Melanoma Detection using Statistical Texture Distinctiveness Segmentation

Adheena Santy
Dept. of Computer Science & Engineering
Mar Baselios College of Engineering &
Technology
Trivandrum, India

Robin Joseph
Dept. of Computer Science & Engineering
Mar Baselios College of Engineering &
Technology
Trivandrum, India

ABSTRACT

Melanoma is the most dangerous form of skin cancer. It must be detected in the initial stage to increase the survival rates. In medical field, Melanoma detection is usually done by clinical analysis and biopsy tests. These methods are time consuming, expensive and have many side effects. Thus, an automated melanoma detection system is better to assess a patient's risk of melanoma in the initial phase with high accuracy. Existing automated melanoma detection systems make use of thresholding, statistical region merging and Otsu's method for segmentation. These segmentation methods do not include texture analysis, so the accuracy is less. Accuracy of segmentation and melanoma detection can be improved by examining the textural features of skin lesion. Computer aided melanoma detection system using image processing techniques is proposed for accurate and early detection of melanoma. This system has different stages which include preprocessing for image enhancement, segmentation of skin lesion using textural features to improve accuracy, feature extraction and classification. The input image is preprocessed using contrast stretching for image enhancement. The enhanced image is segmented using Texture Distinctiveness Lesion Segmentation (TDLS) algorithm to extract the lesion area from the background skin. Feature extraction is done using graylevel cooccurrence matrix. The system is trained with the extracted features using a good classifier to classify the lesion as malignant or benign melanoma. Accuracy of the proposed system is computed and compared with other segmentation and classification algorithms.

General Terms

Melanoma Detection, Segmentation, Classification.

Keywords

Melanoma; Skin Cancer; Statistical region Merging; TDLS.

1. INTRODUCTION

2. Skin cancer is a major disease that is commonly seen now a day's which occurs due to the abnormal growth of skin tissues. There are various types of skin cancers like Melanoma, Basal Cell Carcinoma and Squamous Cell Carcinoma out of which melanoma is the deadliest form of skin cancer [1]. Skin cancer that forms in melanocytes (skin cells that make pigment) is called melanoma [1]. Incident rates of melanoma have been increasing among non-Hispanic white males and females, but survival rates can be increased if detected early [2]. By extracting the cancerous tissue, melanoma can be cured completely in the initial stage itself. Otherwise it might go to the secondary stage in which recovery becomes impossible. Proper melanoma detection mechanism is needed to detect and diagnose melanoma in the initial stage. Clinical analysis and biopsy tests are commonly used for melanoma detection and diagnosis. Clinical analysis is done using a dermatoscope by trained dermatologists. A dermatoscope is an optical device used by the dermatologists to get a magnified and enhanced view of skin structure using skin surface reflection. Here melanoma is detected based on a visual examination. Visual examination by a dermatologist is based on ABCDE [3] rule where A denotes asymmetry, B denotes border, C denotes color and E denotes evolving structures.

- **A- Asymmetry**: If a line is drawn through this mole, the two halves will be different.

- **B-Border:** For an early melanoma, borders tend to be irregular with scalloped or notched edges.

- **C-Color:** Occurrence of different colors is another warning signal for malignant melanoma. Different shades of brown and tan or black could appear.

- **D-Diameter:** Malignant Melanomas are larger in diameter (1/4 inch or 6 mm), but occurs smaller when detected in initial phase.

- **E-Evolving:** Change in size, shape, color, elevation, or another trait, or any new symptoms like bleeding, itching or crusting.

