IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CLASSIFICATION OF MELANOMA AND NEVUS USING SVM AND KNN ALGORITHMS

Dr. Nancy Jasmine Goldena and M.Subashini

Department of Computer Applications, Sarah Tucker College, Tirunelveli-7.

ABSTRACT

Melanoma is considered a fatal type of skin cancer. However, it is sometimes hard to distinguish it from Nevus due to their identical visual appearance and symptoms. The mortality rate because of this disease is higher than all other skin related consolidated malignancies. The number of cases is growing amongst young people but if it is diagnosed at its earlier stage then the survival rates become very high. The cost and time required for the doctors to diagnose all patients for Melanoma are very high. In this work, we propose an intelligent system to detect and distinguish Melanoma from Nevus by using state of the art image processing techniques. At first, Gaussian Filter is used for removing noise from the skin lesion of the acquired images followed by the use of fuzzy c-mean clustering to segment out the lesion. A distinctive hybrid super feature vector is formed by the extraction of textural and color features from the lesion. Support Vector Machine (SVM) is utilized for the classification of skin cancer into melanoma and nevus. Our aim is to test the effectiveness of the proposed segmentation technique, extract the most suitable features and compare the classification results with the kNN techniques present in the literature. The proposed methodology is tested on DERMIS dataset.

1. INTRODUCTION

- To develop a complete automated computer-aided system to detect melanoma cancer accurately.
- Design of an improved K-Mean approach for computationally efficient segmentation.
- Utilization of hybrid features incorporating both texture and color of the lesion.
- Scope
- With the advent of computer-aided diagnostic systems, researcher mainly emphasizes on the automatic detection and classification of skin disease.
- Medical images in the form of textural features, geometric features, color features and in a combination have been used to identify and classify skin disease diseases.
- However, it is still a challenging task to identify most discriminative features for identifying melanoma at its initial stage.
- Our research work aims to achieve high accuracy results in identifying and classifying skin disease,

2. LITERATURE REVIEW

1. In today's world Skin disease (Melanoma) has become a very common disease. Melanoma is the cancer cells which exhibit as the abnormal cells from skins that can be developed in any other parts of the body. Over exposure to UV rays is the main cause for Melanoma. Other way that causes Melanoma is the tanning beds. Skin diseases are classified into three different types as follows: Basal-Cell skin disease (BCC), squamous-cell skin disease (SCC) and melanoma. Melanoma causes enormous and irreparable damage. Symptoms incorporate a mole that can change in its volume, contour, and color. The effective way to prevent Melanoma is by decreased exposure to UV rays and tanning beds. The stipulated approach followed by dermatologist is rightful supervision of the Skin. As this approach is time and power consuming, a new feature based classification for the detection of skin in various features of images as described in this paper. This approach reduces the professionals work. The morphological operation is used to differentiate the cancerous cell from the image .The skin texture features are obtained from processed image and used for classification of images as malignant and non-malignant.

2. The majority of pigmented skin lesions can be diagnosed correctly on the basis of clinical criteria; however, there remain a surprisingly high number of small pigmented lesions in which the distinction between melanocytic and non-melanocytic and benign and malignant lesions, and thus between melanoma and non-melanoma, is difficult or impossible to make with the naked eye. Epiluminescence microscopy is a non-invasive technique that, by use of oil immersion, makes sub-surface structures of skin accessible for in vivo microscopic examination and thus provides additional criteria for the diagnosis of pigmented lesions. The technique of epiluminescence microscopy is reviewed, and the significant improvement in the clinical diagnosis of pigmented skin lesions and, in particular, melanoma by this technique is documented

3. To assess, by means of meta-analysis techniques for diagnostic tests, the accuracy of dermoscopic (also known as dermatoscopy and epiluminescence microscopy) diagnosis of melanoma performed by experienced observers vs. naked-eye clinical examination. MEDLINE, EMBASE, PASCAL-BIOMED, and BIUM databases were screened through May 31, 2000, without any language restrictions. Original studies were selected when the following criteria were met: spectrum of lesions well described, histologic findings as standard criterion, and calculated or calculable sensitivity and specificity. Eight of 672 retrieved references were retained. Three investigators extracted data. In case of disagreement, consensus was obtained. Summary receiver operating characteristic curve analysis was used to describe the central tendency of the studies, and to compare dermoscopy and clinical examination. Selected studies represented 328 melanomas, mostly less than 0.76 mm thick, and 1865 mostly melanocytic benign pigmented skin lesions. For dermoscopic diagnosis of melanoma, the sensitivity and specificity ranges were 0.75 to 0.96 and 0.79 to 0.98, respectively. Dermoscopy had significantly higher discriminating power than clinical examination, with respective estimated odds ratios of 76 (95% confidence interval, 25-223) and 16 (95% confidence interval, 9-31) (P =.008), and respective estimated positive likelihood ratios of 9 (95% confidence interval, 5.6-19.0) and 3.7 (95% confidence interval, 2.8-5.3). The roles of the number of lesions analyzed, the percentage of melanoma lesions, the instrument used, and dermoscopic criteria used in each study could not be proved. For experienced users, dermoscopy is more accurate than clinical examination for the diagnosis of melanoma in a pigmented skin lesion.

4. Background: There is a need for better standardization of the dermoscopic terminology in assessing pigmented skin lesions. Objective: The virtual Consensus Net Meeting on Dermoscopy was organized to investigate reproducibility and validity of the various features and diagnostic algorithms. Methods: Dermoscopic images of 108 lesions were evaluated via the Internet by 40 experienced dermoscopists using a 2-step diagnostic procedure. The first-step algorithm distinguished melanocytic versus nonmelanocytic lesions. The second step in the diagnostic procedure used 4 algorithms (pattern analysis, ABCD rule, Menzies method, and 7-point checklist) to distinguish melanoma versus benign melanocytic lesions. κ Values, log odds ratios, sensitivity, specificity, and positive likelihood ratios were estimated for all diagnostic algorithms and dermoscopic features. Results: Interobserver agreement was fair to good for all diagnostic methods, but it was poor for the majority of dermoscopic criteria. Intraobserver agreement was good to excellent for all algorithms and features considered. Pattern analysis allowed the best diagnostic performance (positive likelihood ratio: 5.1), whereas alternative algorithms revealed comparable sensitivity but less specificity. Interobserver agreement on management decisions made by dermoscopy was fairly good.

5. Malignant melanoma is one of the most common and the deadliest type of skin disease. In Australia, it represents 10% of all cancers and its incidence is four times higher than in Canada, the UK and the US. The worldwide steady increase in incidence of melanoma in recent years, its high mortality rate and the massive medical cost has made its early diagnosis a continuing priority of public health. Melanoma survival rate depends highly on its stage and thickness; advanced and thickened melanoma is still incurable, whereas early detection of thin melanoma and immediate surgical excision of the lesion shows effective prognosis. However, early diagnosis of melanoma is not trivial even for experienced dermatologists. Despite the use of advanced imaging technologies such as dermoscopy and existing clinical diagnostic algorithms such as the ABCD rule of dermoscopy, clinical diagnosis of melanoma is still challenging and its accuracy has been an issue of concern (estimated to be about 75–85%) especially with equivocal pigmented lesions

6. Background: Dermoscopy is one of the major imaging modalities used in the diagnosis of melanoma and other pigmented skin lesions. Due to the difficulty and subjectivity of human interpretation, computerized analysis of dermoscopy images has become an important research area. One of the most important steps in dermoscopy image analysis is the automated detection of lesion borders. Methods: In this article, we present a systematic overview of the recent border detection methods in the literature paying particular attention to computational issues and evaluation aspects. Conclusion: Common problems with the existing approaches include the acquisition, size, and diagnostic distribution of the test image set, the evaluation of the results, and the inadequate description of the employed methods. Border determination by dermatologists appears to depend upon higher-level knowledge, therefore it is likely that the incorporation of domain knowledge in automated methods will enable them to perform better, especially in sets of images with a variety of diagnoses.

