Computer Aided Diagnosis of Melanoma Skin Cancer using Clinical Photographic images

Mai S. Mabrouk1*, Mariam A. Sheha2, Amr A. Sharawy2
1 Biomedical Engineering, Misr University for Science and Technology (MUST University), 6th of October, Egypt.
Email: msm_eng@yahoo.com

2 Systems and Biomedical Engineering, Cairo University, Giza, Egypt.
Email: mariam_sheha@hotmail.com, amrarsha@gmail.com

ABSTRACT
Melanoma is considered as one of the most malignant, metastatic and dangerous form of skin cancer that may cause death. The curability and survival of this type of skin cancer depends directly on the diagnosis and removal of melanoma in its early stages. The accuracy of the clinical diagnosis of melanoma with the unaided eye is only about 60% depending only on the knowledge and experience that each doctor has accumulated. The need to the Computer-Aided Diagnosis system (CAD) is increased to be used as a non-invasive supporting tool for physicians as a second opinion to increase the accuracy of detection, as well contributing information about the essential optical characteristics for identifying them. The ultimate aim of this research is to design an automated low cost computer aided diagnosis system of melanoma skin cancer to increase system flexibility, availability. Also, investigate to what extent melanoma diagnosis can be impacted using clinical photographic images instead of using dermoscopic ones, regarding that both are applied upon the same automatic diagnosis system. Texture features was extracted from 140 pigmented skin lesion (PSL) based on Grey level Co-occurrence matrix (GLCM), effective features are selected by fisher score ranking and then classified using Artificial Neural Network (ANN), the whole system is processed through an interactive Graphical User Interface (GUI) to achieve simplicity. Results revealed the high performance of the proposed CAD system to discriminate melanoma from melanocytic skin tumors using texture analysis when applied on clinical photographic images with prediction accuracy of 100 % for the training phase and 91 % for the testing phase. Also, results indicated that using this type of images provides high prediction accuracy for melanoma diagnosis relevant to dermoscopic images considering that photographic clinical images are acquired using less expensive consumer which exhibit a certain degree of accuracy toward the edges of our field of view.

Indexing terms/Keywords
Skin cancer, GLCM, CAD, Melanoma, Texture analysis

Academic Discipline And Sub-Disciplines
Biomedical Engineering

SUBJECT CLASSIFICATION
Medical Image Processing: Computer Aided Diagnosis.

TYPE (METHOD/APPROACH)
Research Paper.
1. INTRODUCTION

Melanoma accounts less than 5% of skin cancer cases but it is responsible for majority of skin cancer deaths. According to the World Health Organization Report, about 46,000 out of melanoma-related deaths occur worldwide each year. Fortunately, if skin cancer is detected early, the curability is very high and over 92% [1]. Accurate depended procedure of biopsy involves some expense and morbidity; therefore, flexible automatic early detection techniques are essential.

In the last two decades, a strong impulse has been given to develop automated systems to assist physicians in this task, mainly because they have the desire of quantizing a number of anatomical and functional parameters, valuable for making correct diagnoses, which can be evaluated only qualitatively by human beings such as texture features, that may not be readily perceived by human eyes.

Improving diagnosis accuracy is the main aim of any biomedical research but, toward achieving a reliable melanoma diagnosis system it’s important to concern with three main points; cost reduction, flexibility and availability. Easy use system with high accuracy range and least price can be widely available and increase survival rate. It’s that what we aimed to discuss through the presented paper.

Although, dermoscopic images minimized color differences between images occurred due to differences in ambient lighting during their photographic process but photographic images is used for database that can increase system availability and reduce its cost. Also, the proposed system is presented through an interactive graphical user interface dedicating time and efforts to improve the early screening process and increase system flexibility.

Towards our goal, the system at hands reflects a high simplicity as it avoids segmentation process at pre-processing step that is considered as a fatal problem due to tumor shape irregularity, color and size variation. Texture analysis is used in order to characterize regions in the image by their texture content to discriminate between melanocytic nevi and malignant melanoma. Features are extracted based on the Grey Level Co-occurrence Matrix (GLCM). For classification, Artificial Neural Network (ANN) has been reported to be an accurate tool for the evaluation of pigmented skin lesions (PSLs) [2-4]. The ANN classification procedure used is MLP, represent the classes of benign common nevi and malignant melanoma. Also, the prediction accuracy of the proposed system using photographic images is compared to the prediction accuracy of the same system using dermoscopic image as in [5].

