A Novel Approach For Explicit Skin Lesion Detection From Digital Images

Authors

N.Ezhilventhan¹, S.Sathiya²

¹Assistant Professor, Department of EEE. Arunai Engineering College, Tiruvannamalai, India
²PG Scholar, Department of EEE. Arunai Engineering College. Tiruvannamalai, India
Email: sathiarasu91@gmail.com, eeeezhilventhan@gmail.com

ABSTRACT:

This paper proposes a novel approach for detection and classification of skin lesion images taken with digital camera. Due to the increased cost and subjectivity of training to users in case of dermatoscope make it difficult in clinical purposes and headed the research to computerized analysis by using digital image processing techniques. The main challenge is to locate the skin lesion explicitly from digital images. In this paper our work comprises of preprocessing, segmentation, feature extraction and classification of digital images. The results clearly differentiating the benign and malignant with desired accuracy, specificity and sensitivity.

Keywords: Adaptive mean filter, k-means, Gray level co-occurrence matrix (GLCM), Support vector machine (SVM), Segmentation.

1. INTRODUCTION:

Skin cancer is the most form of cancer all-inclusive rating of 50% around the world affected. This is noxious one and has been increase at an alarming rate of 3% per year. Cancer can be defined as a disease in which there is a fractious growth of cells aggressively and disorderly manner. Cancer can be classified based on tissues from which the cancerous cells originate. Skin cancer can usually begin with skin lesions and is divided into melanoma and non-melanoma. Melanoma often occurs as an irregularly bordered, pigmented macular, numerous shades of color, ranging from tan to brown to jet-black. Here the melanocytes reproduces a melanin at a high, abnormal rate. They leave melanin deposits there and thus changing the nature of skin coloration. Melanoma is a lethal form of cancer which causes high death rate. This malignant cancer if detected and treated early survival rates are high. Detection of malignant melanoma in its early stages substantially lower the mortality and morbidity rate. Skin cancer can be cured at high rates if and treated
early, conventionally dermatoscope is used to detect the skin lesion and later diagnosis can be done. Dermoscopy is a non-invasive technique or a bloodless clinical test to detect the skin lesions accurately. But due to high cost and lack of training to the dermatologists, it handling is difficult. So a new computerized approach for the disclosure of skin lesions from digital images has been made. The proposed method comprises at first preprocessing the images to remove impulse noise using adaptive median filter. Secondly extract the lesion from the digital image by fuzzy k-means clustering algorithm. Thirdly extracting the textural features by Gray Level Co-occurrence Matrix (GLCM) from segmented skin lesion image. Finally classify the lesion as a benign or malignant by using Support Vector Machine (SVM) classifier. The performance of the proposed method is figured out with accuracy, specificity and sensitivity.

2. PROPOSED METHOD

2.1 PRE-PROCESSING

The digital form of images are degraded by optic lens in a digital camera. This leads to compilation in human perception, as the information contents of digital images are very complex. Efficient image evolution techniques essentially needed and so developed to help physician to make a correct and accurate diagnosis. Image pre-processing makes an acquired image suitable for the particular application. It basically involves improvement on enhancement of image, which includes noise removal, edge headlining, making acute, unsmudging, illuminating, change in image divergence, masking, hair removal, cropping or resizing. The pre-processing step clear away the unwanted parts, emphasize the image, regulate the image skew and expel noise from the image. Adaptive median filter is used to eliminate the impulse noise. It performs well in case of spatial noise density. Here the noisy pixels values were replaced by median pixel values. The filter regions are defined by m×n rectangular window $S_{xy}$.

From the conventional filters it differs by made $S_{xy}$ varying depending on

$Z_{\text{min}} = \text{Minimum gray level value in } S_{xy}$

$Z_{\text{max}} = \text{Maximum gray level value in } S_{xy}$

$Z_{\text{med}} = \text{Median gray level value in } S_{xy}$

$Z_{xy} = \text{Gray level co-ordinate } (x,y)$

$S_{\text{max}} = \text{Maximum allowed size of } S_{xy}$
The scope of LEVEL 1 in Fig.1 is to resolve whether the median filter output $Z_{med}$ is impulse output or not. If LEVEL 1 performs filtering operation to find an impulse output then that would make a move to branch to LEVEL 2. Then the size of the window is raised and rerun LEVEL 1 and continues until it finds a median value that is not impulse or the maximum window size is reached, the algorithmic steps return back the value of $Z_{xy}$.

Every time the algorithm outputs a value, the window $S_{xy}$ is moved to the next location in the image. The algorithm is then reloaded and applied to the pixels in the current location. The median value can be renovated iteratively using only the new pixels values, thus reducing the complication overhead.

2.2. SEGMENTATION

In image analysis the problem of robust, proper and optimal data grouping in the form of image data segmentation admitting the most significant stage upon the quality of image analysis systems is primarily dependent. Image segmentation presents process of subdividing the image data into disjoint regions that exhibit within group resemblance according to some prefixed norm and exhibit between group dissemblance. It confers the complex combinatorial problem and exact optimal segmentations are of high computational cost that is practically not attainable in real time applications. So image segmentation is based on some predefined criteria that make them doable.

