ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2021.0329

SKIN CANCER DETECTION USING SUPPORT VECTOR MACHINE WITH HISTOGRAM OF ORIENTED GRADIENTS FEATURES

G. Neela Krishna Babu[1] and V. Joseph Peter[2]
[1]Department of Computer Science, Manonmaniam Sundaranar University, India
[2]Department of Computer Science, Kamaraj College, India

Abstract

This research work proposes an efficient skin cancer detection technique based on Support Vector Machine (SVM) with Histogram of Oriented Gradients (HOG) features. In this, skin cancer images from ISIC 2018 (International Skin Imaging Collaboration 2018) dataset are converted into gray scale and pre-processed using the median filter. The image resampling technique is then applied to rebalance the class distribution. The HOG features are extracted from these preprocessed images. After, the Radial Basis Function (RBF) kernel based SVM classification method is used to classify these extracted HOG features for detecting cancer class labels. These predicted class labels are compared with original labels for performing the evaluation. This proposed method is tested using and achieves 76% accuracy, 85% specificity, 84% precision, 76% recall and 75% F1-score.

Keywords:

Skin Cancer, Detection, Machine Learning, Support Vector Machine, Histogram of Oriented Gradients.

1. INTRODUCTION

Skin is the body's largest organ and it protects from injury, heat and infection. Skin cancer is a growing abnormal cell within the skin that produces metabolic changes in our human body [1]. Normally, the skin contains three different layers namely, innermost layer (hypodermis), outermost layer (epidermis) and middle layer (dermis) [2]. The skin cancer begins from the group of cell division is also known as lesion [3]. The major source of skin cancer is an excess exposure of UV (Ultra Violet) rays from the sun [4]. Mostly, the skin cancer images are acquired by using a device called dermatoscope [5]. This research work is implemented by skin cancer dataset named International Skin Imaging Collaboration-2018 (ISIC-2018) [6] [7].

Early detection of skin cancer from dermatoscopic images is essential in medical imaging because most cancer types are curable at early diagnosis [8]. In earlier days, biopsy based methods were used for identifying skin cancer. In the biopsy method, a tissue sample or a piece of tissue portion is removed from the skin of affected patients [3]. This method requires a long time to complete the diagnosis process for physicians. It is one of the painful processes and it has a high possibility rate of disease spreading from one part to other parts of the body.

In recent days, semi-automatic methods like K-means, Particle Swarm Optimization (PSO), Support Vector Machine (SVM), Random Forest (RF) and K-Nearest Neighbor (KNN) methods are used for skin cancer identification process [9]. Among these methods, SVM based methods are performing better in skin cancer detection than other methods [10]. This research work proposes SVM with HOG based method for finding cancer type in ISIC 2018 database [6]. In this research, HOG features are extracted from skin cancer images and classified using SVM to detect the cancer class labels of an input image. The predicted class labels are compared with original class labels for finding an evaluation.

This research article is organized as follows: Section 2 gives a detailed discussion related to skin cancer detection technique; Section 3 explains the proposed methodology; Section 4 provides the comparative analysis of the proposed method and Section 5 concludes the paper.

2. RELATED WORKS

An abnormal appearance and development of a lesion in the skin layer are defined as skin cancer [11]. An efficient skin cancer detection technology is essential for disease diagnosis and cancer treatment planning [12]. In earlier days, biopsy based manual detection methods were used for identifying skin cancer. This method was painful and it required long time to complete the diagnosis process [13]. To mitigate the limitations, semi-automatic based machine learning methods are used for detecting the cancer types in skin images. Initially, Fuzzy-C-Means (FCM) method is used to identify the cancer types. This method has limited performance in real-world problems [14].

K-Means is another popularly used cancer detection method in medical image diagnosis. The initial cluster point selection problem arises while using this k-means clustering method [15]. Many researches have used Conditional Random Field (CRF) and Markov Random Field (MRF) for detecting cancer [16]. These two methods need the highest level of programmer interaction and it has less accuracy in multi-class problems. KNN is also machine learning based supervised algorithm which uses a similarity index for identifying cancer types in skin cancer images [17]. Praveen et al. [14] have used K-Means and PSO based cancer identification methods. RF is another popularly used cancer detection method that created decision trees for random data point selection.

Murugan et al. [5] have proposed KNN and RF method for skin cancer identification. Greedy fusion is another performing method in the field of cancer detection. These methods are poor in performance in the real world skin cancer identification problems and are time consuming and less accurate. Recently, SVM based methods are performing better in skin cancer identification. Murugan et al. [5], Mane et al. [8] and Patel et al. [18] have proposed SVM based detection techniques. Fusion based SVM methods also have competitive performance in skin cancer identification. These methods take the whole image as an input and process using SVM fused architecture, which needs higher computation than other methods [19]. To avoid these limitations, HOG feature extraction with SVM classification method is proposed for detecting skin cancer types in dermatoscopic images.

