

Classification Of Skin Cancer Images Using Discrete Wavelet Transform Features And Support Vector Machine

G.Neela Krishna Babu ¹ and Dr.V.Joseph Peter ²

¹Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India

E-mail: mdtgnk@gmail.com

²Research and PG Department of Computer Science, Kamaraj College, Thoothukudi, Tamilnadu, India

E-mail:peterchristy_in@yahoo.com

ABSTRACT

In this research, an efficient skin cancer detection technique based on Discrete Wavelet Transform (DWT) with Support Vector machine (SVM) is proposed. In this, all skin cancer images from ISIC 2018 dataset (International Skin Imaging Collaboration 2018 dataset), are pre-processed using a random sampling technique. The significant cancer features from these pre-processed images are extracted using DWT. This technique yields four significant features in the form of frequency sub-bands, namely High-Low (HL), Low-High (LH), Low-Low (LL) and High-High (HH). Here, the LL sub-band is having cardinal pixel information of an input image. So, the LL sub-band is further processed using SVM classifier to detect the type of skin cancer in dermatoscopic images. The detected results are compared with the existing results, for performing an evaluation using Accuracy, Precision, F1-score, Recall or Sensitivity and Specificity. The performance of our proposed method achieves 25% higher accuracy than the existing Random Forest (RF), K-Nearest Neighbor (KNN), Naïve Bayes and SVM classification.

Key Words: Skin Cancer, Detection, Discrete Wavelet Transform (DWT), Random Sampling, Machine Learning, Support Vector Machine (SVM).

1. INTRODUCTION

Skin cancer is the abnormal growth of cells caused by DNA damage [1]. It is appeared in three kinds of layers: outermost layer (epidermis); middle layer (dermis); innermost layer (hypodermis) [5]. Based on mortality rate and prevalence, skin cancer is divided into two types, namely malignant and benign [17]. The malignant cancer spreads across the skin layer and it has a high mortality rate. Benign cancer causes weighing pounds and it has a low mortality rate [16]. The illuminated and magnified images of skin cancer are acquired by an efficient imaging technique named as dermatoscope [6]. This technique is used to enhance the visual effect of the Region of

Interest (ROI) for obtaining detailed and very deeper levels of lesions through the surface reflection [12]. Fortunately, if skin cancer is identified in its early stage, then the proper diagnosis and treatment can ensure a complete recovery [18].

The detection of cancer types from dermatoscopic skin images is a difficult task because all cancer images contain volatile pixel intensity deviations [8]. The images from ISIC 2018 dataset are used for a proposed algorithm [21]. This dataset contains seven cancer types, namely MEL (melanoma), NV (melanocytic nevus), BCC (basal cell carcinoma), AKIEC (actinic keratosis), BKL (benign keratosis), DF (dermatofibroma) and VASC (vascular Lesion) [9]. The image samples from this dataset are pre-processed using random over and under sampling to rebalance the seven cancer class labels [23]. Then, the cancer identification is done on these pre-processed images. This identification technique is done traditionally by biopsy based manual method [19]. The accuracy and rater variability of this task depends on the physician's knowledge [11]. To avoid these drawbacks, machine learning methods like K-means, Random Forest (RF), Markov Random Field (MRF), Particle Swarn Optimization (PSO), K-Nearest Neighbor (KNN), Conditional Random Field (CRF) and Fuzzy-C-Means algorithms are used to detect cancer in dermatoscopic skin images [15]. These methods require some handcrafted features to detect skin cancer that requires large computation and time-consuming. These methods need a high end level of programmer interaction to initialize the parameters which are having limited accuracy on real-time cancer images.

Recently, the Support Vector Machine (SVM) method shows significantly higher accuracy in cancer detection process. In this research, the Discrete Wavelet Transformation (DWT) based features like High-Low (HL), Low-High (LH), Low-Low (LL) and High-High (HH) have been extracted in pre-processed skin cancer images [2]. In this, the LL feature contains 50% of significant pixels from the original image and it's enough to identify the cancer type's effectively using SVM algorithm. This proposed method needs less storage and computation; it achieves greater performance than the existing state-of-art detection methods.