If any trace of melanoma is found in the initial clinical screening using dermatoscope, the patient is referred for further biopsy tests. Biopsy test is usually done in an outpatient department by a trained pathologist where all or part of the skin lesion will be extracted for further analysis. The clinical analysis and biopsy tests are very expensive and time consuming. Clinical analysis sometimes gives false positive or false negative results which may lead the patient to the secondary stage and further biopsy tests. The biopsy tests also have many side effects like scarring and other health issues. To overcome the disadvantages, an automated system is needed to detect and diagnose melanoma in the initial stage itself so that clinical trials and biopsy tests can be eliminated. More accurate results will be obtained in the initial stage itself without going for expensive dermatological screening and biopsy tests and the treatment can be immediately started if detected early and thereby reducing the death rates. An automated screening system is needed to diagnose a patient's risk of melanoma by using their skin lesion images which can be developed by applying image processing techniques consisting of pre-processing stage, segmentation, feature extraction and classification. Pre-processing stage is used for illumination correction, enhancement and for removing artifacts like hair and other small particles on the skin surface. Accurate diagnosis requires locating the lesion and extracting the lesion area from the background skin. Before extracting features & classifying the lesion, location of the lesion border must be identified. Segmentation methods can be used to

locate the skin lesion from the digital images. Then features from the segmented lesion area can be extracted using different feature extraction techniques. By using the extracted features, lesion can be classified as malignant or benign melanoma using various classifiers. After classification, melanoma images must be categorized as Superficial Spreading Melanoma (SSM), [4] Nodular Melanoma (NM), Lentigo Maligna Melanoma (LLM).

2. LITERATURE REVIEW

In medical field, Melanoma detection is done using clinical screening by a trained dermatologist using the device called dermatoscope and it is done by analysing the ABCDE [3] rule. If any trace of melanoma is found in the initial screening, then the patient is referred for further biopsy tests which are done by a trained pathologist. The conventional methods are time consuming, expensive and have many side effects due to biopsy tests. Lack of trained dermatologists is a major issue in conventional melanoma screening methods. Existing automated melanoma detection systems can be used for dermoscopic images. Segmentation is the important step in automated melanoma detection systems to get an accurate result by analysing only the segmented lesion area.

Existing Melanoma detection systems make use of Statistical Region Merging (SRM) [5], Iterative Stochastic Region Merging [6], Multilevel Thresholding [7] and Color Enhancement and Iterative Segmentation [8]. SRM can be used only for dermoscopic images. It does not consider any texture information and it is noise sensitive. Iterative Stochastic Region Merging method is suitable only for macroscopic images which are robust to noise and artifacts, structural irregularities, illumination variations but the procedure is more complex. This method of segmentation does not consider texture analysis so this method is less accurate. Multilevel Thresholding is suitable for both dermoscopic and digital images but does not include textural information during segmentation and it is computationally more complex. In Color Enhancement and Iterative Segmentation, texture based segmentation is not used. So it is less accurate and complexity is high due to iterative segmentation.

Existing Melanoma detection systems are less accurate and do not include texture analysis for segmentation of skin lesion from the background skin. Texture based segmentation enables accurate segmentation of skin lesion and feature extraction. Classification of the segmented images must be done to accurately classify the images as melanoma or not after extracting various features of skin lesion. Textural based segmentation gives accurate classification results and improves accuracy of the overall melanoma detection system.

3. PROPOSED SYSTEM

Computer aided melanoma detection system using texture based segmentation and classification is proposed in this work. The proposed framework is shown in Fig.1.

This system consists of different modules which include preprocessing, segmentation, feature extraction and classification. Preprocessing is done using histogram equalization. The preprocessed image is segmented using Texture Distinctiveness Lesion Segmentation [2] (TDLS) method to extract the lesion area from the background skin. The extracted lesion area is used for feature extraction which is done by graylevel cooccurrence matrix (GLCM). Using the extracted features, the system is trained using ANN classifier inorder to classify the images as malignant or benign

melanoma. The melanoma images are again categorized as Superficial Spreading Melanoma, [4] Nodular Melanoma, Lentigo Maligna Melanoma.
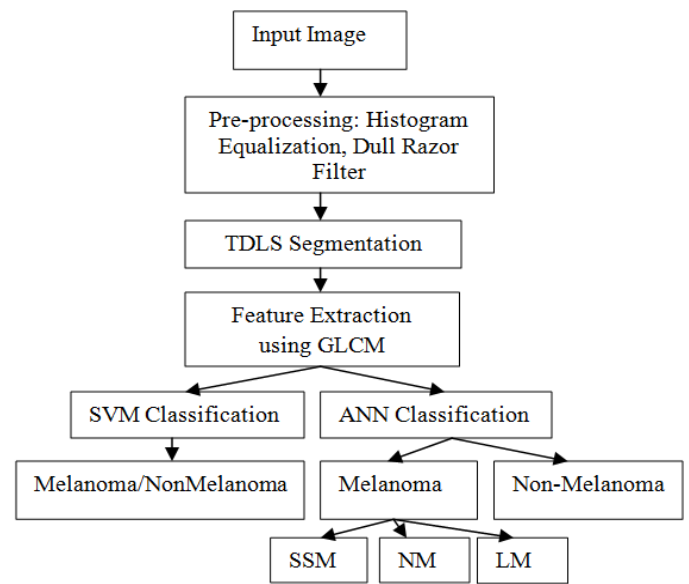

