmail.com, [#]sunita.kulkarni@mitcoe.edu.in

**Abstract** **The detection of melanoma skin cancer in the premature stage is crucial for effective treatment. Recently, it is well known that the most dangerous form of skin cancer in other types of skin cancer is melanoma, since it is more likely to spread to other parts of the body that were not originally diagnosed or treated for uncertainty. Non-invasive medical computer vision or medical image processing plays an increasingly important role in the clinical diagnosis of various diseases. These technologies provide an automated image analysis tool for accurate and rapid assessment of wear. The next article will contain a brief method of skin cancer, in particular, melanoma cancer is detected and classified in various ways using a strong method.**

***Keywords —skin, detection, diseases, melanoma, segmentation.***

## I. INTRODUCTION

Skin cancer is a harmful disease. The skin has three main layers. Skin cancer occurs on the outermost layer and consists of the first layer of squamous cells, the second layer of basal cells and the innermost layer or the third layer of melanocytes. Squamous and basal cells are sometimes called non-melanoma.

Non-melanoma skin cancer is always treatable and rarely spreads to other skin tissues. Melanoma is at higher risk than most additional categories of skin cancer [3]. If it is not detected at the beginning, it will quickly penetrate into nearby tissues and spread to other parts of the body. An analytical technique suitable for recognizing skin cancer is a surgical technique. Surgery is a method that removes a piece of tissue or a part of a cell from a patient for analysis in a research laboratory. This is an awkward way. Surgical methods are time consuming for both patients and doctors, because testing takes a long time.

The operation is performed by removing skin tissue (skin cells) that are analyzed in a number of research laboratories. There is a risk of spreading the disease to other parts of the body. The risk is greater. Given all this, SVM is used to recognize skin cancer. This method is used in digital image processing and SVM for classification. This program has inspired the initial detection of skin cancer. To detect skin cancer in image processing, we must first take the affected skin, but we must be careful that no oil is applied to the skin. Thus, it is faster and cleaner way. However, most importantly, due to the high increase, the detection of skin cancer using SVM can prevent the unnecessary removal of completely harmless sputum and skin damage.

## II. LITERATURE REVIEW

The detection of melanoma skin cancer (MSC) using non-invasive methods, such as image processing techniques, has been one of the attractive and demanding studies examined over the years. According to a specific study, a brief method of skin cancer, especially using a strong method for detecting melanoma in various technical ways.

The relevant steps in the study area group information about the dermoscopic image, pre-process, segment the victim's threshold, and use mathematical functions to extract the victim's gray level matrix (GLCM). In many analytical documents, it is used to detect regional units of color and new parameters of texture, asymmetry, border, color, diameter (ABCD), detection of victim image processing, and artificial neural networks.

Also during the study, the following classification methods were used: PCA (main component analysis), KNN (nearest neighbor K) and SVM (support vector machine).

## III. PROPOSED SYSTEM

Skin cancer detection using SVM is basically defined as the process of detecting the presence of cancer cells in an image. The detection of skin cancer is provided by the victim of the GLCM and the support vector machine (SVM). The Gray Level Matrix Matrix (GLCM) is used to extract parameters from images that can be used for classification. SVM is a machine learning method used mainly for classification and multidimensional analysis.

 DOI : 10.35291/2454-9150.2019.0312 

A system based on image processing that basically finds, extracts and classifies lesions from dermatoscopic images is proposed. This system can be very convenient for diagnosing malignant melanoma. In many cases, we aim to use a new technology to extract lesions from digital dermoscopic images, which can be mentioned in the next section, as shown in Figure 1.

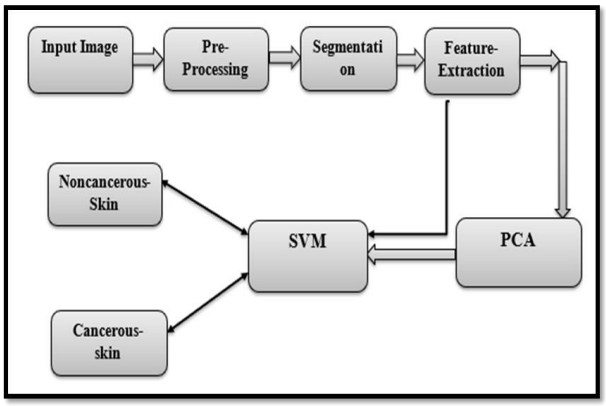


**Figure 1: The projected system block diagram.**

## IV. METHODOLOGY

To classify skin images as cancer or not cancer, we tend to follow the chosen method. Initially, the threshold image is the same pigment processing threshold, and then the ABCD method is applied to the metameric image to extract traits. Finally, the rules of the SVM (Support Vector Machine) algorithm are used to checking whether rous is cancerous or does not support extraction parameters. A pair of graphs shows the overall situation in the system. Photos used cannot be obtained from the quietest and hierarchical datasets.

