Probabilistic Neural Network based Benign and Malignant Skin Cancer Detection

Sathish Parvatham, Jogi Prakash

Department of Electronics and Communication Engineering

Sree Dattha Group of Institutions, Hyderabad, Telangana, India.

Abstract

Skin cancer is leading type of cancer which causes millions of deaths of human beings. Early identification and appropriate medications for new harmful skin malignancy cases are fundamental to guarantee a low death rate as the survival rate. Most of the related works are focusing on machine learning based algorithms, but they failed to provide the maximum accuracy and specificity. Thus, to overcome this problem, the research work is implemented with the Advanced deep learning based probabilistic neural networks (PNN) classification mechanism. Initially, k-means clustering based segmentation approach is used for efficient detection the region of skin cancer. Finally, to archive the maximum efficiency of the system, PNN developed for classification of skin cancer with the gray level co-occurrence matrix (GLCM) based Texture features; discrete wavelet transforms (DWT) based low level features and Statistical Color features respectively. Thus, the research work can be effectively used for classification of Benign and Malignant skin cancers. The simulation analysis shows that the proposed method shows better qualitative and quantitative analysis compared to the state of art approaches.

Keywords: probabilistic neural networks, discrete wavelet transform, gray level co-occurrence matrix, benign, malignant, skin cancer.

1. Introduction

In recent days, skin cancer becomes most effected disease of all the types of cancers and it is divided as benign and malignant. In these two types Malignant is recognized as most deadliest one while comparing with the non-Malignant skin cancers [1].It is known fact that Malignant effects more people year by year wise and early treatment are really important for the survival of the patients. Inspection of malignant needs well experienced dermatologists. These people use computer-assisted system early detection of Malignant [2]. More algorithms in deep learning models were used for diagnosis of skin cancer diagnosis. The accuracy rate of these models is the challenging task are still facing more challenges For achieving the high accuracy rate, models are to be overcome all the drawbacks of conventional models. This paper proposes a novel skin cancer detection approach. Many research papers have utilized image preprocessing for the identification of the Malignant at the initial times, which leads to effective treatment. In this way; it is necessary to broaden the span of such essential diagnostic care by arranging efficient frameworks for skin disease classification. Many research papers have utilized image preprocessing for the identification of the Malignant at the initial times, which leads to effective treatment. Proficient dermatologists have set up the ABCDEs [3-4] (Asymmetrical shape, Border irregularities, Color, Diameter, and Evolution) as the standardized descriptions to help with visualizing standard features of severe Malignant cases. One of the main challenges of classifying harmful skin injuries is due to sheer proportions of varieties over the different skin tones from people of different ethnic backgrounds. Recently, new accomplishments in the improvement of convolutional neural networks (CNN) [5] have permitted computers to beat dermatologists in skin cancer classification tasks. The following phase is to improve the accuracy of

Malignant location further. Our strategy for early diagnosis of skin lesion incorporates deep learning which helps us to enhance the accuracy of automated framework compared methods. In this work, we proposed our custom network for lesion classification. The major contributions of this research as follows:

- K-means clustering based segmentation mechanism is used to identify the cancer region from the input test image.
- Network is trained and tested with the GLCM based Texture features, DWT based low level features and Statistical Color features by using the PNN deep learning model.
- The proposed classification accuracy is compared with the conventional SVM [14] and active contour segmentation methods and gives the better results compared to them.

The remainder of the paper is structured as: Literature survey conducted for the paper is covered in Section II. The Section III covers the proposed skin cancer detection method while Section IV describes the environment in which experiments were conducted. In Section V, the results obtained from experimentation and observations are discussed. Finally, Section VI has the remarks that conclude the outcomes and draws inferences from the presented research work.

