Tushar Bhangale, 10:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Detect Skin Cancer with Support Vector Machines Using Oriented Gradient Function Histograms

Tushar Bhangale
Department of Information Technology
AISSMS IOIT
Pune

Abstract-Detecting melanoma cancer in its early stages can help cure it. In this research paper, we propose an efficient skin cancer detection method based on support vector machines (SVM) with histogram of oriented gradient (HOG) functions. In it, skin cancer images from the ISIC 2018 dataset (International Skin Imaging Collaboration 2018) are converted to grayscale and preprocessed with a median filter. We then apply an image re-sampling technique to readjust the class distribution. HOG features are extracted from these preprocessed images. We then use the kernel-based His SVM classifier Radial Basis Function (RBF) to classify these extracted HOG features and detect cancer class labels. These predicted class tags are compared to the original tags to perform the evaluation. This proposed method has been tested and achieves 76% accuracy, 85% specificity, 84% accuracy, 76% recall, and 75% F1 score.

Keywords- Machine Learning, Skin Cancer, Histogramog Oriented Gradients.

I. INTRODUCTION

The skin is the body's largest organ and protects against injury, heat, and infection. Skin cancer is abnormal cells that grow in the skin and cause metabolic changes in the human body [1]. Skin normally consists of three layers: the innermost layer (hypodermis), the outermost layer (epidermis), and the middle layer (dermis) [2]. Skin cancer begins with clusters of cell divisions, also called lesions [3]. The main cause of skin cancer is overexposure to the sun's ultraviolet (UV) rays. Most often, images of skin cancer are captured using a device called a dermatoscope [5].

Early detection of skin cancer using dermatoscopic imaging is very important in medical imaging because most cancers can be cured if diagnosed early. Skin cancer is found in various types, including melanoma, basal carcinoma, and squamous cell carcinoma, of which melanoma is the most unpredictable. Detecting melanoma cancer in its early stages can help cure it. Melanoma can be detected by dermatological screening and biopsy tests, which are time-consuming, costly, and require a medical professional.

Because it is expensive for a dermatologist to see each patient, an automated system for melanoma detection is needed so thatmortality can be minimized if detected early.

Recently, semi-automated methods such as support vector machines (SVM), random forest (RF), and K-nearest neighbor (KNN) methods have been used for skin cancer detection. Among these methods, the SVM-based method outperforms the others in skin cancer detection. This research paper proposes his SVM to find cancer types in the ISIC 2018 database using HOG-based methods [6]. In this study, HOG features are extracted from skin cancer images and classified using SVM to recognize cancer class labels in the input image. The predicted class labels are compared with the original class labels to get a score.

II. RELATED WORKS

Skin cancer is an abnormal growth of skin cells, which is usually caused by the sun’s harmful rays. Caught early, skin cancer is highly treatable.K-Means is another cancer detection method commonly used in medical imaging. Using this k-means clustering method presents the first problem of cluster point

selection. Many researchers use Conditional Random Fields (CRF) and Markov Random Fields (MRF) to detect cancer. These two methods require the most programmer interaction and are less accurate for multiclass problems.

ANN is also a machine learning-based supervised algorithm that uses a similarity index to identify cancer types in skin cancer images. Praveen et al. use K-Means and PSO-based cancer detection methods. RF is another widely used cancer detection method that creates a decision tree for randomly selecting data points.

III. METHODOLOGY

The proposed skin cancer detection technology follows five main processes. Preprocessing, image re-sampling, HOG feature extraction, SVM classification, and performance evaluation. The main workflow of this proposed skin cancer detection method is shown in Figure 1 and represented as follows:

1. Dataset Information:
The proposed study will be conducted using the ISIC 2018 dataset. This dataset includes melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AKIEC), benign keratosis (BKL), dermatofibroma (DF), and vascular lesions (VASC).

2. Pre-Processing:
All images in the ISIC 2018 dataset are divided into training and test images. Both images are in color format with three-dimensional (3D) pixel values (red, green, blue). These color images are converted to grayscale. 3D pixel values are converted to one-dimensional (1D) values to reduce computational complexity.

After grayscale conversion, a median filter is applied to the skin image to remove unwanted noise. Median filtering is a nonlinear method of removing noise from an image. It is widely used because it is very effective at removing noise while preserving edges. This is especially effective at removing "salt and pepper" noise.