Fig 1: Proposed System

3.1 Preprocessing

The input images of the proposed system can be either dermoscopic images or digital images taken from a standard digital camera. These images may contain contrast variations illumination artifacts and unwanted shading. So these images cannot be used directly for segmentation of the lesion area since it reduces the accuracy of the overall system. Therefore the image must be enhanced before segmentation. Histogram Equalization is used in the preprocessing stage to enhance the image contrast and to produce better quality images. In Histogram Equalization, contrast of the image is enhanced using the image histogram. Image histogram is the graphical representation of the tonal distribution in a digital image in which number of pixels of each tonal value is plotted. In this contrast adjustment, intensities are distributed on the histogram so the low contrast pixel intensities are enhanced to high contrast intensity values by spreading out the most frequent intensity values. The hair particles present in the image are removed using dull razor filter in which morphological closing operation [9] is performed.

3.2 TDLS Segmentation

Segmentation of lesion area from the background skin is the most important step in automatic melanoma detection systems. Segmentation is an image processing method to divide the digital image into different segments which aims to produce the representation of an image into something that is easier for further analysis. The input images contain both skin and lesion area. The lesion area is used for feature extraction and further analysis. If the image contains skin surface, accurate feature extraction cannot be done and it affects the classification results. Inorder to improve the accuracy of the system, lesion area must be properly extracted from the background skin and used for further analysis. Texture Distinctiveness Lesion Segmentation (TDLS) method is used for extracting the lesion from the skin surface.

In TDLS algorithm, the input image in RGB color space is first converted into XYZ color space. For each pixel in the

image, a local texture vector is found out which contains the neighborhood pixels of size *n* centered on a particular pixel of interest *s*. For an image of size N * M, a set of N * M texture vectors are obtained.

$$T = \{\ t_{sj}\ |\ 1 \le j \le N*M\ \} \tag{1}$$

Two step clustering process is used after extracting the texture vectors of each pixel. First k-means clustering is used which divides the image into *k* number of clusters where each cluster contains the different texture data. Next a Gaussian Mixture Model (GMM) clustering is used since k-means clustering does not consider any probabilistic information of the texture. For each cluster, a [2] Gaussian distribution is assumed and model parameters like distribution mean *μ* and distribution covariance Σ are calculated. Next step is to define a metric $l_{j,k}$ given in (2) which represents the similarity of two texture distributions and it is asymmetric since $\Sigma_i \neq \Sigma_j$. It gives the probability that the mean of one texture distribution is a realization of the mean of other texture distribution. Let the mean and covariance of each texture distribution T_j be t_j and Σ_j .The average of $l_{j,k}$ and $l_{k,j}$ gives the measure of similarity $L_{j,k}$ which is defined in (3) [2].

$$l_{j,k} = \frac{1}{\sqrt{|\Sigma j\ |2\Pi^{n \times n \times a}}} exp\left(\frac{-1}{2}\left(t_j - t_k\right)^T \Sigma_j^{-1}\left(t_j - t_k\right)\right) \tag{2}$$

$$L_{j,k} = \frac{1}{2} *(\ l_{j,k} + l_{k,j}\ ) \tag{3}$$

The texture distributions will be different form one another. Next step is to find the distinctiveness of the texture distributions by calculating a metric $d_{j,k}$ defined in (4) which gives the probability of one texture distribution is different from other texture distributions. Let $P(T_k\ /\ I)$ be the probability of occurrence of a pixel [2] being associated with a particular texture distribution T_k. The measure of dissimilarity of a texture distribution T_j from other texture distributions is defined using a Texture Distinctiveness (TD) metric D_j which is given in (5). For normal skin, the dissimilarity of skin texture distributions will be very small so the TD metric of normal skin texture distributions will be small. But the lesion textures are dissimilar from one another so the TD metric will be large.