2. RELATED WORKS

Early detection of skin cancer remains a challenging task in medical image processing [4]. The skin cancer detection task is done in many ways. Traditionally, biopsy-based manual method is used in the brain tumor segmentation process [7]. In this method, the part of the tissue region is removed from an affected area and tested by the physicians. The accuracy of this task depends on the physician's knowledge [22]. To avoid this limitation, machine learning methods have been used to address the solution of cancer identification [14]. Mostly, the authors are using K-Means, Fuzzy-C-Means algorithm for cancer identification. This method requires an input of the initial cluster point to classify the type of skin cancer [20].

In recent years, hierarchical methods, Conditional Random Field (CRF), Random Forest and Markov Random Field (MRF) are used for identifying the type of skin cancer [20]. The CRF and MRF methods are having very poor performance on volatile intensity variations in real dermatoscopic images. K-Nearest neighbor is another popularly used method to detect skin cancer based on the similarity index. In Random Forest (RF), the random data points are selected from decision trees [15]. Praveen et al has been developed K-Means and PSO based detection algorithm [13]. In this, the optimization is done by inheriting the postulated PSO variant and the K-Means classification algorithm is used for identification. These methods contain high-end level of programmer interaction to initialize the parameters

Murugan et al have been implemented KNN and RF base algorithm to classify the type of skin cancer types [10]. This method contains very limited accuracy on poorly illuminated images. Mane et al [11] and Patel et al [18] have been developed SVM for skin cancer detection, which is having significantly greater performance than KNN and RF algorithms. Most of the existing cancer detection methods are using handcrafted features for performing classification. These handcrafted features require larger computation and high storage space. To avoid this, the DWT based features are extracted in this proposed algorithm. These extracted features are then classified using SVM to identify skin cancer in dermatoscopic images.

3. EXPERIMENTAL METHODOLOGY

This research work proposed an accurate skin cancer detection system based on DWT and SVM techniques. It has five main phases: defining skin cancer dataset; pre-processing of cancer images; the DWT feature extraction; classification using SVM and performance evaluation. The overall architecture of this proposed method is depicted in Figure 1 and detailed in the following sub-sections.

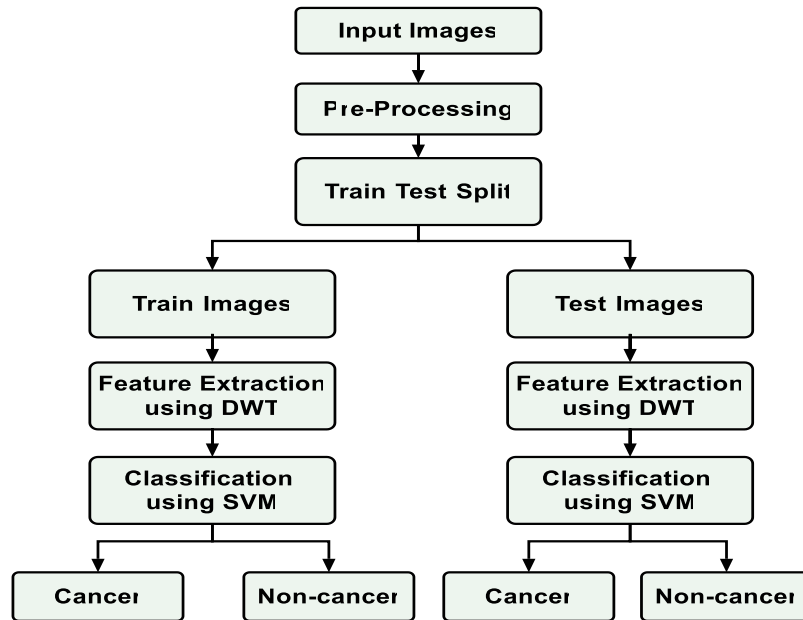


Figure 1: The overall architecture of proposed detection technique.

3.1. Skin Cancer Dataset

The proposed skin cancer detection algorithm is implemented by ISIC 2018 dataset [21]. This dataset consists of 10015 images of seven cancer classes: MEL (melanoma), NV (melanocytic nevus), BCC (basal cell carcinoma), AKIEC (actinic keratosis), BKL (benign keratosis), DF (dermatofibroma) and VASC (vascular Lesion). It is having 1113 images in Mel, 6705 images in NV, 514 images in BCC, 327 images in AKIEC, 1099 images in BKL, 115 images in DF and 142 images in VASC is defined in Table 1.