3. Image Resampling:
Changing the pixel dimensions of an image is called re-sampling. Resampling can degrade image quality. This approach is mainly divided into two types:

- Random Undersampling.
- Random Oversampling.

Images in the ISIC 2018 database have an unequal number of class designations that affect the accuracy of the skin cancer detection process. To overcome these limitations, random image resampling algorithms are applied to imbalanced datasets to rebalance class distributions.

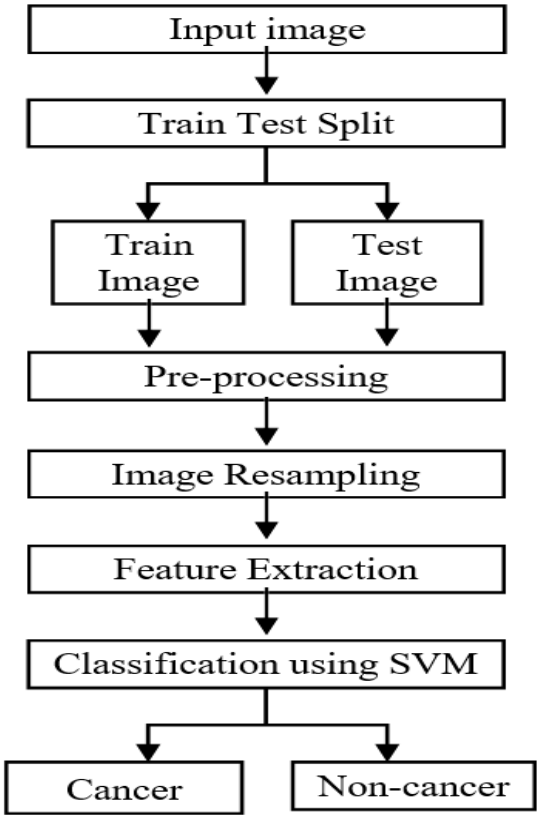

Fig 1. Proposed skin cancer detection architecture

Table1.

Class Name	ISIC 2018 Dataset		
	Initial Sample Size	After Random Under Sampling	After Random Over Sampling
MEL (Class 0)	1113	1113	1431
NV (Class 1)	6705	1431	1431
BCC (Class 2)	514	514	1431
AKIEC (Class 3)	327	327	1431
BKL (Class 4)	1099	1099	1431
DF (Class 5)	115	115	1431
VASC (Class 6)	142	142	1431
Total	**10015**	**4741**	**10017**

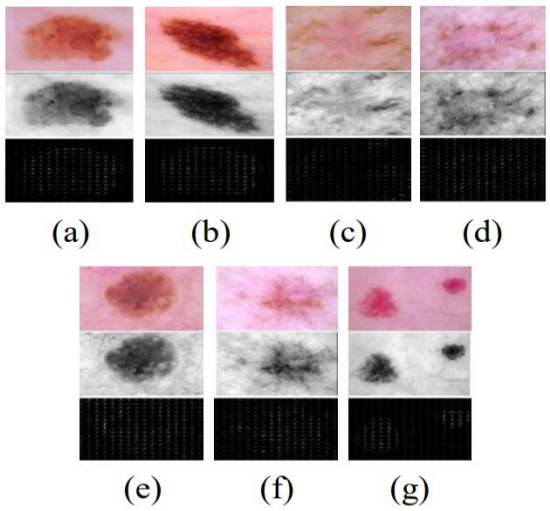

Fig 2. HOG feature extraction of skin cancer images: 1st, 2nd and 3rd row represent the skin cancer, gray scale and feature extracted images respectively for seven cancer types; (a) melanoma; (b) melanocytic nevus; (c) basal cell carcinoma; (d) actinic keratosis; (e) benign keratosis; (f) dermatofibroma and (g) vascular lesion.

4. Feature Extraction:

The HOG features are extracted from pre-processed skin cancer images. In this, the horizontal and vertical gradients of an image are calculated by filtering the horizontal and vertical kernels. Then, the magnitude of the gradient and orientation of the gradient are calculated using Eq.

$$gradient\ magnitude\ (g) = \sqrt{g_x^2 + g_y^2}$$

$$gradient\ magnitude\ (g) = \arctan \frac{g_y}{g_x}$$

Where, *gx*and *gy*are the horizontal and vertical gradients.