7. The aims of this study were to provide a quantitative assessment of the tumour area extracted by dermatologists and to evaluate computer-based methods from dermoscopy images for refining a computer-based melanoma diagnostic system. Dermoscopic images of 188 Clark naevi, 56 Reed naevi and 75 melanomas were examined. Five dermatologists manually drew the border of each lesion with a tablet computer. The inter-observer variability was evaluated and the standard tumour area (STA) for each dermoscopy image was defined. Manual extractions by 10 non-medical individuals and by two computer-based methods were evaluated with STA-based assessment criteria: precision and recall. Our new computer-based method introduced the region-growing approach in order to yield results close to those obtained by dermatologists. The effectiveness of our extraction method with regard to diagnostic accuracy was evaluated. Two linear classifiers were built using the results of conventional and new computer-based tumour area extraction methods. The final diagnostic accuracy was evaluated by drawing the receiver operating curve (ROC) of each classifier, and the area under each ROC was evaluated. The standard deviations of the tumour area extracted by five dermatologists and 10 non-medical individuals were 8.9% and 10.7%, respectively. After assessment of the extraction results by dermatologists, the STA was defined as the area that was selected by more than two dermatologists. Dermatologists selected the melanoma area with statistically smaller divergence than that of Clark naevus or Reed naevus (P=0.05). By contrast, non-medical individuals did not show this difference. Our new computer-based extraction algorithm showed superior performance (precision, 94.1%; recall, 95.3%) to the conventional thresholding method (precision, 99.5%; recall, 87.6%)..

8. simple naked-eye examination. However, even with the use of dermoscopy and dermoscopic algorithms, clinical diagnosis is still challenging and its accuracy is considered to be limit ed, especially with difficult cases (3). Computer aided diagnos is of melanoma provides quantitative and objective evaluation of the skin lesion, as opposed to visual assessment, which is subjective in nature. It allows for reproducible diagnosis by diminishing the inter-observer and intra-observer variab ilities that could be found in dermatologists' examinations. It als o automates the analysis, and thereby reduces the amount of repetitive and

tedious tasks to be done by physicians. A system for the computer-aided diagnosis of melanoma is generally comprised of four major components: skin image acquisition, lesion segmentation, feature extraction, an d lesion classification. Automatic segmentation of lesions i n color skin images, which is the main focus of this paper, is one of the most important steps towards the automated analysis and evaluation of dermoscopy images in the computer aided diagnosis of melanoma..

9. It is highly desirable to identify malignant melanoma, a common cancer, at an early stage. One important clinical feature of this cancer is asymmetrical skin lesions. In this paper, we propose an adaptive fuzzy approach that uses symmetric distance (SD) to measure lesions with fuzzy borders. The use of a number of SD variations and the adoption of a backpropagation neural network enhances the discriminative power of the approach. Digitized images from the Lesion Clinic in Vancouver, Canada, demonstrate the accurate classification of asymmetric lesions at around 80%.
10. In this paper a methodological approach to the classification of pigmented skin lesions in dermoscopy images is presented. First, automatic border detection is performed to separate the lesion from the background skin. Shape features are then extracted from this border. For the extraction of color and texture related features, the image is divided into various clinically significant regions using the Euclidean distance transform. This feature data is fed into an optimization framework, which ranks the features using various feature selection algorithms and determines the optimal feature subset size according to the area under the ROC curve measure obtained from support vector machine classification.

3. METHODOLOGY

- **PREPROCESS**
  - **GRAY**
- **SEGMENTATION**
  - **FUZZY C--MEANS**
  - **BINARY IMAGE**
- **BLOB EXXTRACTION**
  - **BLOB DETECTION**
- **FEATURE EXTRACTION**
- **CLASSIFICATION**
  - **SVM**
  - kNN

PRE-PROCESSING

Medical images are often susceptible to noise mainly due to bad illumination, hair and air bubbles. This inclusion of noise in images results in the formation of artifacts. Due to such artifacts, the segmentation results may get affected causing inaccurate detection results. Therefore, noise removal is a significant step before applying any segmentation or feature extraction technique for an accurate diagnosis. To smoothen the image, Gaussian filter is highly recommended as it removes the speckle noise added during the process of acquisition.

SEGMENTATION

K-mean clustering is a common machine learning technique that is extensively used in many applications such as data mining, image processing, and pattern recognition. K-mean is considered as one of the basic methodologies for grouping and clustering of data-points into K number of clusters [18]. It works by splitting the image into non-overlapping groups
of pixels based on their intensity levels. The process initiates by selecting the centroids from the data-points either randomly or through a certain criterion. The pixels or data-points are clustered based on their minimum distance from the selected centroids. After each iteration,
the mean values of the formed clusters are found and are set as the centroids for the next iteration. The iterative process repeats itself until there is no variation in the successive cluster centroids. The proposed K-Mean initializes its centroids by centroid selection technique as given in equation 2 below

$$\square\square = (1: \square) * \square\ \square + 1$$

Where K is the number of clusters, m is the maximum value of a pixel in the image, and k goes from 1 to K.
The centroid selection technique works by ensuring significant difference among the values of initialized centroids making it more efficient and robust by converging to the final position in a lesser number of iterations. In our proposed system, input images consist of affected lesion surrounded by the background skin. The value of K is taken as 2, such that the foreground lesion is extracted from the background skin.

K-MEANS SEGMENTATION

K-mean clustering is a common machine learning technique that is extensively used in many applications such as data mining, image processing, and pattern recognition. K-mean is considered as one of the basic methodologies for grouping and clustering of data-points into K number of clusters. It works by splitting the image into non-overlapping groups of pixels based on their intensity levels. The process initiates by selecting the centroids from the data-points either randomly or through a certain criterion. The pixels or data-points are clustered based on their minimum distance from the selected centroids. After each iteration, the mean values of the formed clusters are found and are set as the centroids for the next iteration.

The iterative process repeats itself until there is no variation in the successive cluster centroids. The proposed K-Mean initializes its centroids by centroid selection technique as given in equation 2 below. Where K is the number of clusters, m is the maximum value of a pixel in the image, and k goes from 1 to K.
The centroid selection technique works by ensuring significant difference among the values of initialized centroids making it more efficient and robust by converging to the final position in a lesser number of iterations. In our proposed system, input images consist of affected lesion surrounded by the background skin. The value of K is taken as 2, such that the foreground lesion is extracted from the background skin images.

