Malignant Melanoma Detection Based on Machine Learning Techniques: A Survey

Munya A. Arasi, El-Sayed A. El-Dahshan, El-Sayed M. El-Horbaty, Abdel-Badeeh M. Salem

1,3,4Dept. of Computer Science, Faculty of Computer and Information Sciences
Ain Shams University, Abbassia, Cairo, Egypt
2Egyptian E-Learning University (EELU), Eldoki, Giza, Egypt
munya_arasi@yahoo.com, seldahshan@eelu.edu.eg, shorbaty@cis.asu.edu.eg, absalem@cis.asu.edu.eg

Abstract

Skin cancer is one of the most growing types and dangerous cancer in the world; the important of these cancers are malignant melanoma. The early diagnosis of malignant melanoma is a critical issue for dermatologists. In this paper, we present an overview of recent the state of the art in Computer-aided detection/diagnosis (CAD) systems in identifying and diagnosing malignant melanoma of dermoscopy images and describe its steps starting with image acquisition, preprocessing; and finishing with malignant melanoma classification of dermoscopic images. The comparative study shows that the most common methods for features extraction are the Discreet Wavelet Transform (DWT) and the method which combines both texture and color features resulting in output of very high accuracy. The methods for the classification: K-Nearest Neighbor, Artificial Neural Networks, and Support Vector Machines are very well in the range [%90 – % 97, 5].

Keywords: Skin cancer, Malignant melanoma, Machine learning, Medical Knowledge-Based systems, Medical imaging, Medical informatics.

1. Introduction

Dermatology is the branch of medicine that concerned with the diseases’ diagnosis of skin, hair and nails, the skin is the most important part in the human body which protects the internal parts from the outside world. The skin cancer is the most important of these diseases; it can be growing at any part in the body and occurring from non-pigmented cells [1]. Skin cancer is one of the most growing types and the most dangerous in the world of cancer; the important of these tumors is malignant melanoma, the rates of melanoma have been rising for at least 30 years.

The main risk of melanoma is could be spread entire the body by lymphatic vessels and blood vessels [2], thus the main strategy is the early diagnosis of melanoma and removal of thin melanoma; it is the most common cancer if diagnosed at an early stage can be cured without complications. Therefore, the early diagnosis of cancer malignant melanoma is a critical issue and the main challenges for dermatologists to reduce mortality and morbidity [3]. It is very difficult the diagnosis of melanoma by naked eye using the features that recognize a benign from malignant melanoma when using the clinical practice.

Nowadays the computer technology in medical decision support is used widespread and pervasive across a wide range of medical area, such as cancer research, dermatology[4], thus
the computerized methods analysis of dermoscopy images allows overcome various issues which help the dermatologist to take less time and high efficiency in diagnosis of skin cancer. It is important to develop of various computerized methods for clinical diagnosis using computer aided diagnosis (CAD) systems, which these systems give to dermatologist different analysis steps such as border detection, features extraction and diagnostic by using classification methods [5]. Studies have shown that CAD systems for melanoma diagnosis is still complex problem, the problem of lighting effects the resulting from the reflection surface of the skin, this affecting on all steps of diagnostic [6,7] in addition there is difficulties for information contents of dermoscopy images which they are not clear due to the noises and presence the hair. Therefore, it has become the extraction of the features a complex issues, hence there are many researchers are interested to the development of systems for automated diagnosis of malignancy in skin lesions.

The rest of the paper is organized as follows: the generic methodology of Malignant Melanoma diagnosis “CAD” scheme is presented in section 2. In section 3 the review of CAD systems for Malignant Melanoma Detection is presented. Section 4 provides the discussion of image acquisition and preprocessing; and finishing with skin lesions classification. Finally, conclusions and future work are given in section 5.

2. Generic methodology of Malignant Melanoma diagnosis (CAD) scheme

The goal of CAD is to automatically decide Type of lesion, i.e. melanoma or benign by examining various features of lesion and examining these features by using machine learning techniques that help the dermatologist to take less time and high efficiency in diagnosis of skin cancer and enhancement the diagnostic accuracy of physicians and reduce the overall rate of misdiagnosis [8]. The system [9] uses image processing techniques for improving and segmentation the images for detection the melanoma in early stages without the need for biopsy.

The main advantages of CAD systems that the features extracted from lesion based on computer provide high accuracy than the features extraction based on dermoscopy algorithms. The general scheme of a CAD system for the skin lesions is shown in figure 1. It consists of four main components steps: Image acquisition and preprocessing, feature extraction, classification and evaluation.