3.2. Pre-processing

The ISIC 2018 dataset contains an unequal number of image samples in each cancer class. It influences the results of the cancer detection process. To avoid this limitation, the random resampling technique is used to rebalance the class distribution from an imbalanced dataset [24]. There are two main approaches that are mostly used to randomly resample the images from an imbalanced dataset: random oversampling and random undersampling. First, the Average Sample Size (ASS) is computed by adding the cardinality sample values of all classes is given in Equation 1 and Equation 2.

$$ASS = \sum_{i=0}^n (\#C_i) / n \quad \dots (1)$$

$$ASS = (\#C_0 + \#C_1 + \#C_2 + \dots \#C_n) / \dots (2)$$

where, # is the cardinality value and C is the class label. The sample size of each class C is compared with ASS, in which the greater sizes of class labels are called majority class and lesser labels are called minority class. In Table 1, class 1(NV) is considered as majority class and the remaining six classes are coming under minority class. The random undersampling technique is applied over the majority class (NV) to delete some similar samples or images, which turns the size of majority class into ASS. Further, the random oversampling technique is applied over the six minority classes. It is used to add some duplicate image samples in each minority class for getting ASS size. The sample sizes of seven classes are now equal to ASS are detailed in Table 1.

Table 1. The 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
Class 0 (MEL)	1113	1113	1431
Class 1 (NV)	6705	1431	1431
Class 2 (BCC)	514	514	1431
Class 3 (AKIEC)	327	327	1431
Class 4 (BKL)	1099	1099	1431
Class 5 (DF)	115	115	1431
Class 6 (VASC)	142	142	1431
Total	10015	4741	10017

3.3. The DWT Feature Extraction

The pre-processed skin cancer images are processed using DWT to extract cardinal features through the dimensionality reduction. In this research, the Haar-based DWT technique has been used for extracting the coefficient of wavelet by localizing frequency information [3]. In Haar DWT, Each skin cancer image is row-wise decomposed by low-pass and high-pass filter to yield L (Low) and H (High) frequency subbands respectively. These two subbands are then column-wise decomposed to produce four frequency subbands: Low-Low (LL); Low-High (LH); High-Low (HL) and High-High (HH). The detailed image decomposition of skin cancer images using Haar DWT is visualized in Figure 2.

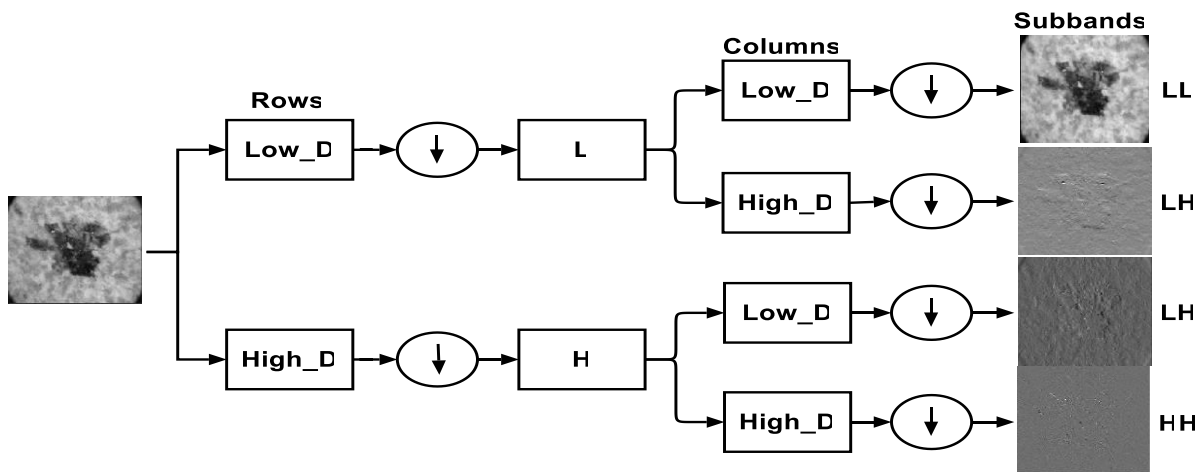


Figure 2: The DWT feature extraction in skin cancer images.