2. Literature Survey

There have been several systems developed for detecting Malignant as early as possible using the dermoscopic images. The dermatologist's asses the skin lesions using the "ABCD Rule". Based on this rule, many methods have been devised to classify dermoscopic images. Researchers have used extracted features and attempted to train diverse machine learning classifiers such as k-NN, SVM [6]. In [6-7] authors used very deep and machine learning residual networks to classify the images. In order to cope with degradation and over fitting, first machine learning is applied. Then, Radial basis function network (RBFN) is constructed so that skin lesion segmentation can be accurate. Then, this RBFN and deep residual networks that are used to classify the images are taken together to make a two-stage framework. In paper [8], images have been obtained by Epi luminescence Microscopy, which enhances the chances of early recognition of Malignant as malignant or benign. Binary mask is done, and shape and radiometric features are extracted to detect how malignant a lesion is. After that, the CNN classifier is deployed for classifying images as malignant or benign.

In paper [9], automatic border detection is performed and then shapes are extracted from these borders. Texture features are then computed using GLCM and Euclidean distance transform. Images are then classified using the SVM classifier. In Paper [10-11] uses fractional coefficients of cosine transformed skin image, which results in better space complexity and optimum performance in Malignant skin cancer identification. Ensemble of "SVM- AD Tree-Random Forest" gave the superior performance amongst all the classifiers used. Researchers are now focusing more on deep learning concepts as there have been significant advancements in deep learning. They are using the Neural Network Ensemble model, Very Deep Residual Networks, Artificial Neural Network. But they might have certain drawbacks like more processing power is required or more data is required which might get difficult to find as such datasets are not readily available. In paper [12], an overview of the most important implementations of Malignant detection is given and then comparison of the performance of numerous classifiers on the classification of dermoscopy images as benign or malignant is presented. All the existing approaches[13] of skin cancer detection can be grouped in three streams as Malignant detection with Machine Learning models using spatial domain features, Malignant detection with machine learning models using transform domain

features and Malignant detection using unsupervised neural network models. The transform domain feature based machine learning models of malignant detection are time complexity wise heavier. The SVM models are more complex and do need heavy hardware as well as huge dataset for getting trained in Malignant detection. The spatial domain feature-based machine learning models are simple, faster and applicable to any size of skin dermoscopy images.

3. Proposed Method

The proposed research work majorly focusing on detection of following skin cancers such as Malignant and Benign respectively. The detailed operation of the skin cancer detection and classification approach is presented in figure1.

3.1 database training and testing

Database is trained from the collected images of "International Skin Imaging Collaboration (ISIC)" Archive. ISIC is one of the biggest available collections of quality controlled dermoscopic images. The dataset consisted of 15 benign and 15 malignant images. All the images are trained using the PNN network model with GLCM features, statistical and texture features. And random unknown test sample is applied to the system for detection and classification respectively.

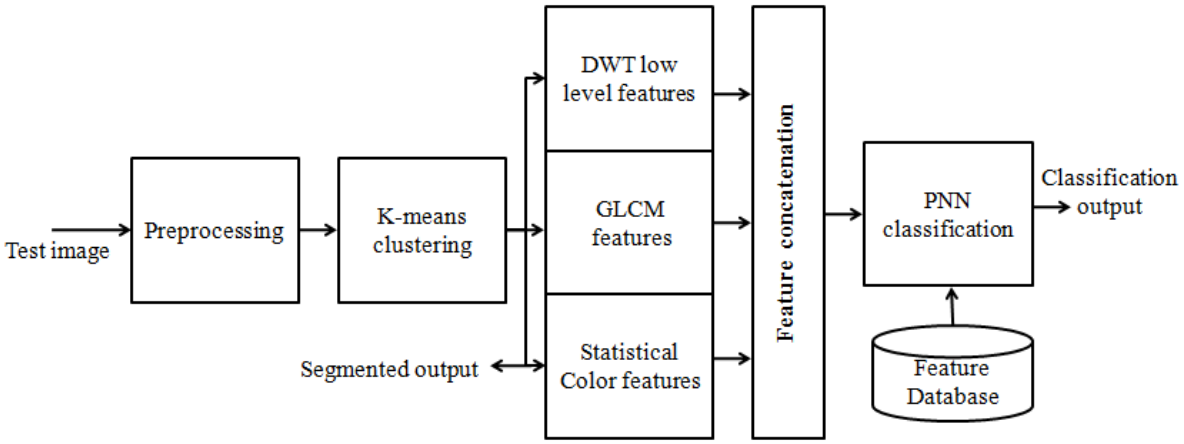

Fig 1: skin cancer detection and classification.

