Automatic Detection and Analysis of Melanoma Skin Cancer using Dermoscopy Images

Mahmudul Hasan, Mohammad Mohsin, Md. Kamal Hossain Chowdhury

Abstract: Skin cancer is known as one of the most risky types of cancer. Several kinds of skin cancer, such as melanoma, basal and squamous cell carcinoma, etc., are available. The most unpredictable cancer is melanoma. If we can detect melanoma skin cancer at an early stage, the chances of recovery will be good and we can save many valuable lives. But if we fail to detect early, melanoma can disperse to the different parts of the body and chance of recovery will become difficult. This research presents a developed system to do melanoma diagnosis by using several dermoscopy images. In this research, we preprocessed the images to remove hairs and noises by using some filter techniques such as dull razor technique, median filtering, etc. After that, we segmented the image to find the infected area using some segmentation method and we choose the method that will give us the best results. Then we post-process the images and choose the most infected lesion. After segmentation of the skin lesion, we checked the segmentation accuracy concerning some basic criteria. We compared the segmented skin lesions with the marked skin lesions by a dermatologist. Then we extracted the features of the images of different criteria, such as Asymmetry, Border irregularity, Color variance, Diameter which have the acronym as ABCD. We also analyzed the texture of the lesions and extracted the geometrical features. Finally, we choose decision tree classification methods that gave us the best results.

Keywords: Feature extraction, Image segmentation, Object recognition, Pattern clustering, Pattern matching.

I. INTRODUCTION

Skin cancer is widely known as one of the most common types of cancer. By building blocks of tissues cancer cells development begins. Human body is the combination of different cells. Cells normally grow and develop when it is needed. After growing older or being damaged, the cells normally die and replaced by the new cells. But sometimes it may be imbalanced. New cells may be formed but the body doesn’t need them. On the other hand, the old cells become older day by day but not die as they should be. So, body cannot replace the old cells by new and developed extra new cells create an imbalance in the organs which is known as tumor. Cells growth on the skin is benign which isn’t responsible for cancer and malignant which causes melanoma. Malignant growths are so much harmful than benign growths. It may cause infection on the other parts of the body and damage organs and tissues. Among skin cancers, malignant melanoma is one of the most well-known types. It is one the most fatal type of skin cancer. The outbreak is considerably increasing. Though less than one percent of skin cancer is accounted as melanoma cancer, the vast majority of deaths are cause by this cancer type. Every year almost 68000 Americans are diagnosed with melanoma [1]. Every hour (approximately 54 minutes) one person dies on account of melanoma [2]. The World Health Organization approximates that every year more than 70230 people die from too much sun, mostly melanoma skin cancer [3]. Risks of melanoma cancer double with the probability of sunburns [4]. The fatality rate of men is more than women [5]. Melanoma is particularly deadly than all types of skin cancer. Though only 4 percent of all skin cancer is caused by melanoma, 75% of skin cancer death is the cause of melanoma [6]. Early detection and treatment has a higher probability of recovery. If we failed to detect early it may spread and once melanoma spread. Dermoscopy is also known as Dermatoscopy. By using non-invading diagnosis technique like light microscopy can be used to detect pigmented skin lesion. It is a new type of imaging technique which can use to examine skin lesion attached with a piece of equipment called dermatoscopy. Although analysis of dermoscopy images plays an important role to detect malignant melanoma in the early stage, this traditional method is subjective and time-consuming. Due to these limitations, computer aided diagnosis system is an urgent need for the dermatologists for the clinical evaluation to detect early the risk factor of melanoma.

Even though computer is not so much intelligent like human, it may be able to extract information accurately and quickly which may not be realized by human eyes. To analyze dermoscopy images several algorithms such as the seven-point checklist, ABCD rule, and the Menzie’s method can be used. [7] and it can improve the diagnostic performance of the clinicians. Different segmentation process are used in the most of the proposed techniques that may cause a problem due to skin lesion irregularity. A variety of patterns is shown in the skin in the dermoscopy view. It may be difficult to detect different patterns by automatic segmentation, and cannot be done in an ordinary approach. So, automatic detection and analysis of melanoma skin cancer are needed.

