Received: March 1, 2017

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Automatic Detection and Classification of Skin Cancer

Akila Victor [1]* Muhammad Rukunuddin Ghalib[1]

[1]*Vellore Institute of Technology, Vellore, Tamilnadu, India*
* Corresponding author's Email: akilavictor@vit.ac.in

Abstract: Cancer is a deadly disease in today's world. Various types of cancers are spreading for which skin cancer becomes a very common cancer nowadays. Skin cancer can be of two types namely melanoma and non melanoma cancer. The objective of this paper is to detect and classify the benign and the normal image. Benign meaning the normal image, melanoma the cancerous one. And more over compare the various classification algorithms. Detection of skin cancer in earlier stages can be a life saving process. The detection of skin cancer includes four important stages namely Pre-processing, Segmentation, Feature Extraction and Classification. Detection can help in curing the cancer and hence detection plays a very vital role. In this paper, pre-processing the first and the foremost part of image processing which helps in noise removal is done by means of the median filter where the output of the median filter which is fed as an input to the histogram equalization phase of the pre-processing stage, then the input of the histogram equalized image is fed as an input to the segmentation phase where Otsu's thresholding is done to separate the foreground and the background. The segmentation helps to identify the region of interest, Now using the area, mean, variance and standard deviation of the extracted output from the segmentation phase the calculations for feature extraction is carried and the output is fed into classifiers like Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Decision tree(DT) and Boosted Tree(BT).Comparison of the classification is done. The algorithm shows the accuracy of the classification rate of KNN is 92.70%, SVM is 93.70%, Decision tree (DT) is 89.5% and finally the boosted tree (BT) is 84.30%.

Keywords: SVM, KNN, Decision Tree, Boosted Tree

1. Introduction

Cancer is increasing day by day in today's world. Most of the people are suffering from cancer. The treatment given during those periods are really painful. There are various stages of cancer. Detecting cancer in earlier stages can have a chance to cure the disease. Later stages are very difficult .The treatment for cancer is very painful. They have to undergo a series of chemotherapy followed by a surgery if required and then a radiation. Cancer can be in any part of the body. There are various types of cancer namely internal organ and external organ cancer, comparatively the cure rate of external cancer is more when compared to that of the internal cancer. The types are breast, brain, colon, lung, uterus, bladder, cervical, skin, kidney, liver pancreatic, thyroid and lot more. This paper focuses much on to the skin cancer and detecting of skin cancer from the images. Fig.1 explains the various stages of detection namely pre-processing, segmentation, Feature extraction and classification. There are various types of skin cancer like basal cell carcinoma, squamous cell carcinoma. These are skin cancers that need treatment just high above the minor treatments that are being given. The most crucial thing is whatsoever the types of cancer it might be there are various stages that involves namely stage 1, stage 2, stage 3 and stage 4. The stage 1 to stage 3 have higher rate of cure, whereas the stage 4 comparatively has a lesser chance when compared to that of the other stages. Skin cancer can have another type namely melanoma, melanoma is actually the one which causes mostly the deaths. Our aim focuses on finding out or detecting the cancer and classifying the same as if it is a melanoma or not. Skin cancers are caused usually to people who expose too much to the sunlight and

also a melanoma can also be occurring on the basis on the moles, it can develop from the moles. Basal cell carcinoma and squamous cell carcinoma fall under the category of non melanoma cancer. The proposed method can work as a complete flow from pre-processing to the classification phase as it describes all the steps which are involved in the entire identification of cancer cells, and moreover it also includes the stages from pre-processing to the classification which helps to find out the entire flow of the work that is to be performed for the detection of cancerous area. The textural features help in increasing the accuracy of the classification algorithms. And moreover the comparison of various classification algorithm helps in finding out which could be a suitable classification algorithm for the above performed sequence of stages for detection of the melanoma or not.

As the proposed method helps in a good classification accuracy using the support vector machine the features considered can be good, but still this could be improved by using various feature selection methodology and by considering more features other than area, mean, standard deviation and variance. And as a continuation to this the upcoming methodology can include more number of features to improve the classification rate further.

1.1 Signs and symptoms

The signs and symptoms are very important a small mole that is changing in shape, colour and texture. The signs are given below and explained in detail and more over signs and symptoms have to be consulted with a physician as early as found so that proper treatment will be given. In case of carelessness the risk will become higher. Sooner the better always. The cancer is actually a disease which can spread to all parts of the body and finally land up in a very big trouble taking a patient to the most crucial state of his entire life.