3. PROPOSED METHODOLOGY

The proposed skin cancer detection technique follows five main processes; pre-processing, image resampling, feature extraction using HOG, classification using SVM and performance evaluation. The main workflow of this proposed skin cancer detection methodology is presented in Fig.1 and depicted by the following subsections.

3.1 SKIN CANCER DATASET

The proposed research work is implemented using ISIC 2018 dataset. This dataset comprises 10015 images of seven skin cancer types namely, Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis (AKIEC), Benign Keratosis (BKL), Dermatofibroma (DF) and Vascular Lesion (VASC) are detailed below.

Step 1: MEL skin cancer is caused by melanocytes and it appears like brown or black color. This cancer cell occurs especially in women's neck, men's trunk, face and legs.

Step 2: NV cancer cells are naturally present at birth or appear after birth due to melanocytes.

Step 3: BCC cancer is produced from the sun rays and it appears in the exposed skin portion of the body [2].

Step 4: AKIEC is a scaly, rough patch on our skin that develops from Deoxyribonucleic Acid (DNA) damage and it is commonly found in ears, lips, face, forearms, scalp and neck.

Step 5: BKL is a common harmless skin disorder and it appears like scaly brown or black color spot.

Step 6: DF skin cancer caused by the insect bite and it frequently develops on the arms and legs. This cancer type is not producing any symptom, but sometimes it causes itching and painful.

Step 7: Vascular lesion is a common skin abnormality that caused by cluster of blood vessel growth found in children and infants.

3.2 PRE-PROCESSING

All images from ISIC 2018 dataset [6] are split into training and testing images. Both images are present in the form of color having 3 Dimensional (3D) pixel values (red, green, blue). These color images are converted into gray scale i.e. converted from 3D pixel value to 1 Dimensional (1D) value for reducing the computational complexity. After gray scale conversion, the median filter is applied over the skin images to remove an unnecessary noise. In this median filter, all pixel values of an image are replaced by the neighboring pixel's median value is defined in Eq.(1).

$$f(x,y) = \text{median}\{g[x,y], (x,y)\in W\} \quad (1)$$

where, $f(x,y)$ is an output function and w is the neighborhood pixel value. The median filter reduced the intensity variance of an image by utilizing the neighborhood's median value.

3.3 IMAGE RESAMPLING

The images from ISIC 2018 database are having unequal number of class labels, which affected the accuracy of skin cancer detection process. To overwhelm these limitations, the random image resampling algorithm is applied over an imbalanced dataset to rebalance the class distribution [20]. This approach is mainly divided into two types: random undersampling; random oversampling. The Average Sample Size (ASS) value is calculated using the addition of cardinality sample values from all classes is mentioned in Eq.(2) and Eq.(3).

$$ASS = \sum_{i=0}^{n} \frac{(\#C_i)}{n} \quad (2)$$

$$ASS = \frac{(\#C_0 + \#C_1 + \#C_2 + ... + \#C_n)}{n} \quad (3)$$

where, C is the class label and # represents the cardinality value of all classes. If the sample size of class C is greater than ASS means majority class else it has been considered as minority class.

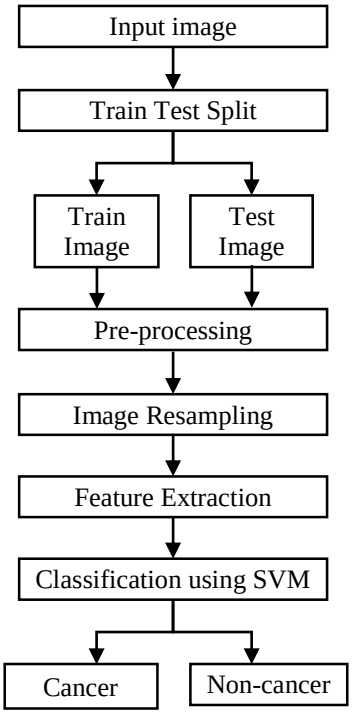
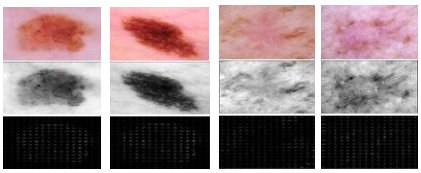


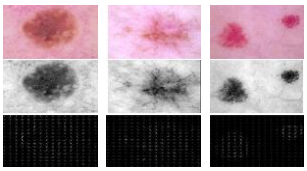
Fig.1. Proposed skin cancer detection architecture

Table.1. Pre-processing of ISIC 2018 dataset images using Random Sampling Technique.