The Basic techniques in CAD systems for Malignant Melanoma Detection in previous studies that published during 2001–2015 are highlighted in figure 2, which presents clearly all the techniques that used in this field. The inputs to the computer aided system are dermoscopy images; in the first phase preprocessing of image is done that allows reducing the noise effects and various artifacts like hair that may be present in the dermoscopic images and improve the image quality such as image cropping [10], gradient operation, morphological operation, scaling color space transformation [11, 12], color quantization, contrast enhancement [6,7];[13,14], Filters [3],[15, 16, 17, 18,19,20,21], it is the most important step for a successful feature extraction and diagnosis.
Figure 1. General scheme of a CAD system for malignant melanoma diagnosis

Figure 2. Various Techniques in Malignant Melanoma diagnosis
After preprocessing the image, the features are extracted for classification of lesion. Feature extraction focuses on analysis the properties of skin lesions for a single skin lesion, there are many techniques, e.g. the textural features [14],[17],[19],[22, 23, 24], shape features [16], [18], color features [16, 17], [20],[22], [25, 26], [18, 19], [24]Asymmetry (A), Boundary (B), Color (C), and Diameter of a lesion (D) (ABCD rule)[6,7], [14], DWT features [15] [27, 28, 29], Gray Level Co-occurrence Matrix(GLCM)[7], [12], [21], histogram analysis [30], and High-level intuitive features (HLIF) [31].After the feature extraction step for extraction the attributes of skin lesions, it is important to classify the attributes of these regions for proper recognition of melanoma. The aim of the classification step is to distinguish benignity and malignancy by making use of extracted features. This step uses machine learning algorithms to reach a decision[32], the most recent published classification techniques, e.g. supervised techniques such as k-NN [10], [12] support vector machine[7],[13, 14];[17]; [31],[35], Naïve Bayes (NB) [3], [20], Artificial Neural Networks [3],[12],[15, 16, 17],[27],[33],Multilayer Perceptron (MLP) [12],[16], Logistic Model Tree (LMT) [28], Hidden Naïve Bayes (HNB) [28] and Decision Tree [11], [24 25],and unsupervised classification techniques such as Clustering [27] and fuzzy C-means. Also, hybrid intelligent systems using soft computing techniques are used for diagnosis of melanoma in dermoscopy images like Neuro-fuzzy [20]. It gives high accuracy %91.26. The evaluation step measures the performance of classification techniques for the Malignant Melanoma Detection based on the Parameters like Accuracy (AC), Sensitivity (SE), and Specificity (SP).

3. Review of CAD systems for Malignant Melanoma Detection

Most studies revolve around analysis of skin lesion images taken using dermatoscope, table1 gives the summary of many different techniques are used for developing Computer-aided diagnosis for Malignant Melanoma, generally CAD systems are included these stages: image acquisition, image pre-processing, feature extraction, classification and Diagnostic Evaluation.

Many existing techniques have been employed in recent years for the prediction of skin cancer at initial stage (see table 1). It gives an overview the classification and image processing techniques that published during 2001–2015, which used in the field for diagnosis Malignant Melanoma. The results obtained from these methods are used to recognize the patterns which are aiming to help the doctors for classifying the malignant and benign cases. Hence, the Malignant Melana diagnostic problems are basically in the scope of the classification problems.
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Dataset</th>
<th>Preprocessing</th>
<th>Features Extraction</th>
<th>Classifier</th>
<th>Diagnostic Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Elbaum et al., 2001)</td>
<td>from 4 Clinical Centers.</td>
<td>NA</td>
<td>Statistical Features</td>
<td>Linear or nonlinear Classifiers</td>
<td>SE =95% SP = 70%</td>
</tr>
<tr>
<td>(Dreiseitl et al., 2001)</td>
<td>Dermatology, University of Vienna Medical School.</td>
<td>Gray level thresholding to</td>
<td>Global and Local Features</td>
<td>k-NN Log Regression ANN Decision Trees SVM</td>
<td>AC=%0.933 AC=%0.967 AC=%0.968 AC=%0.885 AC=%0.970</td>
</tr>
<tr>
<td>(Rubegni et al., 2002)</td>
<td>Dermatology of Siena University</td>
<td>Filtering, Detection of Borders</td>
<td>Geometry, Color, Texture</td>
<td>ANN</td>
<td>AC=%93</td>
</tr>
<tr>
<td>(Barzegari et al., 2005)</td>
<td>Images of pigmented skin lesions</td>
<td>NA</td>
<td>Software of Visio med AG</td>
<td>ANN</td>
<td>SE=%83 SP=%96</td>
</tr>
<tr>
<td>(Yuan et al., 2006)</td>
<td>from University of Texas</td>
<td>Resizing, Cropping, Hair Removal</td>
<td>Texture Feature</td>
<td>SVM</td>
<td>AC=% .70</td>
</tr>
<tr>
<td>(Grammatikopoulos, 2006)</td>
<td>NA</td>
<td>Histogram Equalization, Prewitt Method.</td>
<td>ABCD rule</td>
<td>TDS Calculation</td>
<td>NA</td>
</tr>
<tr>
<td>(Stanley, Stoecker, &amp; Moss, 2007)</td>
<td>EDRA interactive Atlas</td>
<td>Histogram Quantization</td>
<td>Color Feature</td>
<td>Thresholds Discriminate</td>
<td>AC=% .87.7</td>
</tr>
<tr>
<td>(M. Emre Celebi et al., 2007)</td>
<td>Atlas of Dermoscopy</td>
<td>Median Filter</td>
<td>Shape, Color and Texture Features</td>
<td>SVM</td>
<td>SP=% .92.34 SE=% .93.33</td>
</tr>
<tr>
<td>(M. Emre Celebi et al., 2008)</td>
<td>Atlas of Dermoscopy</td>
<td>Manual Border Determination</td>
<td>Color Features and Texture Features</td>
<td>Decision Tree</td>
<td>SE=% 69.35 SP =%89.97</td>
</tr>
<tr>
<td>(M Emre Celebi et al., 2009)</td>
<td>public dermoscopy image set</td>
<td>Gaussian Filter(GF), Median Filter(MF)</td>
<td>Color and Texture Features</td>
<td>Euclidean Distance Transform</td>
<td>Higher Diagnostic Accuracy.</td>
</tr>
<tr>
<td>(Garnavi et al., 2010)</td>
<td>Interactive Atlas of Dermoscopy</td>
<td>the lesion borders were obtained manually</td>
<td>Wavelet-based Texture Analysis</td>
<td>SVM (RBF kernel), LMT HNB</td>
<td>AC=%88.24 by LMT</td>
</tr>
</tbody>
</table>
### Follow Table 1. Classification performance for different studies

