Skin Cancer Diagnostic using Machine Learning Techniques - Stationary Wavelet Transform and Random Forest Classifier

S. Mohan Kumar, J. Ram Kumar, K. Gopalakrishnan

Abstract*: *The abnormal growth of skin cells in the skin is known as skin cancer. It develops in skin due to exposure of the sun. Skin cancer spreads into other parts of the body easily. The death rate of skin cancer is increasing day by day, so, early diagnosis of skin cancer is necessary. In this paper, an efficient method for skin image classification using Stationary Wavelet Transform (SWT) based entropy features and Random Forest (RF) classifier is presented. The input skin images are decomposed by SWT. The skin image features are extracted by the entropy of decomposed skin images and classified using RF classifier. The performance of the system is evaluated in terms of accuracy, sensitivity and specificity. The results show the better classification accuracy of 91.5% at the 3rd level of SWT decomposition based on entropy features and RF classifier, and also their sensitivity and specificity are 90 % and 93 %..

Keywords* : *Skin cancer images, SWT, entropy features, RF classifier

I. INTRODUCTION

The largest organ in the human system is skin. Skin cancer is the most common among all cancers. The premature diagnosis is important for the survival of skin cancer. Segmentation of dermoscopic image by grabcut and k-means algorithms is discussed in [1]. Initially, the dermoscopic images are preprocessed by image normalization, removal of hair and illumination correction. The segmentation is made by flood fill algorithm, grabcut segmentation and k-means clustering. Detection and classification of skin lesions by advanced processing techniques are discussed in [2]. Statistic and deterministic feature types make the feature extraction. The specific classifier is used for classification.

Melanoma classification using Convolutional Neural Network (CNN) algorithm with activation function is described by [3]. The feature extraction and classification is made by nine layers of CNN. The first four convolutional and two pooling layers are used to extract the features, and the last three fully connected layers are used for classification. Skin disease detection in image analysis model is discussed in [4].

Initially, the preprocessing is made by Gaussian filter for smoothing the image, and median filter is used to remove

noise. Then the image is converted into a grayscale before employing k means clustering, and fuzzy c means for segmentation. Color and texture features are used as features. Support Vector Machine (SVM) classifier is used for classification .

Effective classification and feature selection technique for cancer diagnosis is discussed in [5]. At first, the input skin images features are extracted by the brute force algorithm. The segmentation of images is made by fuzzy clustering means algorithm. Finally, classification is made by the k-nearest neighbor algorithm. Skin cancer classification system by using curvelet and wavelet transform is discussed in [6]. The statistical feature is used for feature extraction, and the features are reduced by principal component analysis. The artificial neural network with backpropagation neural algorithm is used for classification.

A feature selection and neural network classification for skin lesions are discussed in [7]. Initially, the feature extraction is made by shape and color features. Classification is made by neural network classifier. Wavelet-based skin cancer classification is described in [8]. The feature extraction is made by wavelet transform. The classification is made by the probabilistic neural network.

Skin cancer classification based on contourlet transform is described in [9]. The input skin images are decomposed by the non-sub sampled version of contourlet transform. Naïve Bayes classifier is used for classification. Combination of three different features for skin cancer classification is discussed in [10]. Shape, color, histogram of oriented gradients and fractal features are combined and extracted. The classification of skin images is made by SVM classifier.

Skin lesion classification using an ensemble of deep neural networks is discussed in [11]. The layers in the neural network such as googlenet, alexnet and vggnet are used to extract the features and classification is done by a fully connected layer. Color feature and texture based on skin

Revised Manuscript Received on December 05, 2019.
Dr. S. Mohan Kumarr , Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India. Email: drsmohankumar@gmail.com
Dr. J. Ram Kumar, Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India.Email: jrkumar@iitk.ac.in
Dr. K. Gopalakrishnan, New Horizon College of Engineering, Bangalore, Karnataka,India. Email: xyz3@blueeyesintlligence.org

cancer classification system is described in [12]. Color, texture features and gray level co-occurrence matrix and are extracted. Classification is made by SVM classifier.

In this work, a skin cancer classification system using dermoscopic images by the use of SWT based entropy features and RF classifier is presented.

The organization of this paper is as follows: The methods and materials of skin cancer classification system are described in section 2. In section 3, results and discussion is described in terms of accuracy, sensitivity and specificity. The last section concludes the skin cancer classification system.

II. MATERIALS AND METHODS

In this paper en efficient method for skin cancer image classification using SWT and RF classifier is discussed. Figure 1 demonstrates the workflow of the classification of skin cancer images. The input skin image features are decomposed by SWT and the decomposed features are extracted by entropy features. These extracted features are the input to the classification stage. The classification is made by RF classifier.

