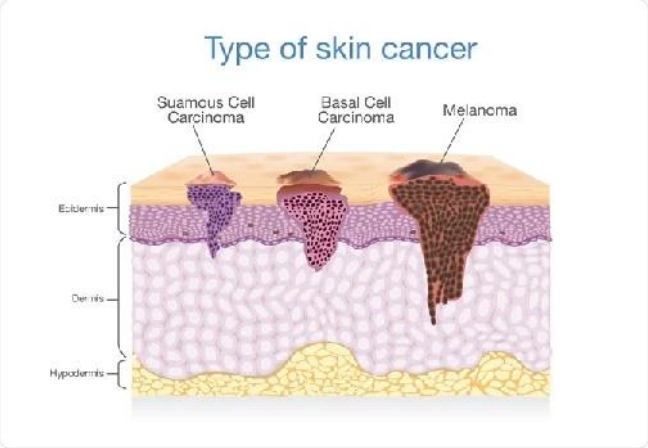
Skin Cancer Detection Using Machine Learning And Image Processing

Hariharan Kumar, Rohit R Nagarahalli, Srinivasa MS, Sangey Srivastava, MaqdumShariff
Dept of Computer Science
Atria Institute of Technology

Abstract- *Skin cancer is a major health issue in the present day especially melanoma skin cancer. In general, most of the skin cancers are curable if they are detected and treated in the early stages. With the speedy growth of skin cancer, there is a need for an automated computerized diagnosis mechanism of skin cancer in the early stage which is required. Various skin cancer images have similar visual characteristics and representations. It is very important to extract the features from the skin cancer images. The automated computerized diagnosis mechanism helps to improve the accuracy of skin diseases which helps the dermatologists to provide the diagnosis and better treatment for the patients. This paper depicts the comparative study on various traditional image processing and current technologies of different image processing techniques for skin cancer image classification, pre-processing techniques, Feature extraction/abstraction.*

I. INTRODUCTION

Skin diseases are mostly found in animals as well as human beings. Skin diseases is a particular kind of disease which is caused by bacteria or an infection. These diseases mostly include ringworms, yeast infection, brown spots, allergies etc., One among these skin diseases is Skin disease.Skin disease is a condition wherein pores and skin cells building up and shape scales and itchy, dry patches. Skin disease is notion to be an immune device problem. Triggers includes infections, stress and bloodless. Over 10 million cases in step within year are determined in India. Treatment can help, but this situation can't be cured.Skin disease includes some of the symptoms like, rashes or patches of pink, inflamed skin, frequently included with free, silvered-coloured scales. In some instances, the plaques will grow and merge into any other, protecting huge areas. Itchy, painful skin which could crack or bleed.Small regions of bleeding wherein the involved pores and skin is scratched. Problem with finger nails or toe nails might also start to disintegrate or detach from the nail mattress. Scaly patches at the scalp is another example.Skin disease additionally have some of the treatments that targets to eliminate scales and stop skin cells from growing so quick. Topical ointments, light remedy and medicinal drug can offer relief. Photodynamic therapy, steroid, vitamin A derivatives, anti-inflammatory immunosuppressive drug and vitamin. Stress control, petroleum jelly, mild therapy.



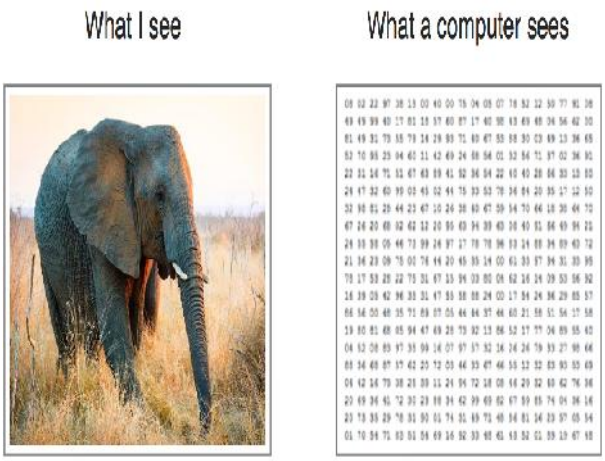
Skin cancer is very common these days. The statistics of American Cancer Society, Inc, Surveillance Research in 2020estimated new Melanoma skin cancer cases are 100,350 among 60,350 are male cases and 43,070 are female cases. The estimated death rate by Skin cancer is 6,850 out of which 8,030 are male and 3,450 are female, it will increase almost by 2 percent. Generally, three types of skin cancer are Basal Cell Carcinoma (BCC): It grows from the bottom of the epidermis in the long-term exposure area to sunlight. The growth rate of skin cancer is slow, so diagnosis is very easy. Basal Cell Carcinoma can visualize as tiny, shiny, smooth, waxy or pale lump, red with rough, dry, or scaly patches. Squamous Cell Carcinoma (SCC): It is another type of skin cancer. It develops at the outer most layer of the skin that is known asBasal Cell Carcinoma. It spread to the other skin areas, tissues and even bones during its early stage. It is the main difference between BCC and SCC. Squamous Cell Carcinoma can visualize as tiny, smooth, small lumps with real or brown Malignant Melanoma (MM). It is the third type and the most dangerous skin cancer disease. It is formed in the melanocytes. Melanoma skin cancer are generally as an asymmetry in shape with irregular borders with unnatural color. The growth of skin cancer is rapidly increased by time. Melanoma which is the most dangerous type of skin cancer is rapidly increasing. The death rate of melanoma skin cancer is high compared to other types of skin cancer diseases. Our system is used to

identify the skin disease using ML, deep neural network and deep CNN and compare the performance

Convolutional Neural Network

A convolutional neural network (CNN) is a special architecture of artificial neural networks, suggested by Yann LeCun in the yearof 1988. CNN uses few features from the visual cortex of the image. One of the most required use of this architecture is image classification. Forexample Instagram uses CNN for automatic tagging algorithms, Flipkart—for generating product recommendations and Google for search through among users' photos.

The main goal ofthe image classification is the acceptance of the input image and the following definitions and characteristics of its class. This is a skill and talent that people learn from their birth and are able to easily identify that the image in the picture is an elephant. But the computer sees it quite differently which is nothing but in terms of binary.

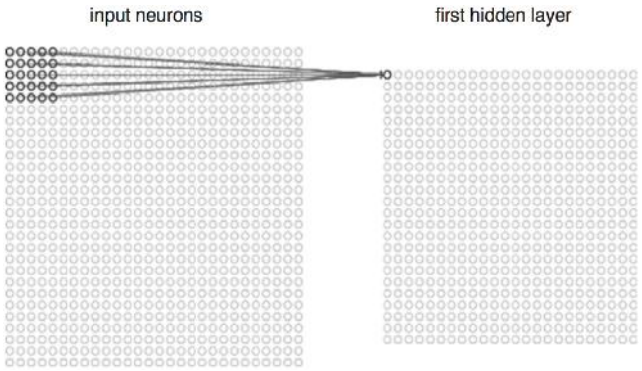


The computer considers to be an array of pixels. For example, if the size of the image is 300 x 300. Then, the size of the array will be as 300x300x3. Where 300 isfor the width, next 300 is for the height and 3 is RGB channel values. The computer assigns a value from 0 to 255 to each numbers. This value is the intensity values of the pixel at each point.

Solution to the problem is,the computer looks for the characteristics from the base level. In human understanding way these characteristics are nothing but the trunk or large ears. For the computer, these characteristics are simply boundaries or curvatures. Later the computer constructs more abstract concepts using CNN.

In more detail: the image is passed through various convolutional, nonlinear, pooling layers and fully connected layers, and then it generates the respective output.

The Convolution layer is always the first step. The image (pixel values consisting of the matrix) is entered into it. Reading of the matrix begins at the top left corner of thheimage. Next the software selects a smaller matrix, which is called a **filter**. Then the filter produces convolution as it moves along way the input image. The filter multiplies its values by the original pixel values. All these multiplications are summed up together to obtain a new value. Filter has read the image only in the upper left corner, it moves further right by 1 unit performing a similar operation. Once the filter is done passing all the positions, a matrix is obtained, which is smaller than the input matrix.