$$d_{j,k} = 1 - L_{j,k} \tag{4}$$

$$D_j = \sum_{k=1}^{K} d_{j,k}P(T_k|I) \tag{5}$$

The next main step in TDLS algorithm is to find and classify regions in the input image as being a part of the lesion based on the texture distributions and associated TD metric. Statistical Region Merging (SRM) [5] is used to divide the lesion image into number of regions .The two main steps in SRM are sorting and merging of pixels. In sorting step, the pixels in image are sorted to find the order in which pixels are compared. In the merging step, [2] the pairs of pixels are combined to form regions based on their similarity. For each texture distribution, a TD metric is computed and it is combined with the contents of each region to find out a regional TD metric, D_R (6) which represents the average TD of a region *R* where probability of a pixel being associated with the *j*th texture distribution of a region *R* is given as $P(T_k\ /\ I)$.

$$D_j = \sum_{k=1}^{K} d_{j,k}P(T_j|R) \tag{6}$$

After computing the D_R metric, regions in the image must be classified (7) as skin or lesion using the average TD metric D_R. The classification of the regions is done using a threshold

τ which divides the set of texture distributions into two categories, normal skin and lesion. It acts as a decision boundary between the normal skin and the lesion class and it is computed using Otsu's threshold method [2]. The segmentation results are shown in Fig.2.

$$y(R)\ =\ \begin{cases} 1, D_R \ge \tau \\ \\ 0, otherwise \end{cases} \tag{7}$$
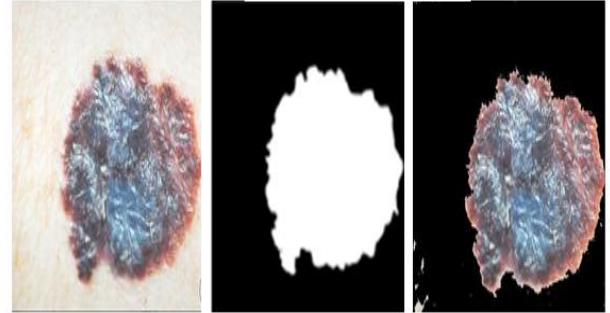

Fig 2: TDLS Segmentation (a) Original image (b) Segmentation map (c) Segmented lesion area

3.3 Feature Extraction

The segmented lesion is used for extracting various features which is given as input for the classification algorithms to classify the lesion as malignant or benign. Textural features such as Angular second moment (ASM), Contrast, Inverse Difference Moment (IDM), Entropy, Variance, Correlation, Cluster shade, [10] Cluster Prominence, Homogeneity, Difference entropy, Dissimilarity and sum average are extracted from the segmented lesion. Gray Level Cooccurrence Matrix (GLCM) is used to extract the textural features of segmented skin lesion. GLCM is a matrix representation of the image where the number of rows and columns in the matrix indicates the number of gray levels in the image. It represents the textural information of the image and it contains the count of how many times a particular texture feature occurs in a particular spatial relation to another texture feature. Sum average gives the average of values in the horizontal direction. Image homogeneity is measured as ASM. Local variations in the image give contrast. IDM also depends on [10] homogeneity of the image. Measure of complexity or disorder of an image gives entropy. Homogeneity measure how close the elements are distributed in GLCM. Joint probability of the specified pairs of pixels is measured as correlation. Image is not symmetric when cluster shade and cluster prominence are high. Dispersion in the [10] gray level difference is measured as variance. Evenness of two groups is measured as dissimilarity. Difference in entropy [10] is the measure of content of the histogram. The set of extracted features are used to train the classifiers.