FUZZY C-MEANS CLUSTERING ALGORITHM

This algorithm is used to cluster/divide the object based on the feature of the leaf in to k number of groups. This is done by using the Euclidean distance metric.

The algorithm of k means

- Initialization: User should select the value of k. k means the number of clusters/groups, i.e. the image is divided in to k number of clusters.
- Every pixel is assigned to its nearest centroid (k).
- The position of centroid is changed by means of data values assigned to the group. The centroid moves to the centre of its assigned points.

Out of these three clusters classification is done for only one cluster which has affected area. k-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$\mathbf{J(V)} = \sum_{i=1}^{c} \sum_{J-1}^{C_i} (|\boldsymbol{xi} - \boldsymbol{vj}|)^2$$

where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centers.

ALGORITHMIC STEPS FOR FUZZY C-MEANS CLUSTERING

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..

4) Recalculate the new cluster center using:

$$\mathbf{V_i} = \mathbf{(1/C_i)} \sum_{J=1}^{CI} \boldsymbol{x_i}$$

where, 'c_i' represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centres.

6) If no data point was reassigned then stop, otherwise repeat from step 3).

ADVANTAGES

1) Fast, robust and easier to understand.

2) Relatively efficient: O(tknd), where n is # objects, k is # clusters, d is # dimension of each object, and t is # iterations. Normally, k, t, d << n.

3) Gives best result when data set are distinct or well separated from each other.

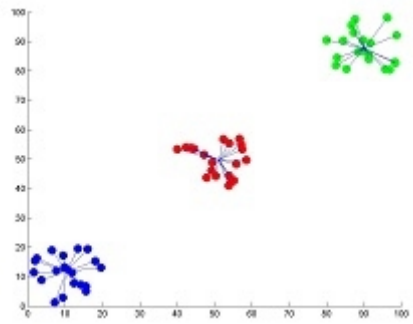


Fig I: Showing the result of k-means for *'N'* = 60 and *'c'* = 3

FEATURE EXTRACTION

Once the lesion is segmented out of the background skin, it is then classified as malignant or benign. For better classification results, it is required to use the best feature descriptors for machine learning modeling.

1. Total mass (number of pixels in a binarized character)

2. Centroid - Center of mass

3. Elliptical parameters

   i. Eccentricity (ratio of major to minor axis)

   ii. Orientation (angle of major axis)

4. Skewness

5. Kurtosis

6. Higher order moments

Skewness

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right.

Skewness= sum ((GLs – mean GL). ^ 3.* pixel Counts) / ((number Of Pixels - 1) * sd^3)

$$S_{k\,(X)}=\left(\frac{1}{mxn}\right)\sum\frac{(f(x,y)-M^3}{SD^3}$$

Kurtosis

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak.

Kurtosis = sum((GLs – mean GL) .^ 4 .* pixel Counts) / ((number Of Pixels - 1) * sd^4)

$$SVMK_{urt\ (X)}=\left(\frac{1}{mxn}\right)\frac{\sum(f(x,y)-M^4}{SD^4}$$

Support Vector Machine – SVM

SVM stands for **Support Vector Machine**. It is a **machine learning** approach used for classification and regression analysis. It depends on **supervised learning models** and trained by learning algorithms. They analyze the large amount of data to identify patterns from them. An SVM generates parallel partitions by generating two parallel lines. For each category of data in a high-dimensional space and uses almost all attributes. It separates the space in a single pass to generate flat and linear partitions. Divide the 2 categories by a clear gap that should be as wide as possible. Do this partitioning by a plane called hyper plane.

An SVM creates hyper planes that have the largest margin in a high-dimensional space to separate given data into classes. The margin between the 2 classes represents the longest distance between closest data points of those classes.

SVM

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a no probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. The main advantage of the SVM network used as a classifier is its very good generalization ability and extremely powerful learning procedure, leading to the global minimum of the defined error function.

Inputs

The complete MRI database consists of 30 patients: 10 Normal, 10 MCI and 10 AD. In this project SVM using non linear algorithm for classification.

Applications

SVMs can be used to solve various real world problems:

- Classification of images can also be performed using SVMs.
- SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly.
- Hand-written characters can be recognized using SVM.

Advantages

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels

kNearest Neighbour

We can implement a KNN model by following the below steps:

1. Load the data
2. Initialize the value of k
3. For getting the predicted class, iterate from 1 to total number of training data points

    - Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.
    - Sort the calculated distances in ascending order based on distance values
    - Get top k rows from the sorted array
    - Get the most frequent class of these rows
    - Return the predicted class

Dataset

The datasets used for the Experiments and to calculate the performance evaluation are shown below.