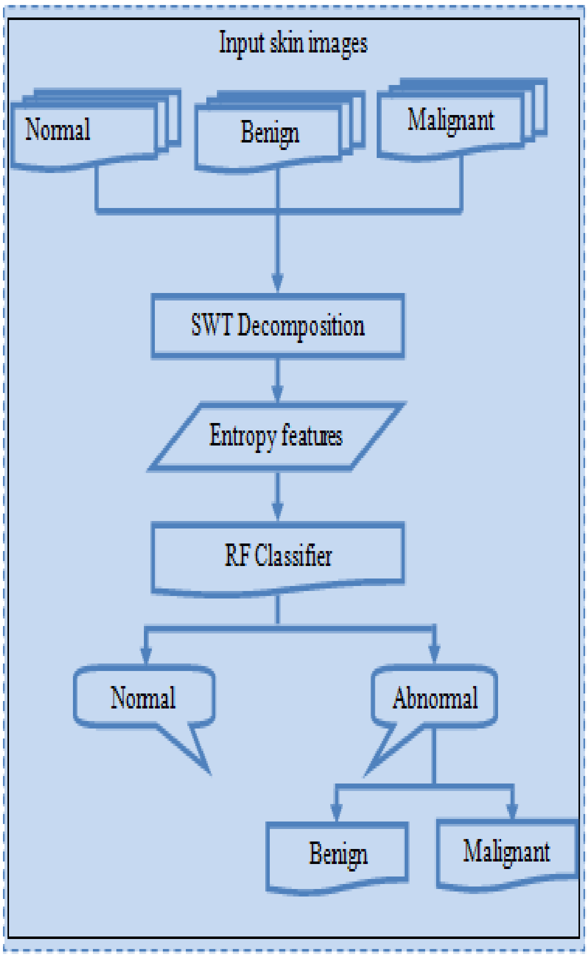


Fig. 1. Skin cancer image classification system - work flow

A. SWT Decomposition

SWT is the extension of DWT that has the shift-invariant property. SWT overcomes the lack of translation invariance in DWT. SWT achieves this by removing down samplers and up samplers in DWT. The upsampling factors are in the jth level. SWT contains the output of each level SWT includes the same number of input samples. The inherently redundant scheme for the decomposition of K levels and there is a redundancy for M in the wavelet coefficients. SWT is also used in image resolution enhancement [13] and microarray data classification [14]. Skin image features are decomposed by SWT and produce subband coefficients. Figure 2 and 3 shows the three levels of SWT filter bank and SWT filters.

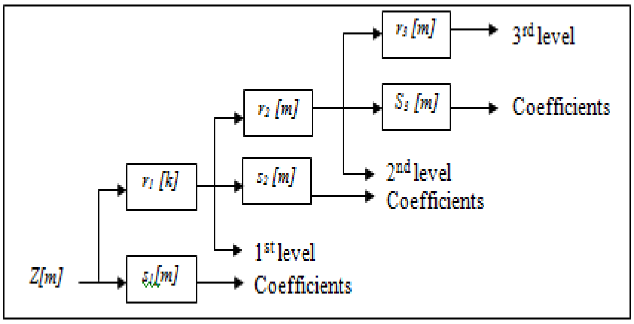


Fig. 2. Three levels of SWT Filter bank

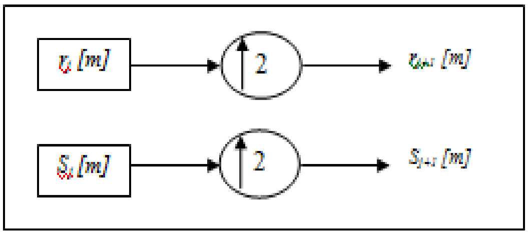


Fig. 3. SWT Filters

B. SWT based entropy features

The quantity of an image is defined by the entropy of an image. The entropy features are computed from all SWT sub-bands of skin images. The SWT based entropy features are defined by,

$$Entropy_{Ey} = -\sum_{i=1}^{k} R(z_j)\log_2 R(z_j) \qquad (1)$$

All where $R(z_j)$ is the probability distribution, $-\log_2 R(z_j)$ be the information amount of sub-bands. Entropy features is also used in glaucomatous digital fundus image classification [15], Lung sound detection [16] and palm print recognition [17]. In this study, entropy features are used which are extracted from the decomposed skin images by SWT.

