Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8, July 2021: 108-116

An Introductory Approach of Recognition of Melanoma Skin Cancer Recognition in early stages using Support Vector Machine with Various Kernel

[1] Rahul Thakur ,[2] Dr. Vaibhav Panwar
[1]M.TECH , Department of Computer Science, IILM Greater Noida
[2]Associate Professor, Department of Computer Science, IILM Greater Noida

Abstract
In today's world, skin cancer is seen as the most prevalent cause of death in people. This type of cancer is not consistent or patchy, but it can occur everywhere in the body, with skin cells that usually arise on particular sections of the body that are more likely to be exposed to light. When identified early, most skin cancers can be treated. This means that a patient's life is saved by discovering skin cancer early and readily. Modern technologies have made it feasible to early detect skin cancer at an early stage.The biopsy procedure [1] is a method for diagnosing skin cancer that is done in a systematic manner. It is accomplished by harvesting skin cells, which are then sent to various laboratories for analysis. It's a long and unpleasant procedure (in terms of time). We suggested a skin cancer detection method based on svm for primitive detection of skin cancer disease. It's more beneficial to the patients. The identification approach employs a variety of image processing techniques as well as the supervised learningalgorithm Support Vector Machine (SVM).Microscopy of epiluminescience is done using an image and in particular with numerous pre-processing procedures used to reduce sound objects and improvise image quality. Certain thresholding techniques, such as OTSU, are used to segment data. To erase particular visual features, the GLCM approach must be utilised. As input, these properties are sent into the classifier.To identify data sets, the supervised learning model (SVM) is used. It assesses whether or not a photograph is malignant using various different kernel with their accuracies.

Keywords: Super Vised Learning Approach,Thresholding,Feature Extraction, GLCM, Skin cancer, Classifier, Linear, RBF, Gaussian Kernel

1.Introduction

Childhood cancer is an illness that threatens life. There are three fundamental layers of skin.. The outermost cover, which consists of the first-school squamous cells, the second-schicht basal cells and the innermost layer of the melanocyte cells, is at the onset of skin cancer. Squamous and basal cell carcinomas are malignant non-melanoma. Skin cancer that isn't melanoma is usually curable and seldom spreads to other parts of the body. Melanoma is a type of skin cancer that is more lethal than most others [3]. The surrounding component or structure can readily be penetrated and extended to other areas of the body, unless it is early discovered. Biopsy is a formal tool for detecting skin cancer. A biopsy involves taking a small portion of the tissue or cell sample from the body of a patient for laboratory examination. It's an unpleasant process. The process of biopsy takes time for both the patient and the doctor since it takes a long time to conduct research.Biopsy is carried out through skin scraping (tissues) and testing in a laboratory[1]. The sample is subjected. The sickness is likely to spread to other sections of the body. It's more risky. Given the above situations, detection of skin cancer by svm is suggested. This methodology uses SVM and

digital imaging techniques for classification. The primal identification of skin cancer has resulted from this procedure, since it does not require the application of skin oil to create clear, improved photographs of your moles. This is a straightforward and secure process. In particular, Skin Cancer Identification is more precise due to the higher magnification. The SVM can help prevent the removal, otherwise eliminated, of totally harmless lumps and skin blemishes.