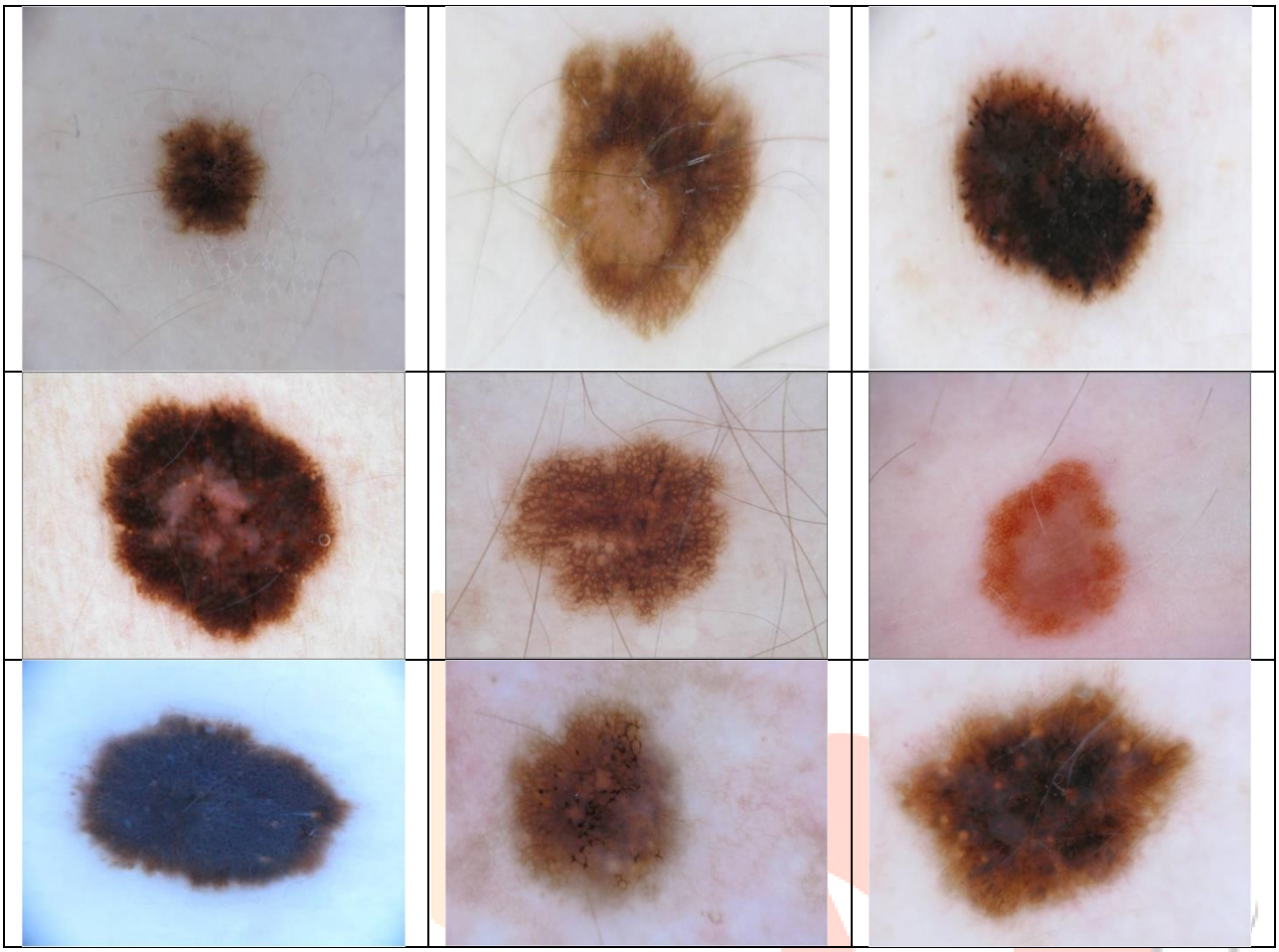


Fig 3.2 Datasets used in experiments

Skin cancer images were collected for the experiment; Segmentation take place after the input get by hair remove method of image processing.

Performance Analysis

Here several performance metrics are used to check the segmentation.

Segmentation results of an image and ground truth of an image are compared to evaluate the performance.

1. Accuracy
2. Precision
3. Recall

Accuracy

An alternative metric to evaluate a semantic segmentation is to simply report the percent of pixels in the image which were correctly classified. The pixel accuracy is commonly reported for each class separately as well as globally

across all classes. When considering the per-class pixel accuracy we're essentially evaluating a binary mask; a true positive represents a pixel that is correctly predicted to belong to the given class (according to the target mask) whereas a true negative represents a pixel that is correctly identified as not belonging to the given class.

Accuracy=TP+TN/TP+TN+FP+FN

TP True Positive **TN** True Negative **FP** False Positive **FN** False Negative

Precision

Precision effectively describes the purity of positive detections relative to the ground truth.

Precision=TP/TP+FP

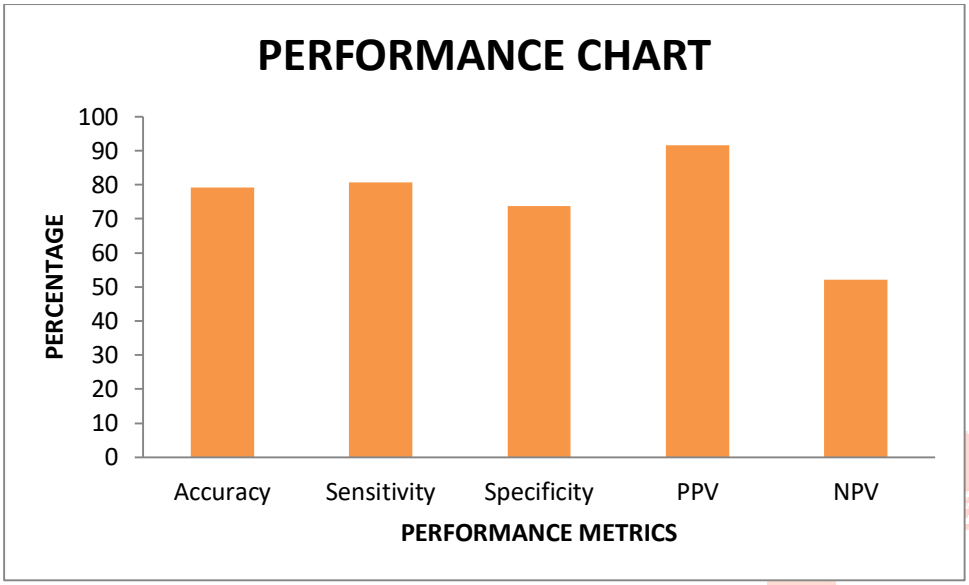
Recall

Recall effectively describes the completeness of positive predictions relative to the ground truth.
Recall=TP/TP+FN

4.2.5 Performance Analysis

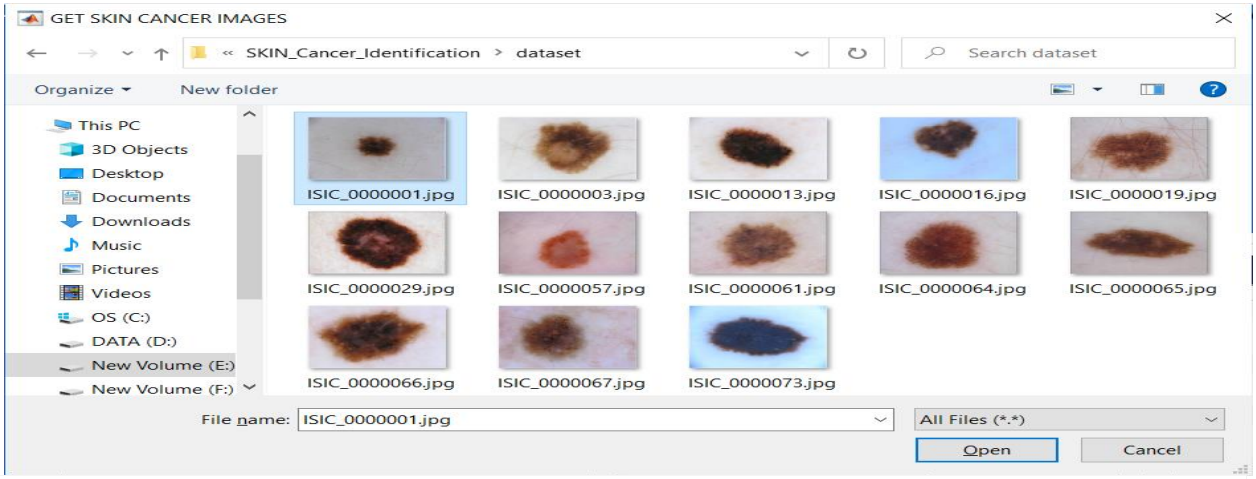
The performance Analysis values of global threshold are shown in Table. 4.2
Table 4.2 Performance Evaluation Values of global threshold

