Detection And Classification Of Human Malignant Melanoma Using Cad Approach

Akila Victor, Muhammad Rukunuddin Ghalib, VIT University, Vellore

Abstract: Cancer is one of the dreadful diseases in today’s world. Skin cancer is one among those. Manual methods have been available, still a contribution towards computer aided diagnostic in the field of medicine plays a challenging role. The objective of the paper is to detect and classify the skin image and find if it is a melanoma or non melanoma images. This paper includes the method for image enhancement as adaptive median filter, segmentation as a hybrid method and the feature extraction to be a hybrid method and finally the KNN classifier is used to classify the images. The total of 600 images is taken and the efficiency is said to be 96.8%.

Keywords: Adaptevmedianfilter, Hybrid, KNN

1. INTRODUCTION
Cancer is highly complex to and is of various types’ internal cancers and external cancers. There can be n number of cancer types out of which few are named here. They are skin cancer, oral cancer, breast cancer, thyroid cancer, stomach cancer, colon cancer, ovarian cancer, throat cancer, lung cancer, bladder cancer, cervical cancer, colorectal cancer, liver cancer, gall bladder cancer, bone cancer, pancreatic cancer, thyroid cancer, blood cancer and uterian cancer. skin cancer can be of various types as basal cell carcinoma, squamous carcinoma and melanoma. The skin cancer occurs when there is an abnormal or extended growth in the cells in the layers. Sometimes the skin cancer can develop from an existing mole. Melanoma is the most life threatening disease. Melanoma is the most dangerous cancer out of the three which actually can spread to all other parts of the body as well. It occurs exposing oneself into too much of sunlight. Melanoma or in general skin cancer can occur in any part of the body so there are no specific area to be mentioned for the occurrence of skin cancer in one’s body. The stages can be of primary stage and secondary stage. Primary stages are curable whereas a secondary stage is quite hard to handle.

1.1 Melanoma
The most dreadful form of skin cancer. When not treated properly can cause death. It is the most furious of all skin cancers. Melanoma develops from the melanocytes. In men they occur in back neck and other area. For women it occurs on legs. Rarely does it occur on mouth for anyone. Have to be very careful with the moles. The main cause is too much exposure to sunlight. So its always better to be away from sunlight that affects us largely and that is one of the way to wear sunglasses and so on

1.2 Dermoscope
The dermoscope is an instrument which is used to capture the dermoscopic images. The dermoscope is a non instrument device.

The input image uses the image captured from the dermoscope. The figure shows the clinical view and the dermoscopic view of the image. The dermoscopic image is the input image for the entire process.

Figure 1 The clinical view and the dermoscopic view (courtesy web)

2. EXISTING SYSTEM
The CAD systems can be relatively giving much contribution rather the manual method. We are going to review on the basic methodology that a CAD could use such as Image acquisition, image preprocessing, image segmentation, image feature extraction and classification. The next stage of image processing is the image enhancement phase that is to glow the image or to make the image visibly good than the input images. The next level for computer aided diagnostic design and classification is segmentation. The segmentation explains about splitting the region into many parts and finding out the region of interest. The Feature extraction is something which helps in finding out how the features can be extracted from the image. It reduces the redundant information that is in an image. The size can be reduced as well. Classification can be as to classify if the desired result is yes or no. The skin image classification helps to find if the input image is melanoma or not. The classification is the final step which helps us to detect the accuracy. Preprocessing that helps as a initial screening phase. Table 1.2 analyzes the preprocessing done in recent years. That helps in reducing noise. Processing of the input image before any further process is done. The preprocessing method defines on edge detector used to calculate the first order derivative (Eltayef et al. 2017) and stretching the image to two extreme range to normalize (Munya et al. 2016) and making multiple image into single image that is by image
registration that can include rotation skewing etc (Sharma et al. 2016), and by using the various operators of morphology (Jafari et al. 2016) and then to enhance the low contrast images the adaptive sigmoidal function (Fazli et al. 2016) are used as the preprocessing methodologies. The Region based segmentation is used to extract the region of interest that are needed for segmentation (Gogoi et al. 2016). Then the segmentation helps in detecting cells, filling up the gaps, and reducing the gap, extracting the border and enlarging the image (Wadhwani et al. 2017). And then the clustering algorithm for grouping of the pixels and the Markov random model using random assignment is obtained to segment images (Eltayef et al. 2017). Gabor wavelet filter (Hu et al. 2016) has the ability to model the orientation so as to improve the accuracy. The vertical and the horizontal overlap can be used in local binary pattern extraction (Singh et al. 2016). The statistical features, color (Agaian et al. 2016), shape (Sharma et al. 2016), intensity, texture (Munya et al. 2016, Sumithra et al. 2015) features are also concentrated. There are various classification algorithm defined to predict the accuracy in an image. And so Classification using the convolution neural network (Li et al. 2017, Yu et al. 2016) Artificial neural network (Sharma et al. 2016, Antony et al. 2016, Sasikala et al. 2005) Back propagation network (Abdul et al. 2013) works with a input, weight and the bias to produce a classification. The clustering using the fuzzy algorithm (Munya et al. 2016) also works on classification.