3.2 Preprocessing:

The query image is acquired from image acquisition step, which includes background information and noise. Pre-processing is required and necessary to remove the above-mentioned unwanted portions. The pre-processing stage is mainly used for eliminating the irrelevant information such as unwanted background part, which includes noises, labels, tape and artifacts and the pectoral muscle from the skin image. The different types of noise occurred in the mammogram images are salt and pepper, Gaussian, and speckle and Poisson noise. When noise is occurred in an image, the pixels in the image show different intensity values instead of true pixel values.

So by choosing the perfect method in the first stage of preprocessing, this noise removal operation will perform effectively. Reduction of the noise to a great extent and avoiding the introduction visual artifacts by the analysis of pixels at various scales, sharpening and smoothing filter denoising efforts to eradicate

the noise presented in the pixel, as it conserves the image uniqueness, despite of its pixel satisfied. These filters can effectively detect and remove noise and thin hairs from the image; then we perform top-hat transform for removing the thick hairs. Contrast limited adaptive histogram equalization CLAHE is also performed on the skin lesion to get the enhanced image in the spatial domain. Histogram equalization works on the whole image and enhances the contrast of the image, whereas adaptive histogram equalization divides the whole image and works on the small regions called tiles. Each tile is typically 8*8 pixels, and within each tile histogram is equalized, thus enhancing the edges of the lesion. Contrast limiting is applied to limit the contrast below the specific limit to limit the noise.

3.3 Image Segmentation:

After the preprocessing stage, segmentation of lesion was done to get the transparent portion of the affected area of skin. On transformation, K means clustering method is applied to the image to segment the skin lesion area based on thresholding. In the K means clustering algorithm, Segmentation is the initial process of this work, at the cluster centers, cost junction must be minimized which varies with respect to memberships of user inputs. Image segmentation is the process of dividing the image into multiple clusters based on the region of interest presented to detect the skin cancer. Regions of interest are portion of skin images, which are used by radiologists to detect abnormalities like micro classifications (benign and malignant).

K means clustering is used in the proposed procedure for segmentation to a certain extent than Active counter clustering approach because of its speed of operation with maintaining the highest accuracy. K means clustering procedure combines the properties of jointly possibility and K means clustering approaches as shown in figure 2.

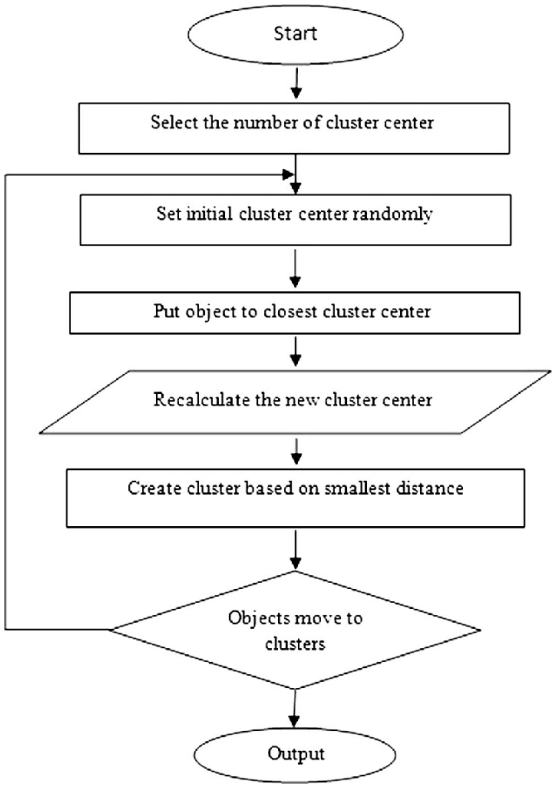

Fig 2: K-means Clustering.