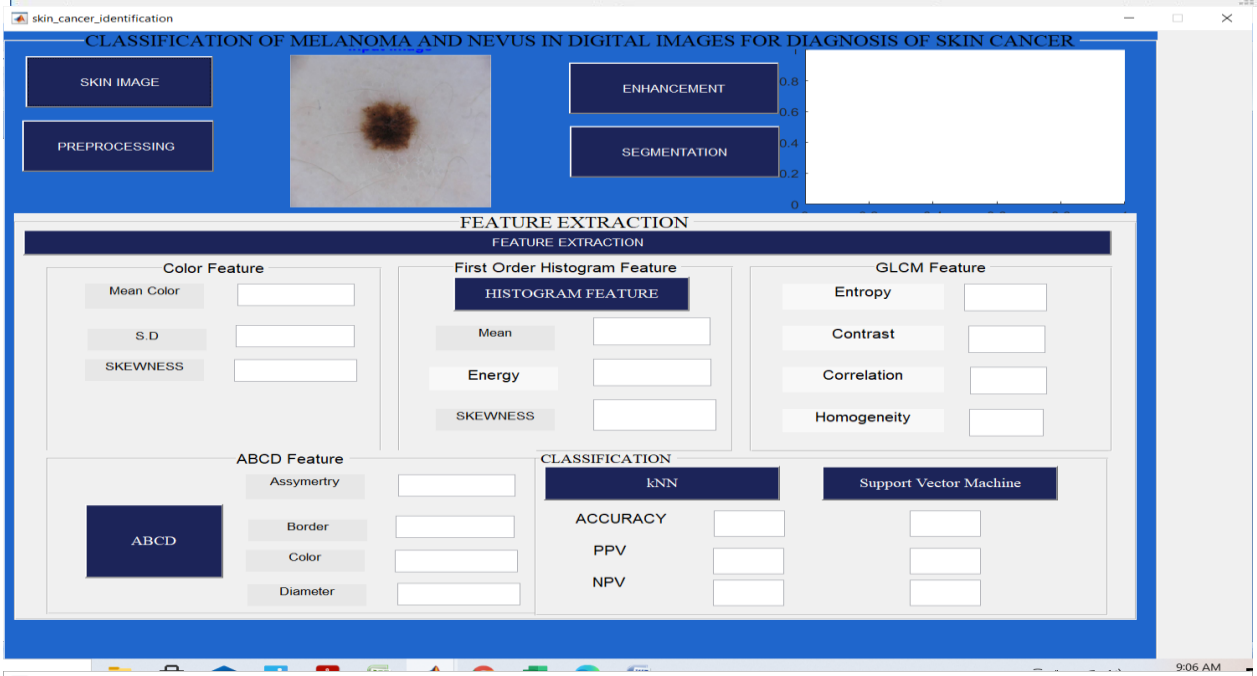
Accuracy	Sensitivity	Specificity	PPV	NPV
79.13	80.67	73.73	91.51	52.09

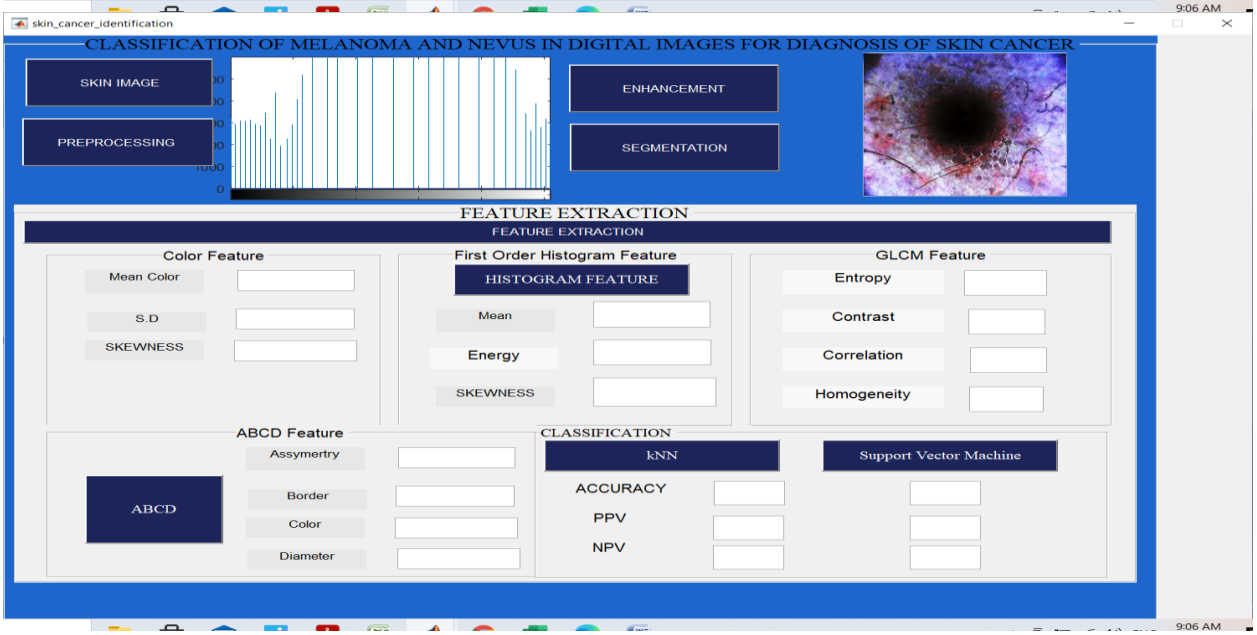


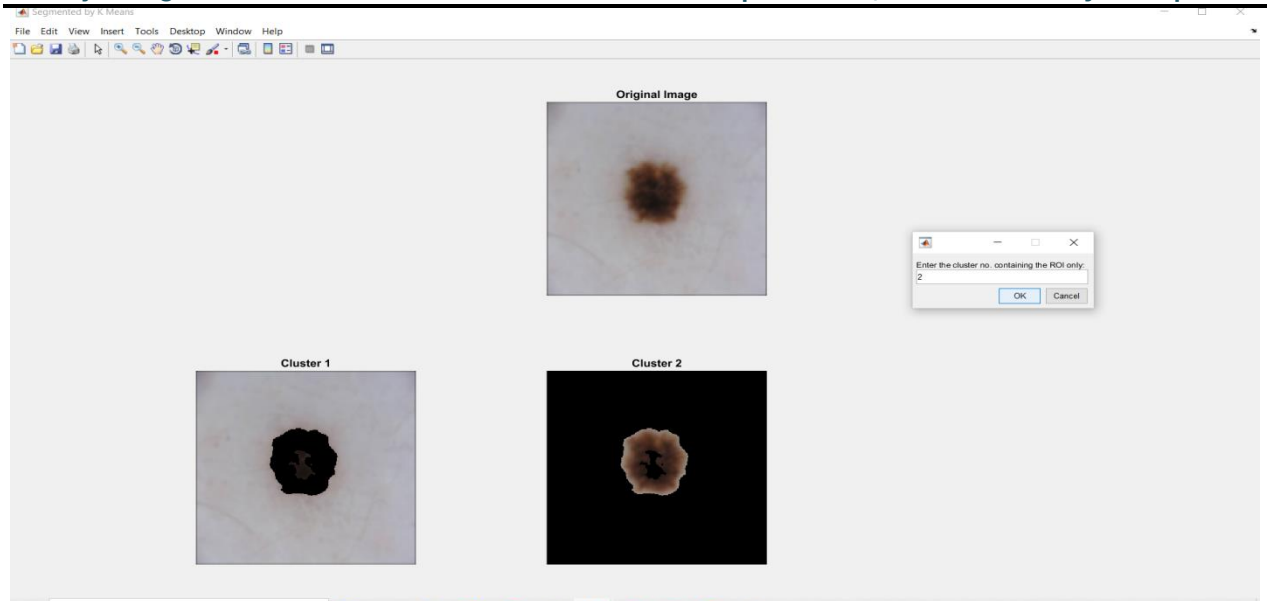
The above chart depicts the performance of the Multilayer Perceptron algorithm. MLP obtains 79.13 percentage of accuracy value and 91.51 of positive prediction value.