C. RF Classifier

RF classifier was built by the combination of decision trees $\{AB(y,\Psi_m)\}_{m=1}^{P}$, where y is input vector, Ψ_m denoted as a random split of independent vectors with the trees in forest with equal distribution $\Psi_1, \Psi_2, \Psi_3, ... \Psi_{m-1}$. P denoted as a bootstrap to training the data. Trees are built by the different bootstrap sample. Let us consider, the sample consisting of n samples of the training set. A number v in each node the total number of predictions V

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

is randomly chosen. The nodes are split by *v*. Then value v is constant throughout the forest. The RF algorithm is defined by,

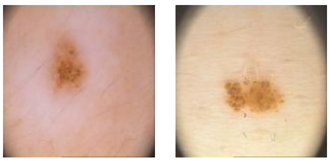
$$H(k) = 1 - \sum_{m=1}^{D} f^2(m/k) \quad (2)$$

where $f^2(m/k)$ for $m = 1,2,3,......D$ is the class probabilities estimation after splitting the node.

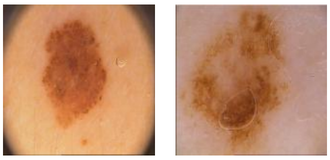
RF classifier is used in other fields like short video detection [18] and cellular image segmentation [19]. The skin image classification is made by RF classifier.

III. Results and Discussion

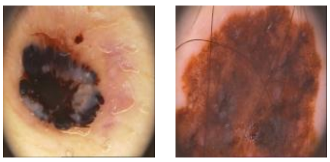
The performance of the skin image classification system is evaluated by using the PH2 database [20]. It contains 100 dermoscopic images with lesions. The resolution of the image is 768x560 pixels. Figure 4 shows some of the sample images in PH2 database.



(a) Normal images



(b) Abnormal images (Benign)



(c) Abnormal images (Malignant)

Fig. 4. Sample normal and abnormal images (a), (b) and (c) in PH2 database

The input skin images in PH2 database are decomposed by SWT transform. Then entropy features are extracted from the SWT decomposed skin images. The RF classifier is used for classification. The performance is evaluated by the terms of accuracy, sensitivity and specificity. Table 1 shows the accuracy, sensitivity and specificity of SWT based entropy features and RF classifier.

Table 1 Performance of SWT based entropy features and RF classifier

SWT based entropy features levels	Performance of RF classifier		
	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	77.5	75	80
2	83.5	79	88
3	91.5	90	93
4	87	82	92
5	89	90	89

From table 1 it is observed that the highest accuracy is 98.5 % in the 3rd level of the SWT decomposition. Also, it is observed that their sensitivity and specificity are 98 % and 99 %. The graphical representation of SWT based entropy features and RF classifier is shown in figure 5.

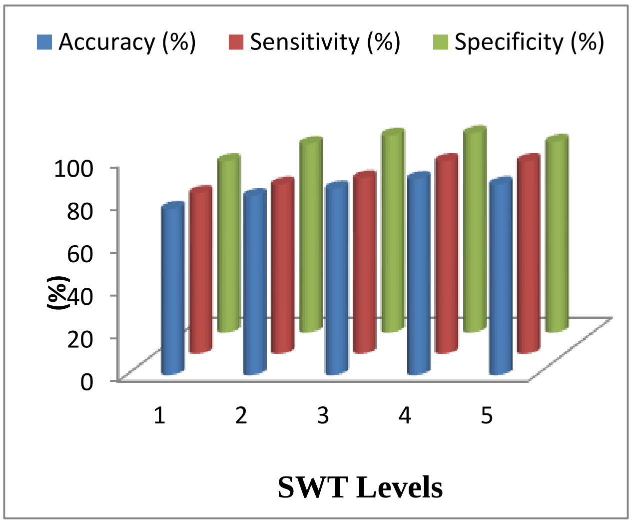


Fig. 5. Graphical representation of SWT based Entropy features and RF classifier

IV. CONCLUSION

An efficient method for skin cancer image classification is presented based on SWT and RF classifier. The entropy features are used as features and for classification RF classifier is used. Skin image features are decomposed by SWT and entropy features are extracted. These extracted features are inputs to RF classifier for classification. Result shows better classification accuracy at the 3rd level of SWT decomposition with RF classifier. The entropy features produces 98.5 % accuracy and the obtained sensitivity and specificity by the same entropy features are 98 % and 99 %.