2.Literature Review

For patients, early detection of skin cancer is vital, and most of them experienced confusion to distinguish between skin cancer and eruptions.S.Subha[1] is intended to enable someone seek the assistance needed or prevent concern about the non-cancerous rash, while learning about the differences between eruptions and skin disease.The Otsu thresholding, a function of grey level co-occurrence matrix (GLCM), and an artificial nerve network classification, was proposed by J Abdul Jaleel[2013] to identify skin cancer through static filters known as maximal entropy. (ANN). (ANN). The classification is carried out using the Back Propagation Neural Network (BPN). [2]
The ABCD rule is the scoring technique, with paper extracted asymmetry, colour, border and diameter [6]. Rule and chain technique for identifying and detecting skin diseases is utilised in this article[8]. The equipment offers users the possibility of accurately recognising skin conditions for toddlers and adults and providing helpful internet information. Researchers have employed different data mining cluster algorithms such as MLP, Naïve Bayes and AdaBoost to predict and diagnose skin illness. This treatment is responded by just three skin conditions (Eczema, Impetigo, and Melanoma) [8]. The extracted feature may be transferred directly onto the neural network, the success rate reached by GLCM is 95.83 percent.To distinguish between cancerous and noncancerous the figure, (GLCM) or gray level co-occurrence matrix and the classifier (ML) also known as multilayer perception were extracted using the characteristics [9].
A deep Sparse auto encoder extracts features. For assessment, we use photos from the PH2 dataset to test the suggested methodology. In terms of specificity, sensitivity and precision, results demonstrate remarkable performance in this paper[5].The identification of insulin resistance can be done without clinical processes with the development of machine learning technologies. In this work, insulin resistance is discovered by non-invasive ways using machine-learning methodologies for people with triglycerides and HDL-c ratios proposed in this paper[13].For the clustering of keys and removing false matches, DBSCAN (Density-Based Spatial Clustering Noise Applications) is employed. In the end, the PSNR and the morphology processing are employed for the exact location of the disturbed areahad been proposed by M.Kumar in paper[14].

A comparative analysis of the principal component analysis (PCA), using the Eigen faces, Hidden Markov (HMM) model using Single Value Decomposition and the Artificial Neuronal Network (ANN) using the Gabor filters, was presented in paper[15]. The researchers used three commonly known facial recognition techniques. The frame segmentation is carried out using fluid C-means, which are clustered into individual homogeneous picture regions. Different filters are used to upgrade the qualities of the image, while the other features are evaluated by implementing rgb colour space, Local Binary Pattern (LBP) and GLCM approaches. The simulated results suggest that the approach described effectively identifies cancer of the skin and results in a paper precision of 97.4% in paper [16].M. Kumar has devised a technique to detect forgeries in the light angle α derived from a light source and a regular surface. This approach is evaluated under specific established parameters for outdoor and indoor photos. The new feature of our methodology is the use of many light source detection procedures in paper [17] for the identification of picture modification.

3.Proposed System

SVM is mostly known as skin cancer detection to detect the appearance of cancer cells. To make a right decision limit, you might define several kernel functions. It works well on a dataset with multiple characteristics. GLCM is used to identify skin cancer with the Vector Support Machine (SVM). (GLCM) is utilised both for the extraction of

statistical features of the second order and for the extraction of the classification picture characteristics. SVM is a sort of machine learning technology frequently used for the analysis and classification of regression. SVM training is faster than any other nucleus with a linear kernel. Only the optimization of the C Regularisation parameter is needed for training the SVM with a Linear Kernel. On the other side, when training with various kernels, we will also propose optimising the Ț parameter which normally takes more time to explore the grid.

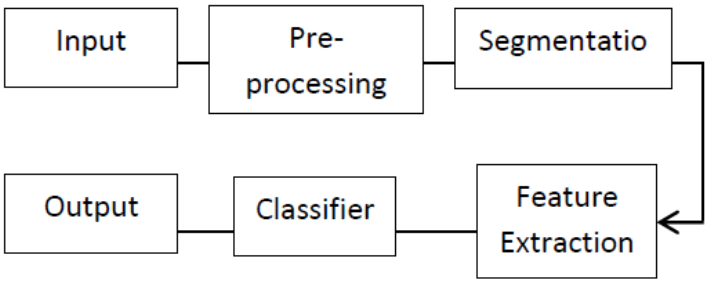


Fig:1-Block Diagram

Implementation Details

Fig 1. Block Diagram of Image Recognition

a) Input image

Dermoscopic images are employed as inputs to the proposed system, which are images captured using a dermatoscope. It is a lupin used to take pictures of skin blemishes (body part). It is a portable instrument which makes it much easier to diagnose skin problems.