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Dataset</th>
<th>Preprocessing</th>
<th>Features Extraction</th>
<th>Classifier</th>
<th>Diagnostic Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Salah et al., 2011)[20]</td>
<td>NA</td>
<td>Blure filter</td>
<td>Area and Color Features</td>
<td>Neuro-fuzzy</td>
<td>AC = %91.26</td>
</tr>
<tr>
<td>(Sheha et al., 2012)[12]</td>
<td>102 dermoscopy Atlasses</td>
<td>Resizing and Color space</td>
<td>GLCM</td>
<td>MLP</td>
<td>AC = %92</td>
</tr>
<tr>
<td>(Kumar Jain &amp; Jain, 2012)[27]</td>
<td>From different sources</td>
<td>Image contour Tracing Algorithm</td>
<td>DWT</td>
<td>Clustering, PNN</td>
<td>AC = %0.92, AC = %0.95</td>
</tr>
<tr>
<td>(Ramteke &amp; Jain, 2013)[29]</td>
<td>from available digital camera</td>
<td>Watershed method</td>
<td>DWT</td>
<td>Fuzzy System</td>
<td>AC = %90</td>
</tr>
<tr>
<td>(Elgamal, 2013)[15]</td>
<td>from a digital camera with dermoscope</td>
<td>Gaussian, Median Filter</td>
<td>PCA, DWT</td>
<td>FP-ANN, k-NN</td>
<td>AC = %95, AC = %97.5</td>
</tr>
<tr>
<td>(Scharcanski et al., 2014)[31]</td>
<td>DermISand DermQuest</td>
<td>Illumination Correction Algorithm</td>
<td>HLIFs</td>
<td>Linear soft-Margin SVM</td>
<td>High Accuracy</td>
</tr>
<tr>
<td>(Science &amp; Engineering, 2014)[34]</td>
<td>Health care Centers in Kottayam</td>
<td>NA</td>
<td>21 Medical Attributes</td>
<td>NB</td>
<td>High Accuracy</td>
</tr>
<tr>
<td>(Li et al., 2014)[3]</td>
<td>spectroscopic system</td>
<td>Median filter and</td>
<td>Statistical Variables</td>
<td>ANN and NB</td>
<td>AC = %88.4, AC = %89.2</td>
</tr>
<tr>
<td>(Gajbar &amp; Deshpande, 2015)[21]</td>
<td>National Cancer Institute.</td>
<td>Median Filter, FCM</td>
<td>GLCM</td>
<td>SVM</td>
<td>High Accuracy</td>
</tr>
<tr>
<td>(Jaiswar, Kadri, &amp; Gatty, 2015)[6]</td>
<td>Using the technique Dermoscope</td>
<td>Histogram Equalization</td>
<td>ABCD rule</td>
<td>TDS Calculation</td>
<td>High Reliable and Robust.</td>
</tr>
<tr>
<td>(Mengistu, 2015)[35]</td>
<td>DermQuest, Dermnet</td>
<td>Median Filtering</td>
<td>GLCM and Color features</td>
<td>SOM and RBF</td>
<td>AC = %96.15</td>
</tr>
<tr>
<td>(Amarathunga, Ellawala, Abeysekara, 2015)[16]</td>
<td>Using the technique Dermoscope</td>
<td>Median Filtering</td>
<td>Color and shape features</td>
<td>MLP J48</td>
<td>AC = %85, AC = %85</td>
</tr>
<tr>
<td>(Kaur, 2015)[7]</td>
<td>NA</td>
<td>Histogram Equalization</td>
<td>ABCD rule and GLCM</td>
<td>Boosting Algorithm</td>
<td>Good Accuracy</td>
</tr>
<tr>
<td>(Immagulate &amp; Vijaya, 2015)[22]</td>
<td>Dermnet Dermofit.</td>
<td>Image resizes</td>
<td>Color and Texture Features</td>
<td>SVM, ASVM and PSVM</td>
<td>AC = %86, AC = %92, AC = %93</td>
</tr>
</tbody>
</table>
4. Discussion

