Detection of Skin Cancer Using Machine Learning

Tulasi Nakka

Abstract: **In recent days, skin cancer is seen as one of the most Hazardous forms of the Cancer found in Humans. Skin Cancer is classified into various types such as Melanoma, Basal and Squamous cell Carcinoma among which Melanoma is the most unpredictable. The detection of Melanoma cancer in an early stage can be helpful to cure it. Computer vision can play an important role in Medical Image diagnosis and it has been proved by many existing systems. In this paper, we present a computer aided method for the detection of Melanoma Skin Cancer using Image Processing tools. The input to the system is the skin lesion image and then by applying novel image processing techniques, it analyses it to conclude about the presence of skin cancer. The Lesion Image analysis tools perform the Pre-Processing, Segmentation, Edge detection and Feature extraction of the lesion image. The extracted feature parameters are used to classify the image as either a normal skin or Melanoma cancer lesion.**

Keywords: Computer vision, Edge Detection, Feature Extraction, Image processing tools, Image Segmentation ,Melanoma, Pre-Processing, Skin Cancer .

1. Introduction

Skin cancer is the out-of-control growth of abnormal cells in the epidermis, the outermost skin layer, caused by unrepaired DNA damage that triggers mutations. These mutations lead the skin cells to multiply rapidly and form malignant tumours. The main types of skin cancer are basal cell carcinoma (BCC), squamous cell carcinoma (SCC), melanoma and Merkel cell carcinoma (MCC).

Basal cell carcinoma is the most common cancer in humans. Over 1 million new cases of basal cell carcinoma are diagnosed in the U.S. each year. There are several different types of basal cell carcinoma, including the superficial type, the least worrisome variety; the nodular type, the most common; and the morph form, the most challenging to treat because the tumours often grow into the surrounding tissue (infiltrate) without a well-defined border. **Squamous cell carcinoma** accounts for about 20% of all skin cancers but is more common in immunosuppressed people. In most instances, its biologic behaviour is much like basal cell carcinoma with a small but significant chance of distant spread.

Melanoma mostly consists of various colours from shades of brown to black. A small number of melanoma are pink, red or fleshy in colour; these are called amelanotic melanoma and tend to be more aggressive. Warning signs of malignant melanoma include change in the size, shape, color or elevation of a mole. Other signs are the appearance of a new mole during adulthood or pain, itching, ulceration, redness around the site, or bleeding at the site. An often-used mnemonic is "ABCDE", where A is for "asymmetrical", B for "borders" (irregular: "Coast of Maine sign"), C for "color" (variegated), D for "diameter" (larger than 6 mm – the size of a pencil eraser) and E for "evolving."

Nowadays people are suffering from skin diseases, more than 125 million people are suffering from skin cancer and it's rate is rapidly increasing over the last few decades specially Melanoma is most diversifying skin cancer. Dermatophytosis rate is high especially in rural areas. If skin diseases are not treated at an earlier stage, then it may lead to complications in the body including spreading of the infection from one individual to the other. The skin diseases can be prevented by investigating the infected region at an early stage. The characteristics of the skin images are diversified, so that it is a challenging job to devise an efficient and robust algorithm for automatic detection of the skin disease and its severity. Skin tone and skin colour plays an important role in skin disease detection. Colour and coarseness of skin are visually different. Automatic processing of such images for skin analysis requires quantitative discriminator to differentiate the diseases. Proposed system is a combo model which is used for the prevention and early detection of skin cancer, basically skin disease diagnosis depends on the different characteristics like colour, shape, texture etc. there is no accepted treatment for skin diseases Different physicians will treat differently for the same symptoms. Key factor in skin diseases treatment is early detection and then further treatment is reliable. In this proposed system we implemented for the diagnosis of skin lesions using statistical and GLCM parameters. Statistical analysis parameters include some properties like mean, standard deviation, median etc. Standard database of images is used; this data does not have any mathematical expression, it has only two sets of images data with benign and melanoma. The GLCM features have the texture properties of images like the energy, entropy, correlation etc. To analyse the skin lesion we have considered both statistical and GLCM properties in our systems.

The main objective of proposed system is to:

ISSN 2229-5518

i)To develop an algorithm for efficient extraction and classification of pigmented skin lesions.

ii)To ease diagnosis and treatment of skin patients by means of automation and provide for a cost effective way of treatment.

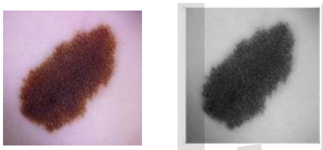
iii)To improve the speed of diagnosing pigment skin lesions in different types of skin cancers.

Figure 1: Benign and Malignant Melanoma Image of Skin

2. Components of Methodology

2.1 Pre-Processing:

Pre-processing techniques are needed on colour, grey-level or binary document images containing text and/or graphics. In character recognition systems most of the applications use grey or binary images since processing colour images is computationally high.