Clustering can be contemplated to be the most important unsupervised learning problem. It deals with determining a structure in a combination of unlabeled data. The main requirements for a clustering algorithm are scalability, handling with different traits, examining the clusters with an arbitrary shape, high dimensionality, interpretability.
To accommodate all the obligations a Fuzzy k-means clustering algorithm is impended. Basically fuzzy set speculation assumes that data entity do not corresponding to one group, concept or other notion but that they may participate in certain number of groups or concepts. The major objective of K-Means algorithm is to describe K centroids in the multidimensional feature space, one for each cluster.

This centroid should be located as far from each other as feasible. Then each object of the whole image is affiliated with the nearest centroid, where the distances are measured in the multidimensional feature space. When all the points have be assigned centroids, the K centroids are reestimating as the average centres of each cluster.

Then a new assembling has to be done between the like data set points and nearest new centroid. This process is repeated up to no more changes are happening or until a given ultimate number of iterations has been attained.

Finally the algorithm minimizes an objective function. The objective function is

$$ J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(j)} - c_j||^2 $$

Where $||x_i^{(j)} - c_j||^2$ is a distance measure between a data point $x_i^{(j)}$ and the cluster centres $c_j$ is an indicator of the distance of the n data points from their respective cluster centres.

Algorithmic steps:

Step1: Place K points into the space represented by the objects that are being clustered. These points serve as basic group centroids.

Step2: Assign each object to the group that has the nearest centroid.

Step3: When all objects have been assigned, recomputed the positions of the K-centroids.

Step4: Repeat steps 2 and 3 until the centroids no longer more. This produces a segregation of the objects into groups from which the metric to be minimized can be calculated.

2.3. Feature Extraction

We computed the spatial dependency among pixels (i.e. Co-occurrence based texture feature) using the Gray-Level Co-occurrence Matrices (GLCM) defined by Haralick. GLCM is a matrix that shows the frequency of adjacent pixels with greyscale values i and j. For example, let matrix I be the gray scale values of image I, and (i, j) denotes a possible pair of the horizontally adjacent pixels i and j.

$$ I = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 1 & 0 \end{bmatrix} \quad \text{where} \quad (i, j) = \begin{bmatrix} (0,0) & (0,1) & (0,2) \\ (1,0) & (1,1) & (1,2) \\ (2,0) & (2,1) & (2,2) \end{bmatrix} $$

GLCM represents the frequency of all possible pairs of adjacent pixel values in the entire image. For instance, in the GLCM for image I (i.e. GLCMI), there is only one occurrence of two adjacent pixel values both being 0 (i.e. (0, 0)), whereas the frequency of having (0, 2) pixel values in image I is two, so and so forth.
From the resulting GLCM, we estimated the probability of having a pair of pixel values \((i, j)\) occurring in each image (i.e. \(P(i, j)\)). For example, the probability of having a pair of pixel values \((0,0)\) in image I is 1/12, and the probability of having pixels \((0,2)\) is 2/12.

Using \(P(i, j)\) estimated from the GLCM of each animal face, several features can be extracted. Following, we extracted three features GLCM contrast, GLCM homogeneity, and GLCM energy.

\[
GLCM \text{ Contrast} = \sum_i \sum_j (i - j)^2 P(i, j)
\]

\[
GLCM \text{ Homogeneity} = \sum_i \sum_j \frac{P(i, j)}{1 + |i - j|}
\]

\[
GLCM \text{ Energy} = \sum_i \sum_j P^2(i, j)
\]

GLCM contrast measures the variance in gray scale levels across the image, whereas GLCM homogeneity measures the similarity of gray scale levels across the image. Thus, the larger the changes in gray scale, the higher the GLCM contrast and the lower the GLCM homogeneity will be. Finally, GLCM energy measures the overall probability of having distinctive gray scale patterns in the image.

Brightness: Given an image \(I(x, y)\), we defined brightness as its average gray scale value:

\[
c_{\text{mean}} = \frac{1}{n^2} \sum_{x=1}^{n} \sum_{y=1}^{n} I(x, y)
\]

Size: To represent the size of an image, we counted the number of pixels above a threshold \(T\) \((T = 157)\):

\[
p = \sum_x \sum_y B(x, y); \text{ where } B(x, y) = \begin{cases} 0 & I(x, y) < T \\ 1 & I(x, y) \geq T \end{cases}
\]

2.4. Classification

Support vector machine, is a supervised learning technique that seeks an optimal hyper plane to separate two classes of samples. Kernel functions are used to map the input data into a higher dimension space where the data are supposed to have a better distribution, and then an optimal separating hyper plane in the high dimensional feature space is chosen. The Support Vector Machine (SVM) technique is well suited to search for an optimal binary classifier, which classifies the skin cancer as a benign or malignant. The results can be tested for accuracy, specificity and sensitivity.

Their formulas are given in where \(TP\) is the number of true positive pixels, \(FP\) is the number of false positive pixels, \(TN\) is the number of true negative pixels, and \(FN\) is the number of false negative pixels,

\[
\text{Sensitivity} = \frac{TP}{TP + TN}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]
3. SIMULATION RESULTS

**Fig. 2.** Original image (Input)

**Fig. 3.** Filter image

**Fig. 4.** Enhanced image

Accuracy = \( \frac{TP + TN}{TP + TN + FP + FN} \)
4. CONCLUSION

This work presented a novel approach for detecting, segmenting and classifying the skin lesion from digital images. This technique pre-processed the image with adaptive median filter, clustered and segmented with
k-means clustering algorithm and extracting the feature using Gray-level co-occurrence matrix and finally classified with Support vector machine. The output image analysis gives the lesion accuracy.

REFERENCE


