METHODS

Skin cancer recognition using SVM image processing technique

Rahul Chand Thakur and Vaibhav Panwar

Department of Computer Science, IILM, Greater Noida, Uttar Pradesh, India

According to some estimates, skin cancer is the leading cause of death for people in the modern period. Although it can happen anywhere on the body, the majority of the time, this sort of cancer targets particular body regions that are more prone to being exposed to light. Skin cells proliferate in an erratic or patchy manner as a symptom. The majority of skin cancers can be treated if they are discovered early. Therefore, a patient's life can be saved by promptly and precisely diagnosing skin cancer. Modern technology has made it feasible to detect skin cancer in its earliest stages. Skin cancer can be diagnosed systematically using the biopsy procedure (1). Skin cells are removed and supplied as a sample, following which they are evaluated in several laboratories. It's an extremely unpleasant and time-consuming process. For primitive detection of skin cancer disease, we proposed a skin cancer detection system based on support vector machine (SVM). It is more helpful to patients. Various methods of image processing and the supervised learning algorithm called SVM are used in the identification process. In particular, epiluminescence microscopy is utilized using an image and several preprocessing techniques which are used in the reduction of sound artifacts and improve the quality of images. Segmentation is done by using certain thresholding techniques like Otsu. The gray level co-occurrence matrix technique must be used to remove certain image features. These characteristics are fed into the classifier as input. The supervised learning model called SVM is used to distinguish data sets. It determines whether a picture is cancerous or not.

Keywords: Otsu thresholding, supervised learning model, GLCM, skin cancer, classifier, SVM

Introduction

Skin has a high fatality rate compared to other diseases. There are three basic layers to the skin. The outermost layer of the skin, which is made up of three layers, is where skin cancer starts: the innermost layer, which is made up of melanocytes, is made up of basal cells, which are the second layer, and the first layer, which is composed of squamous cells. Examples of cancers that are not melanoma include squamous cell carcinoma and basal cell carcinoma. Non-melanoma skin cancer is largely curable and rarely metastasizes to other body organs. A type of skin cancer called melanoma is riskier than the majority of other skin malignancies (2). If it is not discovered right away, it quickly spreads to other parts of the body and infiltrates neighboring tissues. A biopsy is an accepted official diagnostic procedure for detecting skin cancer. During a biopsy, a small sample of the patient's tissue

or cells is taken and sent to a laboratory for analysis. It is a cumbersome procedure. Since research takes a long time, the biopsy process is time-consuming for both the patient and the doctor. During a biopsy, skin tissues are removed, and the sample is then subjected to a number of tests in the laboratory (1). There is a possibility that the infection will spread to other body regions. It is more dangerous. In light of the aforementioned scenarios, skin cancer detection via support vector machine (SVM) is proposed. For classification, this approach employs SVM and digital image processing techniques. This method has led to the primitive detection of skin cancer because it does not involve applying oil to the skin to obtain clear, enhanced images of your moles. It is a simple and safe process this way. Most notably, because of the higher magnification, skin cancer identification is more precise. SVM can help to avoid removing perfectly benign moles and skin lesions that might otherwise be excised.
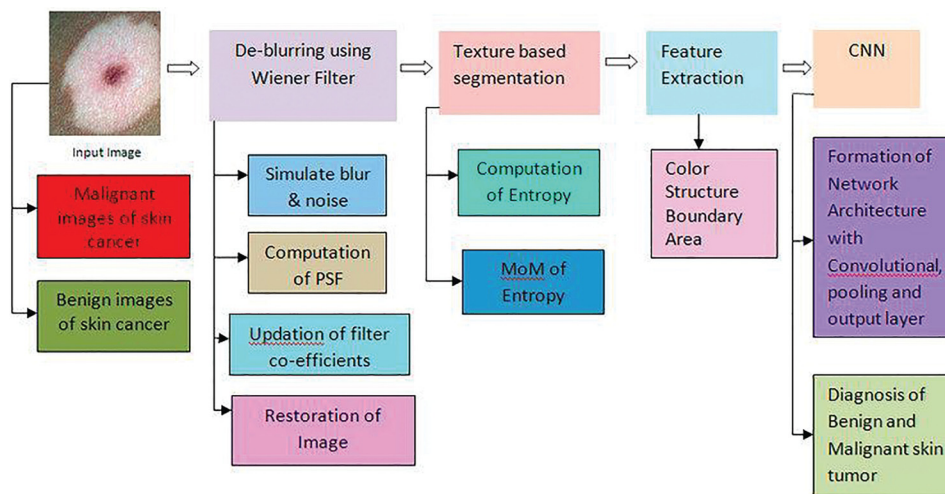
Literature review