The Haar wavelet function $f(p,q)$ of pixel dimension $M \times N$ of four frequency subbands are given in equation (3)-(7).

$$LL = W_{\phi}(j_o, m, n) = \frac{1}{\sqrt{MN}} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} f(p, q) \phi_{j_o, m, n}(p, q) \quad \dots (3)$$

$$LH = W_{\psi}^H(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} f(p, q) \psi_{j, m, n}^H(p, q) \quad \dots (4)$$

$$HL = W_{\psi}^V(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} f(p, q) \psi_{j, m, n}^V(p, q) \quad \dots (5)$$

$$HH = W_{\psi}^D(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} f(p, q) \psi_{j, m, n}^D(p, q) \quad \dots (6)$$

where, $\psi_{j, m, n}^l(p, q)$ and $\phi_{j_o, m, n}(p, q)$ are the wavelet and scaling function of an input image is defined in equation (7) and (8).

$$\phi_{j_0,m,n}(p, q) = 2^{j/2} \phi(2^j p - m, 2^j q - n) \quad \dots (7)$$

$$\psi_{j,m,n}^I(p, q) = 2^{j/2} \psi(2^j p - m, 2^j q - n), \quad I = \{H, V, D\} \quad \dots (8)$$

where, H is the horizontal decomposition direction, V is the vertical decomposition direction and D is the diagonal decomposition direction. Here, the LL sub-band is having cardinal pixel information of an input image. So, the LL sub-band is further processed using SVM classification to detect the skin cancer types in dermatoscopic images. The feature extraction of seven cancer types using Haar DWT is visualized in Fig. 3.

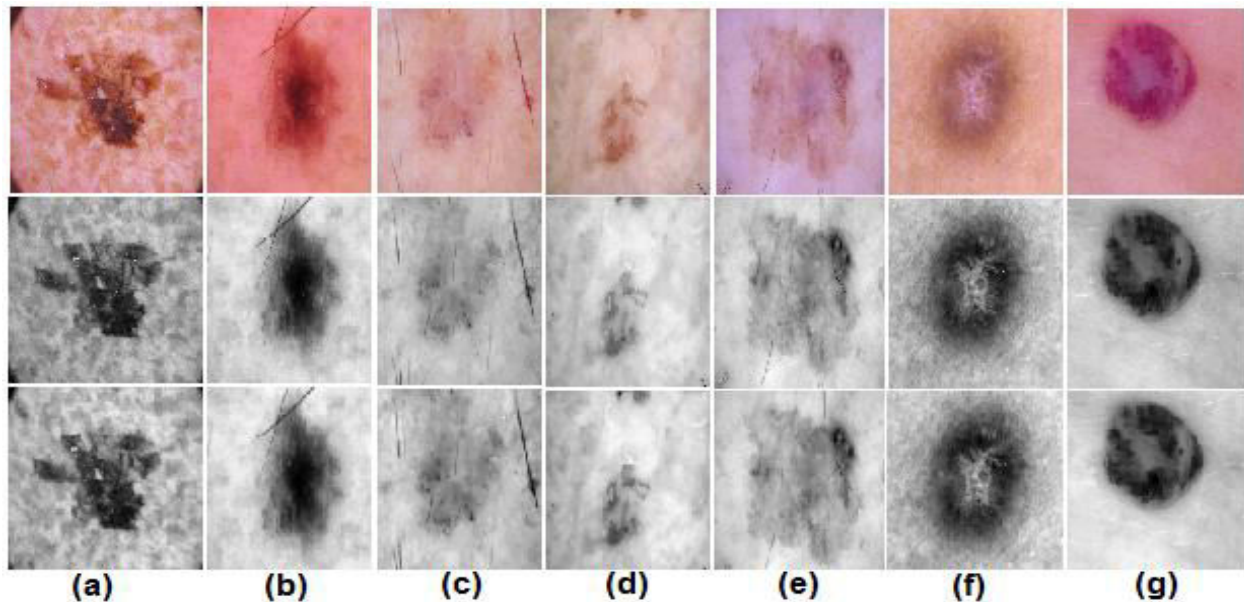


Figure 3: The feature extraction of seven cancer types using Haar DWT: 1st, 2nd and 3rd row represents the skin cancer, gray-scale and Haar DWT feature extracted images of seven skin cancer types; (a) MEL (melanoma); (b) NV (melanocytic nevus); (c) BCC (basal cell carcinoma); (d) AKIEC (actinic keratosis); (e) BKL (benign keratosis); (f) DF (dermatofibroma) and (g) VASC (vascular Lesion)