2. RELATED WORK

Popular scoring systems and algorithms have been proposed to improve diagnosis performance. Such as TDS score which is based on ABCD rule, the seven-point checklist, and the Menzies method [6], [7]. This diagnosis systems had developed over the last 20 years to have diagnosis accuracy around 73% to 98% [8].

In 2002, Bayes rule as one of the skin lesions classification method was published with rather inconclusive results [9]. In 2006, Iris Cheng builds up an algorithm using relative color features; algorithm was tested upon a database of 160 melanoma, 42 dysplastic nevi, and 80 nevus images. The result showed that the algorithm is able to classify 86% of the malignant melanomas successfully compared to an approximately 75% success rate of dermatologists [10]. In 2009, Ho Tak Lau, Adel Al-Jumaily performed automated skin cancer classification system across different types of neural network. Recognition accuracy of the 3-layers back-propagation neural network classifier is 89.9% and auto-associative neural network is 80.8% in the image database that include both dermoscopic and photographic images [8].

In 2011, Daniel Ruiz, Vicente Berengue, Antonio Sorianoand Belén Sánchez proposed types of ANN classifiers, a Bayesian classifier, K nearest neighbors and also including the algorithm of the MLP. These methods work independently and also in combination making a collaborative decision support system. The classification rates obtained are around 87% [11]. In 2012, part of our work was published investigating melanoma diagnosis based on texture analysis and classified by ANN build up upon total database 102 dermoscopic image, 51 images for each malignant melanoma and benign nevi. Results show that algorithm is able to classify malignant melanoma with 92 % accuracy range for test set [5].

3. MATERIAL

3.1 Database:

Development of non-invasive tools to improve early diagnosis by automated systems, primary concerned with image acquisition approaches. There are two types of images used as a database for CAD systems. Image acquired by dermoscopy and standard camera resulted ‘dermoscopic image’, other type of image is photographic or clinical image acquired by non-standard camera [12]. As shown, in “Figure 1” the difference between the two types of image.

Dermoscopy is a non-invasive examination technique using incident light and oil immersion to illustrate the visual examination of skin sub surface structures. Although, Dermoscopic images allow more control over lighting and skin color variations, which contribute to lesion misclassification but it mainly cause unavailability of system specially that dermoscopy is not widely spread among countries not familiar with skin cancer. Also, having extra step for image acquisition will increase diagnosis fesses. On the other hand, clinical photographic image is digital image that acquires using standard camera and can be easily captured. For these reasons, the main aim of this research is to design a high performance CAD system with a simple and fixable GUI with high prediction accuracy and low cost. Also, recommend using clinical photographic images captured in visible light by standard digital cameras instead of conventional dermoscopy.
In this study, a number of 140 clinical photographic images of different cases were randomly sampled from the skin diseases Atlases and doctors’ clinics. Image atlases of reference pictures are widely available [13]-[15]. The database is subdivided to two categories 70 images each, common nevi and malignant melanoma. All images are resized to be [512*512] pixel and then divided for two classes 75% for learning set and 25% for test set.

3.2 Implementation

The automated high performance and low cost image diagnosis system is proposed with a good visualization that is essential for a successful system. Leveraging the rich graphics functionality of MATLAB the proposed system is implemented and tested on the WINDOWS platform with MATLAB version of 2010 to provide a pictorial interface through a simple and flexible graphical user interface (GUI) which makes the system program easier to use by providing them with a consistent appearance and with intuitive controls like pushbuttons, list boxes, sliders, menus, and so forth [16]. The proposed system is designed through three stages; preprocessing, Feature extraction and selection and Classification. In the preprocessing stage, images are loaded and resized. In the feature extraction and selection stage, two text files are obtained containing the extracted co-occurrence matrix features for both malignant melanoma and benign nevi each in a separate file and the most significant features are selected and then classified to discriminate between benign and nevi tumors in the classification stage which is the last stage of the proposed CAD system.