There are various signs and symptoms to which a skin cancer can be found early and detected so that it can be helpful for people to get cured to. The symptoms can be pale patch to the skin, brownish scar, may bleed as day's move on, changes in shape or size of existing moles and it includes the concept of ABCDE where A stands for Asymmetry, B for Border, C for color, D fro diameter and E for Evolving size shape or color.

It is always good to have a regular check on your skin so that the skin cancer if occurred could be found at earlier stages. It also has the chance of spreading to all parts of the body if not treated properly.

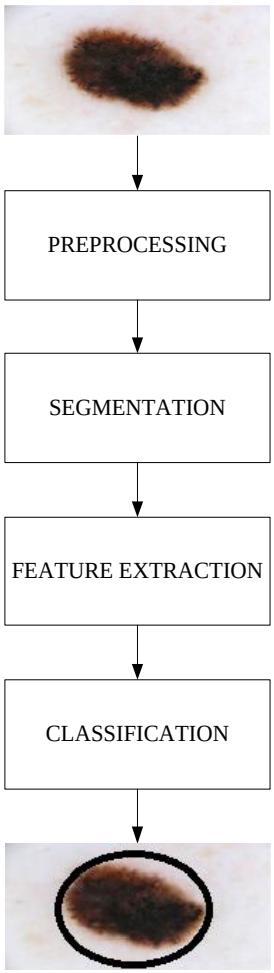

Figure. 1 Basic steps for detection of skin cancer that are pre-processing segmentation, feature extraction and classification

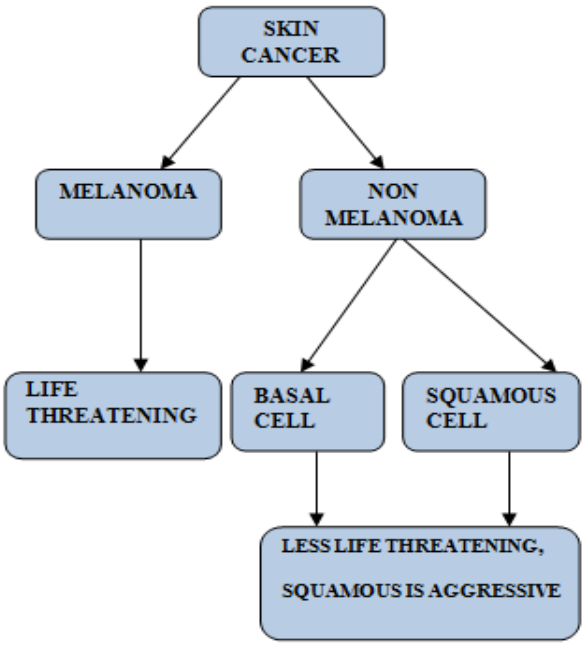

Figure.2 Various types of skin cancer like melanoma, basal and squamous basal cells. And the significance of the types are represented.

1.2 Types of skin cancer

There are various types of skin cancer as discussed below. The skin cancer in Fig.2 can be classified into many types namely melanoma and non-melanoma. The melanoma type cancer are said to be the mostly deadly one and a life threatening one, whereas the non melanoma namely basal and squamous are cancers that do not threat life and can still be cured .The skin melanoma is typically the most dreadful one which could be a life threatening one. The melanoma has to be identified at the very early stage for easy curing of the disease. Therefore the strong need of detection and classification of the cancer cells is very much needed since it helps in finding out the melanoma and the benign classes. The initial image is the image that is actually an image with melanoma or benign image. Now they are subjected to the method of pre-processing with the median filter and the output of the median filter is fed as an input for the histogram equalization method. The median filter in addition with the histogram equalization is used to pre-process the image. The pre-processed image is then segmented using the most commonly used thresholding method called as otsu's thresholding. In this method we separate the foreground and the background image with the help of Eq. (1). Then the required features are collected as area,mean ,variance and standard deviation(SD) which can also be called as textural features are used to extract the necessary features that are fed as an input to the classifiers namely support vector machine(SVM) , K-Nearest Neighbour (KNN), Decision Tree(DT) and Boosted Decision Tree(BT). The propose method is a good method and it proves to have a better accuracy and detection rate as per the study in researches. Accuracy is the key.In proposed method, the performance of the classifiers of measured using sensitivity, selectivity and accuracy.

