International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

SKIN CANCER DETECTION BY USING PHYTON

Mr. S.Mohan Rao[1], Gogula Chaithanya Sree[2], L. Vishnupriya[3], K. Chandra Sekhar[4], Jagarlamudi Vishnu Priya[5], Konatham Praveen Kumar[6]

[1]Assistant Professor, Dept. of ECE, S V College of Engineering, Tirupati, A.P, India.

[23456]B.Tech Students, Dept. of ECE, S V College of Engineering, Tirupati, A.P, India.

ABSTRACT

Among the most dangerous cancer in the world is skin cancer. If not diagnosed in early stages it might be hard to cure. The aim of this work is to present a study of skin segmentation, features selection and classification approaches. In the segmentation stage, we will present the result of the use of a pre-processing based on a multiscale decomposition model where geometrical component is used to get a good segmentation. The features are firstly extracted using the texture component and color of the lesion, and then we will present a comparative study of some features selection approaches that select the relevant ones. In feature classification we will compare between the most and good classifiers used in literature.

Keywords: Skin cancer, Segmentation, PDE multi-scale decomposition, features extraction, features classification, SVM, KNN, machine-learning.

INTRODUCTION

Skin cancer is one of the most common form of cancer and its incidence is rapidly increasing in all the world. At a current rate skin cancer can be developed in one of 10 people during their life. Melanoma and non melanoma are two major groups of skin cancer. The non-melanoma usually begins in the squamous cells or basal cells, which are in the external layer of the skin. The extravagant sun exposition of skin produces an irregularity of melanocyte cells that is considered the main factor behind the development of the most cancer cells. Melanoma is the most aggressive and dangerous compared to others types due to its high level of metastasis. Early detection of this skin cancer can help its curability. Images acquisition, pre-processing, segmentation, features extraction and classification, are the essential steps that can be found in computational systems for skin lesions diagnosing. Many methods in imaging have been employed to help dermatologist, which can give a close look at suspicious skin lesion. Dermoscopic and Macroscopic images are widely employed in diagnosing pigmented skin lesions by computational methods. Dermoscopic images are obtained from the dermoscope devices or particular cameras to visualize well the pigmentation pattern on the skin surface. On the contrary, macroscopic images normally recognized as clinical images, are used in analysis of skin lesion. That are obtained from ordinary cameras or mobile devices. Nevertheless, macroscopic images are frequently obtained from different distances or under various illumination conditions; also the images may contain poor resolution, which may lose important information of the lesion like contour, color and diameter of the lesion. Another problem the both dermoscopic and macrocsopic images may happen, which is related to the presence of artifacts like reflections, hair, shadows and skin lines, which may thwart adequate analysis of the imaged skin lesions. The aim of identification of the lesions images is to help in the operation of classification. Pre-processing methods are normally applied to decrease the impacts of unwanted artifacts. Such methods can be based on illumination correction, artifact removal, contrast enhancement and color space transformation; these methods are used to have a better segmentation. Segmentation allows extracting the region of interest (ROI) from images. Many computational suggestions have been proposed to extract features from pigmented skin lesions in images to classify them according to certain criteria. The features used generally are based on the one that dermatologist used to visually identify a lesion as melanoma or not. They all relay on the ABCD (Asymmetry, Border, Color, and Diameter) and texture. So, researcher proposes to use those features to build a model that can easily classify a lesion as benign or malignant. Generally we will have almost 17 features and the question is: are all those features significant for building a good machine learning model. The key point in classification algorithms is to choose the best and relevant features.

Fig. 1: Examples of skin lesions

The objective of this work is to identify the best practices when dealing with skin lesion classification.. Thereafter, features are extracted from the segmented image using a sets of texture and color features. Finally, the features are employed as an input to different machine learning algorithms to classify the skin lesion.

LITERATURE RIVEW

To help dermatologists, computer aided diagnosis (CAD) systems have been used for an early evaluation for skin cancer. Which consists generally on four steps: preprocessing, segmentation, features extraction and classification. In these systems, the step of preprocessing are generally used to increase the accuracy of the segmentation step. Many filters have been applied to remove noise and artifact and to smoothen the skin lesion images, without removing pertinent information about the lesion and to facilitate the detection of lesion border[1], [2]. The segmentation is the harder step and can affect the precision of classification. Here is some of the difficulties that are encountered in skin image segmentation:

(1) the low contrast from the lesion and the healthy skin, (2) the irregularity of the border of the lesion. Therefore, many segmentation algorithms have been applied to resolve these problems. A simple algorithm that have been used on large scale is the Otsu algorithm, which can segment the lesion automatically and separate the region of interest (ROIs) of the input images. The algorithms give a better result when the contrast between the lesion and skin is good. Active contour algorithms have been suggested for the segmentation of skin lesion in images[3]. Region based algorithms, like region growing, splitting and merging method have also been used to segment skin lesions in images. They assemble similar and close pixels, or sub-regions, into more wide identical regions, these algorithms can perform even with complex variation of illumination and color. Many algorithms based on artificial intelligence have been used, which is justified by the advantages they offer, like the possibility of learning from sample case provided by artificial neural networks[4] or also with deep neuronal network[5]. Some disadvantages might occur by algorithms based on AI concerning the complexity of the implementation and the presence of many steps, which demands high computational efforts. Clustering techniques