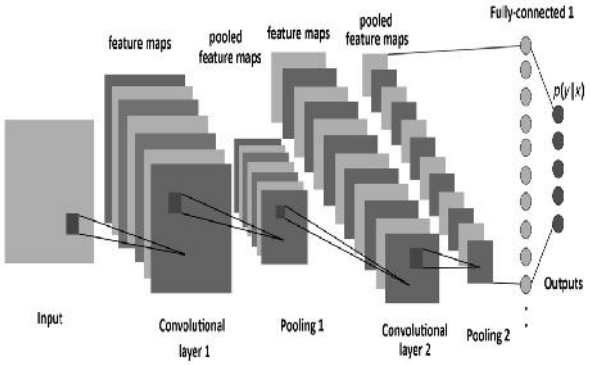
From a human perspective, this operation is analogousin identifying boundaries and simple colours of the image. To recognize the properties of a higher level like the trunk or large ears the whole network is required.

The network consists of various convolutional networks mixed with nonlinear and pooling layers. When the image passes through one of the convolution layer, the output of the first layer becomes the input for the second layer. And this happens with every further upcoming convolutional layer.

The nonlinear layer is added after every convolution operation. It has an activation function that develops nonlinear property. Without this property, the network would not be sufficiently high and will not be able to model the variable (as a class label).

The pooling layer uses the nonlinear layer. It works with the width and height of the image and performs a down sampling operation to those images. As a result the image volume is diminished. This means if some features and characteristics have already been identified in the previous

convolution operation, then a detailed image is no longer required for further processing..



After completion of these convolutional, nonlinear and pooling layers, it is important to attach a fully connected layer. This layer takes the output from these convolutional networks. Attaching a fully connected layer at the end of the network results in an N dimensional vector, where N are the number of classes.

II. METHEDOLOGY AND DESIGN

The proposed work can be grouped into two sections:

Image Segmentation

Pigmented skin lesion segmentation is to separate the lesion from the context is an variant procedure before starting with the feature abstraction/extraction in order to organize the three different types of lesion (benign, malignant and atypical). Image Segmentation To removing the world of enthusiasm from the given picture, the division is employed. The locale of enthusiasm containing every pixel comparative properties. Here the best entropy thresholding is employed. Most significantly we'd like to vary over the primary picture into greyscale and afterward compute a histogram of dark scale picture. then with the help of most extreme entropy separate forefront from the inspiration. The double picture is acquired within the wake of applying the foremost extreme entropy. Highlight extraction assumes a fundamental job. It separates the info present during a given picture. Here GLCM is employed for surface picture investigation. Spatial reliance between picture pixels is caught with the help of GLCM even as catch most

Feature Abstraction

Feature Abstraction is the procedure of computing parameters that symbolize the appearance of the input image.Whose output will have an uninterrupted and robust impact on the enactment of the organization systems.After extracting the lesion in the segmentation stage,the predefined features will be extracted forclassifying the features. The required selected features are shape, color and different texture features. These images have few statistical texture features, we use one of the common algorithm to extract those features which is Gray Level Co-Occurrence Matrix(GLCM). We combined these features to get a good classification results for distinguishing thebenign from the malignant skin lesions. The feature abstractionprocessincludes4phasesasfollows:

1. Phase1:

We consider the original image in RGB format, which contains three channels of colors namely Red, Green and Blue. Color features are extracted by calculating the density of specific colors in the lesion image.

2. Phase2:

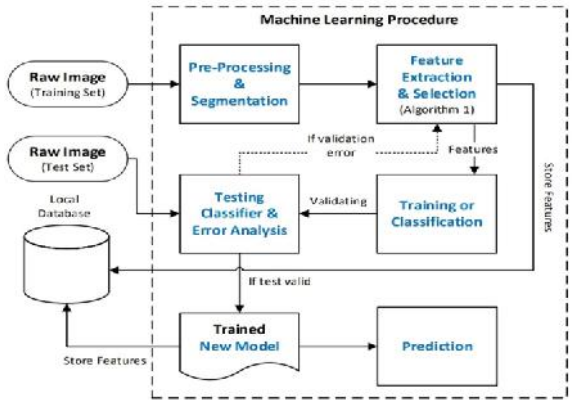
Binary is considered, where the features of Asymmetry, border irregularity, and circulation are obtained from the binary image. TDS features are calculated with parameters used as Asymmetry, Border irregularity, color and diameter

3. Phase3:

A lesion image in gray scale image. Energy, correlation, homogeneity and contrast features are obtained by applying gray-level co-occurrence matrix (GLCM) on the graylevel image of the lesion image.

4. Phase4:

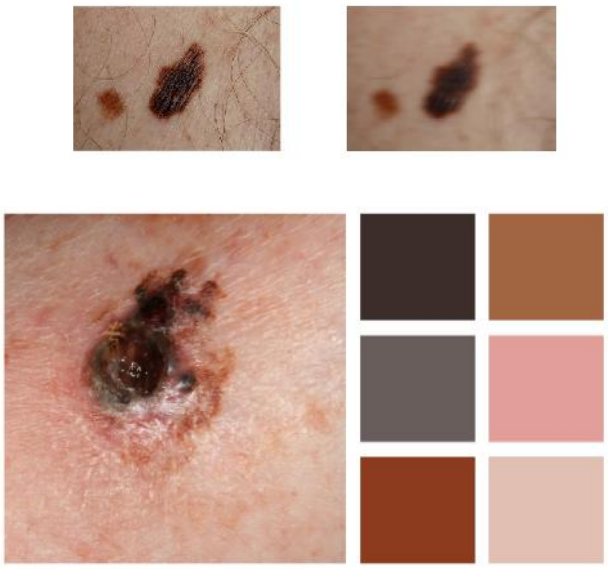
Histogram equalizedimage is considered, where the features of entropy, skewness,kurtosis and mean are obtained as it is shown inFigure.3-G.



III. ALGORITHMS

HAIR DETECTION AND EXCLUSION

In dermoscopy images, if hair exists on the skin, it will show up unmistakably in the dermoscopy images. Therefore, hair can discourage solid sore location and highlight extraction, bringing about inadmissible characterization comes about. This area acquaints a picture preparing strategy with distinguish and prohibit hair from the dermoscopy images as a fundamental advance. The outcome is a perfect hair cover which can be utilized to portion and evacuate the hair in the picture, setting it up for encourage division and investigation. To identify and prohibit the hair from the injury, in the first place, the hair is sectioned shape the sore. To achieve this task, 84 directional channels are used.



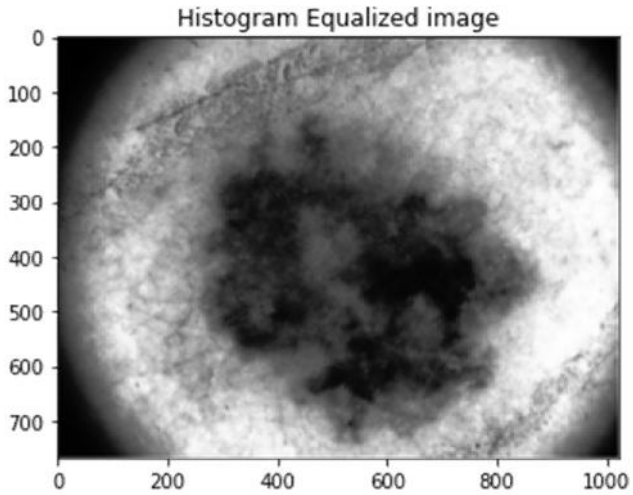
B. IMAGE SEGMENTATION

Pigmented skin lesion segmentation to separate the lesion from the contextual is an indispensable procedure before beginning the feature abstraction in order to organize the dissimilar types of lesion (benign, melanoma, and atypical). The segmentation step follow as: First, RGB dermoscopy image is read and transformed to a gray scale image. It is done by forming a weighted sum of the R, G, and B constituents as 0.2989 x R + 0.5870 x G + 0.1140 x B. Then, a two dimensional Gaussian low-pass filter is produced by Equations 2 and 3.