Here the membership functions are generated in the probability-based manner to gets better detection. Among those detected tumors, the highest accurate cancer regions considered as ROI. The automatic extraction of ROI is difficult. So, ROIs are obtained through possibility cropping, which are based on location of abnormality of original test images. Here the membership functions are generated in the probability-based manner to gets better detection. Among those detected Cancer regions, the highest accurate Cancer region is considered as ROI.

3.3 Feature extraction: Several features can be extracted from the skin lesion to classify the given lesions. We extracted some of the prominent features which help us in distinguishing the skin lesions, those are GLCM based Texture features; DWT based low level features and Statistical Color features respectively.

GLCM is a Texture technique of scrutinizing textures considering spatial connection of image pixels. The texture of mage gets characterized by GLCM functions through computations of how often pairs of pixels with explicit values and in a particular spatial connection are present in image. GLCM matrix can be created and then statistical texture features are extracted from the GLCM matrix. GLCM shows how different combinations of pixel brightness values which are also known as grey levels are present in image. It defines the probability of a particular grey level being present in the surrounding area of other grey level. In this paper, the GLCM is extracted first from the image for all three-color spaces i.e. RGB, CIE L*u*v, and YCbCr. Then the GLCM matrix is calculated in four directions which are 135^o, 90^o, 45^o, and 0^o degrees as shown in figure 3. In the following formulas, let a, b be number of rows and columns of matrix respectively, $S_{a,b}$ be the probability value recorded for the cell (a, b), and number of gray levels in image be 'N'. Then several textural features can be extracted from these matrices, extracted textural features are as shown in following equations:

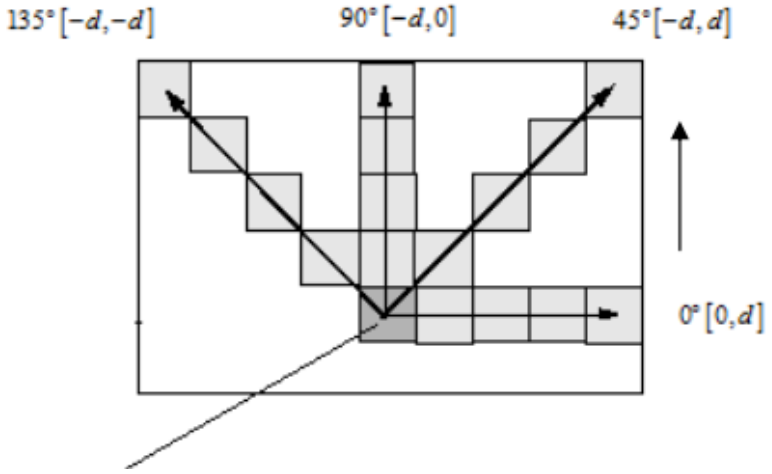

Fig 3: Orientations and distance to compute GLCM.

GLCM features used are

Contrast = $\sum_{a,b=0}^{N-1} S_{a,b}\,(a-b)^2$

Homogeneity = $\sum_{a,b=0}^{N-1} \frac{S_{a,b}}{1+(a-b)^2}$

Correlation = $\sum_{a,b=0}^{N-1} S_{a,b} \left[\frac{(a-\mu_a)(b-\mu_b)}{\sqrt{(\sigma_a^2)(\sigma_b^2)}} \right]$

Angular Second Moment (ASM) = $\sum_{a,b=0}^{N-1} s_{a,b}^2$ and Energy = $\sqrt{ASM}$

Then, 2 level DWT is also used to extract the low-level features. Initially on the on the segmented output DWT is applied, it will result the output as the LL1, LH1, HL1 and HH1 bands respectively. Then entropy, energy and correlation features are calculated on the LL band. Then, on the LL output band again DWT is applied, and results the output as LL2, LH2, HL2 and HH2 respectively. Again entropy, energy and correlation features are calculated on the LL2 band respectively as shown in figure 4.