Skin cancer detection is based on static filters known as maximum entropy, Otsu thresholding, feature extraction by gray level co-occurrence matrix (GLCM), and classification by artificial nerve networks (ANN).

Abdul Jaleel (1) proposed the concept of ANN. For classification, a back propagation neural network (BPN) is employed.

Chaitanya Krishna (3) endorsed that the ABCD (asymmetry index boundary color index diameter) method is used as a clustering technique to extract features from the segmentation.

According to Amarathunga (4), rule- and chain- based strategy is used to identify and detect the skin diseases. The proposed machine allows users to recognize children's and adults' skin-related diseases and provide helpful medical advice through the internet.

To predict and diagnose the skin disease, researchers used various data mining classification algorithms, such as MLP, Naïve Bayes, and AdaBoost. However, just three skin diseases (eczema, impetigo, and melanoma) respond to this treatment (4).

In his article, Kawar Ahmed (5), used a variety of data preprocessing techniques, disease diagnosis, a maximum frequent item algorithm for planning, and segmentation is done using K-Means Clustering algorithm, and important consistent patterns for classification. Mariam (6), describe in their paper a melanoma diagnosis approach based on a series of digital images. The multilayer classifier (ML), often known as GLCM, also known as the co-occurrence matrix for gray levels, were obtained to discriminate between cancerous and non-cancerous tissues.

Proposed system

Skin cancer diagnosis is the practice of using SVM to recognize the appearance of cancer cells in an image. The GLCM and SVM are used to detect skin cancer. GLCM is used to extract image features that can be used for classification as well as second-order statistical features. Machine learning techniques like SVM are often used for classification and regression analysis.

Implementation details

Input image

Dermoscopic images, which are images taken with a dermatoscope, are used as input to the proposed system. It's a magnifier that's used to photograph lesions on the skin (body part). It's a portable device that makes diagnosing skin diseases a lot easier.

Preprocessing

Preprocessing aims to enhance some crucial image features and remove some unnecessary errors to improve the image data for further image processing. There are three major aspects of image preprocessing: (1) Conversion to gray scale, (2) Noise reduction, and (3) Enhancement of the image.

Gray scale conversion

Brightness is the only element of detail in a gray scale image. In a gray scale image, every pixel represents a specific quantity or amount of light. The brightness gradient in a gray scale image can be distinguished. All that is evaluated
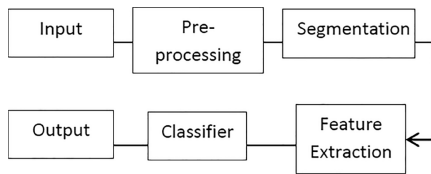
FIGURE 1 Block diagram.

FIGURE 2 Input image.

FIGURE 3 Gray scale image.

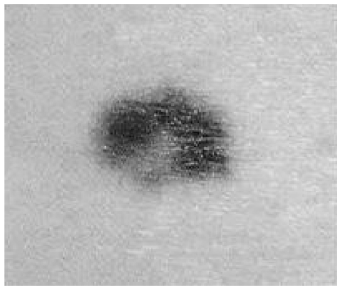**FIGURE 4** Image without noise.