3.4 Classification

Feature extraction step is followed by the classification of images. The set of extracted features are given as input to the classification stage. Based on the extracted features, the system must classify the images as Melanoma or Non-Melanoma. Melanoma images must be again categorized as SSM, NM, LMM. Two classifiers, Support Vector Machine (SVM) and Artificial Neural Network (ANN) [11] are used in the proposed system. Using the extracted feature set, SVM classifier is trained to classify the images as Melanoma or not.

ANN is also used to categorize the melanoma images as SSM, NM or LMM. Classification stage is shown in Fig.3

SVM is a good binary linear classifier which uses supervised learning techniques. It classifies the data using the best hyperplane which is used to separate the data points that belongs to one class from the other class. The hyperplane which has the largest margin between two classes is chosen as the best hyperplane. After training, SVM classifies the images as melanoma or Non-Melanoma. ANN uses supervised learning technique for classification. Neural Network [11] is made of neurons which are the processing units and the structure of neural network consists of input layer, Hidden layer and Output layer. Feed forward network [12] is used in the system which is trained using Back Propagation (BPN) algorithm. Features Extracted are used to train the network. In BPN algorithm, the signal will flow in the forward direction. First a pattern is presented at the input layer. This pattern is transformed when it passes through network layers till it reaches the output layer. The outputs of the network are then compared with the actual outputs. An error signal is generated if the output does not match with the desired output. The error will be propagated backwards and it is reduced by adjusting the weights. If the procedure is performed [12] once for every pattern and category pair in the data set, 1 epoch of learning is completed. After many epochs, the neural network will be able to remember these pattern-category pairs. After the training phase, neural network can classify the images as melanoma or non-melanoma and the category (SSM, LMM, NM) of melanoma image.
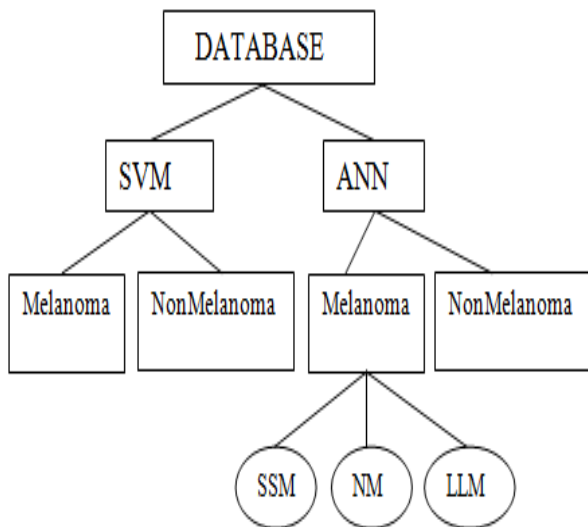

Fig 3: Classification Stage

4. RESULTS AND DISCUSSIONS

In the proposed system, digital images of melanoma and non melanoma were collected from internet and hospitals. Preprocessing is done to enhance the quality of all the images. TDLS algorithm is implemented to segment the lesion area from the collected images. Features are extracted from the segmented lesion area and they are given for classification of melanoma images. SVM and ANN classifiers were used to classify the images as Melanoma or Non-Melanoma. The system is implemented in MATLAB on a computer with an Intel Core i3-380M CPU (2.5 GHz, 4GB RAM). For performance analysis, Feature Extraction is done for both raw images and segmented images and then given to train the classifiers. 12 features are extracted and given to train the

classifier .Classification done using features extracted from the segmented image gives better accuracy than classification done using features from raw images without segmentation. A set of 126 images from the Dermquest database [13] were used for training and testing the classifiers and classified as Melanoma and Non-Melanoma. TDLS algorithm is compared with 3 lesion segmentation algorithms such as L-SRM [5], Otsu-R [14] and Otsu-RGB [15]. Performance of the segmentation algorithms are compared by calculating accuracy and the formula is given in (8) where TP denotes number of true positive pixels, FP denotes the number of false positive pixels, TF denotes the number of true negative pixels, and FN denotes number of false negative pixels [3].

$$Accuracy = (TP+TN) / (TP+FN+TN+FP) \qquad (8)$$

Table 1. Accuracy Results Of Segmentation

Segmentation	Accuracy
L-SRM [5]	92.3%
Otsu-R [14]	84.9%
Otsu-RGB [15]	80.2%
TDLS Segmentation	**96.8%**