3.4. Classification using SVM

The extracted DWT features are processed using SVM for detecting cancer types in dermatoscopic skin images. This SVM technique is used to create a decision boundary between class labels. There are two kinds of SVM techniques: linear and non-linear SVM. Linear SVM technique is used to separate linear data points of two class labels using a hyperplane. Non-linear SVM technique is used for multi-label classification in which third dimension z is added for

separating multiple classes using a hyperplane. In this research, the circular RBF (Radial Basis Function) kernel-based non-linear SVM technique is processed over the DWT feature extracted images for detecting the type of skin cancer. The classification of class labels using linear, non-linear and non-linear circular SVM techniques is presented in Figure 4. The RBF kernel is used to map the cancer class labels is given in equation 9.

$$k(y, y') = \exp\left(\frac{\|y-y'\|^2}{2\sigma^2}\right) \dots \tag{9}$$

where, $\|y-y'\|^2$ is squared Euclidean distance of two data points y and y' in an image. It makes a decision boundary between classes to detect the type of skin cancer in dermatoscopic images.

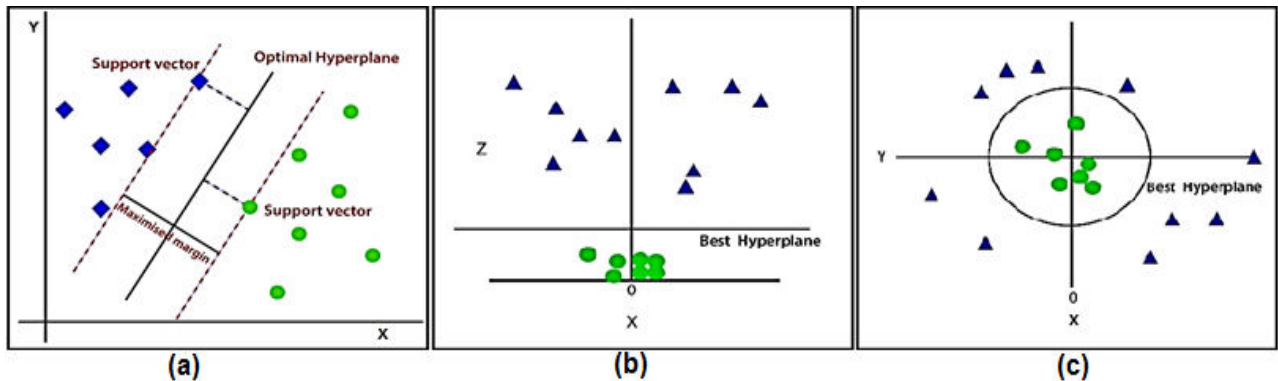


Figure 4: The architecture of SVM technique: (a) Linear SVM; (b) Non-linear SVM and (c) Circular SVM.

3.5. Performance Evaluation

The detected cancer labels of the proposed method are compared with original cancer labels from ISIC 2018 dataset to evaluate performance using Accuracy, Precision, F1-score, Recall or Sensitivity and Specificity. Formulas of these evaluation metrics are detailed in Table 2.

Table 2. Formulas of evaluation metrics.

S.No	Metrics	Formula
1	Accuracy	$Acc = (t_p + t_n)/(t_p + f_p + f_n + t_n)$
2	Precision	$Pre = (t_p)/(t_p + f_p)$
3	F1-score	$F1 - score = (2t_p)/(2t_p + f_p + f_n)$
4	Recall or Sensitivity	$Sen = (t_p)/(t_p + f_n)$

5 Specificity

$$Spe = (t_n)/(t_n + f_p)$$

4. EXPERIMENTAL RESULTS AND DISCUSSION

The efficiency of proposed SVM classification with Haar DWT feature extraction technique is implemented using ISIC 2018 dataset. This dataset consists of 10015 images of seven cancer classes. In this, 75 % of images are divided as training and the remaining 25% of images are considered as testing images. All testing and training images are first converted as gray-scale. Then, the approximation co-efficient of LL (Low-Low) frequency subband is extracted using Haar DWT. These features are further processed using non-linear RBF kernel-based SVM for detecting the type of skin cancer. The detected skin cancer labels of proposed method are compared with original cancer labels from ISIC 2018 dataset to evaluate performance. The performance of training and testing images over seven cancer labels using the proposed method is depicted in confusion matrix of Figure 5.