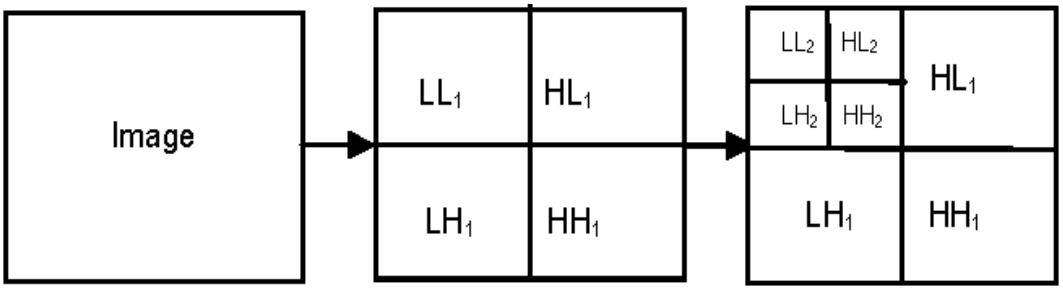

Fig 4: 2 –level DWT coefficients.

And finally, Mean and standard deviation based Statistical Color features are extracted from the segmented image. They are

Mean (μ) = $\frac{1}{N^2} \sum_{i,j=1}^{N} I(i,j)$

Standard Deviation (σ) = $\sqrt{\frac{\sum_{i,j=1}^{N} [I(i,j)-\mu]^2}{N^2}}$

Then all these features are combined using array concatenation and results the output as hybrid feature matrix.

3.5 Classifications of cancer

Neural networks have been effectively applied across a range of problem domains like finance, medicine, engineering, geology, physics and biology. From a statistical viewpoint, neural networks are interesting because of their potential use in prediction and classification problems. PNN is a method developed using emulation of birth neural scheme. The neurons are connected in the predefined architecture for effectively performing the classification operation. Depending on the hybrid features, the weights of the neurons are created. Then, the relationships between weights are identified using its characteristic hybrid features. The quantity of weights decides the levels of layers for the proposed network. Figure 5 represents the architecture of artificial neural networks. PNN basically consists of two stages for classification such as training and testing. The process of training will be performed based on the layer-based architecture. The input layer is used to perform the mapping operation on the input dataset; the hybrid features of this dataset are categorized into weight distributions.

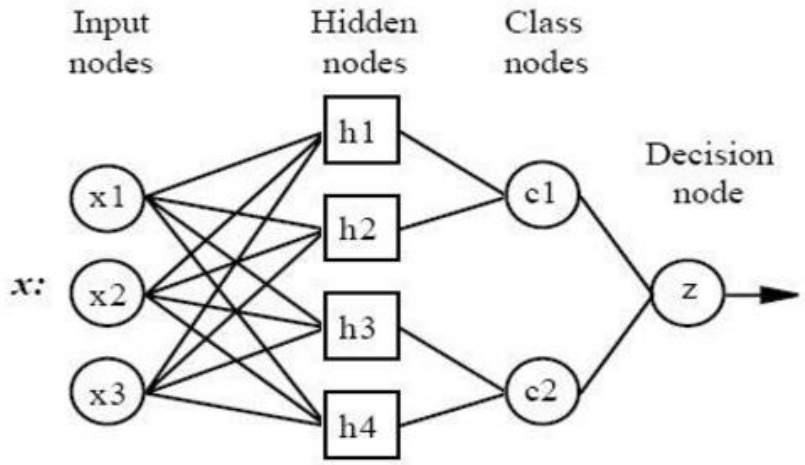

Fig 5: Layered architecture of PNN model.