4. METHODOLOGY:

The approach to overcome the problem of noise presence, variability of biological shapes and tissues has always been to simplify the objective of the analysis and to exploit some kinds of a priori information about the imaged structures. The general and common approach for the automated diagnosis system to skin lesion early detection is configured of three stages as presented in “Figure 2”. Systems usually differ from each other according to the used methods for Preprocessing, features extracted and selected and the classifier used to discriminate between malignant and benign lesions.

4.1 Pre-Processing:

Generally, Pre-processing is to perform some processes on the loaded image after image acquisition, a primary image processing is necessary on original image to obtain the enhanced image recalling the main goal that is to provide a simple method towards accurate diagnosis. Preprocessing shows simplicity compared to other methods of detection. To suppress large variations within the lesion and background, and also to reduce the effect of different skin color variations, the original color RGB image is converted into intensity (grey scale)[17]. Then image is resized to 512*512 pixels.
4.2 Features extraction and selection

Features determine the gathered information from the pigmented skin lesion. Texture Analysis is used to quantify texture notions such as “fine,” “rough and irregular” and to identify, measure, and utilize the differences between them [18]. Also, the statistical approach to texture analysis is always considered to be useful increasing the performance of the classification. For this reason, the extracted features are based on Grey level Co-occurrence matrix (GLCM).

The quantitative measures of texture that describes salient lesion features based on grey level co-occurrence matrix. GLCM considered being good example for higher order statistical features, depending on pixel values and their spatial inter-relationships. It is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image [19] and it gives the joint probability of occurrence of grey levels of two pixels. Extracted features from the lesion’s objects are 23 feature vectors were generated based on GLCM proposed by Haralicks [20]. “Figure 3” illustrates the extracted GLCM features displayed in proposed GUI. A specific number of features according to fisher score method are selected to provide the ideal performance.

![Figure 3: The extracted GLCM Features displayed in the GUI.](image)

Simplicity of extracting features by just one click directly met paper aim having simplest way for automated system. According to the previous figure, images introduced to the system from load normal/abnormal image buttons above, then features are extracted by pressing feature extraction button. On both sides, there is a window and list box, the left one shows list of melanoma images and window display the chosen image from that list and on the right the same but for nevi with the resulted texture features of the last selected image.

In this stage also features of the lesion are selected to feed into the classifier. From a co-occurrence matrix one can draw out some important statistical features for texture classification. The range of grey level values within a given image determines the dimensions of a co-occurrence matrix. To increase detection accuracy, Fisher score ranking method calculated according to equation (1) is used to select the most utilized features, which have a good discrimination power between malignant melanoma and other benign nevi images. Where, $R_i$ is the rank of feature ‘i’ according to the computed mean ($\mu$) and standard deviation ($\sigma$).

$$R_i = \frac{\mu_i - \mu_n}{\sigma_i \sigma_n}$$

(1)

As a comparison between dermoscopic and clinical photographic images, the most significant features selected by fisher score ranking using photographic images differs from those selected using demoscopic images as proposed in [5]. In this study, using clinical photographic images and fisher score ranking, 11 features are selected versus 12 features using demoscopic images as in [5], where the highest score features seems to be the most effective, these features are Correlation: matlab, Correlation: [1, 2], Cluster Prominence, Cluster Shade, Energy, Entropy, maximum probability, Sum of entropy, Information measure of correlation 1, Information measure of correlation 2 and Inverse difference homogenous (INV).

Among all previous studies using texture analysis, It was noticed that there are some common texture features that is used among all literatures and may be used by more than one form as entropy and sum of entropy also homogeneity and inverse difference homogenous. “Table 1” present the Explanation and formula of those significant features.
4.3 Classification:

Classification process is considered as the final significant task in building and developing the proposed reliable automated systems for melanoma diagnosis. There are different classification models, which can be grouped into general classes [21]: Global Models such as Linear Models, Polynomial Models, Neural Networks (MLP), and Local Models such as k–Nearest–Neighbors. One of the AI techniques was mentioned in this survey for classification is Artificial Neural Networks (ANN). A common example of such a network was applied to classify the extracted features of lesions into either malignant or benign, the Multilayer Perceptron (MLP). The MLP is a feed forward network, capable of generating nonlinear boundaries. It has been successfully applied to solve some difficult and diverse problems [22]-[24].