As it is mentioned in Table 1, the step of image acquisition, dermoscopy images are collected from different sources for decoding the characteristics of human skin. The preprocessing and feature extraction techniques of the automatic system are proposed in previous works must be improved in order to become an effective tool and robust in the diagnosis of skin lesions.

The classification methods that have been used in computer-aided diagnosis of melanoma in this study involve K -Nearest Neighbour, Decision Tree, Support Vector Machine, Artificial Neural Network (ANN), Neuro-Fuzzy, Fuzzy C-Mean (FCM), Naïve Bayes and Clustering. For comparison purpose, the best three classifiers with excellent results like (k-NN, artificial neural networks, and support vector machines) When implementing support vector machine and its variants such as proximal support vector machine (PSVM), it based classification model yields a better performance and effectual when compared to other models, the predictive accuracy of PSVM is 93% [22].

Several studies have shown that the diagnostic accuracy of artificial neural networks is improved. From the results the hybrid techniques are robust and effective, it gives a higher accuracy is 91.26% which achieved by using of Area Features and Color Features and Neuro-fuzzy classifier. That means the results of hybrid classifiers are obtained very well, due to the integration of their performance and combining their advantages. It was also showed that the GLCM and Color feature were used together the classification accuracy was increased. The best classification accuracy 96.15% for Melanoma was obtained using combining SOM and RBF [35]. The single layer perceptron (SLP) type of artificial neural network was designed to estimate the probability of melanoma risk, which obtained a maximum accuracy in distinguishing melanoma from benign lesions is 93% [17]. The Multilayer Perceptron (MLP) is a feed forward network, capable of generating nonlinear boundaries. The accuracy 100% and 92% for training and testing respectively, in this study shows that combination between co-occurrence matrix and ANN is a promising technique for discrimination between malignant melanoma and melanocytic nevi dermoscopy images [12].

Probabilistic Neural Network (PNN) is a feed forward neural network, which was derived from Bayesian network and the training consists essentially of incorporating the training cases into the pattern layer. The classification using Probabilistic Neural Network is better than other types of artificial neural networks (ANNs) proposed in the same domain and have shown excellent classification performance 95% as compared to clustering Classifier [27]. It has got the same accuracy 95% by combination between discrete wavelet transforms DWT and back-propagation (BP) algorithm whereas the combination between discrete wavelet transforms DWT and k-nearest neighbor algorithm has 97.5% of accuracy for the same domain [15].

5. Conclusions and Future Work

The most common methods for the features extractions are Discrete Wavelet Transforms (DWT), the combination between texture feature and color feature providing very high accuracy. The results show that the methods for the classification k-NN, Artificial Neural Networks, and Support Vector Machines are very well in the range [%90 – % 97, 5].
Using Naïve Bayesian Classification is effective model for diagnosis Malignant Melanoma, but the decision tree algorithm is not well suited for this domain. The classification using Probabilistic Neural Network is better than the Clustering classifier. This is a significant improvement as compared to the earlier techniques proposed in the same domain. The PNN, back-propagation (BP), and the combining SOM and RBF classifiers show excellent classification performance of Artificial Neural Networks (ANNs).

In future we are looking for constructing a better framework for development a CAD system automatically accurate diagnosis of melanoma in dermoscopy images with extract the statistical and texture features from 2D wavelet transform. We hope that will get the valuable information and overcome the drawbacks and enabling well delivery of medical fields.

References


[27]. Y. KumarJain and M. Jain, “Comparison between Different Classification Methods with