The PNN architecture has four hidden layers with weights. The first convolutional2D hidden layer of the net takes in 224 * 224*3 pixels skin lesion images and applies 96 11×11 filters at stride 4 pixels, followed by class node activation layer and decision normalization layer.

Then the classification operation was implemented in the two levels of class nodes hidden layer. The two levels of hidden layer hold individually normality and abnormalities of the skin cancer characteristic information. Based on the segmentation criteria, it is categorized as normal and abnormal classification stage. These two levels are mapped as labels in output layer. Again, the hidden layer also contains the abnormal cancer types separately; it is also holding the benign and malignant cancer weights in the second stage of hidden layer. Similarly, these benign and malignant weights are also mapped as label into output layer. When the test image is applied, its hybrid features are applied for testing purpose in the classification stage. Based on the maximum feature matching criteria utilizing Euclidean distance manner it will function. If the feature match occurred with hidden layer 1labels, then it is classified as normal skin image. If the feature match occurred with hidden layer C1 labels with maximum weight distribution, then it is classified as benign effected cancer image. If the feature match occurred with hidden layer C2 labels with minimum weight distribution, then it is classified as malignant affected cancer image.

4. Results

4.1 dataset

The experiments are done using the MATLAB Programming language, and classification is done using the MATLAB R2018a tool. ISIC is one of the biggest available collections of quality controlled dermoscopic images. For the implementation of the proposed method, spatial domain, and frequency domain of 30 dermoscopic skin lesion images (15-benign and 15-Malignant) have been obtained respectively by applying rotations at different angles. Train images of each label have been used to train the PNN architecture with fifty Epochs, whereas rest twenty percent is used for testing. The features extracted by GLCM; DWT future network are used to train PNN classifier to classify the images into its respective classes. The efficiency of the model can be computed using various performance metrics.