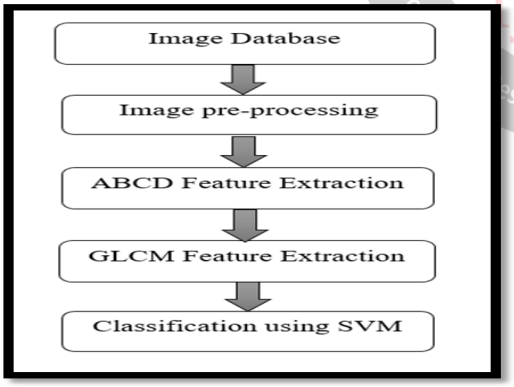


**Figure 2: Methodological flow of proposed system**

### A. Image Database

The database was generated by collecting images from different websites with known category (Normal/Melanoma). These websites are specified for melanoma skin cancer.

### B. Pre-processing

This step involves converting the captured RGB image to a grayscale image, enhancing the contrast, modifying the histogram, and filtering noise. Contrast enhancement and histogram modification are proposed, because some of the resulting images are not uniform due to improper lighting during image acquisition.

Although a histogram modification method, such as histogram alignment, is used to increase image contrast, segmentation has been made more accurate. Noise filtering using a median filter is implemented to reduce the effect of hair coating on the skin in the final image used for classification.

### C. Image Segmentation

Image segmentation is a necessary step in image analysis, since it distinguishes between intact skin and related lesions. The images that make up our dataset are unchanged and retain their original size and resolution. The process of segmentation of images is not as simple as there are many skin types, lesion forms and borders.

In our proposed technique, the threshold value is used for segmentation, first obtaining the binary form of the original image (a), and then transforming it into shades of gray (b). Then we extract the edge of the lesion, as shown in (c). The final step is to get the associated component representing the failure. The latter allows us to obtain an image (d), and then use it to select objects in Figure 3.

Certain geometrical features of an extraction lesion feature may indicate the presence of skin melanoma. After segmentation of the cover image, we can extract four different attributes that belong to the ABCD function.

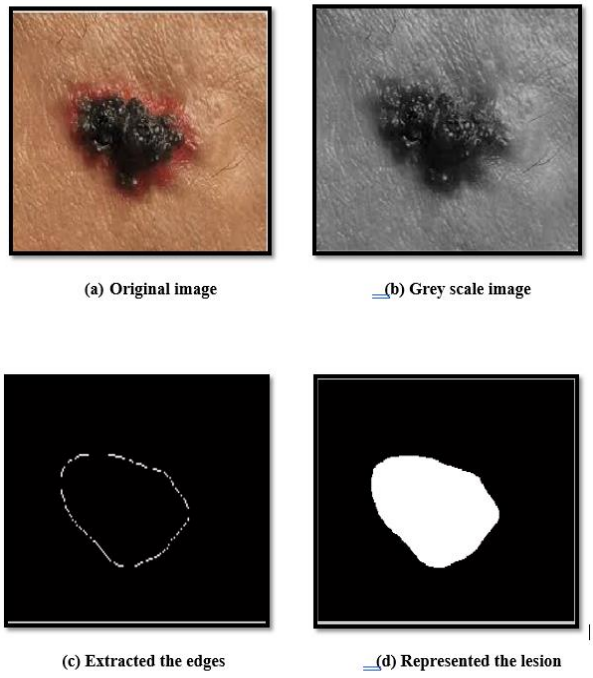


**Figure 3: Methodological flow of proposed system**

### D. ABCD Feature Extraction

ABCD is an abbreviation for Asymmetry. The irregularity of the border changes color and diameter. Features of the selection of melanoma to detect skin cancer. In melanoma

 DOI : 10.35291/2454-9150.2019.0312 

skin cancer, the lesion has asymmetry and has an irregular border with red skin, which is further complicated by red, blue, black, etc. D. And with preservation of a diameter of more than 6 mm when melanoma skin cancer is detected This feature of the extraction is very important. Features are image statistics. GLCM is a method of extracting completely different options, gray and binary images. Remove the square measure in the prediction method after the GLCM option [3]

### A. Contrast

Contrast measures the original changes in the gray-level co-occurrence matrix.