This paper is organised as follows section 2 presents the review of the related work and the section 3 describes the pre-processing of skin cancer images , the hair removal algorithm, the segmentation methodology, the feature extraction on the segmented image and finally the classification done by four different classifiers. Section 4 describes the experimental results that are found by the above steps. And finally section 5 draws a conclusion.

2. Related works

The paper [1] describes on the classification methodology of k-law Fourier non linear technique. It is done for Red, Blue Green channels where green band is chosen. Binary mask is used to multiply, Fourier transform is applied to the result, if positive take those values else print zero. The pre-processing method is done in a frequency domain. Ref. [2] speaks on various pre processing methods enhancement, restoration and hair removal methodologies. Enhancement deals with scaling, contrast stretching and restoration focuses on removing noise and removing the blur and the hair removal includes morphological methods. Ref. [3] proposes on classification using back propagation algorithm. It helps to classify the cells using melanoma and non melanoma. Classification accuracy is 100% in the work. Ref. [4] points on image in frequency domain, and is on Fourier spectral analysis. The main focus is on the Fourier analysis. Ref. [5] suggests a neural network process in addition to fuzzy inference system. The features chosen here are area and colour and a basic neural network is applied for classification. Ref. [6] pre-processing here is done by KL transform histogram equalization, Region of interest for segmentation using thresholding and statistical region merging(SRM) out of which SRM is comparatively better based on the results mentioned, Feature extraction supports wavelet decomposition and classification explains feed forward network. Ref. [7] explains wavelet transformation for pre-processing and also ABCD [8-10] and fuzzy inference for feature extraction and classification of picture colour severity.

The proposed method has the advantage of improving the classification rate. The features extracted improve the classification accuracy. The performance depends on the dataset size as well. The execution time is a crucial one. The algorithms chosen so far in the proposed method are relatively better. And the classification algorithms that are used are compared to give an accuracy rate. And that can be finalised and distilled to give a best classification algorithm.

3. Proposed methodology

The proposed methodology discusses on the input data to be taken as the malignant or the benign image that is 2 classes that are discussed throughout ,which is then fed for the pre-processing using the median filter and the histogram equalization methodologies for further enhancement of the result. Then the output is send as an input to the hair removal operation which is by means of the morphological operators. Then the image is further send for segmentation using the otshu's thresholding

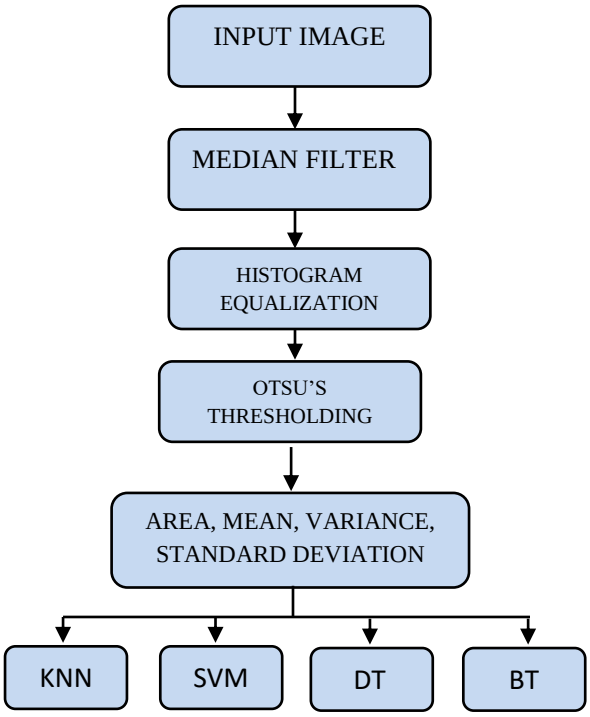

Figure.3 The proposed methodology that includes filters for noise reduction and otsu's threholding for segmentation and features are considered and various classifiers are used to compare.

Table 1. The pre-processing parameters like peak signal to noise ratio, signal to noise ratio and mean squared error are calculated and tabulated.