$$h_g(n_1, n_2) = e^{\frac{-(n_1^2 + n_2^2)}{2\sigma^2}} \quad (2)$$

$$h(n_1, n_2) = \frac{h_g(n_1, n_2)}{\sum_{n_1} \sum_{n_2} h_g} \quad (3)$$

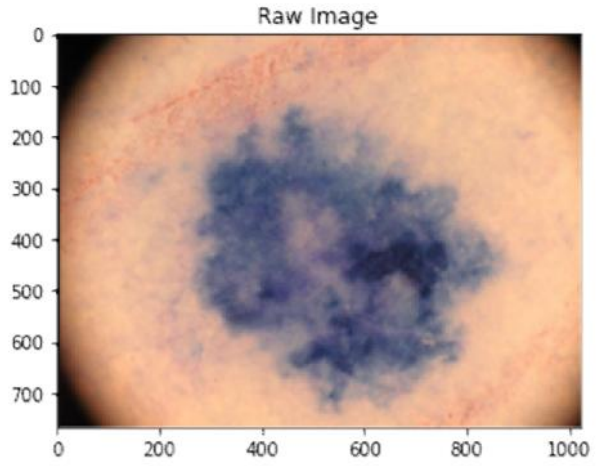
where*h* is a 2-D filter of size *n*1, *n*2 9 x 9, and sigma is 0.5. The filtered image is provided in the figure. After the Gaussian filter is applied, a global threshold is calculated by Otsu's technique to be used to transform an intensity image to a binary image. Otsu's technique pick out the threshold to diminish the intra-class variance of the contextual and foreground pixels. This directly deals with the problem of assessing the goodness of thresholds. A threshold value is selected by the discriminant criterion which is said to be optional. The subsequent image is given in figure..



D. FEATURE ABSTRACTION

Feature abstraction is the procedure of computing parameters that symbolize the appearances of the input image, whose output will have an uninterrupted and robust impact on the enactment of the organization systems. In this case study, five diverse features are to be considered. These are 2-D Fast Fourier Transform, 2-D Discrete Cosine Transform,

Complexity Feature Set, Color Feature Set and Pigment Network Feature Set. In addition to the five feature sets, the subsequent four features are also considered: Lesion Shape Feature, Lesion Orientation Feature, Lesion Margin Feature and Lesion Intensity Pattern Feature.



1)2-D Fast Fourier Transform

The 2-D Fast Fourier Transform (FFT) feature set is decided. The 2-D FFT feature set embraces the first coefficient of FFT2.

2) 2-D Discrete Cosine Transform

A 2-D Discrete Cosine Transform (DCT) articulates a predetermined series of data points in terms of a sum of cosine functions oscillating at diverse rate of renewal. The 2-D DCT feature set comprises the first coefficient of DCT2, the first coefficient of the cross-correlation of the first 20 rows.

3) Complexity Feature Set

The complexity feature set includes the mean (Equation 4), standard deviation (Equation 5), and mode built on the intensity value.

$$\overline{M} = \frac{\sum_{i=1}^{n} I_i}{n}, \tag{4}$$

$$\sigma = \frac{\sqrt{\sum_{i=1}^{n} (I_i - \overline{M})^2}}{n} \tag{5}$$

Where, *M* is the mean, α is the standard deviation, *Ii* is the intensity value of pixel *i*and *n* is the pixels count.

4) Color Feature Set

Color features are very crucial. Average pictures comprise of three-shading channels that are red blue and green. Utilization of shading is another strategy to evaluate melanoma dangers. Normally, melanoma sores tend to change shading seriously making the influenced area to be unpredictable. With a specific end goal to get the 2-D shading histogram from the 3-D shading histogram, all esteems in the enlightenment hub are amassed. Thus, 8 x 8 = 64 shading receptacles are created, each considered as one element.

5) Pigment Network Feature Set

Shade organize is created by melanin or melanocytes in basal keratinocytes. The shade arrange is the most vital structure in dermoscopy. It shows up as a system of thin dark colored lines over a diffuse light darker foundation. Thick shade rings (the system) are because of projections of rete pegs or edges. The shade arrange is found in some atypical and melanoma injuries. In a few locales the system is extended. It doesn't need to involve the entire sore. To remove the shade organize highlight, to begin with, the system is sectioned from the sore, to achieve this assignment; an arrangement of 12 directional channels is constructed.

6) Lesion Shape Feature Extraction

For any pixel p1(x1; y1) on the injury limit, an intersection pixel p2(x2; y2) on the best-fit circle is found by a bar that begins at the focal point of best-fit circle. The variety between p(x1; y1) and p2(x2; y2) is then ascertained by their separation:

$$D(P) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{11}$$

At that point, the sore shape include is computed by the variety between the sore shape and the best-fit circle as the accompanying:

$$L(S) = \frac{\sum D(P)}{LB_N} \tag{12}$$

Where, *P* is any pixel on the lesion boundary and *LBN* signifies the overall number of pixels on the lesion boundary.

7) Lesion Orientation Feature

The orientation of the lesion is measured by the angle of the main axis of the above best-fit ellipse as shown in Figure 8. The range of lesion orientation is defined between 0 and $\frac{\pi}{2}$ as the following:

$$L(O)=\begin{cases}\theta & \text{if} \quad 0\le\theta\le\dfrac{\pi}{2}\\ \pi-\theta & \text{if} \quad \dfrac{\pi}{2}<\theta\le\pi\\ \theta-\pi & \text{if} \quad \pi<\theta\le\dfrac{3\pi}{2}\\ 2\pi-\theta & \text{if} \quad \dfrac{3\pi}{2}<\theta\le 2\pi.\end{cases} \tag{13}$$

8) Lesion Margin Feature

The distance map is used to seizure the ripple and the pointed appearances. For any pixel *P*(*x*; *y*) in the ROI, its eight neighbors are distinct as the following:

$$\begin{aligned}N_8(P)=\{&(x-1,y-1),(x,y-1),(x+1,y-1),\\&(x-1,y),(x+1,y),(x-1,y+1),\\&(x,y+1),(x+1,y+1)\}\end{aligned} \tag{14}$$

The distance from *P* to the lesion boundary is recursively defined as the following,

$$D(P)=Min\{D(N_8(P))\}+1 \tag{15}$$

Where, *Min*f*D*(*N*8 (*P*)) g denotes the least possible of known distance in *P's* eight neighbors.

9) Lesion Intensity Pattern Feature

The average gray intensity of the lesion is used to denote the intensity pattern feature. The average gray intensity pattern *L* (*IP*) is defined as:

$$L(IP)=\frac{\sum_{P\in ROI} I(P)}{N_{ROI}} \tag{16}$$

Where, *I* (*P*) is the gray intensity of lesion pixel *P* and *NROI* denotes the number of lesion pixels.

10) Lesion Variation Pattern Feature

The variation on pixel *P* is assessed by the gradient magnitude. According to the definition of neighbors in Equation 8, the Sobel gradients on *x*-direction and *y*-direction of pixel *P*(*x*; *y*) are respectively defined as:

$$\begin{aligned}G_x(P)=&I(x-1,y-1)+2I(x-1,y)+I(x-1,y+1)\\&-I(x+1,y-1)-2I(x+1,y)-I(x+1,y+1)\end{aligned} \tag{17}$$

And

$$\begin{aligned}G_y(P)=&I(x-1,y-1)+2I(x,y-1)+I(x+1,y-1)\\&-I(x-1,y+1)-2I(x,y+1)-I(x+1,y+1)\end{aligned} \tag{18}$$

Then, the gradient magnitude on *P* is defined as:

$$G(P)=\sqrt{G_x(P)^2+G_y(P)^2} \tag{19}$$

Finally, the average variation pattern is designed.

HU MOMENTS ALGORITHM:

Hu moments (hu second invariants) are a hard and fast of 7 numbers calculated the usage of significant moments which might be invariant. The six moments were proved to be invariant to translational, scale, rotation and reflection. While the 7th moments sign modifications for mirrored image.