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

have also been utilized in the image segmentation of skin lesion such as the fuzzy cmeans(FCM) algorithm[6]. In the features extraction step which is based on the identification of the regions of interest (ROIs) in the image, authors in literature, proposed the use of approaches that are similarly used by dermatologists. Asymmetry, Border, Color and Diameter characteristics (ABCD) [7]and the Gray Level Co-Occurrence Matrix (GLCM)[8] approaches are the most used. For the process of classification[3], more than one method have been evaluated to attain the best outcome. The performance depends on several issues like the quality of the segmentation images and the extracted features and the classification algorithms used. Therefore, the output might be binary or multi-class, and concern different classes according to the classification goal.

PROPOSED METHOD

In this part, an approach to classify the pigmented skin lesions is presented, to provide information that may help dermatologists in their diagnosis. Fig explains the given approach, which involves the following steps.

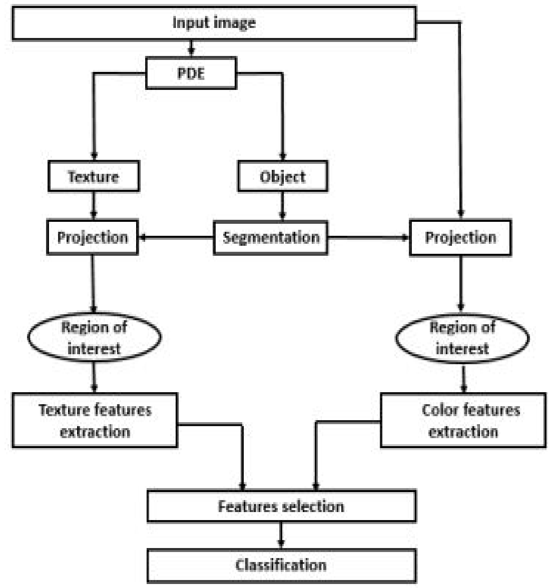

Fig.2: Flow chart of the proposed approach

METHODS OR TECHNIQUES USED

A. Preprocessing: In this section, we will elaborate on the step of preprocessing, which facilitate the border detection, because one of the difficulties of image processing is to remove the artefact like (contrast, hairs, blood vessels, skin lines) therefore we use different methods. The methodology suggested here is to use a multi-scale decomposition model to extract the texture form object. Much progress has been proposed and made for perplexed image and vision problem using partial differential equations (PDE). The objective is to divide the image into two components u and v such as the first component u is well structured and has the geometric information

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

of the image, the second component, v, includes oscillating information both textures and possible noise[9]. Several decomposition models have been proposed to solve the problem of Meyer, in our work we are interested in Aujol model: This decomposition is given by minimizing a functional f where Ȟ1 and Ȟ2 are two functions in R2Í] 0; 1] generalizing the idea of adaptive regularization coefficient according to the image area in which there is. The functions v and w can be seen both as oscillators, taken respectively in Gʒ1 and Gʒ2. The function u will be taken in the space BV.

B. Segmentation: Segmentation is a very important parameter and has a direct influence on the classification result. A poor segmentation will lead to a poor classification result. Even if the best classifier in the literature will perform badly. A correct segmentation will certainly lead us to an acceptable classification rate even with a very basic classifier. In our approach, we do not need a complex algorithm to make the segmentation because of the fact that preprocessing with the PDE has allowed us to eliminate any type of noise. To distinguish between melanoma from benign lesions and get a good classification, we must isolate the lesion from the healthy skin that surrounds lesion. Just a threshold applied on the object give a good result. The Otsu segmentation techniques reduce the gray level image to a binary image, with two classes foreground and background [10].

C. Featuresengineering: Features used in our approach to characterize the skin lesion are texture and color as in the ABCD rule used by dermatologists. Therefore, 17 features can be identified but not all of them are relevant. Features selection procedure should be called to identify the relevant features needed to represent efficiently the lesion instead of using all the 17 features. At the end, a normalization step is required to make equal the features range of values.

1) Texture features: In order to quantify the texture present in a lesion, a set of statistical texture descriptors were employed. These statistics features are:

• Contrast: Sum of Square Variance is the name of Contrast. It defers the calculation of the intensity contrast-linking pixel and its neighbor over the whole image. At constant image contrast value is 0.

• Correlation: It calculate the correlation of a pixel and its neighbor over the whole image means it figures out the linear dependency of gray levels on those of neighboring

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

pixels. On behalf a perfectly positively or negative correlated image, the correlation value is 1 and -1.

• Entropy: it is a statistical measure of randomness that can be used to characterize the texture of the input image.

• Inverse Difference Moment (IDM): It is a measure of image texture. IDM ranges from 0 for an image that is highly textured to 1 for an image that is untextured.