Input Image	PSNR	SNR	MSE	Class
1	28.2229	21.102	82.4637	2
2	39.965	37.671	71.7346	2
3	28.2229	21.102	82.4637	2
4	44.33	39.5529	17.7792	2
5	39.6647	36.4872	79.1467	2
6	34.3022	29.4766	56.6169	2
7	34.1117	29.5092	51.526	2
8	40.8652	36.6337	35.6609	2
9	31.7477	25.3591	75.6119	2
10	45.807	40.9855	17.9299	2

method and then the features are selected and the classification is performed by means of K-Nearest Neighbourhood (KNN) method. And it is then compared with support vector machine (SVM), Decision Tree and Boosted Tree classification methodology. The proposed methodology uses the dataset which include 1000 plus images which helps in calculating the features with various fields like age of the patient, sex of the patient etc. The classification accuracy can also be increased by adding up few more images to the dataset. The proposed methodology in detail discusses about how well the segmentation is done and features are considered so as to improve the classification accuracy.

3.1 Pre-processing

The first and foremost step in image processing is pre processing. The proposed method speaks of two techniques on enhancement namely median filter and histogram equalization from Fig.3. The median filter picks up the input image arrange the pixels in ascending order and then pick the centre value and replace the existing pixel by the newly formed pixel value. The output of the median filter is now fed as an input to the histogram equalization enhancement technique both are done in spatial domain. The output of median filter is arranged on various pixel levels the running sum is calculated and is divided by the total number of the running sum and each element is multiplied by the total gray level values and the result is obtained based on that. The above table tabulates the peak signal to noise ratio (PSNR), Signal to noise Ratio (SNR), Mean Square Error (MSE).

3.1.1. Hair removal

The histogram processed image is fed as an input for the hair removal pre-processing phase. As the skin in any human has hairs present all over the body whatever the input is, it is subjected to the hair removal process. The hair removal can be done by means of morphological operators to the input image and the bottom hat transform is applied to the image and the next step is identifying the long and thin objects and finally using marker concepts the final result is obtained. Hair is removed from the image.

3.2 Segmentation

The next step after the pre- processing phase is the segmentation phase, as segmentation helps to identify the region of necessity and hence otsu's thresholding is done here. otshu explains with both the background and the foreground image by means of the histogram that is calculated and we consider o to 7 intensity level calculations and assumes for 7 total values with the formula given below

$$\sigma_W^2 = W_B\sigma_B^2 + W_F\sigma_F^2 \quad (1)$$

This algorithm helps us to find the background and foreground pixels in an image and equate to the

no of pixels found against the total no of background and foreground pixels. Then the mean and the variance of the corresponding foreground and the background is estimated. Eq. (1) explains the weight of the background image W_B and the background level multiplied with the variance of the background σ^2_B.Then the weight is found for foreground W_F and the corresponding level is multiplied with the variance of the foreground σ^2_F. These two are summed together to form the threholding value. Then with the help of the weight and the variance and with various levels of thresholding the value is calculated.

3.3 Feature extraction

The feature extraction is the third important step of the detection. The features have to be selected, the corresponding feature that is selected in our work are area, mean, variance and standard deviation. The mean, variance and the standard deviation can be calculated using the standard formula and the area as well. The features are very important as of classification is concerned and the feature extraction helps in the detection as well. Figure 4 explains on the various attributes that are considered for extraction of the features. Here we include area, mean (mea) variance (Var) and the standard deviation (SD) for our methodology which could be calculated by standard calculations. There are various feature extraction methodology the GLCM, Gabor filter etc. we use the basic statistical method and in future we would try to take more number of features to further improve the classification accuracy. The extracted features are very important for the classification accuracy. The below table 2 explains on how the features are extracted and the extracted features are fed into the classifiers.

The table 2 explains about various features that are been considered those features are area, variance, standard deviation and mean. And more over the mean is calculated by the standard calculations that is mean is the total value added up and divided by the total number of elements present in it. Similarly the variance and also the standard deviation and the area is calculated. They are fed as an input to the classifiers, thus the texture analysis can be made by checking the attributes listed below. The analysis that is carried out help us to find a better classification rate for all the classifiers those are used and the results are compared.