In our system we used hu moments to extract more than one feature from the given input image which helps in detecting the Skin disease disease.

```
# feature descriptor-1: Hu Moments 7 feature
def fd_hu_moments(image):
    image=cv2.cvtColor(image,cv2.COLOR_BGR2GRAY)
    feature=cv2.HuMoments(cv2.moments(image)).flatten()
    return feature
```

HARALICK TEXTURE FEATURE ALGORITHM

Haralick's texture features were calculated the usage of the kharalick( ) feature of the cytometry Toolbox the premise of this toolbox is used to convert the RGB photograph to Gray scale picture.

Haralick applies Gray Level Co-Incidence Matrices (GLCM).This matrix is square with dimensions Ng, where Ng is the wide variety of gray levels in the photo.

Elements[i , j] of the matrix is generated by counting the number of times a pixel with cost i is adjoining to a pixel with cost j and then dividing the complete matrix.

Each entry is consequently take into account to be opportunity that a pixel with cost i may be found adjacent to a pixel of feej.

Gray scale images are composed of shades of gray varying from black to weakest intensity to white on the strongest intensity. The cost might also variety from 0 to 255.

```
#feature descriptor-1:

deffd_haralick(image): #glcm feature

#convert the image to grayscale

   gray=cv2.cvtColor(image,cv2.COL
OR_BGR2GRAY)

# compute the haralick texture
feature vector

   haralick=mahotas.features.haralic
```

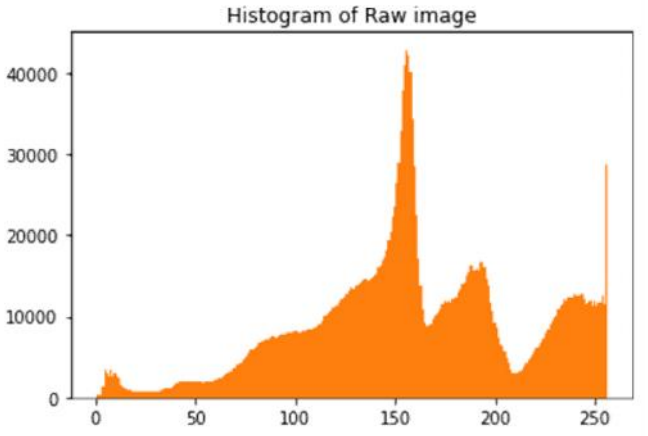
HISTOGRAM EQUALIZATION

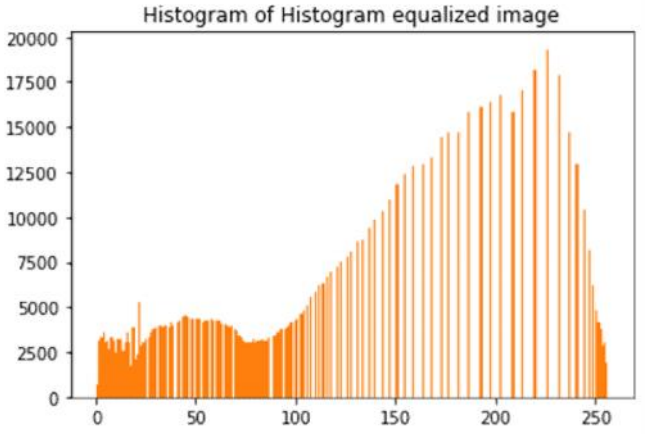
Histogram is a chart that suggests frequencies for intervals of values of a metric variable.

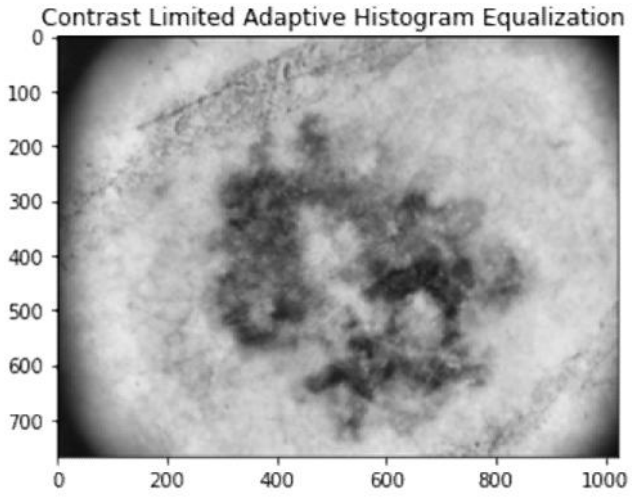
Histogram of an image may be drawn plotting pixel depth v/s frequency of the pixel intensity or possibilities of the pixel depth.

Histogram equalization is aunique technique to adjust the image intensities to enhance the contrast of the image.

```
#feature-descriptor-3: Color Histogram


deffd_histogram(image,mask=None):

   #convert the image to HSV color-space

   image=cv2.cvtColor(image,cv2.COLOR_BGR2HSV
)

   #compute the color histogram

   hist=cv2.calcHist([image],[0,1,2],None,[bins,bins,
bins],

                [0,256,0,256,0,256])
```







IV. CONCLUSION AND FUTURE WORK

In the proposed work, an unsupervised deep learning technique is used in psoriasis detection. The images are first pre-processed to remove digitization noise.

The proposed model has achieved an accuracy of up to 91.6% in detecting psoriasis disease.

Deep neural network and deep belief network are used as feature extractors and feature detectors in efficiently identifying the psoriasis disease.

Experimental result indicates that DNN features easily outperform DBN in terms of better accuracy.
In the future the datasets will be increased for better performance.

In future, we can also add random forest for better results. Since random forest has some methods in it and it randomly selects one process and gives more accuracy

V. RESULT

A result is the final consequence of actions or events expressed qualitatively or quantitatively. Performance analysis is an operational analysis, is a set of basic quantitative relationship between the performance quantities.

METRICES

	precision	recall	f1-score	support
BENIGN	1.00	0.25	0.40	48
MALIGNANT	0.55	1.00	0.71	44
micro avg	0.61	0.61	0.61	92
macro avg	0.78	0.62	0.55	92
weighted avg	0.78	0.61	0.55	92
samples avg	0.61	0.61	0.61	92

OUTPUT

```
test_image = cv.resize(cv.imread(full_data),  (256,256))
test_image = np.array(test_image).reshape( 1,256, 256, -1)
pred=model.predict(test_image)
pos,val=np.where(pred==1)
print(pred)
skin_data[int(val)]
```

```
Out[80]: 'MALIGNANT'
```

ACCURACY

```
from sklearn.metrics import accuracy_score
accuracy=accuracy_score(test_labels, predictions)
print('Accuracy:',accuracy*100)
```

```
Accuracy: 100.0
```

REFERENCES

[1] Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell 100:57– 70.
[2] Liotta L, Petricoin E. 2000. Molecular profiling of human cancer. Nat. Rev.Genet. 1:48–56.
[3] Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. 2002. Clinical proteomics: translating bench side promise into bedside reality. Nat. Rev. Drug Discov. 1:683–95.
[4] A.W.Kopf, T.G. Salopek, J. Slade, A.A. Marghood, R.S. Bart, Techniques of cutaneous examination for the detection of skin cancer, Cancer Supplement 75 (2) (1994) 684-690.
[5] I. Papamichail, N. Nikolaidis, I. Glava, K. Lentzas, Marmagkiolis, et al. Merkel cell carcinoma of the upper extremity: case report and an update, J. Surg. Oncol. 6 (32) (2008).
[6] M.E. Celebi, H.A. Kingravi, B. Uddin, et al., A methodological approach to the classification of dermoscopy images, Comput. Med. Imaging Graph. 31(2007) 362-373.
[7] Santosh Achakanalli, G.Sadashivappa, Journal publication of ABCD system of melanoma detection, June 2014.

[8] ShefaliLaddha, KomalPaliwal, Journal publication of Skin Cancer detection using medical image processing ,Dept of Computer Engineering & IT, IJARCEE, September 2016.
[9] A. Ahmad, A. B. Mansoor, R. Mumtaz, M. Khan, and S. H. Mirza,"Image processing and classification in diabetic retinopathy: A review,"
[10] Yin, F.F.; Giger, M.L.; Vyborny, C.J.; Schmidt, R.A. Computerized detection of masses in digital mammograms: automated alignment of breast images and its effects on bilateral-substraction technique.