• Smoothness: It measure the grey level contrast that can be used to establish descriptor of relative smoothness.

. **2) Color features**: In order to calculate the color that the lesion contains, according the literature the melanoma are described by the presence of six different colors that are, white, red, light brown, dark brown, blue-gray and black. So, to know the number of color, we performed the calculation R is the red, G is green and B is blue component in the RGB representation. We add all the color to have the number of color that contain the image.

3) Feature selection: Feature selection is a process to select the least number of features needed to represent the data precisely. This task is very important to eliminate some irrelevant attributes and to give us the most useful features to improve the score of classification algorithms. In this study we will focus on 4 features selection methods.

• ReliefF: this algorithm tries to determined the nearest neighbors from a number of samples that are selected randomly from the data set. For every selected sample, values of the features are compared with those of the nearest neighbors and the scores for each feature are updated. The idea is to estimate the quality of attributes according to how well their values distinguish between samples that are near to each other.

• Correlation based Feature Selection (CFS): this algorithm tries to find a set of features that individually correlate well with the class but have little intercorrelation.

• Recursive Feature Elimination (RFE): a multivariate mapping method based on the idea to repeatedly construct a model and choose either the best or worst performing, RFE can detect the discriminative features. The basic principle of RFE is to include initially all voxels of a large region, and to gradually exclude voxels, that do not contribute in discriminating features from different classes.

4) Feature normalization: In classification tasks, the features that characterize the models often have different ranges. Many classifiers need that the features be normalized because their values fall within a specified range. One of the most common normalization methods is the standard score (z-score) transformation: Z=(x-ʒ)/ı (8) Where ʒ is the mean (average) and ı is the standard deviation from the mean, this method gives a range of values between 0 and 1.

D. classification: At lot of machine learning approaches for classification of skin cancer currently exist, but few comparison of different machine learning have been done for the same data sets. On our approaches, we compare four classifiers: Logistic regression, Decision Trees, K-nearest neighbors and Support vector machine with different kernel.

1) Logistic regression: The logistic regression is an algorithm that build a separating hyperplane between two data using the logistic function to define distance from the hyperplane as a probability of the class member.

2) Decision Trees: The decision tree model constructs classifiers by dividing the data set into uniform and smaller groups, based on measuring the entropy. It can be divide the data into two groups by identifying a threshold and a variable in the domain, which minimizes the disparity in the resulting groups.

3) Support vector machine:Support vector machine is a machine learning an algorithm based on statistical learning theory. This algorithm calculate the separating hyperplanes that maximize the margin between two data set. By using Lagrange multipliers, the problem can be formulated in such a way that the only operations on the data points are the calculation of scalar products. While the basic training algorithm can only construct linear separators, kernel functions can be used to calculate scalar products in higher- dimensional spaces.

CONCLUSION

In order to help dermatologists in their diagnosis, many solutions in image processing and analysis have been proposed. The major work that we performed in this approach in this paper is to provide a comparison of different classifiers, and this help us improve the classification process. This approach is based on the model decomposition for structure and texture identification, Otsu for segmentation, texture and color for features extraction, ReliefF, Correlation based Feature Selection (CFS), Recursive Feature Elimination (RFE), chi2 method as a features selection and SVM with quadratic kernel as a classifier. The proposed approach have been implemented and tested on medical images. This later achieved good segmentation results as well as good classification accuracy.

FUTURE SCOPE

As future work, concerning the segmentation and classification of skin lesion, we intend to use new algorithms and methods to develop more our systems of a higher quality diagnosis on dermatology images like deep learning algorithms.

REFERENCES

1. L. Smith and S. MacNeil, "State of the art in non-invasive imaging of cutaneous melanoma," Ski. Res. Technol., vol. 17, no. 3, pp. 257–269, 2011.
2. R. Garnavi, M. Aldeen, M. E. Celebi, G. Varigos, S. Finch, A. Bhuiyan, and C. Dolianitis, "Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels," Conputerized Med. Imaging Graph., vol. 35, no. January, pp. 105– 115, 2011.
3. R. B. Oliveira, N. Marranghello, A. S. Pereira, and J. M. R. S. Tavares, "A computational approach for detecting pigmented skin lesions in macroscopic images," Expert Syst. Appl., vol. 61, pp. 53– 63, 2016.
4. S. Mathew and D. Sathyakala, "Segmentation of skin lesions and classification by neural network," In Technology, Communication and Knowledge (ICTCK), 2015 International Congress on, pp. 337-341. IEEE, 2015. vol. 4, no. 2, 2015.
5. J. Ko, S. M. Swetter, H. M. Blau, A. Esteva, B. Kuprel, R. A. Novoa, and S. Thrun, "with deep neural networks," Nature, pp. 1–11, 2017.
6. M. A. Sabri, A. Ennouni, and A. Aarab, "Automatic estimation of clusters number for K-means," 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt). Pages: 450 - 454, DOI: 10.1109/CIST.2016.7805089. Electronic ISSN: 2327-1884. 24-26 Oct. 2016.