Skin Image Pre-processed Image

Figure 2: Pre-processed skin Image

2.2 Segmentation:

In computer vision, **Image segmentation** is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. The popular techniques used for image segmentation are: **thresholding method, edge detection based techniques, region based techniques, clustering based techniques, watershed based techniques, partial differential equation based and artificial neural network based techniques etc**.

Segmentation Edge Detection

Figure 3: Segmented and Sobel Edge detection of skin image

2.3 Feature Extraction:

Feature extraction is a method of capturing visual content of images for indexing and retrieval. Feature extraction is used to denote a piece of information which is relevant for solving the computational task related to a certain application. There are two types of texture features measured, they are first order and second order. In the first order, texture measures are statistics calculated from an individual pixel and do not consider pixel neighbour relationships. The intensity histogram and intensity features are first order calculation. In the second order, measures consider the relationship between neighbour relationships.

a)GLCM Features

A grey level co-occurrence matrix (GLCM) or co-occurrence distribution (less often co-occurrence matrix or co-occurrence distribution) is a matrix or distribution that is defined over an image to be the distribution of co-occurring values at a given offset. A GLCM is a matrix where the number of rows and columns is equal to the number of grey levels, G, in the image.

Contrast: It returns a measure of the intensity contrast between a pixel and its neighbour over the inter image. Range = [0 (size (GLCM, 1)-1) ^2] Contrast is 0 for a constant image.

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j)$$

Correlation: It returns a measure of how correlated a pixel is to its neighbour over the inter image. Range= [-1 1]; If Correlation is 1 positive relation if correlation -1 negative relation if correlation is 0 there is no relation between pixels of image.

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \ \sigma_j}$$

Energy: It returns the sum of squared elements in the GLCM. Range = [0 1] Energy is 1 for a constant image

$$\text{Energy} = \sum_{i,j} p(i,j)^2$$

Homogeneity: It returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Range= [0 1]. Homogeneity is 1 for a diagonal GLCM.

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

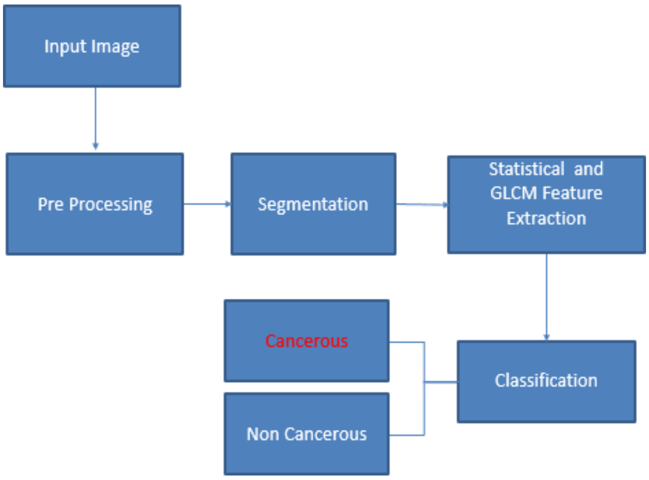
b)Statistical Features

We have extracted the **mean, median, mode , variance and standard deviation** features of a skin image.

2.4 Classification:

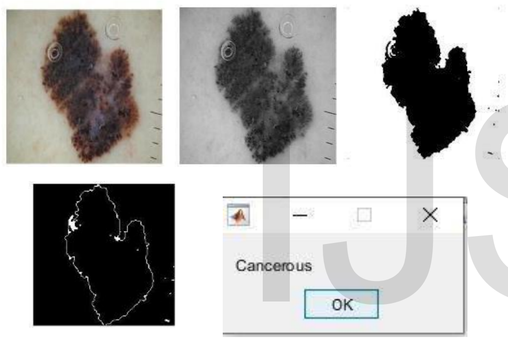
The images have been trained with a designed Machine Learning Model where many classification algorithms have been used to predict the best classification model.

3. Proposed System

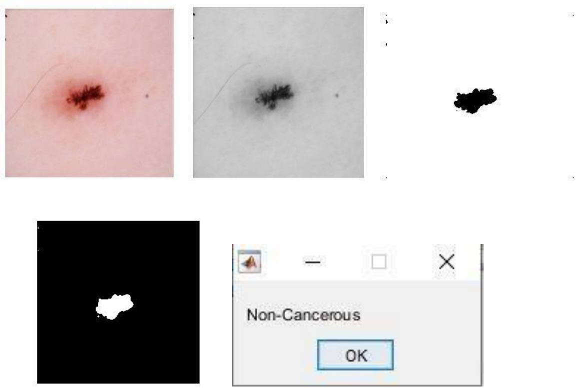


4. Results and Outcomes

Case 1: Processing of Melanoma Image



Case 2: Processing of Non-Melanoma Image



5.Conclusion

The Proposed model takes input as the skin image. The image is then processed through various Image Processing Techniques like grayscale conversion, contrast adjustment, filtering etc through MATLAB software. Hence the blob area/part is detected in the given image through the edge detection applied to the segmented image. In this System we have used five Machine Learning Algorithms like Support Vector Machine, Random Forest, Gradient Boosting, Naïve Bayes, Decision Trees to develop a skin cancer detection system, which includes the models built using the above mentioned Algorithms.