Figure 1: Skin lesions: (a) Benign (Not cancer) (b)Malignant

Revised Manuscript Received on September 15, 2019

* Correspondence Author
Mahmudul Hasan, Asst. Prof., Dept. of CSE, Comilla University, Comilla, Bangladesh. Email: mhasanraju@gmail.com
Mohammad Mohsin, Dept. of CSE, Comilla University, Comilla, Bangladesh. Email: mohssinn009@gmail.com
Md Kamal Hossain Chowdhury, Asst. Prof., Dept. of CSE, Comilla University, Comilla, Bangladesh. Email: kamal.cou@gmail.com

Retrieved Number: C4561098319/190BEIESP
DOI:10.35940/ijrte.C4561.098319

Published By: Blue Eyes Intelligence Engineering & Sciences Publication

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019
II. LITERATURE REVIEW

Stocker et al. [15] proposed a formula to compare the results of the formula with a certain threshold. If the calculated pixel value was larger than the certain threshold, the lesion would be classified as Malignant cancer or non-cancer (benign). They achieved a sensitivity of 80%. But new lesion could not be classified by the experiments. Moreover, Erkal et al. [16] proposed a classifier where it extracts the features of asymmetry, border irregularity, and color features. After that, the classifier trains a feed-forward neural network using these extracted features. But due to the limited number of extracted features, they only achieve 70-80% classification accuracy. On the other hand, Abbas et al. [17] used a new approach for skin lesions segmentation. At first, they convert the RGB dermoscopic image into CIE Lab color space. Then a region of interest (ROI) is detected by using a hill-climbing algorithm of color channels (L, a, b). After that the lesion area is detected by adaptive threshold method. But the computational load of an adaptive threshold is high that is why it should not use where time is an important factor. Besides, Jaleel et al. [18] proposed a classifier that extracts features of the image from the 2-D wavelet transformation. But they use a few images as a sample. Moreover, they did not mention the sensitivity or accuracy of the system. Even though Erkol et al. [19] described a gradient vector flow (snake method) for skin lesion segmentation, it performs poorly when the boundary is not well defined. Similarly, Naheed R. Abbasi, Helen M. Shaw, et al. [20] proposed ABCDE method which uses asymmetry, border irregularity, color variegation, diameter > 6 mm features to detect the medical properties. They include “Evolution” medical properties but there should not be any change in the lesion at that time.

Another similar work is done by Stanley et al. [21]. The paper proposed a skin lesion determination algorithm by using color histogram analysis technique. Though it is helpful for early prediction, only one color histogram based work does not workable in different color changes in advance stages. Amran et al. [22] also proposed a technique that extracts lesion features. They did not mention sensitivity, accuracy. Moreover, they had no way of learning from experiences. Additionally, Aurora et al. [23] describes the characteristics of melanoma thickness and provide a way to distinguish from benign lesions. M. Emre celebi et al [24] describes all these research methods used in this filed in recent years and gave guidance for future image analysis. Gniadecka et al. [25] described a classifier that used neural networks. They used Raman spectrometers to examine skin lesions. But Raman spectrometers are very costly and dermatologists do not use it at all. Moreover, Jeffrey et al. [26] proposed a technique for segmentation of skin lesion using joint statistical texture distinctiveness. They compare the results of these methods to other segmentation algorithms. Besides, Othman et al. [27] proposed a single parametric fuzzy algorithm. Different types of filtering techniques are used.

Perreault et al. [28] proposed a median filtering technique that filters images in constant time complexity. Foirese et al.[29] created a tool to remove hair from the dermoscopic image. They used a top-hat filter to identify individual hairs and then performed a morphological operation. Moreover, Nayaz, Chaya et al.[30] used the Weiner filter to remove noise from the images. Weiner filter is useful when the input image has lower variance even if it minimizes the mean square error. They used a distance regularized level set method for segmentation. Gray Level Co-occurrence Matrix (GLCM) can be used to extract features from the image. They used SVM for classification. Gopinath et al. [31] used the Gabor filter for extracting features from the images. It is a linear filter for edge detection in input images. They use Support Vector Machine and Artificial neural network for classification.