Table 2. Performance Comparison Of Classification After TDLS Segmentation

Classifier	Accuracy
TDLS with ANN	90.63%
TDLS with SVM	**96.87%**

5. CONCLUSION

Melanoma detection in the initial stage is very important since it can reduce the death rates to a great extend. Early detection can be done only with the help of trained dermatologists and proper diagnosis. Due to the lack of trained dermatologists, expensive and time consuming diagnosis procedures, automated system for melanoma detection is required. An automated melanoma detection system using texture based lesion segmentation, feature extraction using GLCM and classification of images using SVM and ANN classifiers is implemented. Using ANN classifier, melanoma images are again classified into 3 categories namely Superficial Spreading Melanoma, Nodular Melanoma and Lentigo Maligna Melanoma. This proposed system gives better diagnosis and accuracy than conventional clinical screening and biopsy tests since it makes use of texture based analysis and classification. The proposed framework gives 90.63% accuracy on ANN and 96.87% accuracy on SVM. This automated detection system can be used for both dermatological images and digital images taken by a standard camera. So the dermatological screening can be eliminated. By extracting more features from the segmented image, the accuracy of the overall system can be further improved and these feature set can be used to classify the images to find out the stages of melanoma skin cancer.

6. REFERENCES

[1] Details of skin cancer. [Online]. Available at: http://www.cancer.gov/cancertopics/types/skin

[2] J. Glaister, A. Wong, and D. A. Clausi, "Segmentation of Skin Lesions From Digital Images Using Joint Statistical Texture Distinctiveness," IEEE Trans. Biomed. Eng., vol. 61, no. 4,pp. 1220–1231, Apr. 2014

[3] Melanoma Available at: http://www.skincancer.org/skin-cancer-information/melanoma

[4] Types of Melanoma. [Online]. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3253944.

[5] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies, "Border detection in dermoscopy images using statistical region merging," Skin Res. Technol., vol. 14, no. 3, pp. 347–353, 2008.

[6] J. Glaister, A. Wong, and D. A. Clausi , " Automatic Skin Lesion Segmentation via Iterative Stochastic Region Merging," IEEE trans. Biomed. Eng.,vol15,No.6,Nov 2011

[7] J.Humayun , A.S.Malik , N.Kamel , "Multilevel thresholding for segmentation of pigmented skin lesions, " IEEE Int. Conf. Imaging Systems and Techniques, pp. 310-314,May. 2011

[8] G.Schaefer , M. Rajab , M. Emre Celebi and Hitoshi Iyatomi , "Skin lesion extraction in dermoscopic images based on colour enhancement and iterative segmentation," ICIP ,IEEE Journal,pp.3361-3364,2009

[9] Dull Razor Filter. [online]. Available at: http://www.dermweb.com/dull_razor/

[10] N.Hema Rajini , R.Bhavani, "Computer aided detetcion of ischemic stroke using segmentation and texture features" ELSEVIER, Measurement 46, pp.1865-1874,2013

[11] Guoqiang Peter Zhang, "Neural Networks for Classification", IEEE Transactions On Systems, Man, And Cybernetics-Part C: Applications And Reviews, Vol. 30,2000

[12] Feed Forward Network. Available at: http://www.fon.hum.uva.nl/praat/manual/Feedforward_neural_networks_1__What_is_a_feedforward_ne.html

[13] DermQuest Image Database [online]. Available at: https://www.dermquest.com/results/?q=Malignant%20melanoma

[14] P. G. Cavalcanti, J. Scharcanski, andC. B. O. Lopes, "Shading attenuation in human skin color images," in Advances in Visual Computing, G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. Chung, R. Hammoud, M. Hussain, T. Kar-Han, R. Crawfis, D. Thalmann, D. Kao, and L. Avila, Eds., (ser.Lecture Notes in Computer Science), vol. 6453 Heidelberg, Germany: Springer, 2010, pp. 190–198.

[15] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," Comput.Med. Imag. Graph., vol. 35, no. 6, pp. 481–491, Sep. 2011.