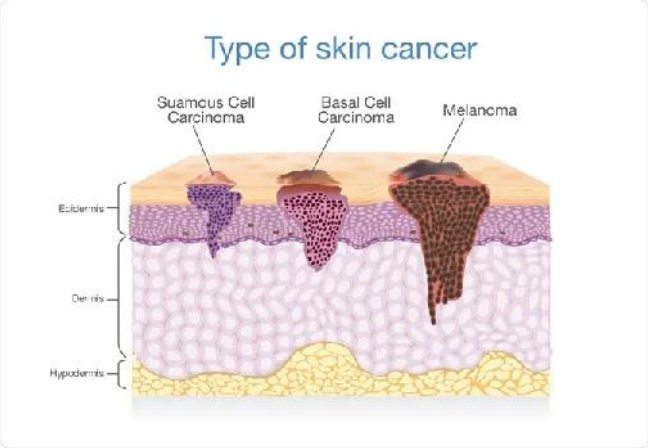
Skin Cancer Detection Using Machine Learning And Image Processing

Hariharan Kumar[1], Rohit R Nagarahalli[2], Srinivasa MS[3], Sangey Srivastava[4], MaqdumShariff[5]
[1, 2, 3, 4, 5]Dept of Computer Science
[1, 2, 3, 4, 5]Atria Institute of Technology

Abstract- *Skin cancer is a major health issue in the present day especially melanoma skin cancer. In general, most of the skin cancers are curable if they are detected and treated in the early stages. With the speedy growth of skin cancer, there is a need for an automated computerized diagnosis mechanism of skin cancer in the early stage which is required. Various skin cancer images have similar visual characteristics and representations. It is very important to extract the features from the skin cancer images. The automated computerized diagnosis mechanism helps to improve the accuracy of skin diseases which helps the dermatologists to provide the diagnosis and better treatment for the patients. This paper depicts the comparative study on various traditional image processing and current technologies of different image processing techniques for skin cancer image classification, pre-processing techniques, Feature extraction/abstraction.*

I. INTRODUCTION

Skin diseases are mostly found in animals as well as human beings. Skin diseases is a particular kind of disease which is caused by bacteria or an infection. These diseases mostly include ringworms, yeast infection, brown spots, allergies etc., One among these skin diseases is Skin disease.Skin disease is a condition wherein pores and skin cells building up and shape scales and itchy, dry patches. Skin disease is notion to be an immune device problem. Triggers includes infections, stress and bloodless. Over 10 million cases in step within year are determined in India. Treatment can help, but this situation can't be cured.Skin disease includes some of the symptoms like, rashes or patches of pink, inflamed skin, frequently included with free, silvered-coloured scales. In some instances, the plaques will grow and merge into any other, protecting huge areas. Itchy, painful skin which could crack or bleed.Small regions of bleeding wherein the involved pores and skin is scratched. Problem with finger nails or toe nails might also start to disintegrate or detach from the nail mattress. Scaly patches at the scalp is another example.Skin disease additionally have some of the treatments that targets to eliminate scales and stop skin cells from growing so quick. Topical ointments, light remedy and medicinal drug can offer relief. Photodynamic therapy, steroid, vitamin A derivatives, anti-inflammatory immunosuppressive drug and vitamin. Stress control, petroleum jelly, mild therapy.



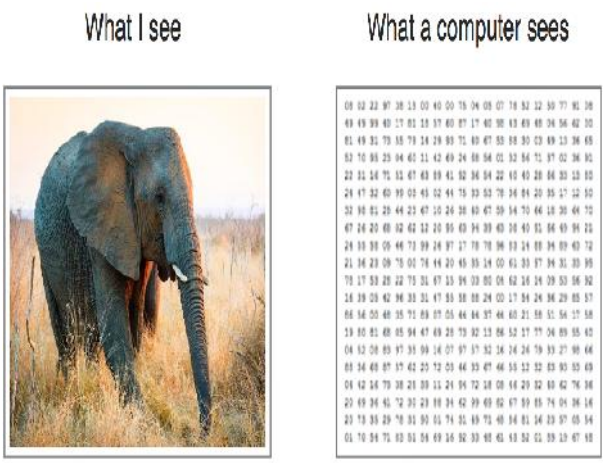
Skin cancer is very common these days. The statistics of American Cancer Society, Inc, Surveillance Research in 2020estimated new Melanoma skin cancer cases are 100,350 among 60,350 are male cases and 43,070 are female cases. The estimated death rate by Skin cancer is 6,850 out of which 8,030 are male and 3,450 are female, it will increase almost by 2 percent. Generally, three types of skin cancer are Basal Cell Carcinoma (BCC): It grows from the bottom of the epidermis in the long-term exposure area to sunlight. The growth rate of skin cancer is slow, so diagnosis is very easy. Basal Cell Carcinoma can visualize as tiny, shiny, smooth, waxy or pale lump, red with rough, dry, or scaly patches. Squamous Cell Carcinoma (SCC): It is another type of skin cancer. It develops at the outer most layer of the skin that is known asBasal Cell Carcinoma. It spread to the other skin areas, tissues and even bones during its early stage. It is the main difference between BCC and SCC. Squamous Cell Carcinoma can visualize as tiny, smooth, small lumps with real or brown Malignant Melanoma (MM). It is the third type and the most dangerous skin cancer disease. It is formed in the melanocytes. Melanoma skin cancer are generally as an asymmetry in shape with irregular borders with unnatural color. The growth of skin cancer is rapidly increased by time. Melanoma which is the most dangerous type of skin cancer is rapidly increasing. The death rate of melanoma skin cancer is high compared to other types of skin cancer diseases. Our system is used to

identify the skin disease using ML, deep neural network and deep CNN and compare the performance

Convolutional Neural Network

A convolutional neural network (CNN) is a special architecture of artificial neural networks, suggested by Yann LeCun in the yearof 1988. CNN uses few features from the visual cortex of the image. One of the most required use of this architecture is image classification. Forexample Instagram uses CNN for automatic tagging algorithms, Flipkart—for generating product recommendations and Google for search through among users' photos.

The main goal ofthe image classification is the acceptance of the input image and the following definitions and characteristics of its class. This is a skill and talent that people learn from their birth and are able to easily identify that the image in the picture is an elephant. But the computer sees it quite differently which is nothing but in terms of binary.

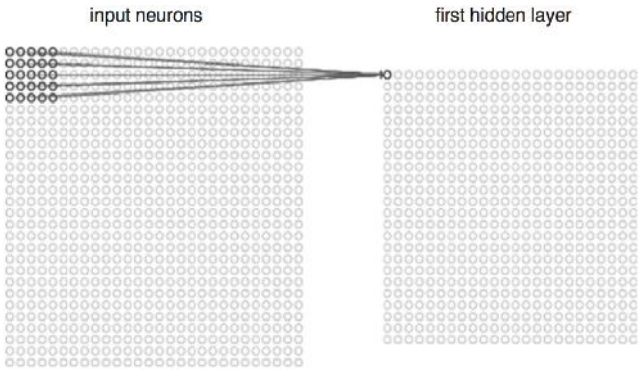


The computer considers to be an array of pixels. For example, if the size of the image is 300 x 300. Then, the size of the array will be as 300x300x3. Where 300 isfor the width, next 300 is for the height and 3 is RGB channel values. The computer assigns a value from 0 to 255 to each numbers. This value is the intensity values of the pixel at each point.

Solution to the problem is,the computer looks for the characteristics from the base level. In human understanding way these characteristics are nothing but the trunk or large ears. For the computer, these characteristics are simply boundaries or curvatures. Later the computer constructs more abstract concepts using CNN.

In more detail: the image is passed through various convolutional, nonlinear, pooling layers and fully connected layers, and then it generates the respective output.

The Convolution layer is always the first step. The image (pixel values consisting of the matrix) is entered into it. Reading of the matrix begins at the top left corner of thheimage. Next the software selects a smaller matrix, which is called a **filter**. Then the filter produces convolution as it moves along way the input image. The filter multiplies its values by the original pixel values. All these multiplications are summed up together to obtain a new value. Filter has read the image only in the upper left corner, it moves further right by 1 unit performing a similar operation. Once the filter is done passing all the positions, a matrix is obtained, which is smaller than the input matrix.