Pablo, Jacob et al. [32] used a novel method for skin lesion segmentation. This method used a Level Set segmentation method which can be used when the boundaries are well defined and do not stick together. An average segmentation error of 16.65% is showed by them. Sadri et al. [33] proposed a novel method for the segmentation of skin lesion based on the wavelet network. Network inputs are the (R, G, B) values of a dermoscopy image. Then, the lesion area is segmented.

III. METHODOLOGY

Image Processing techniques are used in the proposed methodology for Automatic Melanoma Skin Cancer Detection which is shown in Fig. 2.

- Input Image
- Hair removal
- Noise Removal
- Post Processing
- Classification
- Feature extraction
- Segmentation
- Enhancement

Our proposed system consists of the following phase.

- Preprocessing
  - Hair Removal
  - Noise Removal
  - Contrast Adjustment
- Segmentation method
  - Otsu method
  - K-means method
  - Fuzzy c-means clustering
- Post processing
- Feature extraction
  - Asymmetry
  - Border irregularity
  - Color features
  - Diameter
3.1 Preprocessing
Since the image given to the system may be influenced by many lightning conditions. So, before the segmentation steps, the given image needs to be pre-processed. Preprocessing steps include image resizing, hair removal, noise removal, contrast enhancement, etc. Pre-processing steps try to improve the image quality by removing unwanted effects. Normally, it combines three steps:

3.1.1 Removal of hair from images
In our proposed method, we used a dull razor filter for the given images. The basic idea behind this filtering technique is that it replaces the hair pixels by neighboring pixel intensity values.

The procedure of dull razor filter:
Step 1: Gray scale morphological closing operation is used to locate the dark hair pixel locations.
Step 2: Then it verifies the thin and long structured hair pixels.
Step 3: After verification, a bilinear interpolation method is used to replace the verified pixels.
Step 4: Finally, it applies an adaptive filtering technique to smooth the replaced hair pixels.

3.1.2 Removal of noise from images
After hair removal, still some noise may present in the image. Some forms of noise are globules, air bubbles. To remove such noise, a median filter has been used. Different noises such as impulse noise, salt and pepper noise etc. can be removed by median filter. Image edges can be preserved by reducing random noise in an image. Influence of small structures like thin hairs and different isolated islands of pixels can be minimized by using median filter. The median filter makes smooth the images and is thus useful in reducing noise.

The procedure of Median filter:
Step 1: Choose an image with specific window size (say 3*3)
Step 2: Compute the median intensity value of the specified window.
Step 3: Replaces the median intensity value that are being calculated within the window.

3.1.3 Contrast Enhancement
In this step, we adjust the image to remove unwanted parts of the image. Contrast enhancement technique is used to process an image so that it can produce good results for a specific application. Contrast enhancement can sharp the image border. By enhancing image we can get better segmentation results.

3.2 Segmentation
Segmentation is the crucial part of image processing. It is the process of partitioning an image into multiple segments. It is very useful to detect the regions of interest. Sometimes segmentation is very difficult because of the great varieties of lesion sizes and textures, smooth transition between the skin and lesion border. Several algorithms have been used. Some of them are threshold-based, some of them are region-based, edge-based, etc. In this research, three methods of segmentation have been discussed. The methods are:

- Otsu’s method.
- K-means method.
- Fuzzy c-means clustering method.

3.2.1 Otsu’s method
Otsu method is one of the popular segmentation methods which calculate the optimum threshold by separating the classes. For this reason, their interclass variance is minimal and interclass variance is maximal.