3. METHODOLOGY

The adaptive median filter as the name says is adaptive in nature. The mask of the adaptive median filter is adaptive in nature. The adaptive median filter can be explained by an algorithm as follows

Stage 1
\[ H1 = \frac{F_{\text{median}} - F_{\text{min}}}{F_{\text{max}} - F_{\text{min}}} \]
If \( F_1 > 0 \) and \( F_2 < 0 \) GOTO stage 2
Else increase the filter size
If filter size \( < S_{\text{max}} \) repeat Stage 1
Else output \( F_{\text{median}} \)

Stage 2
\[ G1 = \frac{F_{\text{xy}} - F_{\text{min}}}{F_{\text{max}} - F_{\text{min}}} \]
\[ G2 = \frac{F_{\text{xy}} - F_{\text{max}}}{F_{\text{max}} - F_{\text{min}}} \]
If \( G1 > 0 \) and \( G2 < 0 \), display \( F_{\text{xy}} \)
Else display \( F_{\text{median}} \)

\( S_{\text{xy}} \) is the support of the filter centered at \( x, y \).

### Table 1 The PSNR, SNR, MSE of the images for which the adaptive median filter is applied.

<table>
<thead>
<tr>
<th>Input Image</th>
<th>PSNR</th>
<th>SNR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.8177</td>
<td>28.2152</td>
<td>52.1294</td>
</tr>
<tr>
<td>2</td>
<td>34.169</td>
<td>29.8256</td>
<td>24.2677</td>
</tr>
<tr>
<td>3</td>
<td>31.9478</td>
<td>27.7946</td>
<td>26.7344</td>
</tr>
<tr>
<td>4</td>
<td>30.2266</td>
<td>26.05</td>
<td>30.4705</td>
</tr>
<tr>
<td>5</td>
<td>32.6065</td>
<td>25.9566</td>
<td>55.2646</td>
</tr>
<tr>
<td>6</td>
<td>29.6225</td>
<td>25.9313</td>
<td>41.8588</td>
</tr>
<tr>
<td>7</td>
<td>30.0273</td>
<td>26.6006</td>
<td>34.3551</td>
</tr>
<tr>
<td>8</td>
<td>29.9059</td>
<td>25.4595</td>
<td>38.0873</td>
</tr>
<tr>
<td>9</td>
<td>30.329</td>
<td>25.522</td>
<td>38.6791</td>
</tr>
<tr>
<td>10</td>
<td>31.4465</td>
<td>25.8673</td>
<td>24.4702</td>
</tr>
<tr>
<td>11</td>
<td>29.1108</td>
<td>25.1612</td>
<td>30.7278</td>
</tr>
<tr>
<td>12</td>
<td>28.075</td>
<td>25.9492</td>
<td>28.0483</td>
</tr>
<tr>
<td>13</td>
<td>29.085</td>
<td>25.049</td>
<td>53.3695</td>
</tr>
<tr>
<td>14</td>
<td>33.8174</td>
<td>28.9424</td>
<td>23.4498</td>
</tr>
<tr>
<td>15</td>
<td>34.2554</td>
<td>28.9412</td>
<td>22.7558</td>
</tr>
<tr>
<td>16</td>
<td>32.1612</td>
<td>27.9163</td>
<td>31.2433</td>
</tr>
</tbody>
</table>

\( F_{\text{median}} \) is nothing but the median of gray levels at \( S_{\text{xy}} \). \( F_{\text{min}} \) is the minimum gray level at \( S_{\text{xy}} \). \( F_{\text{max}}, F_{\text{min}} \) are the minimum and the maximum gray levels at \( S_{\text{xy}} \). \( F_{\text{xy}} \) is simply the gray level at coordinates \( x, y \). \( S_{\text{max}} \) is the maximum size allowed size of \( S_{\text{xy}} \).

3.1 Segmentation

**ALGORITHM**

Thus the algorithm can be as follows

1. Detect object Boundary. The two contour points are chosen.
2. Energy is computed.
   \[
   \text{Energy} = \text{Internal Energy} + \text{External Energy} + \text{Constraint Energy} 
   \]
   \[
   \text{Internal Energy} = \text{Elastic Energy} + \text{Bending Energy} 
   \]
   \[
   \text{Elastic Energy} = \frac{1}{2} \int_0^1 \alpha(s) \text{mod } c(s)^2 \text{ ds} 
   \]
   \[
   \text{Blending Energy} = \frac{1}{2} \int_0^1 \beta(s) \text{mod } c_{xy}^2 \text{ ds} 
   \]
   \[
   \text{External Energy} = \int_0^1 \text{image } c(s) \text{ ds} 
   \]
3. Shrinks and wraps around the object.
4. Gradient magnitude is calculated.
5. Morphological opening closing by reconstruction is performed.
6. Watershed transform is applied.
7. Final extracted region is obtained.