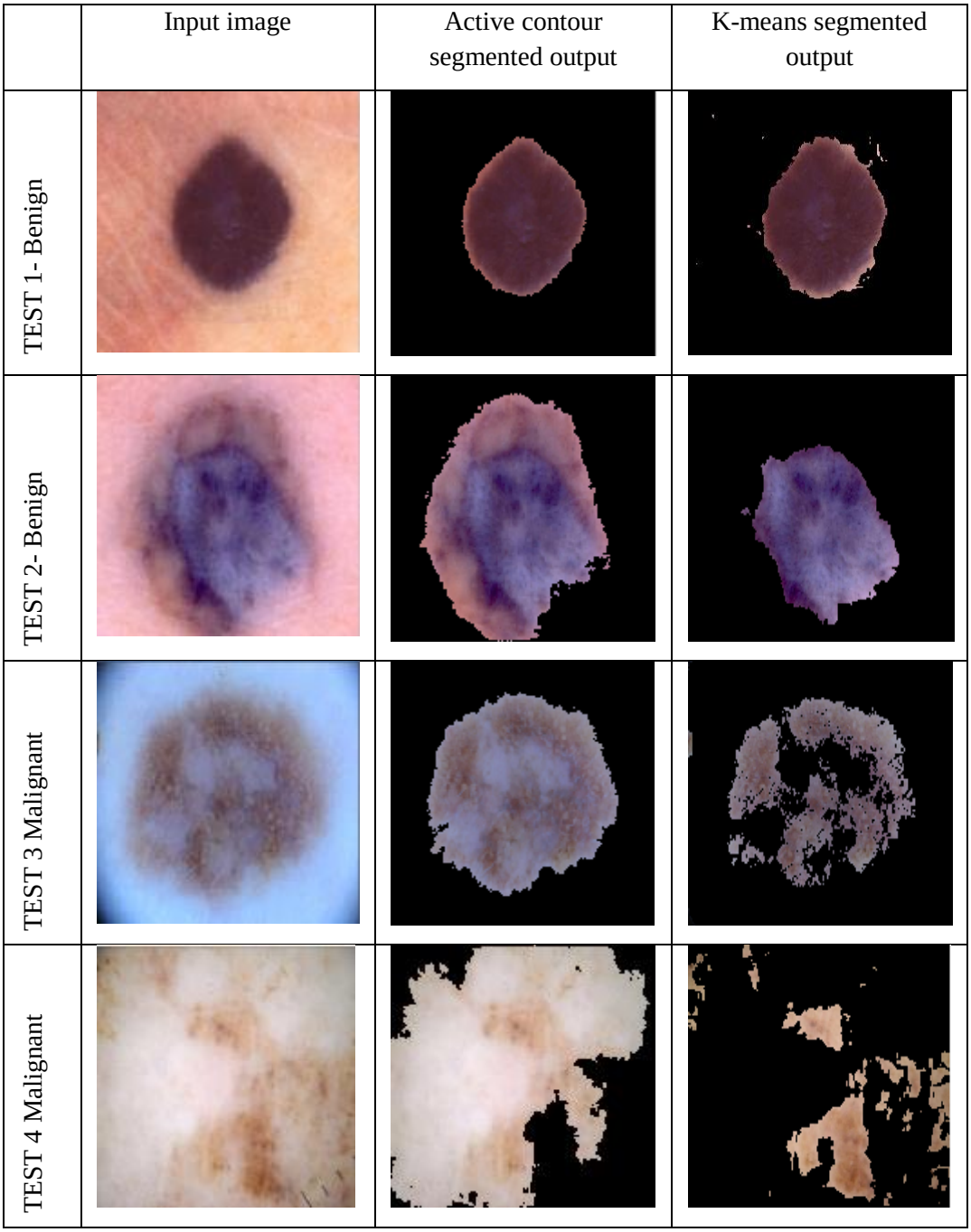

Fig 6: Segmented output images of various methods.

From fig 6, it is observed that the proposed method can be effectively detecting the regions of skin cancers, it indicates the segmentation done very effectively compared to the Active contour approach. Here, TEST-1 and TEST 2 images are considered as the benign and TEST-3 and TEST-4 images are considered malignant type images respectively. For the malignant images, the segmentation accuracy is more.

4.2 Performance metrics

For evaluating the performance measure the proposed method is implemented with the two types of segmentation methods, they are Active contour (AC) and k-means clustering respectively. For performing this comparisons Accuracy, Sensitivity, F measure, Precision, MCC, Dice, Jaccard and Specificity parameters are calculated respectively.

Table 1: Performance comparison.

Metric	method	Test 1	Test 2	Test 3	Test 4
Accuracy	**PNN-AC**	0.9157	0.78099	0.85796	0.47765
	PNN-kmeans	0.99985	0.99715	0.99999	0.99999
Sensitivity	**PNN-AC**	0.70588	0.90024	0.9166	0.83857
	PNN-kmeans	0.99931	0.99198	1	1
F measure	**PNN-AC**	0.82207	0.68494	0.79395	0.44602
	PNN-kmeans	0.99965	0.99381	0.99998	0.99998
Precision	**PNN-AC**	0.98404	0.55275	0.70023	0.30381
	PNN-kmeans	1	0.99852	0.99997	0.99997
MCC	**PNN-AC**	0.7869	0.56857	0.70305	0.1835
	PNN-kmeans	0.99956	0.99198	0.99998	0.99998
Dice	**PNN-AC**	0.82207	0.68494	0.79395	0.44602
	PNN-kmeans	0.99965	0.99381	0.99998	0.99998
Jaccard	**PNN-AC**	0.69789	0.52085	0.65831	0.28702
	PNN-kmeans	0.99931	0.9877	0.99997	0.99977
Specificity	**PNN-AC**	0.99564	0.73812	0.83298	0.35685
	PNN-kmeans	1	0.99956	0.99999	0.99998

From the Table 1 and Fig 6, it is observed that the proposed K-means clustering method along with PNN gives the highest performance for all metrics compared to the Active counter method.

Table 2: Accuracy comparison.