5.Classification Using SVM:

The HOG features of the extracted images were processed using a nonlinear SVM skin cancer classification algorithm based on supervised machine learning. SVM takes an input image and maps it with class labels to predict the type of cancer present in a given image. In this work, we use the RBF (Radial Basis Function (RBF)) kernel to map the class labels used in equation Eq.

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

Where, ||x-x′||2 is squared Euclidean distance of 2 data points x and x′ in an image. This algorithm makes a decision boundary between seven different cancer classes by generating multiple hyper planes. These hyper planes are used to detect the cancer class labels of an input image. The detected class labels are compared with original class labels for evaluating performance.

6.Experimental Evaluation

The performance of skin cancer technique has been evaluated on the basis of accuracy, specificity, precision, recall and F1-score.

- Accuracy = (tp+tn)/(tp+fp+fn+tn)
- Specificity = (tn)/(tn+fp) (
- Precision = (tp)/(tp+fp)
- (Recall = (tp)/(tp+fn)
- F1-Score = (2tp)/(2tp+fp+fn)

(True Positive (tp) is used to count the number of truly identified positive case pixels in detected region. True Negative (tn) is used to count the number of truly identified negative case pixels in the detected region. The number of incorrectly identified negative case pixels in detected region is calculated using False Negative (fn). The number of incorrectly identified positive case pixels in detected region is calculated using False Positive (fp).

IV. EXPERIMENTAL RESULTS

Table 2. Performance of proposed SVM and SVM with HOG Technique.

ISIC 2018 Dataset						
Method		Accuracy	Specificity	Precision	Recall or Sensitivity	F1-Score
SVM	Training	0.80	0.90	0.81	0.81	0.81
	Testing	0.53	0.79	0.54	0.53	0.52
	Average	0.67	0.85	0.68	0.67	0.67
SVM and HOG	Training	0.83	0.80	0.85	0.83	0.83
	Testing	0.68	0.89	0.82	0.68	0.66
	Average	0.76	0.85	0.84	0.76	0.75

V. LIMITATION

- So, all in all HOG is a great feature descriptor that we can use for image recognition. But the images that we use should have very distinguishable gradients; else the HOG feature descriptor may perform poorly.
- Training with non-linear kernel does not come always with good performance. Non-linear kernel sometimes may lead in over training and thus in bad testing performance.

VI. CONCLUSION

Cancer detection from skin imaging is an important task for disease diagnosis and cancer treatment planning. Existing manual and semi-automatic methods are very time consuming, inaccurate and require high computational power. To circumvent these limitations, his HOG feature extraction by SVM classification method for detecting skin cancer class labels in dermoscopy skin cancer images is proposed. These detected class labels are compared with the original labels to perform the evaluation.

This research work will be implemented and tested using the ISIC 2018 dataset. This work achieves 76% accuracy, 85% specificity, 84% precision, 76% recall, and 75% F1 score. This is a 2% and 9% higher sensitivity and specificity value than existing methods,respectively.

VII. REFRENCES

[1]. U. B. Ansari and T.Sarade, "Skin Cancer Detection using Image Processing", International Research Journal of Engineering and Technology, Vol. 4, No. 4, pp. 2875- 2881,2017.

[2]. P. Thsper and A. Singh, "A Survey of Lesion Detection using Dermoscopy Image Analysis", Journal of Gujarat Research Society, Vol. 21, No. 6, pp. 129-143, 2019.

[3]. M.A. Shukran and N.S. Mariam, "Melanoma Skin Cancer Diagnosis Device using Image Processing Techniques", International Journal of Recent Technology and Engineering, Vol. 7, No. 4, pp. 490-494, 2019.

[4]. S. Jaiswar and M. Kadri, "Skin Cancer Detection using Digital ImageProcessing", International Journal of Scientific Engineering and Research, Vol. 3, No. 6, pp. 138- 140, 2018.

[5]. A. Murugan and S. Anu, "Detection of Skin Cancer using SVM, Random Forest andkNN Classifiers", Journal of Medical Systems, Vol. 1, No. 9, pp. 269-275, 2019.

[6]. C.F. Noel, "Skin Lesion Analysis Toward Melanoma Detection: A Challenge", Proceedings of International Symposium on Biomedical Imaging, pp. 345-356, 2017.

.