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j)$$

### B. Homogeneity

Measuring the homogeneity of the distribution of closeness of the components in the GLCM and GLCM diagonal.

$$\text{Homogeneity} = \sum_{i,j} \frac{1}{1+(i-j)^2} p(i,j)$$

### C. Energy

Energy is a measure of homogeneity between pixels.

$$\text{Energy} = \sum_{i,j} p(i,j)^2$$

### D. Entropy

The image entropy may be a randomness of the applied component or element of the mathematical dimension.

$$\text{Entropy} = -\sum_{i,j} p(i,j)\log(p(i,j))$$

### E. Sum of average

The average sum is the sum of the diagonal elements in the image in grayscale.

$$\text{Sum of Average} = \sum_{i=0}^{2(N-1)} i * p_{x+y}(i)$$

Where,

$$P_{x+y}(k) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p_{(d,\theta)}(i,j), \qquad k = i+j$$

### F. Autocorrelation

Autocorrelation is a measure of the relationship between adjacent pixels in an image. This suggests that the image has no associated pixel elements and everything is unique.

$$\text{Autocorrelation} = \sum_{i,j} p(i,j) / \log(p(i,j))$$

### G. Dissimilarity

Dissimilarity is a measure that determines the change in gray level in an image.

$$\text{Dissimilarity} = \sum_{i,j} |i-j| p(i,j)$$

### H. Difference variance

Variance is a set of statistics whose observations differ from each other.

$$\text{Diffvariance} = \left(\sqrt{\frac{\sum|p(i,j) - p(i,j)^2|}{n}}\right)^2$$

### I. Difference entropy

This is the difference in the randomness of the texture of the input image.

$$\text{Difference Entropy} = \sum_{i,j} |i-j|^2 p(i,j)$$

### J. Inverse difference normalized (INN)

Image INN is affected by image uniformity.

$$\text{INN} = \sum_{i,j} p(i,j)/(1+\frac{|p(i,j)|}{i,j})$$

### K. Correlation

The correlation on the sides of a pixel is the ratio of linear correlation to pixel.

$$\text{Correlation} = \sum_{i,j} \frac{(i,j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$$

### L. Inverse difference homogeneity (INV)

INV is given as

$$\text{INV} = -\sum_{i,j} p(i,j)/(1+|(p(i,j))|)$$

### M. Inverse difference moment

IDM is also considered local homogeneity. It is proportional to the gray level in the image.

$$\text{IDM} = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p_{i,j}}{1+(i-j)^2}$$

### N. Maximum probability

This simple statistical record records the maximum value of Pij found in the window in the center pixel of the window.

$$\text{max probability} = \max\left(\sum_{j=0}^{N} p(i,j)\right)$$

Where,

p (i, j) = image pixel to be processed

n = number of pixels in an image.

$\mu_x$, $\mu_y$ = mean of the Px and Py

$\sigma_x$, $\sigma_y$ = standard deviations of Px and Py.

N = number of gray levels

### O. Mean

The average value is obtained by taking the average of all the color components of the used color space.

$$\text{Mean } (\mu) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}$$

### P. Standard Deviation

Standard deviation is a different statistical measure used, which is defined as the number that represents a different

 DOI : 10.35291/2454-9150.2019.0312 

number of color channel members than the mean value of the channel.

$$\text{Standard Deviaition}(\sigma) = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(P_{ij}-\mu)^2}$$

### Q. Skewness

Skewness measures the degree and direction of symmetry or distribution asymmetry. The skewness of the normal or symmetric distribution is zero (0). But in the real world it's hard to get a normal distribution. Therefore, the distribution can have a positive skew (tilt to the right; tilt to the right longer; represented by a positive value) or a negative tilt (tilt to the left; tilt to the left longer; has a negative value).

$$\text{Skewness}(\theta) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(P_{ij}-\mu)^3}{MN\sigma^3}$$