Let, given an image gray levels L {0, 1, 2, 3…L}. The Otsu method divides the image intensity values into two classes. Let, 
\[ C_{0} = \{0, 1, 2, 3, \ldots t\} \]
\[ C_{t} = \{t+1, t+2, t+3, \ldots L-1\} \]

Algorithm:
Step 1: First it computes histogram and find out intensity level probability for each pixel.
Let, \[ n_{th} \] gray level pixels, \[ n = \text{total number of pixel values} \]

The probability of gray level, \[ P_{i} = \frac{n_{th}}{n} \] and \[ \mu_{th}(O) \]
Step 3: For all possible threshold value $t$ from 1 to maximum intensity level.
1. Update weight $w_i$ and mean $\mu_i$ using the following formula.

$$w_i(t) = \frac{1}{\sum_{t=1}^{T} P(t)}$$

$$\mu_i(t) = \frac{1}{\sum_{t=1}^{T} t^{-1} P(t)}$$

2. Compute variance using the following formula

$$\sigma_i^2(t) = w_i(t) \frac{1}{\sum_{t=1}^{T} \mu_i(t) - \mu_i(t)^2}$$

Step 4: Final threshold values are the maximum variance value, that means

$$t = \text{Arg} \{ \max_{0 \leq i \leq 1} \sigma_i^2(t) \}$$

Algorithm:

Let, $X = \{x_1, x_2, ..., x_n\}$ be the set of pixel values.

$C = \{c_1, c_2, ..., c_k\}$ be the set of centers.

Step 1: First, it selects $k$ cluster centroid randomly.

Step 2: Calculate the fuzzy membership value using the following formula.

$$\mu_{ij} = 1 / \sum_{k=1}^{C} \left( \frac{d_{ij}}{d_{ik}} \right)^{2/m}$$

Where, $d_{ij}$ = Euclidean distance between $i$ and $j$ and $m$ = fuzziness index.

Step 3: Compute the centroid $C_j$ using the following formula.

$$C_j = \left( \sum_{i=1}^{m} (\mu_{ij})^m x_i \right) / \left( \sum_{i=1}^{m} (\mu_{ij})^m \right) \forall j = 1, 2, 3, ..., C$$

Step 4: if $\| U^{k+1} - U^k \| < \beta$ then stop. Otherwise, go to step 2.

3.2.2 K-means method

K-means clustering algorithm makes group color pixels values that have like intensity values based on the distance between them. It may be computationally faster than hierarchical clustering. However, k-means clustering requires a predefined set of value, $k$.

Algorithm:

Step 1: Set the number of cluster values (Say k)

Step 2: Choose the K cluster centroid randomly.

Step 3: Measures the distance between each pixel to the center of the clusters.

Step 4: If the distance is near to a particular cluster then move the pixel values to those clusters.

Step 5: Otherwise move to another cluster.

Step 6: Recalculate the cluster centroid.

Step 7: Repeat the process until the centroid of each cluster remains the same.

3.2.3 Fuzzy C-means clustering method

Fuzzy C-means clustering allows each pixel values to belong to more than one cluster. It works considering the distance between the cluster center and data points by assigning each point membership function. So, a high degree of membership will be assign to a data point that is closer to the center of a cluster. On the other hand, far away data points will have a lower degree of membership from the center of the cluster.

Algorithm:

Let, $X = \{x_1, x_2, ..., x_n\}$ be the set of pixel values.

$C = \{c_1, c_2, ..., c_k\}$ be the set of centers.

Step 1: First, it selects $k$ cluster centroid randomly.

Step 2: Calculate the fuzzy membership value using the following formula.

$$\mu_{ij} = 1 / \sum_{k=1}^{C} \left( \frac{d_{ij}}{d_{ik}} \right)^{2/m}$$

Where, $d_{ij}$ = Euclidean distance between $i$ and $j$, $m$ = fuzziness index.

Step 3: Compute the centroid $C_j$ using the following formula.

$$C_j = \left( \sum_{i=1}^{m} (\mu_{ij})^m x_i \right) / \left( \sum_{i=1}^{m} (\mu_{ij})^m \right) \forall j = 1, 2, 3, ..., C$$

Step 4: if $\| U^{k+1} - U^k \| < \beta$ then stop. Otherwise, go to step 2.