FIGURE 5 Enhanced image.

in a gray scale image is light intensity. Red, blue, and green (RGB) colors are coded on 256 levels, starting from 0 to 255. The gray scale conversion is the process of converting a color image into a gray scale image, as shown in **Figure 3**. Compared to processing color photos, processing gray scale images is simple and quick. On a gray scale image, all image processing techniques are used (7).

The following equations describe the process we suggest using the weighted sum technique to turn an RBG image gray level picture matrix to capture the most frequent features, including contrast, mean, energy, and homogeneity, in the foreground and context (8).

$$Grayscale intensity = 0.299R + 0.587G + 0.114B$$

Noise reduction

Noise reduction is the technique of identifying and eliminating undesirable noise from a digital image. The difficulty is to differentiate between the correct aspects that are required for further bifurcation and those that have to be treated as noise. The word "noise" refers to the unpredictability of pixel values.
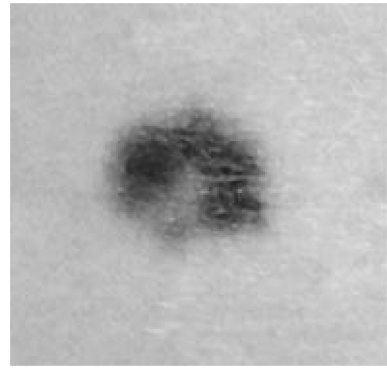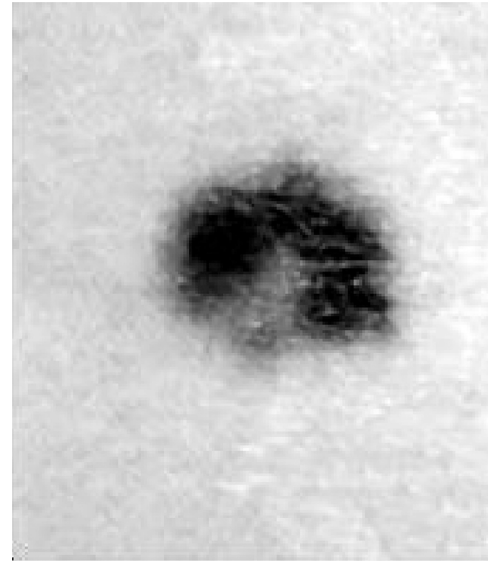
As shown in **Figure 1**, we use a median filter in our proposed method to eliminate unnecessary noise (4). A non-linear filter like a median filter has invariant sharp edges. A sliding window of an odd length is used to apply a median filter (7). To provide filtered output, each sample value is sorted by magnitude, with the center value representing the median of the sample contained within the window.

Image intensification

Image intensification aims to make a photograph's main feature more visible. To obtain a higher quality result in this case, contrast enhancement is used as shown in **Figure 5**.

Segmentation

Segmentation is the method of eliminating an image's region of interest. Each pixel has similar attributes in a region of interest. For segmentation, we use maximum entropy
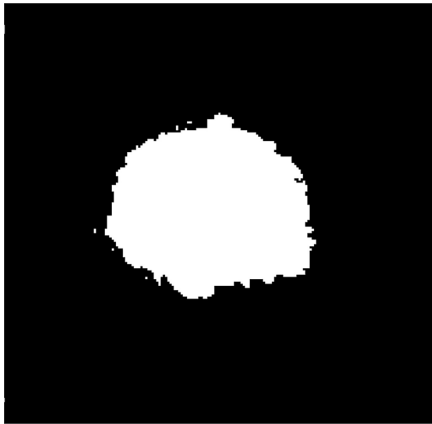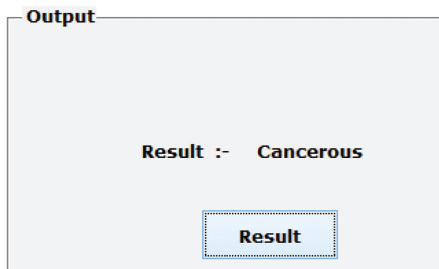
FIGURE 6 Segmented image.