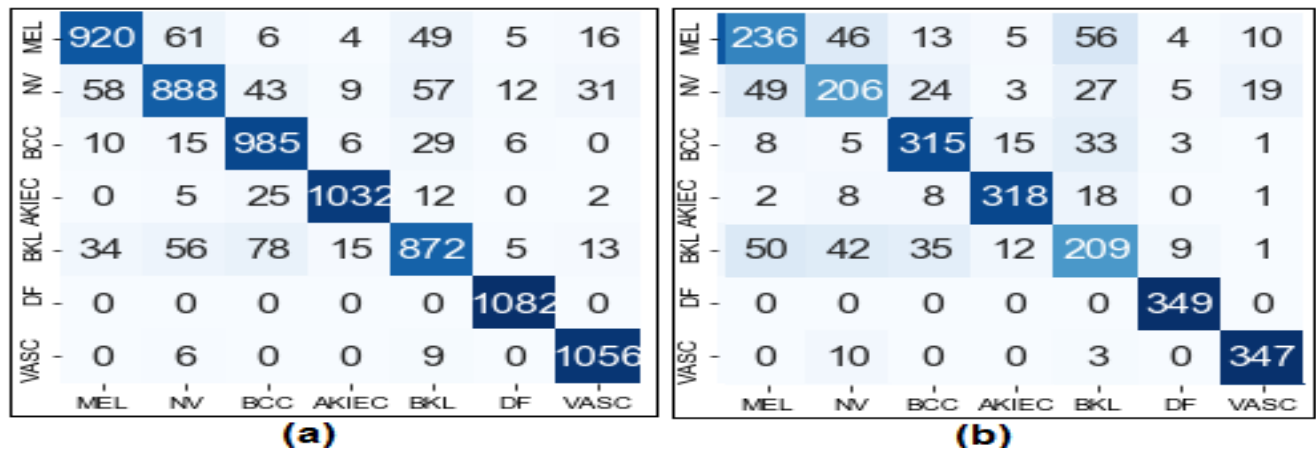


Figure 5: Confusion matrix: (a) Training performance of proposed method; (b) Testing Performance of Proposed method.

The evaluation results of training and testing images using the proposed method are detailed in Table 3. The overall testing and training performance of this SVM with DWT technique achieves 84% of Accuracy, 85% of Precision, 85% of F1-Score, 85% of Recall and 88% of sensitivity values.

Table 3. Performance of DWT-SVM Method.

ISIC 2018 Dataset					
Method	Accuracy	Precision	F1-Score	Recall	Specificity
Proposed Method (Training)	0.90	0.91	0.91	0.91	0.94
Proposed Method (Testing)	0.79	0.79	0.79	0.79	0.82
Proposed Method (Average)	0.84	0.85	0.85	0.85	0.88

Table 4 and Figure 5 show the performance comparison of existing Random Forest (RF), K-Nearest Neighbor (KNN), Naïve Bayes and SVM with the proposed Method. The RF based skin cancer identification method selects random data points using decision trees. These methods contain less performance in volatile intensity variations of real-time cancer images. The KNN method is using a similarity index for cancer detection. Naive Bayes theorem is another popularly used machine learning algorithm for the classification problem, which uses the probability theory of Bayes theorem. This algorithm performs well in small data sizes. The SVM method is having higher accuracy than the RF, KNN and Naïve Bayes, it achieved 67% of Accuracy, 68% of Precision, 67% of F1-score, 67% of Recall and 85% of specificity values. This SVM technique needs higher data storage and larger computation because the ISIC 2018 dataset contains 10015 skin cancer images of pixel dimension 256 x 256. To avoid this limitation, the Haar DWT based SVM technique is proposed. This DWT feature extraction technique yields 10015 images of pixel dimension 128 x 128. These features are classified using SVM for detecting the type of skin cancer. Thus, this proposed method achieves 84% of Accuracy, 85% of Precision, 85% of F1-Score, 85% of Recall and 88% of sensitivity values, which is competitively higher than the existing methods.