F1 Score:

F1 score also known as F-measure to measure segmentation accuracy.

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{Total population}}$$

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu method</td>
<td>92.2%</td>
</tr>
<tr>
<td>k-means method</td>
<td>88.4%</td>
</tr>
<tr>
<td>Fuzzy c-means</td>
<td>94.4%</td>
</tr>
<tr>
<td>clustering</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: F1 measures for segmentation methods

3.3 Post Processing

After segmentation, the segmented images should be processed again to remove some unnecessary regions, filling holes inside the regions, the morphological operation for sharpening the image border.
3.4 Feature Extraction

Feature extraction is one of the important phases in image processing. It can be categorized as color features, shape features, and texture analysis. These extracted features help us to develop a system that can properly distinguish melanoma skin cancer from non-skin cancer (benign).

3.4.1 Color features

Melanoma skin cancer has a variety of colors. Normally, it may have 6 colors: black, red, white, blue-gray, light brown, dark brown. So, color is an important feature in the detection of melanoma skin cancer. We used the following algorithm for detecting the number of colors in an image.

Algorithm:
Step 1: First, segment the image for getting the binary image.
Step 2: Set the complement of that binary image.
Step 3: Then mask with the original RGB images.
Step 4: Overlay with a green outer mask to the resulting image.
Step 5: Calculate the percentage of each color and if the percentage is above from the certain percentage set a score 1 for that color.

3.4.2 Shape Features

To calculate lesion in a binary image different shape features are used which are shown below:

3.4.2.1 Irregularity

Border irregularity is one of the discriminant features of melanoma skin cancer. Melanoma lesions have an irregular border with sharp edges. Benign has a smooth transition between the skin and the lesion border. So, border irregularity is also an important feature. There are different ways to measure border irregularity. One of them is the compact index. The compact index can be calculated using the following formula.

$$\text{Compact index} = \frac{P_L^2}{4\pi A_L}$$

Where, \( A_L \) = Lesion Area,
\( P_L \) = Lesion Perimeter.

The irregularity index tends to be 1 for a perfect circle. For irregular regions, the irregularity index tends to 0.

3.4.2.2 Asymmetry

If first half of an image is equal to the other part of the image, then the image is called symmetric object. For a completely asymmetrical shape, the asymmetry score is 1. For a symmetric pattern, the asymmetry ratio is closer to 0.

Algorithm:
Step 1: Rotate the segmented image along with the major axis.
Step 2: Translate the image.
Step 3: Obtain the flipped image of the translated image.
Step 4: Calculate the intersection of the two images.
Step 5: Calculate the union of the two images.
Step 6: Calculate the ratio of two images obtained from steps 4 and 5.
Step 7: Repeat from step 1 for the major axis and again calculate the ratio of intersection and union of the image along with the major axis.
Step 8: Final output values are the ratio of two values obtained from the major and minor axis.

3.4.3 Texture Features

Texture features of melanoma cancer image shows better discriminant features compared to other features. Contrast, Energy, Entropy, Homogeneity are the texture features that can be computed from the GLCM.

Contrast measures the unevenness of texture, local
Automatic Detection and Analysis of Melanoma Skin Cancer Using Dermoscopy Images

intensity variation. It can be calculated using the following formula:

\[
\text{Contrast} = \sum_{i,j=0}^{N} P_{ij} (i - j)^2
\]

Where \( N \) = number of the total pixels and \( P_{ij} \) = pixels at locations (i,j).

Energy measures the image textural uniformity. The following formula can be used to calculate it:

\[
\text{Energy} = \sum_{i,j=0}^{N} P_{ij}^2
\]

Entropy measures the degree of uncertainty. The basic formula of entropy is as follows:

\[
\text{Entropy} = \sum_{i,j=0}^{N} -\ln (P_{ij})\]

Homogeneity measures the distribution of elements. It can be calculated as follows:

\[
\text{Homogeneity} = \frac{\sum_{i,j=0}^{N} P_{ij}}{1 + (i - j)^2}
\]

IV. CLASSIFICATION

We use a decision tree classifier. Performance can be evaluated by using the expression of sensitivity, specificity, and accuracy.