Table 2. Feature Extraction parameters that includes area, mean, variance and standard deviation (SD)

Image No	Mean	Variance	SD	Area
1	0.7609	0.1819	0.4265	49966.75
2	0.4475	0.2472	0.4972	29358.88
3	0.777	0.1733	0.4162	50948.88
4	0.4475	0.2472	0.4972	29358.88
5	0.7225	0.2005	0.4478	47447
6	0.6297	0.2332	0.4829	41304.13
7	0.674	0.2197	0.4688	44241.88
8	0.5448	0.248	0.498	35775.88
9	0.5136	0.2498	0.4998	33732.88
10	0.8337	0.1386	0.3723	54702.38

3.4 Classification

Classification is a very crucial and important step which is based on feature extraction it should clearly classify if the result is a melanoma or not. So the methodology uses the K-Nearest Neighbourhood (KNN) classification. The KNN classification is considered to be a non parameterized classification algorithm we use the Euclidean distance to calculate the KNN algorithm for classification. SVM classifier uses a hyperplane to classify the pixels. The decision tree decides with the classification in a tree shape. The boosted tree is something which uses functions in each and every stage.

4. Experimental results

The results are discussed below [11]. Figure 5 is the original image that is taken and the original image is then converted to its corresponding gray level values and the values are then applied with the filter where Fig.6 explains the median filter applied and its corresponding histogram that is actually calculated. Figure 7 is the output of the filtered and hair removal methodology applied and then the image is subjected to thresholding where the output is obtained as Fig. 8. Then the feature extraction is done by selecting certain features and then classification is done by comparing SVM, KNN, Decision tree (DT) and boosted tree.(BT).The various classifiers used are compared and then analysed based on the feature extraction data that is provided. The classification that is performed can help in plotting the accuracy components against the classifiers.

There are various other classifiers that can also be used in addition to the ones that are discussed above. For the dataset used the SVM classifier gives

an accurate result. The other classifiers have also given good results but still the results are comparatively lesser and the next better classifier is KNN. And then the classification that yields a result is decision tree (DT) and the final one can be boosted tree. Comparatively SVM can give us a better result.

This is the input image in Fig. 5. Which can be fed as either a melanoma or normal image? The input image is then fed into a pre-processor to remove the noise. The median filter is then used to filter. It helps to enhance the image. Figure 6 explains the histogram equalized image and the filtered image in Fig. 7 are also represented. Thus the image can be pre-processed by the filters and the result obtained from the figures. Figure 8 explains about the segmented image. The segmented image helps to give the region of interest.

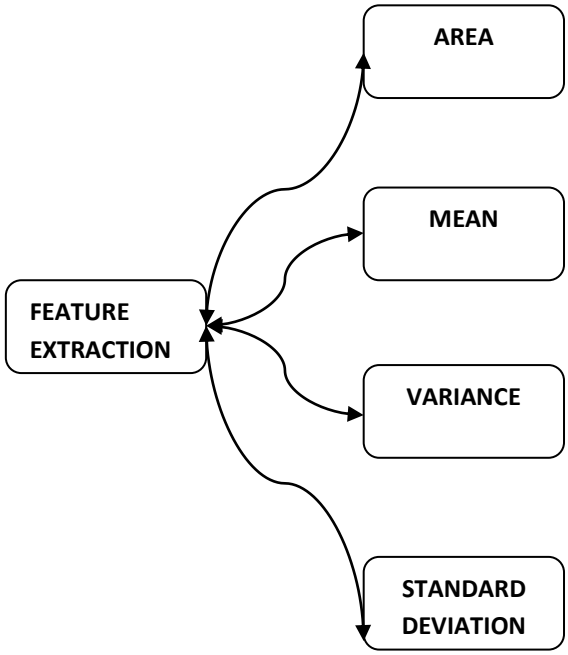

Figure.4 The feature extraction has various factors that are considered like area, variance, standard deviation and mean

Figure.5 The input image is a image with wound which could be a melanoma or not.

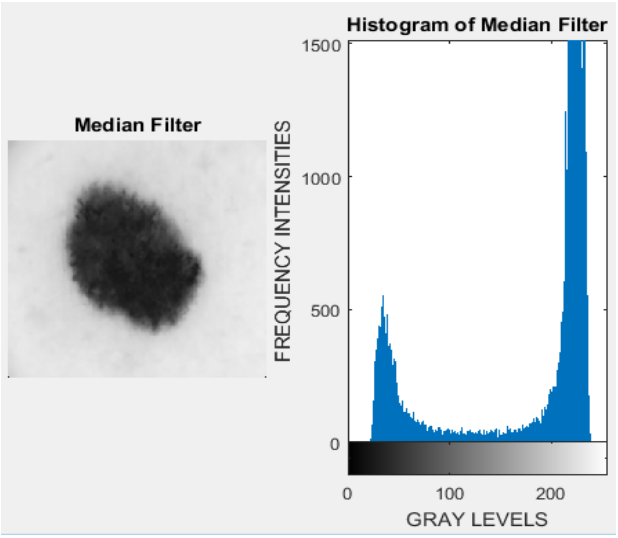

Figure.6 The median filtered image and its corresponding histogram equalization obtained based on the filter.