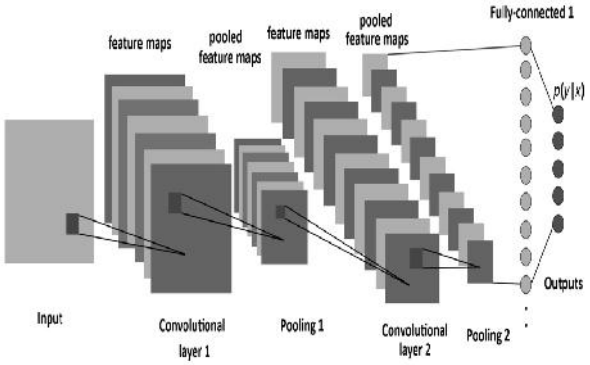
From a human perspective, this operation is analogousin identifying boundaries and simple colours of the image. To recognize the properties of a higher level like the trunk or large ears the whole network is required.

The network consists of various convolutional networks mixed with nonlinear and pooling layers. When the image passes through one of the convolution layer, the output of the first layer becomes the input for the second layer. And this happens with every further upcoming convolutional layer.

The nonlinear layer is added after every convolution operation. It has an activation function that develops nonlinear property. Without this property, the network would not be sufficiently high and will not be able to model the variable (as a class label).

The pooling layer uses the nonlinear layer. It works with the width and height of the image and performs a down sampling operation to those images. As a result the image volume is diminished. This means if some features and characteristics have already been identified in the previous

convolution operation, then a detailed image is no longer required for further processing..



After completion of these convolutional, nonlinear and pooling layers, it is important to attach a fully connected layer. This layer takes the output from these convolutional networks. Attaching a fully connected layer at the end of the network results in an N dimensional vector, where N are the number of classes.

II. METHEDOLOGY AND DESIGN

The proposed work can be grouped into two sections:

Image Segmentation

Pigmented skin lesion segmentation is to separate the lesion from the context is an variant procedure before starting with the feature abstraction/extraction in order to organize the three different types of lesion (benign, malignant and atypical). Image Segmentation To removing the world of enthusiasm from the given picture, the division is employed. The locale of enthusiasm containing every pixel comparative properties. Here the best entropy thresholding is employed. Most significantly we'd like to vary over the primary picture into greyscale and afterward compute a histogram of dark scale picture. then with the help of most extreme entropy separate forefront from the inspiration. The double picture is acquired within the wake of applying the foremost extreme entropy. Highlight extraction assumes a fundamental job. It separates the info present during a given picture. Here GLCM is employed for surface picture investigation. Spatial reliance between picture pixels is caught with the help of GLCM even as catch most

Feature Abstraction

Feature Abstraction is the procedure of computing parameters that symbolize the appearance of the input image.Whose output will have an uninterrupted and robust impact on the enactment of the organization systems.After extracting the lesion in the segmentation stage,the predefined features will be extracted forclassifying the features. The required selected features are shape, color and different texture features. These images have few statistical texture features, we use one of the common algorithm to extract those features which is Gray Level Co-Occurrence Matrix(GLCM). We combined these features to get a good classification results for distinguishing thebenign from the malignant skin lesions. The feature abstractionprocessincludes4phasesasfollows:

1. Phase1:

We consider the original image in RGB format, which contains three channels of colors namely Red, Green and Blue. Color features are extracted by calculating the density of specific colors in the lesion image.

2. Phase2:

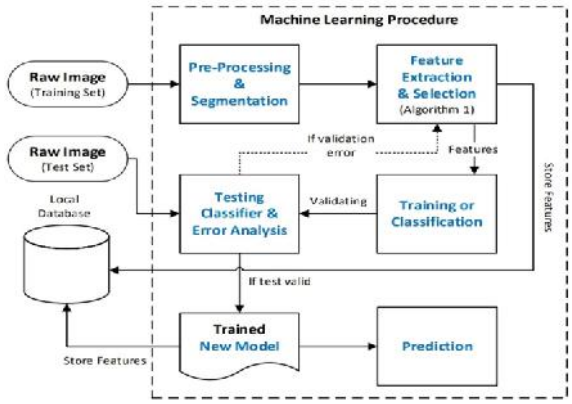
Binary is considered, where the features of Asymmetry, border irregularity, and circulation are obtained from the binary image. TDS features are calculated with parameters used as Asymmetry, Border irregularity, color and diameter

3. Phase3:

A lesion image in gray scale image. Energy, correlation, homogeneity and contrast features are obtained by applying gray-level co-occurrence matrix (GLCM) on the graylevel image of the lesion image.

4. Phase4:

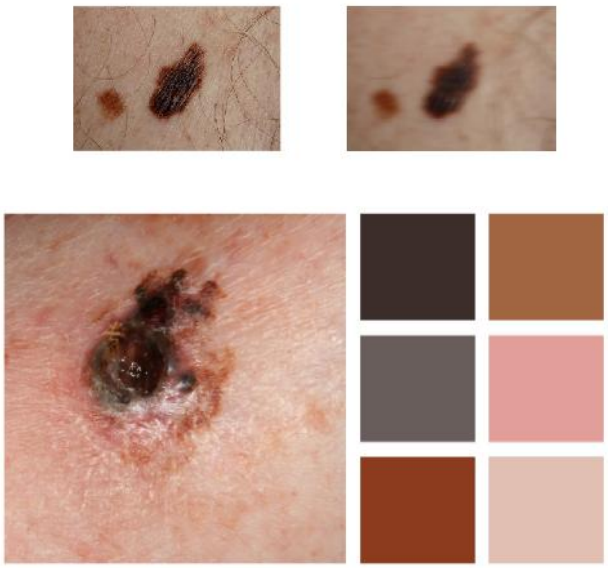
Histogram equalizedimage is considered, where the features of entropy, skewness,kurtosis and mean are obtained as it is shown inFigure.3-G.



III. ALGORITHMS

HAIR DETECTION AND EXCLUSION

In dermoscopy images, if hair exists on the skin, it will show up unmistakably in the dermoscopy images. Therefore, hair can discourage solid sore location and highlight extraction, bringing about inadmissible characterization comes about. This area acquaints a picture preparing strategy with distinguish and prohibit hair from the dermoscopy images as a fundamental advance. The outcome is a perfect hair cover which can be utilized to portion and evacuate the hair in the picture, setting it up for encourage division and investigation. To identify and prohibit the hair from the injury, in the first place, the hair is sectioned shape the sore. To achieve this task, 84 directional channels are used.



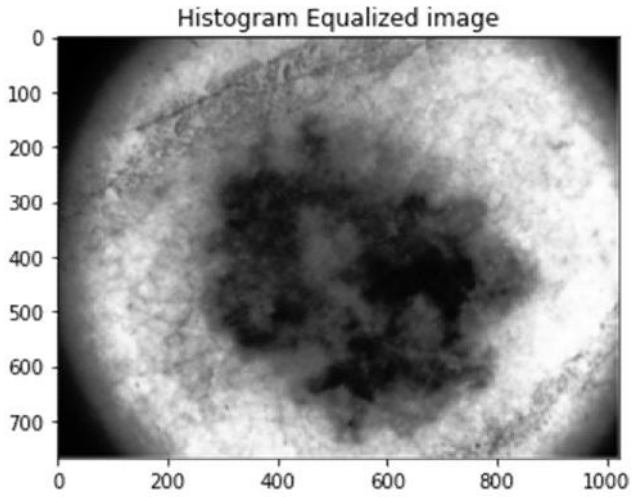
B. IMAGE SEGMENTATION

Pigmented skin lesion segmentation to separate the lesion from the contextual is an indispensable procedure before beginning the feature abstraction in order to organize the dissimilar types of lesion (benign, melanoma, and atypical). The segmentation step follow as: First, RGB dermoscopy image is read and transformed to a gray scale image. It is done by forming a weighted sum of the R, G, and B constituents as 0.2989 x R + 0.5870 x G + 0.1140 x B. Then, a two dimensional Gaussian low-pass filter is produced by Equations 2 and 3.