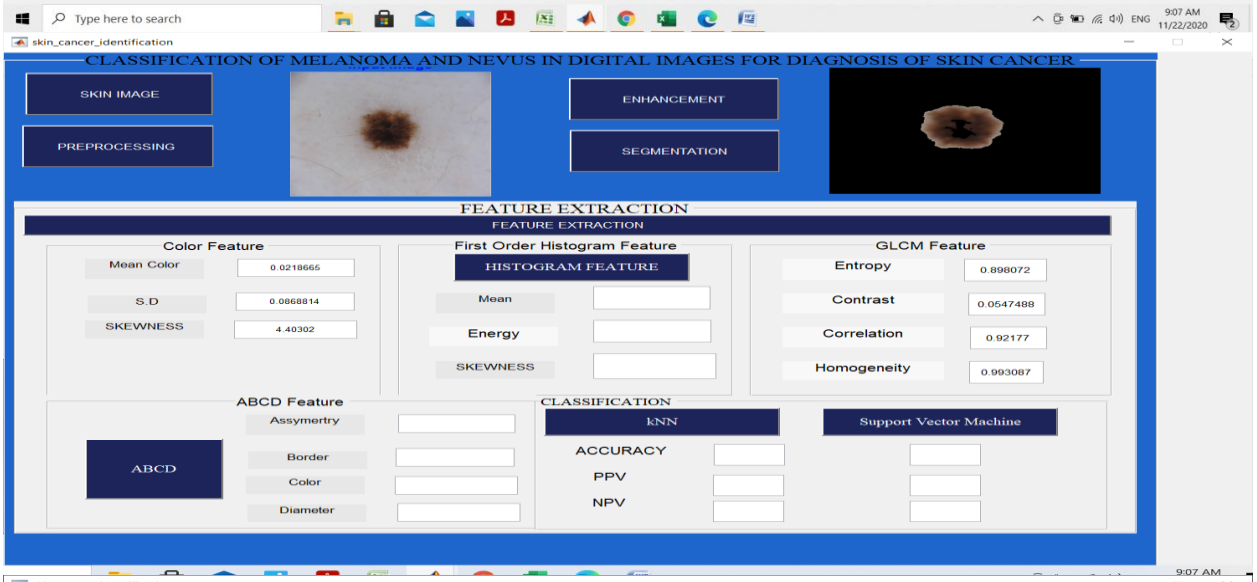
4. RESULTS

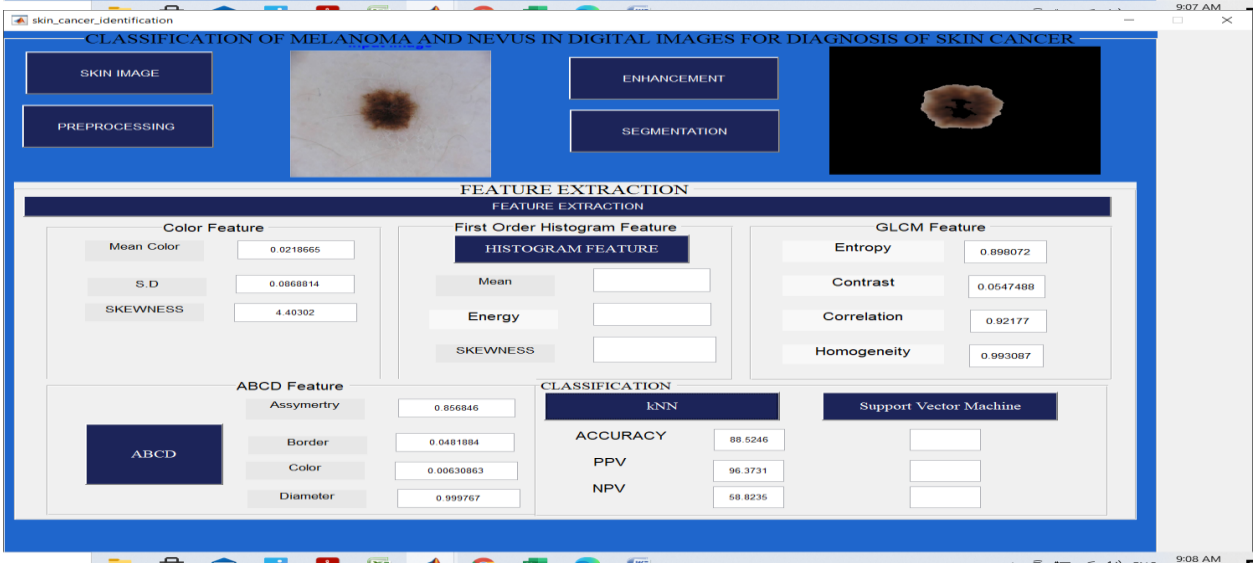

GET SKIN CANCER IMAGES
SKIN_Cancer_Identification > dataset
Search dataset
Organize
New folder
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
OS (C:)
DATA (D:)
New Volume (E:)
New Volume (F:)
ISIC_0000001.jpg
ISIC_0000003.jpg
ISIC_0000013.jpg
ISIC_0000016.jpg
ISIC_0000019.jpg
ISIC_0000029.jpg
ISIC_0000057.jpg
ISIC_0000061.jpg
ISIC_0000064.jpg
ISIC_0000065.jpg
ISIC_0000066.jpg
ISIC_0000067.jpg
ISIC_0000073.jpg
File name: ISIC_0000001.jpg
All Files (*.*)
Open
Cancel