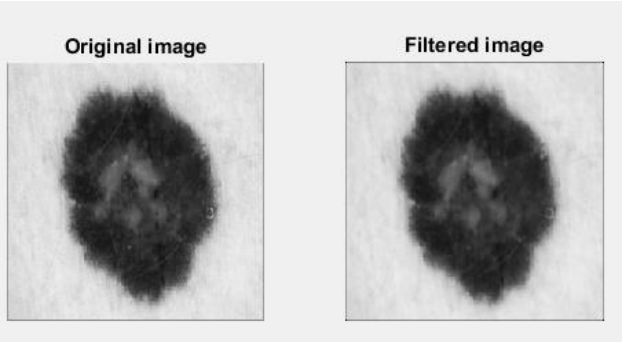

Figure.7 The image obtained after applying both the median filter and the histogram equalization.

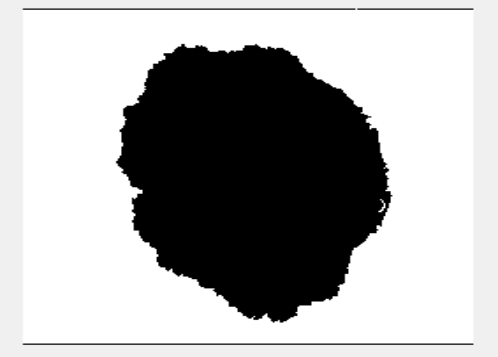

Figure.8 Segmented image after applying the otshu's thresholding.

Table.3 Comparison on classification algorithms based on sensitivity, specificity and the accuracy.

Classifiers	Sensitivity	Specificity	Accuracy
SVM	80.64%	98.09%	93.70%
KNN	93.54%	93.67%	92.70%
DT	67.74%	94.93%	89.50%
BT	96.77%	100%	84.30%

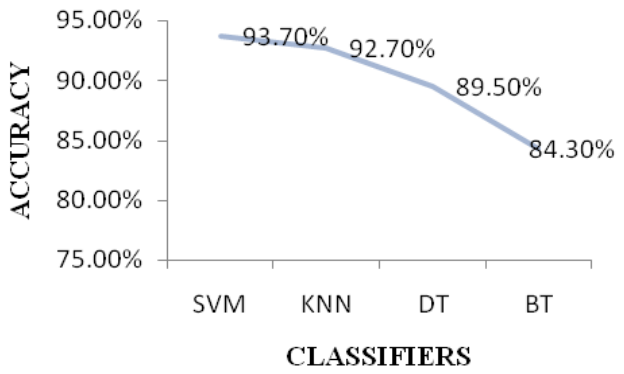

Figure.9 comparison of classification performance of four different algorithms SVM, KNN, DT and BT.

The classification for the above four algorithms are performed and KNN finds an accuracy rate of 92.70%, SVM is 93.70%, Decision tree (DT) is 89.5% and finally the boosted tree (BT) is 84.30%.

Based on the accuracy obtained from the 1011 images that are considered for the proposed methodology the accuracy holds better for the Support vector machine classification. The data are trained and tested to obtain the classification rate for each and every classifier discussed above. Figure 9 is about the classification accuracy of all the four classifiers that we have taken and the true positive and false positive have also been computed. The value of sensitivity and selectivity are calculated by True Positive(TP),True Negative(TN), False Positive(FP), False Negative(FN).There are three important parameters that are derived from the confusion matrix obtained they are accuracy sensitivity and the specificity .The classification accuracy focuses on the specificity and sensitivity. Sensitivity and specificity is given below. The accuracy is calculated as follows:

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

The sensitivity is calculated as follows:

$$SENSITIVITY = \frac{TP}{TP + FN} \quad (3)$$

The specificity is calculated as follows:

$$SPECIFICITY = \frac{TN}{TN + FP} \quad (4)$$

The above Eqs. (2), (3) and (4) explain about how the classification is made if it is a melanoma or not. The values TP and TN finds if the desired output is benign or melanoma. Benign is fed as class 1 and melanoma as class 2. And FP and FN are the classifications which are falsely made.