### R. Kurtosis

Kurtosis measures however peaked a distribution is and therefore the lightness or heaviness of the tails of the distribution. It measures what quantity of the distribution is really placed within the tails

$$\text{Kurtosis}(\gamma) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(P_{ij}-\mu)^4}{MN\sigma^4}$$

| A | B | C | . . . | Kurtosis | Mean | Vari | Label |
|---|---|---|---|---|---|---|---|
| 0.958 | 0.027 | 0.577 | . . . | 2.124 | 127.22 | 4714.3 | 0 |
| 0.9740 | 0.469 | 0.4909 | . . . | 2.5940 | 127.57 | 3934.2 | 0 |
| 0.0496 | 0.490 | 0.5975 | . . . | 2.9669 | 127.09 | 2571.6 | 0 |
| 0.0545 | 0.5024 | 0.4299 | . . . | 2.2455 | 127.28 | 2690.0 | 0 |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . | . |
| 0.2067 | 0.372 | 0.1763 | | 2.6232 | 127.4 | 4419.5 | 1 |
| 0.094 | 0.692 | 0.6248 | . . . | 2.0985 | 127.53 | 3758.7 | 1 |
| 0.0144 | 0.661 | 0.5864 | . . . | 2.1584 | 127.43 | 4351.2 | 1 |
| 0.1743 | 0.6113 | 0.4553 | . . . | 2.0331 | 127.64 | 4627.7 | 1 |

**Table 1: Feature Extraction**

## V. PRINCIPAL COMPONENT ANALYSIS (PCA)

The functions (contrast, asymmetry, kurtosis, energy, mean, standard deviation, cycle, energy, correlation, uniformity, and TDS value) extracted from the above four steps are entered into the PCA. Some features may not be effective for accuracy due to the time required for accurate classification. PCA is used to reduce the number of functions due to different units in the set of functions. The PCA uses a correlation matrix instead of a covariance matrix. After performing this operation and calculating eigenvalues and deviations, a set of basic components are obtained, which are ranked based on their ability to distinguish between benign and malignant lesions.

To determine the number of signs that lead to the best classification results, we save them and check their effectiveness during the classification. Finally, we selected the top 5 features with maximum efficiency, as shown below: TDS, mean, standard deviation, energy and contrast.

## VI. CLASSIFICATION USING SVM

n the classification, we classify cancerous and non-cancerous skin images. There are several classification methods: for example, decision trees, nearest neighbors, support vector machines (SVM), and neural networks. Among them, SVM is the best way to assess melanoma skin cancer. The advantage of SVM is that it is very useful for non-linear common data [12].

Support Vector Machine (SVM) SVM can be a very effective method of regression, classification and general pattern recognition. SVM provides excellent performance without adding a priori information, so it is considered an honest classifier, even if the size of the input area is very large. For a linearly shared data set, the linear classification operation corresponds to a separate hyperplane f (x) passing through the centers of the two categories, and two SVMs are developed separately for binary classification, but it can be quickly expanded for many types of tasks.

In addition to active linear classification, SVM will quickly perform non-linear classification damage, the so-called kernel method, implicitly displaying its input in the spatial domain.

Learning under supervision is not possible when the information appears to be unmarked, requires untrained learning methods, tries to find a natural bunch of information in the team, and then displays new information for the commands of these forms. The bundle rule, which provides improved correlation for the reference vector machine, is called the reference vector package and is commonly used in industrial applications where the information does not appear to be tagged or just some information is marked as preprocessing for the classification delivery [5].

More formally, support vector machines create hyperplanes or sets of hyperplanes in multidimensional or infinite-dimensional spaces that can be used for classification, regression, or other tasks.

It is intuitively clear that a good separation is achieved using a hyperplane (the so-called functional boundary), which has the greatest distance from the most recent points of training data of any class, because, as a rule, the larger the line, the lower the classifier error.

 DOI : 10.35291/2454-9150.2019.0312 

SVM is a binary classifier that separates two classes. Two important aspects of the development of SVM as a classifier are the definition of an optimal hyperplane that best separates the two classes, and the other is the conversion of a non-linearly classified classification task into a linearly separable task. The possibility of linearly separable binary classification problems without erroneous data classification is shown in the figure. Let the set of input feature vectors and class labels be x and y. The vector of the input object and the class label can be expressed as {xi, yi}, where i = 1, 2 ... N and y = ± 1. The separation of hyperplanes is that the kernel is of a different type: linear, rbf, polynomial, etc. The kernel function defines the boundary of the hyperplane. In this system we use the rbf kernel (radial basis function).