Table 4. Performance comparison of DWT-SVM and existing methods

ISIC 2018 Dataset						
Method		Accuracy	Precision	F1-Score	Recall	Specificity
RF	Training	0.70	0.70	0.69	0.70	0.83
	Testing	0.63	0.62	0.61	0.63	0.67
	Average	0.67	0.66	0.65	0.67	0.75
KNN	Training	0.59	0.61	0.59	0.59	0.75

	Testing	0.41	0.41	0.40	0.42	0.69
	Average	0.50	0.51	0.50	0.51	0.72
Naïve	Training	0.30	0.32	0.30	0.30	0.60
Bayes	Testing	0.31	0.34	0.30	0.31	0.62
	Average	0.31	0.33	0.30	0.31	0.61
SVM	Training	0.80	0.81	0.81	0.81	0.90
	Testing	0.53	0.54	0.52	0.53	0.79
	Average	0.67	0.68	0.67	0.67	0.85
DWT-SVM	Training	0.90	0.91	0.91	0.91	0.94
	Testing	0.79	0.79	0.79	0.79	0.82
	Average	0.84	0.85	0.85	0.85	0.88

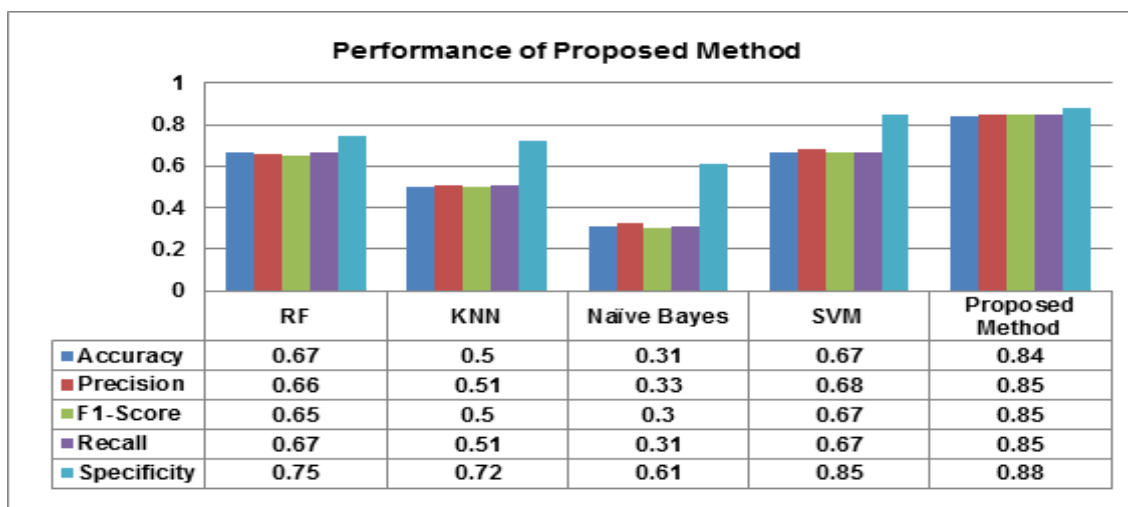


Figure 6: Performance comparison of DWT-SVM and existing methods

5. CONCLUSION

Skin cancer detection plays a vital role in medical image processing. This research work proposed energy-efficient cancer identification system based on the DWT feature extraction and SVM classification technique using ISIC 2018 dataset. First, the random oversampling and undersampling technique is performed over ISIC 2018 dataset to rebalance the samples in seven cancer class labels. The rebalanced dataset is processed using Haar DWT to extract image features and also to reduce the storage and computation load of the detection process. These features are classified using SVM for detecting the type of skin cancer. The performance of this proposed

method achieves comparatively 25% higher detection accuracy than the existing Random Forest (RF), K-Nearest Neighbor (KNN), Naïve Bayes and SVM, classification.