According to Neural Network classification procedure two successive processes are applied training or learning process then testing process. Neural network are trained by experience or known data, when fed an unknown input into neural network, it can generalize from past experience ‘trained data’ and produce a result. For that, MLP were applied by two different approaches Automatic and traditional MLP. Two approaches was proposed to compare the efficiency of having 1000 iteration in each training cycle by using Traditional MLP to the Automatic MLP that stop training cycle early at any number of iterations if the network performance on the validation subset remains the same for six epochs in a row.

The selected 11 GLCM features representing the classes of benign nevi and malignant melanoma constitute the entry stage parameters that fed to neural input layer of neural network classification. Then output layer produces either ‘1’ for benign nevi or ‘0’ for malignant melanoma. Regarding the importance of the decision taken based on the trained data the proposed classifier was trained with 75% and tested by 25% of the total number of image.

5. Evaluation:

When classification is done results could have an error rate, either fail to identify an abnormality, or identify an abnormality that is not present. That was described by terms of true, false positive and negative. According to this study, “Table 2” shows the number of images that correctly diagnosed (TP, TN) and wrongly diagnosed (FP, FN).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Explanation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>It measures the coarseness of texture between pixel and its neighbor over the whole image.</td>
<td>[ \sum_{i,j}</td>
</tr>
<tr>
<td>Energy</td>
<td>It measures the image textural uniformity in the range [0, 1].</td>
<td>[ \sum_{i,j} P_{(i,j)}^2 ]</td>
</tr>
<tr>
<td>Entropy</td>
<td>It measures the degree of disorder or non-homogeneity, in this sense opposite to the energy.</td>
<td>[ -\sum_{i} \sum_{j} P_{(i,j)} \log_2 P_{(i,j)} ]</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>It measures the distribution of element in the GLCM with respect to its diagonal.</td>
<td>[ \sum_{i,j} \frac{P_{(i,j)}}{1 +</td>
</tr>
<tr>
<td>Correlation</td>
<td>Refers to pixel’s neighborhood influence on the whole image surface.</td>
<td>[ \frac{\sum_{i,j} (i-j)^2 P_{(i,j)} - \mu_x \mu_y}{\sigma_x \sigma_y} ]</td>
</tr>
</tbody>
</table>

**Table 2: Confusion matrix for testing**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predict AMLP</th>
<th>Predict TMLP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive (Melanoma)</td>
<td>Negative (Nevi)</td>
</tr>
<tr>
<td>Positive (Melanoma)</td>
<td>15 (TP)</td>
<td>1 (FN)</td>
</tr>
<tr>
<td>Negative (Nevi)</td>
<td>2 (FP)</td>
<td>16 (TN)</td>
</tr>
</tbody>
</table>
- TP is number of true positives: Predict Melanoma as Melanoma.
- FP is number of false positives: Predict Benign Nevi as Melanoma.
- TN is number of true negatives: Predict Benign Nevi as Benign Nevi.
- FN is number of false negatives: Predict Melanoma as Benign Nevi.

Then, system diagnosis performance was evaluated in the form of sensitivity, specificity and accuracy, through the following equations (2, 3, 4). Accuracy is generally regarded with balanced measure; it takes into account true and false positives and negatives, whereas sensitivity deals with only positive cases and specificity deals with only negative case. For that, Accuracy always increase when the system show high measurable of high sensitivity, i.e., ‘high probability of true positive’ and high specificity, i.e., ‘high probability of true negative’ [20].

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

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

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

5. RESULTS AND DISCUSSIONS:

This section overviews the study and details the results obtained using GLCM features upon set of photographic images, after building a suitable classifier for automatic skin cancer detection. Database disposed with about 140 medical photographic images, Images was previously diagnosed to have 70 images for each benign nevi and malignant melanomas. A number of 23 GLCM features are extracted from the proposed dataset, and then 11 features are selected according to fisher score method these input fed to neural input layer to get 1 output layer. In order to evaluate this work, MLP classifier is used and exhibited in two techniques, Automatic MLP and Traditional MLP. The classifier technique by its two approach was trained with 106 photographic image (53 malignant, 53 nevi) and tested by 34 (17 malignant, 17 nevi) photographic image that represent 75% to 25% of the total images for training and test set respectively. At the end of the training process updated weight values are stored. In testing process outputs are calculated using saved weight value. The overall prediction accuracy rates of 91% are calculated for both malignant melanoma and benign nevi images based on GLCM features extracted using photographic images and classified using neural network.