Method	Test 1	Test 2	Test 3	Test 4
SVM-Linear kernel [14]	0.4	0.40	0.7	0.7
SVM-RBF kernel [14]	0.4	0.45	0.55	0.6
SVM-Polynomial kernel [14]	0.4	0.3667	0.50	0.5667
SVM-5 fold cross validation [14]	0.6	0.55	0.60	0.45

Proposed PNN-AC	0.9157	0.78099	0.85796	0.47765
Proposed PNN-K-means	0.99985	0.99715	0.99999	0.99999

From the Table 2, it is observed that the proposed method gives the highest accuracy for both Benign and malignant diseases compared to the various kernels of SVM [14] such as SVM-Linear kernel, RBF kernel; Polynomial kernel and 5-fold cross validation respectively.

5. Conclusion

This article presented a computational methodology for detection & classification of skin cancer from MRI images using PNN based deep learning-based approach. Here, Gaussian filters are utilized for preprocessing, which eliminates any unwanted noise elements or artifacts innovated while image acquisition. Then K-means clustering segmentation is employed for ROI extraction and detection of cancerous cells. Then GLCM, DWT based method was developed for extraction of statistical, colour and texture features from segmented image respectively. Finally, PNN was employed to classify the type of cancer such as either benign or malignant using trained network model. Thus, upon comparing with state of art works, we conclude that PNN is better than conventional SVM method. In future, this work can be extended by implementing a greater number of network layers into the PNN and can also be applied for other type of benign and malignant cancers.

References

[1] Nasiri, Sara, et al. "DePicT Malignant Deep-CLASS: a deep convolutional neural networks approach to classify skin lesion images." *BMC bioinformatics* 21.2 (2020): 1-13.

[2] Munir, Khushboo, et al. "Cancer diagnosis using deep learning: a bibliographic review." *Cancers* 11.9 (2019): 1235

[3] Kadampur, Mohammad Ali, and Sulaiman Al Riyaee. "Skin cancer detection: applying a deep learning-based model driven architecture in the cloud for classifying dermal cell images." Informatics in Medicine Unlocked 18 (2020): 100282.

[4] Akram, Tallha, et al. "A multilevel features selection framework for skin lesion classification." *Human-centric Computing and Information Sciences* 10 (2020): 1-26.

[5] Marka, Arthur, et al. "Automated detection of nonMalignant skin cancer using digital images: a systematic review." *BMC medical imaging* 19.1 (2019): 21.

[6] Gaonkar, Rohan, et al. "Lesion analysis towards Malignant detection using soft computing techniques." *Clinical Epidemiology and Global Health* (2019).

[7] Hekler, Achim, et al. "Superior skin cancer classification by the combination of human and artificial intelligence." *European Journal of Cancer* 120 (2019): 114-121.

[8] Rajasekhar, K. S., and T. Ranga Babu. "Skin Lesion Classification Using Convolution Neural Networks." Indian Journal of Public Health Research & Development 10.12 (2019): 118-123.

[9] Iyer, Vijayasri, et al. "Hybrid quantum computing based early detection of skin cancer." *Journal of Interdisciplinary Mathematics* 23.2 (2020): 347-355.

[10] Roslin, S. Emalda. "Classification of Malignant from Dermoscopic data using machine learning techniques." Multimedia Tools and Applications (2018): 1-16.

[11] Moqadam, Sepideh Mohammadi, et al. "Cancer detection based on electrical impedance spectroscopy: A clinical study." *Journal of Electrical Bioimpedance* 9.1 (2018): 17-23.

[12] Hosny, Khalid M., Mohamed A. Kassem, and Mohamed M. Foaud. "Skin cancer classification using deep learning and transfer learning." *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*. IEEE, 2018.

[13] Dascalu, A., and E. O. David. "Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope." *EBioMedicine* 43 (2019): 107-113.

[14] M. Vidya and M. V. Karki, "Skin Cancer Detection using Machine Learning Techniques," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-5, doi: 10.1109/CONECCT50063.2020.9198489.