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**



(a) (b) (c) (d)



(e) (f) (g)

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

In Table.1, NV (Class 1) is the majority class and the remaining classes are the minority classes. The majority class (NV) is processed using random undersampling, which turns into the size of ASS. The six minority classes are further processed using random oversampling to resample class labels for getting ASS size.

3.4 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.(4) and Eq.(5) respectively.

$$\textit{gradient magnitude}\ (g) = \sqrt{g_x^2 + g_y^2} \tag{4}$$

$$\textit{gradient magnitude}\ (g) = \arctan \frac{g_y}{g_x} \tag{5}$$

where, g_x and g_y are the horizontal and vertical gradients. Further, the histogram of gradients has been calculated to extract the feature descriptor of an image. This feature descriptor holds the significant information of an image. The HOG feature extraction of seven types of skin cancer images is represented in Fig.2.

3.5 CLASSIFICATION USING SVM

The extracted HOG features of an image have been processed using supervised machine learning based non-linear SVM algorithm for classifying skin cancer. SVM takes an input image and maps with class labels to predict which cancer type presents in a particular image. In this research, the RBF (Radial Basis Function (RBF) kernel is used to map the class labels which is defined in Eq.(6).

$$k(x, x') = \exp\left(\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{6}$$

where, $\|x\text{-}x'\|^2$ 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 hyperplanes. 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.

3.6 PERFORMANCE EVALUATION

The performance of skin cancer detection technique has been evaluated by the benchmark metrics namely accuracy, Specificity, Precision, Recall and F1-Score values are mentioned in Eq.(7)-Eq.(11).

$$\text{Accuracy} = (t_p+t_n)/(t_p+f_p+f_n+t_n) \tag{7}$$

$$\text{Specificity} = (t_n)/(t_n+f_p) \tag{8}$$

$$\text{Precision} = (t_p)/(t_p+f_p) \tag{9}$$

$$\text{Recall} = (t_p)/(t_p+f_n) \tag{10}$$

$$\text{F1-Score} = (2t_p)/(2t_p+f_p+f_n) \tag{11}$$

True Positive (t_p) is used to count the number of truly identified positive case pixels in detected region. True Negative (t_n) 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 (f_n). The number of incorrectly identified positive case pixels in detected region is calculated using False Positive (f_p).

4. EXPERIMENTAL RESULTS AND DISCUSSION

The efficiency of this proposed skin cancer detection technique based on HOG with SVM is evaluated using ISIC 2018 dataset. This dataset contains 10015 labeled images of seven cancer types.

MEL	90	5	2	0	9	2	2
NV	6	80	6	2	8	1	2
BCC	0	3	105	1	5	2	0
AKIEC	0	3	2	75	7	1	2
BKL	2	2	6	0	98	4	2
DF	2	1	7	0	4	86	0
VASC	1	5	6	3	2	1	97
	MEL	NV	BCC	AKIEC	BKL	DF	VASC

(a)

MEL	19	0	0	0	8	7	0
NV	0	6	2	0	14	10	1
BCC	0	0	35	0	2	2	0
AKIEC	0	0	3	18	4	5	0
BKL	0	0	0	0	38	1	0
DF	0	0	0	0	3	36	0
VASC	0	0	1	0	7	2	26
	MEL	NV	BCC	AKIEC	BKL	DF	VASC

(b)

Fig.3. Confusion matrix of proposed Technique: (a) Performance using SVM; (b) Performance using SVM with HOG

These images are split into training and testing. Both testing and training images are converted into gray scale and pre-processed using the median filter. The image resampling technique is then applied to rebalance the class distribution. Then, HOG features are extracted in these preprocessed images. Further, RBF kernel based SVM is processed over the features to detect the cancer class labels of an input image. These predicted class labels are compared with original labels for performing evaluation using benchmark metrics namely, Accuracy, Specificity, Precision, Recall and F1-Score. The classification performance over seven skin cancer labels of this proposed method is defined in the confusion matrix of Fig.3.

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**

The Table.2 represent the performance of SVM and SVM with the HOG technique. From this, it is clear that the skin cancer detection results of SVM with HOG features are competitively higher performance than the SVM method.

Table.3. Performance comparison of proposed technique.