**Sensitivity:**

It correctly measures the positive ratio. It is computed as follows:

\[
\text{Sensitivity} = \frac{TP}{(TP+FN)} \times 100\%
\]

**Accuracy:**

It measures the correctness of a diagnostic test. It is computed as follows:

\[
\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\%
\]

**Specificity:**

It is the opposite of sensitivity. It correctly measures the negative ratio. It is computed as follows:

\[
\text{Specificity} = \frac{TN}{(TN+FP)} \times 100\%
\]

Where FN is False Negative that means the diagnosis of positive cases and FP is False Positives to diagnosed negative cases.

TP is True Positives that means the diagnosis is correctly diagnosed positive cases;

TN is True Negative that means the diagnosis is correctly diagnosed negative cases.

<table>
<thead>
<tr>
<th>Image Classified</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2: Training performance measurement

In the training dataset, the melanoma and benign accuracy, sensitivity and specificity are all 100% (Table 2) whereas in the testing performance measurement Sensitivity, Specificity and Accuracy is 94.5%, 92.8 and 91.1% for Melanoma and 89.6%, 95.6% and 92.4% for Benign respectively.

<table>
<thead>
<tr>
<th>Image classified</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma</td>
<td>94.5%</td>
<td>92.8%</td>
<td>91.1%</td>
</tr>
<tr>
<td>Benign</td>
<td>89.6%</td>
<td>95.6%</td>
<td>92.4%</td>
</tr>
</tbody>
</table>

Table 3: Testing performance measurement

V. CONCLUSION

Melanoma skin cancer has been rising for the last twenty years. In the area of melanoma cancer research, different techniques are presented to detect and categorize skin cancer based on the digital image. But most of the works have done to detect whether the mole is cancer or not at a late stage. So, early detection of melanoma skin cancer is of foremost importance now. If we can develop a system that can detect melanoma at an early stage, the skin has one of the highest cure rates and the treatment will be quite simple and will involve excision of the lesion. Moreover, if we can detect early it will remove the necessity of unnecessary biopsy which is economical. On the other hand, at a late stage, it may result in fatal consequences and extremely high costs associated with the necessary treatments. But it will be a challenging task to predict early the risk of melanoma. It is a quite difficult task to determine whether in future a mole will turn to a melanoma skin cancer or not. Some benign lesion may have some features which are very much similar to malignant lesions. On the other hand, some malignant melanoma may also have some confusing features such as they may be symmetric or they may contain smooth border or has few color variation. For that reason, the feature extraction from dermoscopy images sometimes it may be difficult to predict. So, good research is highly needed in this domain. In this research, we proposed new approaches for an automatic system for early skin cancer (melanoma) detection.

ACKNOWLEDGMENT

We express our deep gratitude to Comilla University for facilitates this research work. Without their financial support this research work cannot be done.

REFERENCES

10. Thieu, Luong, Rocchisani, and Viennet, “A convex active contour region-based model for image segmentation,” in...


35. Zhen Ma, Joao Manuel, Tavares ”Segmentation of Skin lesions using level set method”.


37. Sumithra, Mohammad Shuil, Guru,” Segmentation and classification of skin lesion for Disease diagnosis”


AUTHORS PROFILE

Mahmudul Hasan is currently working as a faculty member (Assistant Professor) in the Department of Computer Science and Engineering at Comilla University, Comilla, Bangladesh. His research interest is in Machine Learning, NLP and Speech Signal Processing.

Mohammad Mohsin Completed his Bachelor of Engineering and Master of Science degree from the department of Computer Science and Engineering of Comilla University, Bangladesh. He was a top level ACM coder and participated many international programming contests.

Md. Kamal Hossain Chowdhury is an Assistant Professor of CSE dept. of Comilla University. His research interest is in NLP and Machine learning.