$$h_g(n_1, n_2) = e^{\frac{-(n_1^2+n_2^2)}{2\sigma^2}} \quad (2)$$

$$h(n_1, n_2) = \frac{h_g(n_1, n_2)}{\sum_{n_1}\sum_{n_2} h_g} \quad (3)$$

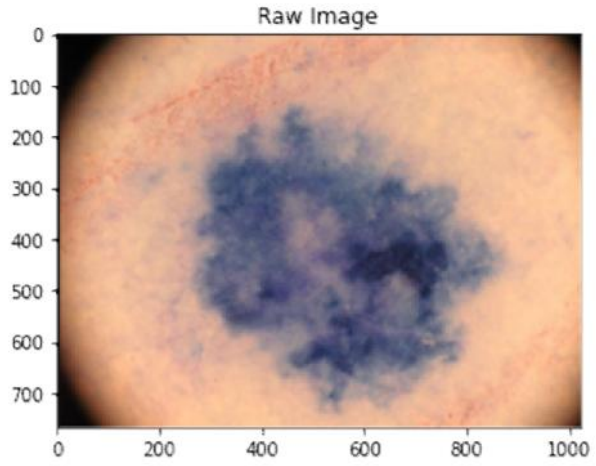
where*h* is a 2-D filter of size *n*1, *n*2 9 x 9, and sigma is 0.5. The filtered image is provided in the figure. After the Gaussian filter is applied, a global threshold is calculated by Otsu's technique to be used to transform an intensity image to a binary image. Otsu's technique pick out the threshold to diminish the intra-class variance of the contextual and foreground pixels. This directly deals with the problem of assessing the goodness of thresholds. A threshold value is selected by the discriminant criterion which is said to be optional. The subsequent image is given in figure..



D. FEATURE ABSTRACTION

Feature abstraction is the procedure of computing parameters that symbolize the appearances of the input image, whose output will have an uninterrupted and robust impact on the enactment of the organization systems. In this case study, five diverse features are to be considered. These are 2-D Fast Fourier Transform, 2-D Discrete Cosine Transform,

Complexity Feature Set, Color Feature Set and Pigment Network Feature Set. In addition to the five feature sets, the subsequent four features are also considered: Lesion Shape Feature, Lesion Orientation Feature, Lesion Margin Feature and Lesion Intensity Pattern Feature.



1)2-D Fast Fourier Transform

The 2-D Fast Fourier Transform (FFT) feature set is decided. The 2-D FFT feature set embraces the first coefficient of FFT2.

2) 2-D Discrete Cosine Transform

A 2-D Discrete Cosine Transform (DCT) articulates a predetermined series of data points in terms of a sum of cosine functions oscillating at diverse rate of renewal. The 2-D DCT feature set comprises the first coefficient of DCT2, the first coefficient of the cross-correlation of the first 20 rows.

3) Complexity Feature Set

The complexity feature set includes the mean (Equation 4), standard deviation (Equation 5), and mode built on the intensity value.

$$\overline{M} = \frac{\sum_{i=1}^{n} I_i}{n}, \tag{4}$$

$$\sigma = \frac{\sqrt{\sum_{i=1}^{n} (I_i - \overline{M})^2}}{n} \tag{5}$$

Where, *M* is the mean, α is the standard deviation, *Ii* is the intensity value of pixel *i*and *n* is the pixels count.

4) Color Feature Set

Color features are very crucial. Average pictures comprise of three-shading channels that are red blue and green. Utilization of shading is another strategy to evaluate melanoma dangers. Normally, melanoma sores tend to change shading seriously making the influenced area to be unpredictable. With a specific end goal to get the 2-D shading histogram from the 3-D shading histogram, all esteems in the enlightenment hub are amassed. Thus, 8 x 8 = 64 shading receptacles are created, each considered as one element.

5) Pigment Network Feature Set

Shade organize is created by melanin or melanocytes in basal keratinocytes. The shade arrange is the most vital structure in dermoscopy. It shows up as a system of thin dark colored lines over a diffuse light darker foundation. Thick shade rings (the system) are because of projections of rete pegs or edges. The shade arrange is found in some atypical and melanoma injuries. In a few locales the system is extended. It doesn't need to involve the entire sore. To remove the shade organize highlight, to begin with, the system is sectioned from the sore, to achieve this assignment; an arrangement of 12 directional channels is constructed.

6) Lesion Shape Feature Extraction

For any pixel p1(x1; y1) on the injury limit, an intersection pixel p2(x2; y2) on the best-fit circle is found by a bar that begins at the focal point of best-fit circle. The variety between p(x1; y1) and p2(x2; y2) is then ascertained by their separation:

$$D(P) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{11}$$

At that point, the sore shape include is computed by the variety between the sore shape and the best-fit circle as the accompanying:

$$L(S) = \frac{\sum D(P)}{LB_N} \tag{12}$$

Where, *P* is any pixel on the lesion boundary and *LBN* signifies the overall number of pixels on the lesion boundary.

7) Lesion Orientation Feature

The orientation of the lesion is measured by the angle of the main axis of the above best-fit ellipse as shown in Figure 8. The range of lesion orientation is defined between 0 and $\frac{\pi}{2}$ as the following:

$$L(O) = \begin{cases} \theta & \text{if} \quad 0 \le \theta \le \frac{\pi}{2} \\ \pi - \theta & \text{if} \quad \frac{\pi}{2} < \theta \le \pi \\ \theta - \pi & \text{if} \quad \pi < \theta \le \frac{3\pi}{2} \\ 2\pi - \theta & \text{if} \quad \frac{3\pi}{2} < \theta \le 2\pi. \end{cases} \tag{13}$$

8) Lesion Margin Feature

The distance map is used to seizure the ripple and the pointed appearances. For any pixel *P*(*x*; *y*) in the ROI, its eight neighbors are distinct as the following:

$$\begin{aligned} N_8(P) = \{&(x-1, y-1), (x, y-1), (x+1, y-1), \\ &(x-1, y), (x+1, y), (x-1, y+1), \\ &(x, y+1), (x+1, y+1)\} \end{aligned} \tag{14}$$

The distance from *P* to the lesion boundary is recursively defined as the following,

$$D(P) = Min\{D(N_8(P))\} + 1 \tag{15}$$

Where, *Min*f*D*(*N*8 (*P*)) g denotes the least possible of known distance in *P's* eight neighbors.

9) Lesion Intensity Pattern Feature

The average gray intensity of the lesion is used to denote the intensity pattern feature. The average gray intensity pattern *L* (*IP*) is defined as:

$$L(IP) = \frac{\sum_{P \in ROI} I(P)}{N_{ROI}} \tag{16}$$

Where, *I* (*P*) is the gray intensity of lesion pixel *P* and *NROI* denotes the number of lesion pixels.

10) Lesion Variation Pattern Feature

The variation on pixel *P* is assessed by the gradient magnitude. According to the definition of neighbors in Equation 8, the Sobel gradients on *x*-direction and *y*-direction of pixel *P*(*x*; *y*) are respectively defined as:

$$\begin{aligned} G_x(P) = &I(x-1, y-1) + 2I(x-1, y) + I(x-1, y+1) \\ &-I(x+1, y-1) - 2I(x+1, y) - I(x+1, y+1) \end{aligned} \tag{17}$$

And

$$\begin{aligned} G_y(P) = &I(x-1, y-1) + 2I(x, y-1) + I(x+1, y-1) \\ &-I(x-1, y+1) - 2I(x, y+1) - I(x+1, y+1) \end{aligned} \tag{18}$$

Then, the gradient magnitude on *P* is defined as:

$$G(P) = \sqrt{G_x(P)^2 + G_y(P)^2} \tag{19}$$

Finally, the average variation pattern is designed.

HU MOMENTS ALGORITHM:

Hu moments (hu second invariants) are a hard and fast of 7 numbers calculated the usage of significant moments which might be invariant. The six moments were proved to be invariant to translational, scale, rotation and reflection. While the 7th moments sign modifications for mirrored image.