Method	**Sensitivity**	**Specificity**
Random Forest (RF)	0.74	0.69
KNN	0.65	0.58
Greedy Fusion	0.58	0.82
Linear SVM Fusion	0.66	0.76
Nonlinear SVM Fusion	0.70	0.78
Proposed Method	**0.76**	**0.85**

The performance of proposed SVM with HOG features are compared with Random Forest (RF), KNN, Greedy fusion, linear SVM fusion and non-linear SVM fusion methods is illustrated in Table.3. In RF, skin cancer identification is performed by creating decision trees using random data point selection. KNN is used to identify class labels using a similarity index. Greedy fusion is another performing method for skin cancer detection. Among these methods, linear and non-linear SVM fusion methods are performing better in skin cancer identification. This SVM fusion method takes the whole image as an input and processes using SVM fused architecture, which needs higher computation than other methods. Contrasting these algorithms, the proposed method based on SVM with HOG features are competitively 2% and 9% higher sensitivity and specificity values than the existing methods.

5. CONCLUSION

Cancer detection from skin images is an important task for disease diagnosis and cancer treatment planning. Existing manual and semi-automated methods are high time consuming and less accuracy, which needs high computational power. To avoid these limitations, HOG feature extraction with the SVM classification method is proposed for detecting skin cancer class labels in skin cancer dermatoscopic images. These detected class labels are compared with original labels for performing the evaluation. This research work is implemented and tested using ISIC 2018 dataset. This work achieves 76% accuracy, 85% specificity, 84% precision, 76% recall and 75% F1-score, which are competitively 2% and 9% higher sensitivity and specificity values than the existing methods.

REFERENCES

[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 Image Processing", *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 and kNN 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.

[7] M.M. Vijayalakshmi, "Melanoma Skin Cancer Detection using Image Processing and Machine Learning", *International Journal of Trend in Scientific Research and Development*, Vol. 3, No. 4, pp. 1-14, 2019.

[8] S.S. Mane and S.V. Shinde, "Different Techniques for Skin Cancer Detection using Dermoscopy Images", *International Journal of Computer Sciences and Engineering*, Vol. 5, No. 12, pp. 159-163, 2017.

[9] M.Z. Alom, "Skin Cancer Segmentation and Classification with NABLA-N and Inception Recurrent Residual Convolutional Networks", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-7, 2019.

[10] A. Hecler, "Superior Skin Cancer Classification by the Combination of Human and Artificial Intelligence", *European Journal of Cancer*, Vol. 120, pp. 114-121, 2019.

[11] Y. Fujisawa and S. Inoue, "The Possibility of Deep Learning-Based, Computer-Aided Skin Cancer Classifiers", *Frontiers in Medicine*, Vol. 6, pp. 1-19, 2019.

[12] K. Muthukumar and P. Gowthaman, "GTCM Based Skin Lesion Melanoma Disease Detection Approach for Optimal

Classification of Medical Images", *International Journal of Recent Technology and Engineering*, Vol. 7, No. 3, pp. 1-12, 2019.

[13] K. Munir and H. Elahi, "Detection Cancer Diagnosis using Deep Learning: A Bibliographic Review", *Cancers*, Vol. 11, pp. 1-36,2019.

[14] R. Kaur and Praveen Kumar, "An Enhanced and Automatic Skin Cancer Detection using K-Mean and PSO Technique", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 8, No. 9, pp. 634-639, 2019.

[15] A. Dascalu and E.O. David, "Skin Cancer Detection by Deep Learning and Sound Analysis Algorithms: A Prospective Clinical Study of an Elementary Dermoscope", *EBioMedicine*, Vol. 43, pp. 107-113, 2019.

[16] E. Okur and M. Turkan, "A Survey on Automated Melanoma Detection", *Engineering Applications of Artificial Intelligence*, Vol. 73, pp. 50-67, 2018.

[17] D.A. Gavrilov and N.N. Shchelkunov, "Deep Learning based Skin Lesions Diagnosis", *Remote Sensing and Spatial Information Sciences*, Vol. 2, No. 12, pp. 81-85, 2019.

[18] I. Patel and S. Patel, "Dermoscopic Image Classification using Image Processing Technique for Melanoma Detection", *International Journal of Research in Advent Technology*, Vol. 23, No. 2, pp. 97-103, 2019.

[19] A. Masood and A. Al-Jumaily, "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms", *International Journal of Biomedical Imaging*, Vol. 2013, pp. 1-22, 2013.

[20] M. Hacibeyoglu and M.H. Ibrahim, "The Effect of Over-sampling and Undersampling Techniques in Medical Datasets", *Proceedings of International Conference on Advanced Technologies, Computer Engineering and Science*, pp. 429-432, 2018.