Then, performance measures are done by computed output of TP, TN, FP and FN as discussed in the upper section. Table 3 shows the computed sensitivity, specificity and prediction accuracy for training and testing of MLP techniques. According to results obtained in this study and comparing them to those obtained in [5] using the same system but with clinical Photographic images, it is found using Automatic MLP provides higher sensitivity, specificity and prediction accuracy values when applied on clinical Photographic images but it’s clear that the second technique; Traditional MLP is more effective for detection of melanoma from benign nevi either by using photographic or dermoscopic images. Figure 3, shows a comparison of the sensitivity , specificity and prediction accuracy values obtained in this study using photographic image and those obtained in [5] using dermoscopic images . The experimental results obtained reveals that the proposed system works very well in the testing phase using photographic images with short processing time reflecting high efficiency of the proposed system.

Table 3: Performance measure of Training and Testing set.

<table>
<thead>
<tr>
<th>Training Performance Measure</th>
<th>Testing Performance Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic MLP</td>
<td>Traditional MLP</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>92.5%</td>
</tr>
<tr>
<td>Specificity</td>
<td>94.2%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>93.5%</td>
</tr>
</tbody>
</table>
CONCLUSION:

As the importance of early detection of skin cancer in treatment, this paper describes a method that can be used to analyze and classify medical photographic images using texture features and neural network classification. The system performs simple image processing operations (image resize, type conversion from RGB to grey), statistical textural features extraction derived from GLCM, feature selection according to fisher score ranking and image classification. During the course of experimentation it was observed that however using dermoscopic or photographic image results were more productive at traditional MLP as a result of expanding training cycle having 1000 Iteration at each. Another significant aspect of this work is to prove the high performance of using photographic images that can be easily captured and available avoiding the extra fesses of dermoscopic image. The study results are consistent with theory that using photographic image instead of dermoscopic images is promising as it provides high accuracy rates at testing phase of 91% relevant to that of dermoscopic images 92%. The second phase of this work considers the misclassification of melanoma that suggests the presence of a considerable overlap according to shape and size variety. This indicates that some additional features and more trail classifiers are required to improve result of the automatic differentiation method. Also, the aim is to improve the automated processing speed and efficiency by implementing all algorithms and procedures on parallel GUI system to be well designed as a substitute of clinician in melanoma diagnosis.

REFERENCES


Mai S. Mabrouk

Mai Said Mabrouk received her BSc in Systems and Biomedical Engineering Department, Cairo University, Giza, Egypt, in 2000. She completed her MSc and PhD in Biomedical Engineering from the same school in 2004 and 2008, respectively. She is an Assistant Professor with the Biomedical Engineering Department, Misr University for Science and Technology (MUST), since August 2008. She has several research papers in the area of image processing and bioinformatics. Her research interests include biomedical image processing, bioinformatics and digital signal processing in addition to genomic signal processing.

Mariam Ahmed Sheha received her BSc in Systems and Biomedical Engineering Department, Misr University for Science and Technology (MUST), 6 of October, Giza, Egypt, in 2008. She is student in MSc in Biomedical Engineering at Cairo University. She is working as a Biomedical and Clinical Engineer in Dar Al Fouad Hospital, since September 2010. She did one research papers in the area of image processing. Her main interest lies in discrete mathematics with applications on digital image.

Amr A. Sharawi

Amr A. Sharawi, PhD, Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, P.O.B. 12211, Giza, Egypt (amransha@link.net). Amr A. Sharawi, PhD, is with Cairo University. He received his BS degree in Electronics and Communication in 1976 from Cairo University. He received his MS degree in Systems and Biomedical Engineering in 1981 from Cairo University. Furthermore, he received his PhD degree in Systems and Biomedical Engineering in July 1991. Since 2001, he has been an associate professor in biomedical engineering at Cairo University, Faculty of Engineering. His research interests include mainly clinical engineering, bioinformatics and medical imaging.