In our system we used hu moments to extract more than one feature from the given input image which helps in detecting the Skin disease disease.

```
# feature descriptor-1: Hu Moments 7 feature
def fd_hu_moments(image):
    image=cv2.cvtColor(image,cv2.COLOR_BGR2GRAY)
    feature=cv2.HuMoments(cv2.moments(image)).flatten()
    return feature
```

HARALICK TEXTURE FEATURE ALGORITHM

Haralick's texture features were calculated the usage of the kharalick( ) feature of the cytometry Toolbox the premise of this toolbox is used to convert the RGB photograph to Gray scale picture.

Haralick applies Gray Level Co-Incidence Matrices (GLCM).This matrix is square with dimensions Ng, where Ng is the wide variety of gray levels in the photo.

Elements[i , j] of the matrix is generated by counting the number of times a pixel with cost i is adjoining to a pixel with cost j and then dividing the complete matrix.

Each entry is consequently take into account to be opportunity that a pixel with cost i may be found adjacent to a pixel of feej.

Gray scale images are composed of shades of gray varying from black to weakest intensity to white on the strongest intensity. The cost might also variety from 0 to 255.

```
#feature descriptor-1:

deffd_haralick(image): #glcm feature

#convert the image to grayscale

   gray=cv2.cvtColor(image,cv2.COL
OR_BGR2GRAY)

# compute the haralick texture
feature vector

   haralick=mahotas.features.haralic
```

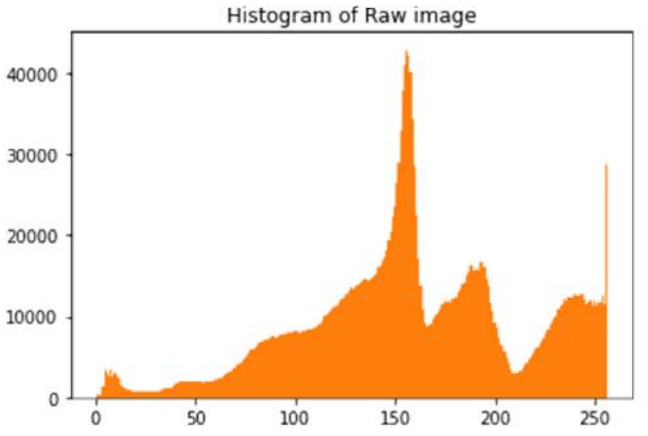
HISTOGRAM EQUALIZATION

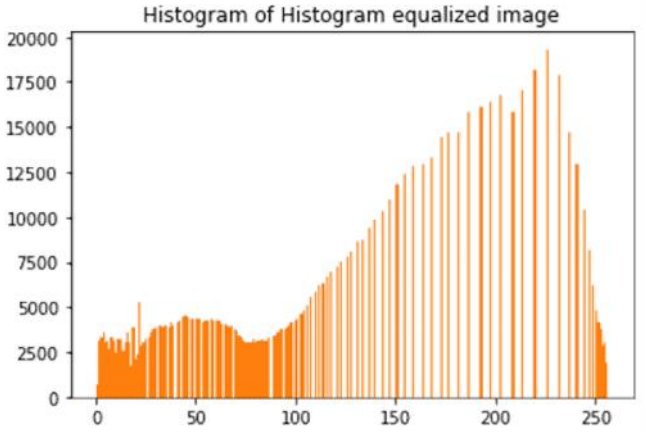
Histogram is a chart that suggests frequencies for intervals of values of a metric variable.

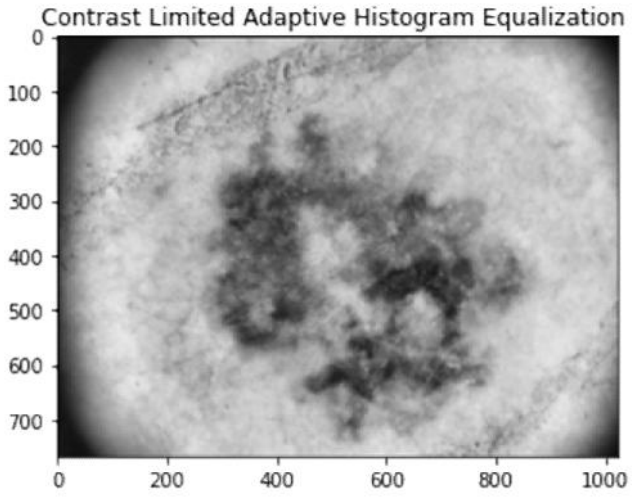
Histogram of an image may be drawn plotting pixel depth v/s frequency of the pixel intensity or possibilities of the pixel depth.

Histogram equalization is aunique technique to adjust the image intensities to enhance the contrast of the image.

```
#feature-descriptor-3: Color Histogram


deffd_histogram(image,mask=None):

   #convert the image to HSV color-space

   image=cv2.cvtColor(image,cv2.COLOR_BGR2HSV
)

   #compute the color histogram

   hist=cv2.calcHist([image],[0,1,2],None,[bins,bins,
bins],

                [0,256,0,256,0,256])
```







IV. CONCLUSION AND FUTURE WORK

In the proposed work, an unsupervised deep learning technique is used in psoriasis detection. The images are first pre-processed to remove digitization noise.

The proposed model has achieved an accuracy of up to 91.6% in detecting psoriasis disease.

Deep neural network and deep belief network are used as feature extractors and feature detectors in efficiently identifying the psoriasis disease.

Experimental result indicates that DNN features easily outperform DBN in terms of better accuracy.
In the future the datasets will be increased for better performance.

In future, we can also add random forest for better results. Since random forest has some methods in it and it randomly selects one process and gives more accuracy

V. RESULT

A result is the final consequence of actions or events expressed qualitatively or quantitatively. Performance analysis is an operational analysis, is a set of basic quantitative relationship between the performance quantities.

METRICES

	precision	recall	f1-score	support
BENIGN	1.00	0.25	0.40	48
MALIGNANT	0.55	1.00	0.71	44
micro avg	0.61	0.61	0.61	92
macro avg	0.78	0.62	0.55	92
weighted avg	0.78	0.61	0.55	92
samples avg	0.61	0.61	0.61	92

OUTPUT

```
test_image = cv.resize(cv.imread(full_data),  (256,256))
test_image = np.array(test_image).reshape( 1,256, 256, -1)
pred=model.predict(test_image)
pos,val=np.where(pred==1)
print(pred)
skin_data[int(val)]
```

```
Out[80]: 'MALIGNANT'
```

ACCURACY

```
from sklearn.metrics import accuracy_score
accuracy=accuracy_score(test_labels, predictions)
print('Accuracy:',accuracy*100)
```

```
Accuracy: 100.0
```

REFERENCES

[1] Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell 100:57– 70.
[2] Liotta L, Petricoin E. 2000. Molecular profiling of human cancer. Nat. Rev.Genet. 1:48–56.
[3] Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. 2002. Clinical proteomics: translating bench side promise into bedside reality. Nat. Rev. Drug Discov. 1:683–95.
[4] A.W.Kopf, T.G. Salopek, J. Slade, A.A. Marghood, R.S. Bart, Techniques of cutaneous examination for the detection of skin cancer, Cancer Supplement 75 (2) (1994) 684-690.
[5] I. Papamichail, N. Nikolaidis, I. Glava, K. Lentzas, Marmagkiolis, et al. Merkel cell carcinoma of the upper extremity: case report and an update, J. Surg. Oncol. 6 (32) (2008).
[6] M.E. Celebi, H.A. Kingravi, B. Uddin, et al., A methodological approach to the classification of dermoscopy images, Comput. Med. Imaging Graph. 31(2007) 362-373.
[7] Santosh Achakanalli, G.Sadashivappa, Journal publication of ABCD system of melanoma detection, June 2014.

[8] ShefaliLaddha, KomalPaliwal, Journal publication of Skin Cancer detection using medical image processing ,Dept of Computer Engineering & IT, IJARCEE, September 2016.
[9] A. Ahmad, A. B. Mansoor, R. Mumtaz, M. Khan, and S. H. Mirza,"Image processing and classification in diabetic retinopathy: A review,"
[10] Yin, F.F.; Giger, M.L.; Vyborny, C.J.; Schmidt, R.A. Computerized detection of masses in digital mammograms: automated alignment of breast images and its effects on bilateral-substraction technique.