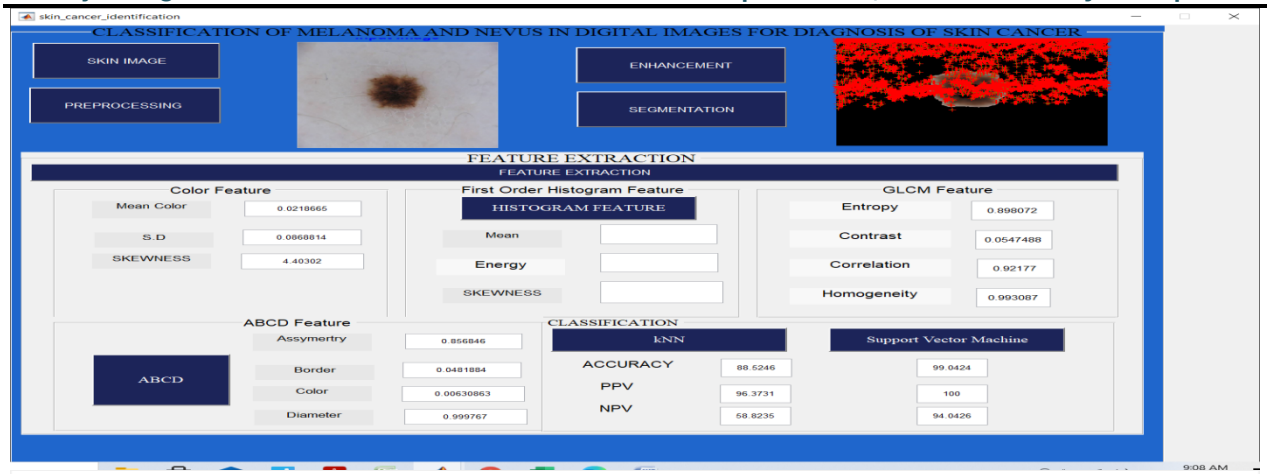

skin_cancer_identification
CLASSIFICATION OF MELANOMA AND NEVUS IN DIGITAL IMAGES FOR DIAGNOSIS OF SKIN CANCER
SKIN IMAGE
PREPROCESSING
ENHANCEMENT
SEGMENTATION
FEATURE EXTRACTION
FEATURE EXTRACTION
Color Feature
Mean Color
S.D
SKEWNESS
First Order Histogram Feature
HISTOGRAM FEATURE
Mean
Energy
SKEWNESS
GLCM Feature
Entropy
Contrast
Correlation
Homogeneity
ABCD Feature
Assymetry
ABCD
Border
Color
Diameter
CLASSIFICATION
kNN
Support Vector Machine
ACCURACY
PPV
NPV
9:06 AM


skin_cancer_identification
CLASSIFICATION OF MELANOMA AND NEVUS IN DIGITAL IMAGES FOR DIAGNOSIS OF SKIN CANCER
SKIN IMAGE
PREPROCESSING
ENHANCEMENT
SEGMENTATION
FEATURE EXTRACTION
FEATURE EXTRACTION
Color Feature
Mean Color
S.D
SKEWNESS
First Order Histogram Feature
HISTOGRAM FEATURE
Mean
Energy
SKEWNESS
GLCM Feature
Entropy
Contrast
Correlation
Homogeneity
ABCD Feature
Assymetry
ABCD
Border
Color
Diameter
CLASSIFICATION
kNN
Support Vector Machine
ACCURACY
PPV
NPV
9:06 AM


Segmented by K Means
File Edit View Insert Tools Desktop Window Help
Original Image
Enter the cluster no. containing the ROI only:
2
OK
Cancel
Cluster 1
Cluster 2


Type here to search
ENG 9:07 AM 11/22/2020
skin_cancer_identification
CLASSIFICATION OF MELANOMA AND NEVUS IN DIGITAL IMAGES FOR DIAGNOSIS OF SKIN CANCER
SKIN IMAGE
PREPROCESSING
ENHANCEMENT
SEGMENTATION
FEATURE EXTRACTION
FEATURE EXTRACTION
Color Feature
Mean Color 0.0218665
S.D 0.0868814
SKEWNESS 4.40302
First Order Histogram Feature
HISTOGRAM FEATURE
Mean
Energy
SKEWNESS
GLCM Feature
Entropy 0.898072
Contrast 0.0547488
Correlation 0.92177
Homogeneity 0.993087
ABCD Feature
Assymertry
Border
Color
Diameter
ABCD
CLASSIFICATION
kNN
Support Vector Machine
ACCURACY
PPV
NPV
9:07 AM


skin_cancer_identification
CLASSIFICATION OF MELANOMA AND NEVUS IN DIGITAL IMAGES FOR DIAGNOSIS OF SKIN CANCER
SKIN IMAGE
PREPROCESSING
ENHANCEMENT
SEGMENTATION
FEATURE EXTRACTION
FEATURE EXTRACTION
Color Feature
Mean Color 0.0218665
S.D 0.0868814
SKEWNESS 4.40302
First Order Histogram Feature
HISTOGRAM FEATURE
Mean
Energy
SKEWNESS
GLCM Feature
Entropy 0.898072
Contrast 0.0547488
Correlation 0.92177
Homogeneity 0.993087
ABCD Feature
Assymertry 0.856846
Border 0.0481884
Color 0.00630863
Diameter 0.999767
ABCD
CLASSIFICATION
kNN
Support Vector Machine
ACCURACY 88.5246
PPV 96.3731
NPV 58.8235
9:08 AM


skin_cancer_identification
CLASSIFICATION OF MELANOMA AND NEVUS IN DIGITAL IMAGES FOR DIAGNOSIS OF SKIN CANCER
SKIN IMAGE
PREPROCESSING
ENHANCEMENT
SEGMENTATION
FEATURE EXTRACTION
FEATURE EXTRACTION
Color Feature
Mean Color
0.0218665
S.D
0.0868814
SKEWNESS
4.40302
First Order Histogram Feature
HISTOGRAM FEATURE
Mean
Energy
SKEWNESS
GLCM Feature
Entropy
0.898072
Contrast
0.0547488
Correlation
0.92177
Homogeneity
0.993087
ABCD Feature
ABCD
Assymetry
0.856846
Border
0.0481884
Color
0.00630863
Diameter
0.999767
CLASSIFICATION
kNN
Support Vector Machine
ACCURACY
88.5246
99.0424
PPV
96.3731
100
NPV
58.8235
94.0426

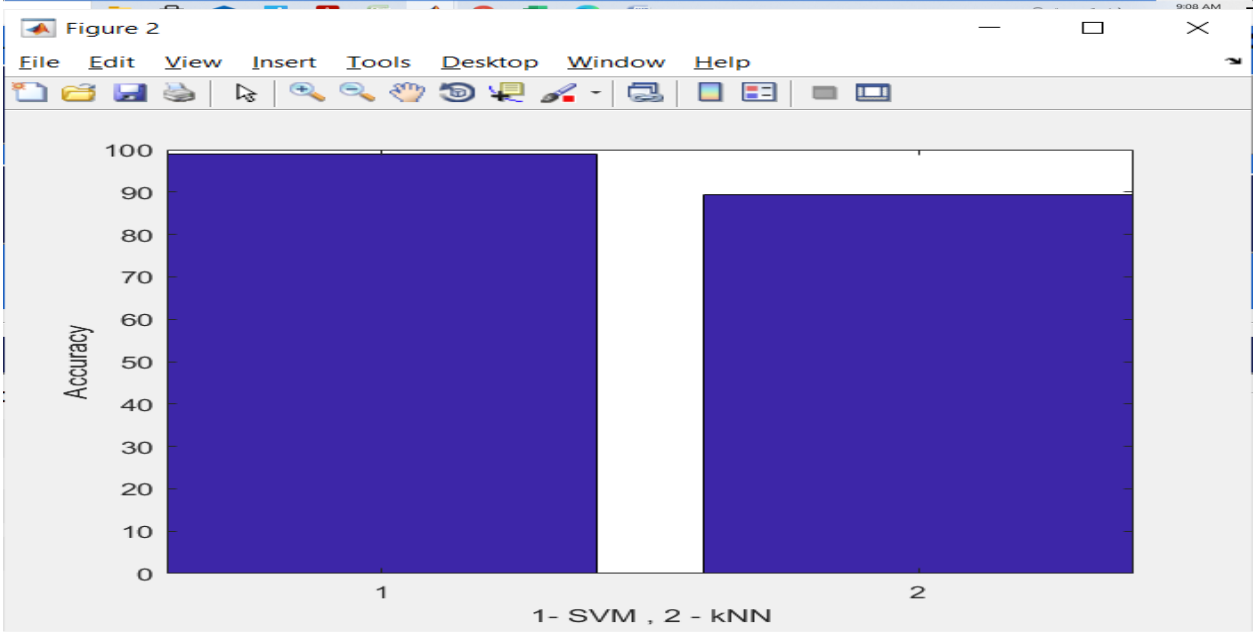


Figure 2
File Edit View Insert Tools Desktop Window Help
Accuracy
1- SVM , 2 - kNN