REFERENCES

1. S. M. Jaisakthi, P. Mirunalini, C. Aravindan . "Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms" *IET Computer Vision*, 12(8), Jul.2018, pp.1088-1095
2. L .Ichim, D. Popescu, O. Chenaru, 2018. "Advanced processing techniques for detection and classification of skin lesions", *International Conference on System Theory, Control and Computing*",Oct.2018, pp. 498-503.
3. Namozov, A. and Im Cho, Y., 2018. "Convolutional Neural Network Algorithm with Parameterized Activation Function for Melanoma Classification", *International Conference on Information and Communication Technology*

Convergence,Oct.2018, pp. 417-419..

4. A. Haddad, S. A.Hameed,"Image Analysis Model For Skin Disease Detection"Framework", *International Conference on Computer and Communication Engineering (ICCCE)*, Sep.2018, pp.1-4.
5. S. Lobo, M. S. Pallavi "Predicting Protein in Cancer Diagnosis Using Effective Classification and Feature Selection Technique", *International Conference on Communication and Signal Processing*,Apr.2018, pp. 156-159.
6. M. Abu Mahmoud, A.AI- "Wavelet and curvelet analysis for automatic identification of melanoma based on neural network classification", *International Journal of Computer Information Systems and Industrial Management (*IJCISIM), 5(1), 2013, pp. 606-614.
7. A. Mikolajczyk, M. Grochowski, A. Kwasigroch, "Optimal Selection of Input Features and an Acompanying Neural Network Structure for the Classification Purposes-Skin Lesions Case Study", *International Conference on Methods & Models in Automation & Robotics*, Agu.2018 pp. 899-904.
8. Y. K. Jain,M. Jain, "Skin cancer detection and classification using Wavelet Transform and Probabilistic Neural Network",Oct.2012, pp.250-252.
9. R. Sonia, "Melanoma image classification system by NSCT features and Bayes classification",*International journal of Advances in Singal and Image Scienc*e,2(2), Dec.2016, pp.27-33.
10. M. Nasir, M. Attique Khan, M. Sharif, I. U. Lali , T. Saba , T. Iqbal, "An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach", *Microscopy research and technique,*81(6), Jun.2018, pp.528-543.
11. B. Harangi, A. Baran, A. Hajdu, "Classification of skin lesions using an ensemble of deep neural networks", *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*,Jul.2018, pp. 2575-2578.
12. E.Almansour, M. A. Jaffar,"classification of Dermoscopic skin cancer images using colour and hybrid texture features. *JCSNS Int J Comput Sci Netw Secur*, 16(4), Apr.2016, pp.135-9.
13. H. Demirel, G. Anbarjafari,"Image resolution enhancement by using discrete and stationary wavelet decomposition", *IEEE transactions on image processing*, 20(5), Oct.2010, pp.1458-1460.
14. X. H. Wang, R. S. Istepanian, Y. H.Song, "Microarray image enhancement by denoising using stationary wavelet transform", *IEEE Transactions on Nanobioscience, 2(4),Dec.2003*, pp.184-189.
15. B. S. Kirar, D. K. Agrawal, "Empirical wavelet transform based pre-processing and entropy feature extraction from glaucomatous digital fundus images". *International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, Oct.2017, pp. 315-319.
16. A. Rizal, R. Hidayat, H. A. Nugroho, "Multilevel wavelet packet entropy: A new strategy for lung sound feature extraction based on wavelet entropy", *International Conference on Robotics, Automation and Sciences*, Nov.2017, pp. 1-5.
17. S. Afsal, J. Jothykumar, S. Ahmed, F. Sayeed, "A novel approach for palm print recognition using entropy information features", International Conference on Wireless Communications, Signal Processing and Networking,Mar.2016, pp. 1439-1442.
18. J. Zhang, J., Jiang, P., Zhang, X. and Wang, F., 201. A Random Forest Concentration Detection Model for Short Video", I*nternational Conference on Cloud Computing and Intelligence Systems (CCIS)*, Nov.2018, pp. 850-854.
19. M. Jin, L. N. Govindarajan, L. Cheng, "A random-forest random field approach for cellular image segmentation", *International Symposium on Biomedical Imaging*, Apr.2014, pp. 1251-1254.
20. PH2 Database Link:https://www.fc.up.pt/addi/ph2%20database.html

AUTHORS PROFILE

Dr. S. Mohan Kumar M.E[CSE]., PhD[CSE] has two decades of experience in the fields of academic administration, research, Quality Assurance and also in educational consultancy. He has six years of experience as Head of the department IT/CSE.He is Editor-in-Chief of International Scientific Journal of Contemporary Research in Engineering, Science and Management.