REFERENCE

1. A. Javaria, S. Muhammad, “Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network”, Pattern Recognition Letters, 129, 115-122, <https://doi.org/10.1016/j.patrec.2019.11.016>, 2020.
2. D. Ravichandran, & N. Ramesh, “Mathematical Representations of 1D, 2D and 3D Wavelet Transform for Image Coding”, International Journal On Advanced Computer Theory And Engineering, 2319-2526, 5, 3, 20-27, 2016.
3. Lu-Ting, K., & Jwu-E, C. (2012), Haar-Wavelet-Based Just Noticeable Distortion Model for Transparent Watermark. Hindawi Publishing Corporation, Mathematical Problems in Engineering, 2012, 635738, 1-14, doi:10.1155/2012/635738.
4. M.Z. Alom et al, “Skin Cancer Segmentation and Classification with NABLA-N and Inception Recurrent Residual Convolutional Networks”, <https://arxiv.org/abs/1904.11126>
5. U. B. Ansari, T.Sarade, “Skin Cancer Detection Using Image Processing”, International Research Journal of Engineering and Technology (IRJET), 4,4, ISSN: 2395 -0056 , 2875-2881,2017
6. P. Thsper, A. Singh, “A Survey of Lesion Detection using Dermoscopy Image Analysis”, Journal of Gujarat Research Society, ISSN: 0374-8588, 21, 6, 129-143, 2019..
7. K.Munir, H.Elahi, “Detection Cancer Diagnosis Using Deep Learning: A Bibliographic Review”, *cancers*, 11, 1-36, 2019, doi:10.3390/cancers11091235.
8. S. Jaiswar, M.Kadri, ” Skin Cancer Detection Using Digital Image Processing”, International Journal of Scientific Engineering and Research (IJSER), 3, 6, ISSN (Online): 2347-3878, 138-140.
9. G. Neela Krishna Babu, V.Joseph Peter, “An Analysis of Skin Cancer Detection using Machine Learning”, DRSR Journal, 10,8, 2347-7180, 2020.
10. A. Murugan, S.Anu, “Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers”, Journal of Medical Systems, 1-9, 43:269, 2019, <https://doi.org/10.1007/s10916-019-1400-8>
11. S.S.Mane, S.V.Shinde, “Different Techniques for Skin Cancer Detection Using Dermoscopy Images”, International Journal of Computer Sciences and Engineering, 5,12, ISSN: 2347-2693, 159-163, 2017.

12. Y.Fujisawa, S.Inoue, “The possibility of Deep Learning-Based, Computer-Aided Skin Cancer Classifiers”, *Frontiers in Medicine*, 6, Article 191, Aug 2019.
13. R.Kaur, Praveen kumar, “An Enhanced and Automatic Skin Cancer Detection using K-Mean and PSO Technique”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8, 9S, ISSN: 2278-3075, 634-639, 2019.
14. Vijayalakshmi M. M, “Melanoma Skin Cancer Detection using Image Processing and Machine Learning”, *International Journal of Trend in Scientific Research and Development (IJTSRD)*, 3,4, ISSN: 2456 – 6470, 2019.
15. M.A.Shukran, N.S. Mariam, “Melanoma Skin Cancer Diagnosis Device Using Image Processing Techniques”, *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, 7, 5S4, 490-494, 2019.
16. K.Muthukumar, P. Gowthaman, “GTCM Based Skin Lesion Melanoma Disease Detection Approach for Optimal Classification of Medical Images”, *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878, 7, 6S3, 2019.
17. E. Okur, M. Turkan, “A survey on automated melanoma detection”, *Engineering Applications of Artificial Intelligence*, 73, 50-67, 2018.
18. I.Patel, S.Patel, “Dermoscopic Image Classification Using Image Processing Technique For Melanoma Detection”, *International Journal of Research in Advent Technology, Special Issue*, ISSN: 2321-9637, 97-103, 2019.
19. D.A. Gavrilov, N.N. Shchelkunov, “DEEP LEARNING BASED SKIN LESIONS DIAGNOSIS”, *Remote Sensing and Spatial Information Sciences*, XLII-2/W12, 81-85, 2019.
20. A. Masood, A. Al-Jumaily, “Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms”, *International Journal of Biomedical Imaging*, Volume 2013, Article ID 323268, 22 pages, <http://dx.doi.org/10.1155/2013/323268>