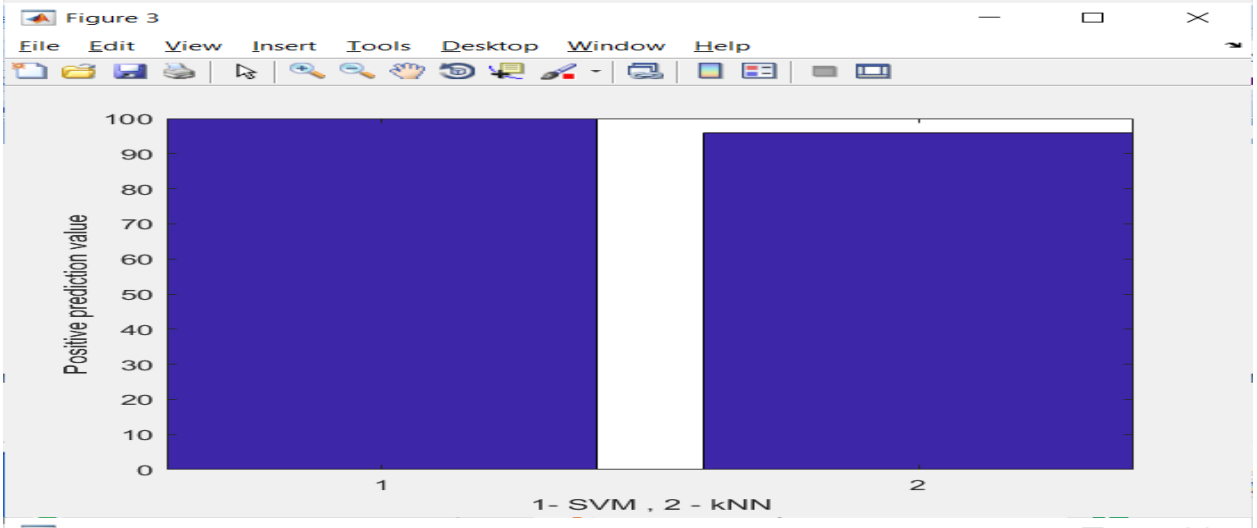


Figure 3
File Edit View Insert Tools Desktop Window Help
Positive prediction value
1- SVM , 2 - kNN

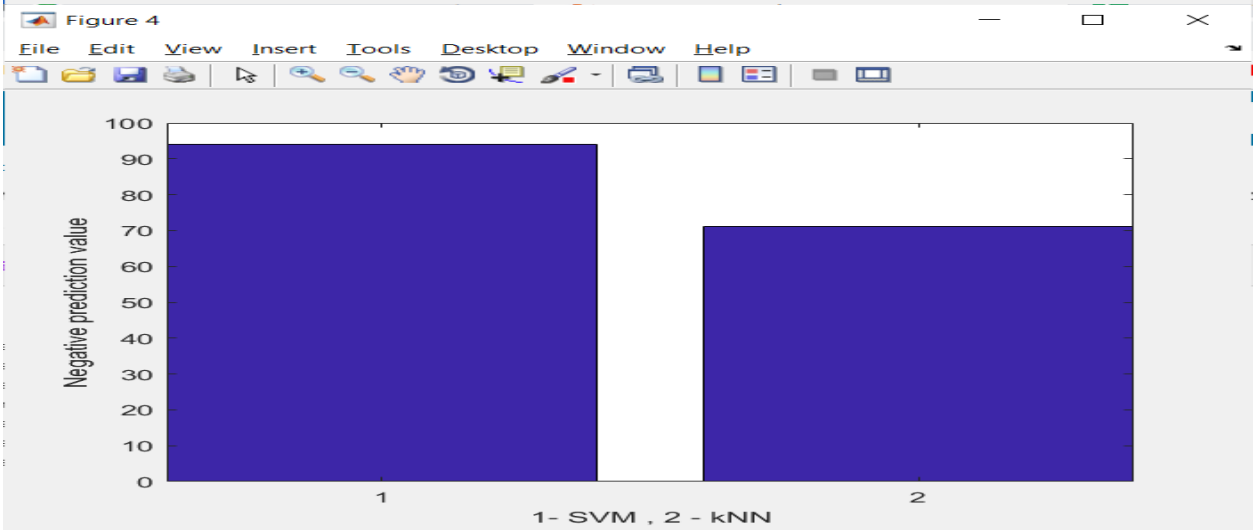


Figure 4
File Edit View Insert Tools Desktop Window Help
Negative prediction value
1- SVM , 2 - kNN

5. CONCLUSION

The aim of this project is to determine the accurate prediction of skin cancer and also to classify the skin cancer as malignant or non-malignant melanoma. To do so, some pre-processing steps were carried out which followed Hair removal, shadow removal, glare removal and also segmentation. SVM and Deep Neural networks will be used to classify. Classifier will be trained to learn the features and finally used to classify. The novelty of the present methodology is that it should do the detection in very quick time hence aiding the technicians to perfect their diagnostic skills.

6. REFERENCE

[1] Bono, S. Tomatis, and C. Bartoli, The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry,border, color, and dimension, "Cancer", vol. 85, no. 1, pp. 72–77, January 1999

[2] Pehamberger H, Binder M, Steiner A, Wolff K. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. J Invest Dermatol,100:356S–62S, 1993

[3] Bafounta ML, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Arch Dermatol,137:13,43–50. 2001

[4] G.Argenziano, H. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, and M. Binder, Dermoscopy of pigmented skin lesions: Results of consensus meeting via the Internet Journal of the American Academy of Dermatology, vol. 48, pp. 679–693, 2003

[5] R. Garnavi, Computer-aided diagnosis of melanoma, Ph.D. dissertation, University of Melbourne, Australia, 2011

[6] M.E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, Lesion border detection in dermoscopy images Computerised Medical Imaging and Graphics, vol. 33, no. 2, pp. 148–153 , 2009

[7] H. Iyatomi, H. Oka, M. Saito, A. Miyake, M. Kimoto, J. Yamagami, S. Kobayashi, A. Tanikawa, M. Hagiwara, K. Ogawa, G. Argenziano, H. P. Soyer, and M. Tanaka, Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system Melanoma Research, vol. 16, no. 2, pp. 183–190, 2006

[8] R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C. Dolianitis, and G. Varigos, Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels International Journal of Medicine and Medical Sciences, vol. 1, no. 2, pp. 126–134, 2010

[9] V. Ng, B. Fung, and T. Lee, Determining the asymmetry of skin lesion with fuzzy borders Computers in Biology and Med., vol. 35, pp. 103– 120, 2005

[10] M. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, and R. Moss, A methodological approach to the classification of dermoscopy images, Computerized Medical Imaging and Graphics.