b)Preprocessing

Pre-processing is intended to improve image data by minimising superfluous distortions and improving certain critical image characteristics for further image processing. The pre-processing of images takes three important components. 1) Grayscale conversion

2) Reduction of noise

3) Image improvement.

c)Grayscale conversion

Only brightness is the detail of a grayscale. Each pixel represents a quantity of light in a grayscale image. The gradient of luminosity can be identified in a grayscale photograph. In a grayscale image only the light intensity is measured. RGB colours are coded from 0 to 255 on 256 levels. The conversion of Grayscale processes the transformation of a colour image into a grayscale image. Gray scale pictures are simply processed and take less time than colour pictures. All processing techniques are used in a grayscale image [4].

Our proposed method uses the following equations to convert a (Red Blue Green) RBG image into a grayscale image using the weighted sum technique.

“Grayscale intensification” = “0.114 B+ 0.587 G + 0.299 R” ---------(1)

d)Noise Reduction

Digital images are known as noise reduction as a procedure for detecting and eliminating unintended noise. The difficulty is to distinguish between the right characteristics, which must be considered as noise for subsequent bifurcation. The term "noise" refers to pixel values being unpredictable.As illustrated in Fig. 1, our strategy for the removal of superfluous noise uses a median filter (4). An invariantly sharp edges non-linear filter like median filter. To enforce a median filter a sliding window with a peculiar length[4] is utilised. The centre value is the median of the sample in the window and each sample values are sorted by magnitude that generates filtered output.

e)Image Intensification

Image strengthening is intended to increase the visibility of a photograph's principal feature. In this situation, an increase in contrast is used to achieve a higher quality outcome.

f)Segmentation

Segmentation is the way to remove an area of interest for an image. In a zone of interest, every pixel has similar properties. The greatest entropy threshold for segmentation is used[5]. In order to start we first have to define the grey level of the original image, and then calculate the histogram of the grey scale image, and then separate the foreground and context using maximum entropy. A binary picture, which is white and black, as shown in figure, is obtained after getting the static filters such as maximum entropy. [5].

g)Feature extraction

A GLCM is a matrix in which the number of rows and columns in the image is equal to G. GLCM characteristics in this study are the following: autocorrelation, contrast, correlation, protagonism of the cluster, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, squares sum, average sum, total variance, sum entropy, difference variance, entropy difference and correlation measurement.

A GLCM feature matrix is constructed that represents an image that uses these features less than one parameter. The GLCM uses the grey image matrix to record the most prevalent characteristics, such as contrast, mean, energy and consistency [2]. The second-order statistical probability values of these matrix members vary based on their grey value. The transient matrix is quite large if the intensity values are broad. This causes a long load of processes [12].

Classifier :

SVM Can be used in better ways if we consider different kernel functions:

Linear Kernel

It is the basic kernel type, generally one-dimensional in nature. When there are many features, it proves to be the best function. For text-classification issues, the linear kernel is chosen in particular because most such classification issues can be straightforwardly separated.

The objective is to identify the perfect line (or hyperplane) which maximum separates the two classes of supported vector machine (SVMs). (SVMs are used for binary rating but can be modified to accommodate multi-class rating). The equation of that decision boundary can be written as a line mathematically.

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$$

Linear Kernel

F(y, yj) = sum(y.yj)

Here, y, yj is the data you want to categorise.For each pd dimensions and N-number training data, the main linear classifier will be given as follows:$\{(xn, yn)|x_n \in R^n, yn \in \{-1, +1\}\}$, n = 1, . . , N

C: This is specified in the tuning parameters section as the regularisation parameter.

Loss Function:

In order to optimise cost function and reach the weights, SVM uses hinge loss as the logistic return utilising the logistic loss function.