FIGURE 7 Snapshot of output.

thresholding (9). To begin with, we must first determine the original image's gray level, then calculate the gray scale image's histogram, and finally, using maximum entropy, separate the background from the context. A binary image–a white and black image–is created after acquiring static filters such as maximum entropy, as illustrated in **Figure 6**.

Feature extraction

Feature extraction is crucial to extract information from a given image. For texture image analysis, we're using GLCM. This perceptual relationship between the image pixels is identified by the GLCM method. GLCM uses the

$$\Sigma_i \Sigma_j (i-j)^2 C(i,j) \tag{1}$$

$$\Sigma_i \Sigma_j C(i,j)^2 \tag{2}$$

$$\Sigma i \Sigma j \frac{C(i,j)}{1+|i-j|} \tag{3}$$

$$\Sigma m \Sigma n C(i,j) m * n \tag{4}$$

An image data set helps in image recognition by measuring specific attributes or values using a suitable technique called feature extraction. A classifier is used to differentiate between cancerous and non-cancerous images. For the sake of consistency, we used a supervised learning model called the SVM. This model examines a collection

TABLE 1 | Support vector machine (SVM) performance.

Parameters	Support vector machine classifiers
True positive value	17
True negative value	04
False positive value	1
False negative value	2
Accuracy rate	

of images and determines if each image falls into the malignant or non-cancerous group. The purpose of SVM is to build a hyperplane that separates the two groups with the least difference (8). Feature extraction (GLCM) is a technique for minimizing the size of an image data set by measuring specific values or attributes that help with picture identification (9).

Results

On the internet, we came across pictures of skin cancer. The images were produced after preprocessing with methods such as gray scale conversion, static filters (median), such as maximum probability, and the GLCM system. SVM was used to differentiate malignant and non-melanocyte non-cancerous images as shown in **Figure 2**.

$$\text{Accuracy Rate} = \frac{\text{True Positive Value} + \text{True Negative Data Value}}{\text{True Positive Value} + \text{False Positive Value} + \text{False Negative Value} + \text{True Negative Value}} \tag{5}$$

Accuracy Rate using above formula = 95%

Conclusion

To swiftly assess whether or not an image has skin cancer, a GLCM and a statistical model learning method known as SVM can be utilized. However, 95% accuracy was achieved for the machine that was developed. In contrast to biopsy, it is a painless and long-lasting treatment. It's more useful to the patients.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

References

1. Abdul Jaleel J, Sibi S, Aswin RB. Computer aided detection 01 skin cancer. *Proceedings of the International Conference on Circuits, Power and Computing Technologies*. (2013).

2. Santosh A, Sadashivappa G. Statistical analysis of skin cancer image–a case study. *Int J Electron Commun Eng*. (2014) 3.

3. Chaithanya Krishna M, Ranganayakulu S. Skin cancer detection and feature extraction through clustering technique. *Int J Innov Res Comput Commun Eng* (2016) 4.

4. Amarathunga AALC. Expert system for diagnosis of skin diseases. *Int J Sci Technol Res*. (2015) 4.

5. Kawar Ahmed. (2013) 12.

6. Mariama S. Automatic detection of melanoma skin cancer. *Int J Comput Appl*. (2012)

7. Gonzlae RC, Wood RE. *Digital image processing" by jayaraman*. (Vol. 4). (2017). p. 270–3.

8. Nageswara Rao C, Sreehari Sastry S, Mahalakshmi KB. Co- occurrence matrix and its statistical features an approach for identification of phase transitions of mesogens. *Int J Innov Res Eng Technol*. (2013) 2.

9. Parker JR. *Algorithm for Image Processing and Computer Vision*. (2013). p. 142–5.

10. Kawsar A, Tasnuba J. Early prevention and detection of skin cancer risk using data mining. *Int J Comput Appl*. (2013) 62 .

11. Anshu B, Sonajharia M. *Discretization Based Support Vector Machines*. Indian Agriculture Statistics Research Institute (2009).

12. Maurya R, Surya KS. GLCM and multi class support vector machine based automated skin cancer classification. *IEEE J*. (2014) 12.