### A. RBF Kernel

A radial basis function is a nonlinear kernel that implements the mapping of nonlinear operators to an input set of functions X. A nonlinear kernel that can share training data is defined as:-

$$f(x) = W^T \emptyset(x) + b \quad \text{(a)}$$

To satisfy this equation. The function f (x) must meet the criteria

$$min\ J(w,\xi) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}\varepsilon_i \quad \text{(b)}$$

$$k(x_i, x_j) = exp\left[\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right] \quad \text{(c)}$$

## VII. CONCLUSION

The following are the main points for identifying and classifying this particular skin cancer melanoma:

• Track the exact location of the pretentious area.

• Professionally detect skin cancer.

• The operator will benefit from automatic detection of skin cancer.

• The system is not expensive and therefore can be used by a large number of people. In addition, it can also be run in rural areas.

Automatic digital image processing systems play a very important role in medical diagnostics, effectively identifying features such as asymmetry, borders, color, and diameter. And using PCA (Basic Component Analysis), select the best function with the highest efficiency, then classify the image according to the SVM (Support Vector Machine) function - cancer or not cancer, with an high efficiency.

## VIII. ACKNOWLEDGMENT

The authors would like to thank Head of School Dr. Arti Khaparde, M.tech coordinator Dr. S. B. Somani for being a source of inspiration and thus providing us with the right approach. Last but not the least, I extend my sincere thanks to all the faculty members who have helped us directly or indirectly for the completion of this project.

## REFERENCES

[1]. Pratik Dubal, Sankirtan Bhatt, Chaitanya Joglekar, Dr. Sonali Patil "Skin Cancer Detection and Classification", *2017 IEEE*.

[2]. Shivangi Jaina, Vandana jagtapb, Nitin Pisea,b"Computer aided Melanoma skin cancer detection using Image Processing" *Procedia Computer Science* 48 ( 2015 ) 735 – 74.

[3]. Rebecca Moussa, Firas Gerges, Christian Salem, Romario Akiki, Omar Falou, and Danielle Azar "Computer-aided Detection of Melanoma Using Geometric Features", *IEEE 3rd Middle East Conference on Biomedical Engineering (MECBME)*,2016.

[4]. Er.Shrinidhi Gindhi, Ansari Nausheen, Ansari Zoya, Shaikh Ruhin "An Innovative Approach for Skin Disease Detection Using Image Processing and Data Mining" Vol. 5, Issue 4, April 2017.pp 8135-8141.

[5]. G.Kesavaraj, Dr.S.Sukumaran "A Study On Classification Techniques in Data Mining" IEEE – 31661 July 4 - 6, 2013, Tiruchengode, India.

[6]. Mr. Sudhir M. Gorade, Prof. Ankit Deo , Prof. Preetesh Purohit "A Study of Some Data Mining Classification Techniques " Volume: 04 Issue: 04 , Apr -2017 IRJET, pp 3112-3115.

[7]. Dalia N. Abdul-wadood, Dr. Loay E. George, Dr. Nabeel A. Rasheed, "Diagnosis of Skin Cancer Using Image Texture Analysis" *International Journal of Scientific & Engineering Research*, Volume 5, Issue 6, June-2014.

[8]. Rahat Yasir,Md. Ashiqur Rahman, and Nova Ahmed "Dermatological Disease Detection using Image Processing and Artificial Neural Network", *8th International Conference on Electrical and Computer Engineering* 20-22 December, 2014, Dhaka, Bangladesh.

[9]. NishaYadav,Virender Kumar Narang ,Utpalshrivastava "Skin Diseases Detection Models using Image Processing: A Survey", *International Journal of Computer Applications* (0975 – 8887) Volume 137 – No.12, March 2016 .

[10]. A.A.L.C. Amarathunga, E.P.W.C. Ellawala, G.N. Abeysekara, C. R. J. Amalraj "Expert System For Diagnosis of Skin Diseases", *International journal of scientific & technology research* volume 4, issue 01, january 2015.

[11]. C.B. Tatepamulwar, V.P. Pawar, K.S. Deshpande,H. S. Fadewar, "Detection and Identification of Human Skin Diseases Using CIE lab Values", June 6,2016.

[12]. Mr. Sudhir m. Gorade, prof. Ankit deo , prof. Preetesh purohit "A Study of Some Data Mining Classification Techniques" International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 04 Apr -2017

 DOI : 10.35291/2454-9150.2019.0312 