Polynomial Kernel Relation:

The grade of polynomial is popular in image processing.

$$k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j} + 1)^d$$

Gaussian Kernel:

It is an all-intention kernel that is utilised when the data is not previously known.

$$k(x, y) = \exp\left(-\frac{||x - y||^2}{2\sigma^2}\right)$$

The polynomials and RBF are particularly helpful when there is no linear separation of data points.

GLCM Features:

The correlation quantifies the grey levels of neighbouring pixels' linear dependence. Digital Image Correlation is an optical technology which uses tracking and image recording techniques to accurately assess changes in the pictures in 2D and 3D.

$$\text{Correlation} = \frac{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}(i,j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$$

Contrast:

$\Sigma_m\Sigma_n (m-n)^2 C(m,n)$ ----------(2)

Energy:

$\Sigma_i\Sigma_j C(i,j)^2$ ----------(3)

Equation Of Homogeneity

$$\frac{\Sigma_m\Sigma_n C(m,n)}{1+|m-n|} \quad \text{----------(4)}$$

Mean (μ)

$$\frac{\Sigma_i^m\Sigma_j^n C(i,j)}{m*n} \quad \text{----------(5)}$$

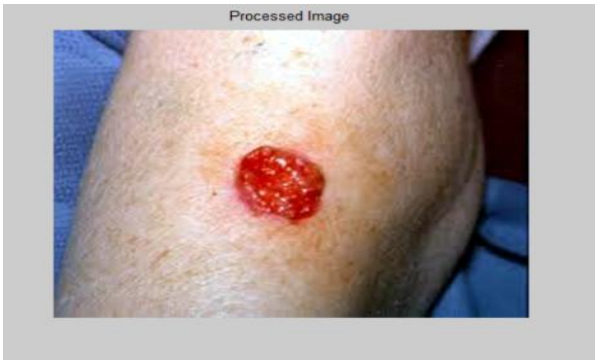
A Picture Dataset helps to identify an image using the appropriate method called Feature Extraction to measure specified properties or values. A classifier is used to distinguish between cancer and non-cancer images. We employed the supervised approach of learning called vector support for consistency. This model analyses a range of photos and checks, to which each image belongs from the two malignant and non-cancerous categories. The aim of SVM is to construct a hyperplane which is the least differentiated from the two groups[2].The function extraction

technique (glcm) is a method for decreasing the measurement of particular values or features that help to identify the image. [5].

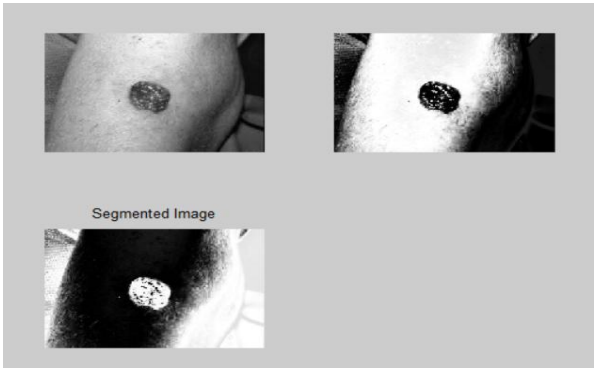
Results

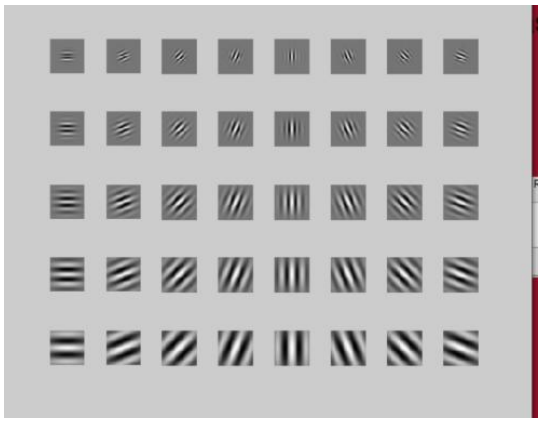
I found pictures of skin cancer on the Internet. They were previously processed using techniques such as gray-scale conversion, static (median) filters such as maximum entropy, and the gray-level coocurence matrix system. The various features were incorporated in the Support Vectors to separate malicious cancer and non-carcinoid non-melanocytes, which produced the picture. Cancer Cancer (as shown in the figure).