From the above proposed method, we infer that the algorithms used such as the pre-processing, segmentation, feature extraction and classification. The basic steps in any of the medical image processing include the above. The noise removal strategy of the algorithm the multiple regions split up the extraction of features and finally to start off with the classification entry. Here in the proposed work, we use the median filter and the histogram equalisation and the median filter can be applied. They are exclusive noise removal filters. The median filter helps in finding the mid pixel and replacing it with the newly formed image. And then the output is fed for otshu's thresholding where otshu value can be obtained by Eq. (1) that is explained by separating the background image from the foreground image the background image can be identified by the number of 0 valued pixels and the foreground image are the number of ones in the image.

The segmented image then can be fed to the features which are mean, area, variance and standard deviation. These above features are calculated here with the help of the mean calculated and the other features that are tabulated in table 2. The feature values are now fed into the various classifiers like K-Nearest Neighbor, Support Vector Machine, Boosted tree, Decision Tree.

5. Conclusion

In this paper an effective detection of skin cancer cells is proposed. Four features are chosen that are trained tested by using various classification techniques like SVM, KNN , Decision tree and Boosted tree have been done. The methodology actually has a good result for the various classifiers but still can be improved. The result discusses that SVM is a better classifier than KNN, DT, BT. SVM has the highest accuracy of 93.70%. The results achieved are relatively good when compared. The future work will focus on improving the number of features selected and then classified to improve the accuracy. Since medical image dataset's classification must be very accurate. The dataset can be increased for future work to give better accuracy.

Acknowledgments

I would take this opportunity to thank my institution VIT University for extending their support and providing me the resources necessary for completing this paper.

References

[1] E. Guerra-Rosas, J. Alvarez-Borrego, and A. Coronel-Beltran, "Diagonosis of skin cancer using image processing", In: *Proc. of the AIP Conference*, Vol.1618, No.1, pp. 155-158, 2004.

[2] A.N. Hoshyar, A. Al-Jumaily, and A.N. Hoshyar, "The beneficial techniques in preprocessing step of skin cancer detection system comparing", *Procedia Computer Science*, Vol.42, pp. 25-31, 2014.

[3] G. Kaur and S. Singla, "Propagation method for detection of skin cancer", *IJIET*, Vol.7, No.4, pp. 284-294, 2016.

[4] J.A. Jaleel, S. Salim, and R.B. Ashwin, "Computer aided detection of skin cancer", In: *Proc. of the 2013 International Conference on Circuits, Power and Computing Technologies*, pp.1137-1142, 2013.

[5] B. Salah, M. Alshraideh, R. Beidas, and F. Hayajneh, "Skin cancer recognition by using a neuro-fuzzy system", *Cancer informatics*, Vol. 10, pp.1-11, 2011.

[6] H.T. Lau and A.A. Jumaily, "Automatically early detection of skin cancer: study based on neural network classification", In: *Proc. of the International Conference on Soft Computing and Pattern Recognition*, pp. 375-380, 2009.

[7] M. Silveira, J.C. Nascimento, J.S. Marques, A.R. Marcal, T. Mendonca, S. Yamauchi, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images", *IEEE Journal of Selected Topics in Signal Processing*, Vol. 3, pp. 35-45, 2009.

[8] S. Sigurdsson, P.A. Philipsen, L.K. Hansen, J. Larsen, M. Gniadecka, and H.C. Wulf, "Detection of skin cancer by classification of Raman spectra", *IEEE Transactions on biomedical engineering*, Vol. 10, pp. 1784-1793, 2004.

[9] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, and A. Huntley, "Segmentation of skin cancer images", *Image and Vision Computing*, Vol. 17, pp. 65-74, 1999.

[10] D.S. Rigel, R.J. Friedman, A.W. Kopf, and D. Polsky, "ABCDE—an evolving concept in the early detection of melanoma", *Archives of dermatology*, Vol. 141, No.8, pp.1032-1034, 2005.

[11] S. Thawkar and R. Ingolikar, "Automatic detection and classification of masses in Digital mammograms", *International Journal of Intelligent Engineering and Systems*, Vol.10, No.1, pp. 65-74, 2017.