3.2 Feature Extraction
The most important phase that is the feature extraction phase which can be called as the heart of the entire diagnostic procedure. The feature extraction extracts the necessary features in the image. The chapter 1 explained the four features that are Area, Mean, Standard Deviation and the variance.

3.3 GLCM
The GLCM takes an input and then converts it to the matrix and then the final calculation can be done in the GLCM matrix. The GLCM has 23 features involved they are
- Auto correlation
- Contrast
- Correlation 1
- Correlation 2
- Cluster Prominence
- Cluster shadowing
- Dissimilarity
- Energy
- Entropy
- Homogeneity
- Homogeneity 2
- Maximum Probability
- Sum of Squares
- Sum Average
- Sum Variance
- Sum Entropy
- Difference Variance
- Difference Variance 2
- Difference Entropy
- Informational measure of correlation
- Informational measure of correlation 2
- Maximum Correlation Coefficient

These are some of the GLCM features that we are focusing on to extract the values so as to get a proper classification in an image since accuracy is a vital key for us. The below equation helps us to find the value of the GLCM features.

3.4 Feature Extraction
The segmented image is the input for the feature extraction. The feature extraction includes 32 features for improving the classification accuracy using support vector machine. The feature extraction is done for Normal statistical method, GLCM and the ABCDE features on skin. Now, the ABCDE features can be discussed as follows. Before getting into that the requirements for calculating these stolz algorithm are area of ABCDE features are that the total number of the entire pixels in the lesion. And the perimeter is nothing but the edge boundaries of the pixel.

3.5 Major Axis Length Or The Greatest Diameter (Gd)
The Major Axis Length can be calculated as a line passing through the lesion centroid and connecting the two boundary points. The output can be given as \((A_c, B_c)\)

\[
(A_c, B_c) = \left(\frac{\sum_{i=1}^{n} A_i}{n}, \frac{\sum_{i=1}^{n} B_i}{n}\right)
\]

Where \(n\) is the number of pixels inside lesion \((A, B)\) are the coordinates of the lesion pixels.

MINOR AXIS LENGTH OR THE SHORTEST DIAMETER (SD)
The Minor Axis Length can be calculated as the line passing through the lesion blob and connecting shortest boundaries.

ASYMMETRY (A)
Asymmetry is something when both the halves of the images are not equal. The figure below shows the asymmetric nature in a skin image.

Asymmetry of a lesion can be calculated as follows

\[
\text{Asymmetry} = (\Delta k/k) \times 100
\]

\(\Delta k\) - Pixel Difference

k- Total count of lesion
Thus the asymmetry is something where the shape of two halves cannot be the same.

BORDER
The next phase of skin cancer is to calculate border. That is it has a irregular border. The below picture depicts how the border irregularity can be showed.

The irregular border can be calculated as

\[
\text{Border} = (\text{Perimeter}^2 | 4 \times 3.14 \times \text{Area})
\]

Thus the border can be calculated with the calculated perimeter upon the values of area and then the irregularities are determined.

Color
The color has various shades to it they are white, red, light brown, dark brown, slate blue and black. The value is considered from 0 to 6 based on the values that are present.

Diameter
The diameter must be greater than 6 mm wide. Though few fit the property but still there can be few exceptions as well.

Total Dermoscopic Score
The total dermoscopic score (TDS) has a standard formula for calculation and is based on asymmetry, Border, Color and Diameter. Asymmetry has totally 4 values they are (0 – symmetry, 1 –axis symmetry, 2- axis asymmetry). Border can be having values 0 to 8. Color has six shades. Diameter has 0-5 score. Any lesion more than 6mm is correctly equal to 5. The TDS is calculated as follows

\[
\text{TDS} = 1.3A + 0.1B + 0.5C + 0.5D
\]

If TDS is less than 4.75 it is benign that is normal image and then if TDS is greater than 4.75 and less than 5.45, it is actually suspicious and then if it is greater than 5.75 it is melanoma or can be a cancerous cell.
The KNN algorithm states that the K value is to be found and the Euclidian distance is calculated and the distance is calculated and the neighborhood is calculated based on the K values.

The classification proves to be 96.8% which is considerably a better one whereas still a betterment can be made by improving the dataset and by improvising the methodology by using deep learning methods which will be in the future.

**REFERENCES**