"Fig 2: Input Image"

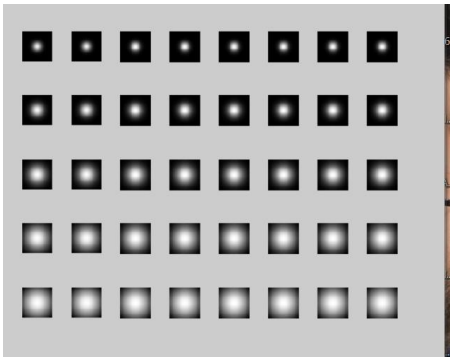


"Fig 3: Preprocessed Image"





"Fig 5: Feature Extraction"



" Fig 6: Magnitude Of Filters"

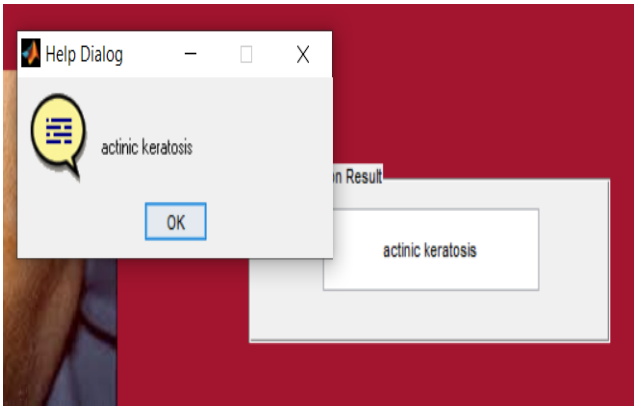


Fig 7: Snapshot of Output

$$\text{Accuracy Rate} = \frac{TP + TN}{TP + FP + FN + TN} \quad \text{---(6)}$$

$$\text{Re}call = \frac{True\_Positive}{True\_Positive + False\_Negative}$$

$$F\_Score = \frac{2 * \text{Pr}ecision * \text{Re}call}{\text{Pr}ecision + \text{Re}call}$$

Parameter	Accuracy	Accuracy	Accuracy	Recall Value	Recall Value	Recall Value	F-measure	F-measure	F-measure
Set 1	90.91	90.89	90.87	0.9991	0.988	0.977	0.951897	0.951769	0.9427
Set2	98.151	97.90	97.56	0.996361	0.992629	0.985	0.991585	0.987449	0.9758
Set3	96.503	95.00	95.231	0.986918	0.975467	0.965	0.992255	0.975467	0.9645
	Proposed Model. (Gaussian kernel)	Previous Model-2 (RBF kernel)	Previous Model-3 (Linear Kernel)	Proposed Model (Gaussian kernel)	Previous Model-2 (RBF kernel)	Previous Model-3 (Linear Kernel)	Proposed Model. (Gaussian kernel)	Previous Model-2 (RBF kernel)	Previous Model-3 (Linear Kernel)

Table 2.1 Skin Cancer Detection Result Comparison.

Conclusion:

With a GLCM and a statistical model of learning called support vector machines, the proposed method of detection for skin cancer may be simply determined to detect whether or not the image is non-cancerous. The accurate rate of construction for the machine is 95%. It is vital to select the proper kernel because if the transformation is inappropriate, the model can produce negative results. In general, validate that you have linear data and apply linear SVM always in the situation (linear kernel). Lineary SVM is a parametric model, however the kernel is not an RBF SVM, which increases SVM's complications by the size of the training data. Not only does the training of an RBF kernel SVM make more costly, you also have to maintain around the kernel matrix, and the projection into this "infinite" higher dimensional space where the data can be linearly separated is also more costly during prediction.