Algorithms Used	Accuracy
SVM	73%
Random Forest	71%
Gradient Boosting	**76%**
Naïve Bayes	64%
Decision Trees	57%

References

1.Shivangi Jain, Vandana jagtap, Nitin Pise,**"Computer aided Melanoma skin cancer detection using Image Processing"**,International Conference on Intelligent Computing, Communication & Convergence (ICCC2015), Elsevier - 2015.

2. VIjayalakshmi M M **"Melanoma Skin Cancer Detection using Image Processing and Machine Learning"**, Assistant Professor, Department of Information Science &Engineering GSSSIETW Mysuru, Karnataka, India.

3. Uzma Bano Ansari,Tanuja Sarode,**"Skin Cancer Detection Using Image Processing"**, International Research Journal of Engineering and Technology (IRJET), Volume: 04,Issue: 04, Apr-2017.

4. Wilson F. Cueva, F. Muñoz, G. Vásquez., G. Delgado, **"Detection of skin cancer ―Melanoma‖ through Computer Vision"**, 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), IEEE 2017.

5. Vedanti Chintawar and Jignyasa Sanghavi **"Computer - Aided Melanoma Skin Cancer Detection using Image Processing"**.

6. Farzam Kharaji Nezhadian, Saeid Rashidi, **"Melanoma skin cancer detection using color and new texture features"**,2017 Artificial Intelligence and Signal Processing (AISP), IEEE 2017.

7. A. Bono, S. Tomatis, C. Bartoli , **"The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry border, color, and dimension Cancer"**, 85 (1) (January 1999), pp. 72-77 View Record in ScopusGoogle Scholar.

8. H. Pehamberger, M. Binder, A. Steiner, K. Wolff In vivo epiluminescence microscopy: **"Improvement of early diagnosis of melanoma"**, J. Invest Dermatol, 100 (1993), pp. 356S-362S CrossRefGoogle Scholar.

9. M.L. Bafounta, A. Beauchet, P. Aegerter, P. Saiag, **"Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests Arch Dermatol"**, 137 (13) (2001), pp. 43-50 View Record in ScopusGoogle Scholar.

10. G. Argenziano, H. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, **"Dermoscopy of pigmented skin lesions"**: Results of consensus meeting via the Internet Journal of the American Academy of Dermatology, 48 (2003), pp. 679-693 ArticleDownload PDFView Record in ScopusGoogle Scholar.

11. R. Garnavi **"Computer-aided diagnosis of melanoma"**, Ph.D. dissertation University of Melbourne, Australia (2011) Google Scholar [6] M.E. Celebi, H. Iyatomi, G. Schaefer, W.V. Stoecker Lesion border detection in dermoscopy images Computerised Medical Imaging and Graphics, 33 (2) (2009), pp. 148-153 ArticleDownload PDFView Record in ScopusGoogle Scholar.

12. H. Iyatomi, H. Oka, M. Saito, A. Miyake, M. Kimoto, J. Yamagami, S. Kobayashi, A. Tanikawa, M. Hagiwara, K. Ogawa, G. Argenziano, H.P. Soyer, M. Tanaka **"Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for anautomatic melanoma diagnostic system Melanoma Research" ,** 16 (2) (2006), pp. 183-190 CrossRefView Record in ScopusGoogle Scholar.

13. R. Garnavi, M. Aldeen, M.E. Celebi, A. Bhuiyan, C. Dolianitis, G. Varigos **"Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels"**, International Journal of Medicine and Medical Sciences, 1 (2) (2010), pp. 126-134 View Record in ScopusGoogle Scholar.

14. V. Ng, B. Fung, T. Lee ,**" Determining the asymmetry of skin lesion with fuzzy borders Computers in Biology and Med".,** 35 (2005), pp. 103-120 ArticleDownload PDFView Record in ScopusGoogle. Scholar

15. M. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, R. Moss, **"A methodological approach to the classification of dermoscopy images Computerized Medical Imaging and Graphics"**, 31 (2007), pp. 362-373 ArticleDownload PDFView Record in ScopusGoogle Scholar.

Author Profile

Tulasi Nakka received the B.Tech degree in Computer Science and Engineering from Vignan's Institute of Technology in 2020.