References

[1] S. Subha, S.Srinivasan, “Detection and Differentiation of Skin Cancer from Rashes”, International Conference on Electronics and Sustainable Communication Systems ICESC 2020.

[2] J Abdul Jaleel, Sibi Salim, “Computer Aided Detection 01 Skin Cancer”, International Conference on Circuits, Power and Computing Technologies, [ICCPCT-2013]

[3] Santosh Achakanalli & G. Sadashivappa ,” Statistical Analysis Of Skin Cancer Image –A Case Study “ , International Journal of Electronics and Communication Engineering (IJECE), Vol. 3, Issue 3, May 2014.

[4] “Digital image processing” by jayaraman. Page 244,254-247,270-273. (gray level, median filter). Vol 4,Issue 4,April 2017

[5] Nadia Smaoui Zghal, Imene Khanfir Kallel,”International Conference on Advanced Technologies for signal and Image processing.20 October 2020

[6] Kasmi, R. and Mokrani, K., 2016. Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Processing, 10(6), pp.448-455.

[7] Goel, Rashi, and Saranjeet Singh. "Skin cancer detection using glcm matrix analysis and back propagation neural network classifier." International Journal of Computer Applications 112, no. 9 (2015).

[8] A.A.L.C. Amarathunga,” Expert System For Diagnosis Of Skin Diseases”, International Journal Of Scientific & Technology Research, Volume 4, Issue 01, 2015.

[9] Mariam A.Sheha ,”Automatic Detection of Melanoma Skin Cancer”, International Journal of Computer Applications, 2012.

[10] Anshu bharadwaj, Sonajharia Minz,” Discretization based Support Vector Machines” Indian Agriculture Statistics Research Institute. 2009

[11] Maurya R, Surya K.S,"GLCM and Multi Class Support Vector Machine based Automated Skin Cancer Classification, "IEEE journal, vol 12, 2014.

[12] Mohanaiah, P., Sathyanarayana, P., &GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. International Journal of Scientific and Research Publications, 3(5), 1.
[13] Chakaradar M., Aggarwal A., Cheng X., Rani A., Kumar M., Shankar A., 2021, “A Non-invasive Approach to Identify Insulin Resistance with Triglycerides and HDL-c Ratio using Machine Learning” Neural Processing Letters, DOI: https://doi.org/10.1007/s11063-021-10461-6

[14] Rani, A., Jain A. K., Kumar M., 2021, “Identification of Copy-Move and Splicing based Forgeries using Advanced SURF and Revised Template Matching”, Multimedia Tools and Applications (MTAP), DOI: https://doi.org/10.1007/s11042-021-10810-6

[15] Aggarwal A., Alshehri M., Kumar M., Sharma P., Alfarraj O., Deep V., 2020 “Principal component analysis, hidden Markov model, and artificial neural network inspired techniques to recognize faces” Concurrency and Computation: Practice and Experience, 2021; vol.33, no-9, e6157, DOI: https://doi.org/10.1002/cpe.6157

[16] Kumar M., Alshehri M., AlGhamdi R., Sharma P., Deep V., 2020, “A DE-ANN Inspired Skin Cancer Detection Approach using Fuzzy C-Means Clustering” Mobile Network and Applications, Vol. 25 (2020), PP. 1319–1329, DOI: https://doi.org /10.1007/s11036-020-01550-2

[17] Kumar M., Srivastava S., Uddin N., 2017 “Forgery detection using multiple light sources for synthetic images” Australian Journal of Forensic Sciences, Vol. 51, No. 3, pp. 243-250, (Taylor & Francis, ISSN: 0045-0618), DOI: http://dx.doi.org/10.1080/00450618.2017.1356871