Detection of Cancerous and Non-cancerous Skin by using GLCM Matrix and Neural Network Classifier

Md. Al-Amin
Department of Information and Communication Technology
Mawlana Bhashani Science and Technology University
Bangladesh

Mohammad Badrul Alam Miah
Department of Information and Communication Technology
Mawlana Bhashani Science and Technology University
Bangladesh

Md. Ronju Mia
Department of Information and Communication Technology
Mawlana Bhashani Science and Technology University
Bangladesh

ABSTRACT
Day by day the use of image processing is increasing. Now a days image processing is the part and parcel of medical science. By image processing many types of cancer are easily detected. Skin cancer is one of them. In this paper the proposed method detects two types of skin one is cancerous skin and another is affected but not cancerous skin. Skin cancers are most common cancer in human. Skin cancers are curable cancer after early detection. The system can distinguish cancerous skin and non-cancerous skin based on some values of features. Some value extracted from Grey Level Co-occurrence Matrix (GLCM). GLCM features include Contrast, Correlation, Energy, Entropy and Homogeneity. Besides those MajorAxisLength, MinorAxisLength, Solidity, Equivdiameter, Perimeter, Mean, Standard Deviation, ConvexArea, Area, Euclidean Distance, Manhattan Distance, Minkowski Distance and Hamming Distance. There are several steps for evaluating the process. The first step is preprocessing, in this step the noise is removed by using filter. The filtered image is segmented into gray level and black and white (BW) image. All the features are calculated on black and white (BW) image. The neural network is used to classify the images. It is an easy system rather than the doctor biopsy procedure. The system consumes less time and gets better result than ordinary systems.

Keywords
Skin cancer, GLCM, Cancer detection, Neural Network

1. INTRODUCTION
Skin: Skin is one of the most amazing organs in the human body. Skin protects body from infection. It protects our body from ultraviolet (UV) radiation. It is storehouse of water and fat. Skin has several layers. Among them epidermis and dermis are main layer.

Epidermis: The epidermis is the outermost layer of our skin. Its primary function is to protect our bodies and provide an effective barrier from the outside world. The thickness of the epidermis varies in different types of skin.

Dermis: The dermis is the middle layer of skin. It stands between epidermis layer and hypodermis layer. The dermis is composed of cells, connective tissue, and ground substance and can contain blood vessels, sweat glands, fat and hair follicles. It ranges from 1-4mm in thickness. It is much thicker than epidermis.

- **Sweat glands:** Sweat glands are used to regulate temperature and remove waste by secreting water, sodium salts and nitrogenous waste (such as urea) onto the skin surface.
- **Fat:** Fat is a macronutrients. It is also known as triglycerides. Fats are solids at room temperature.
- **Hair Follicle:** The hair follicle is a skin organ from which hair grows. There are hair follicles all over the skin, without the lips, palms of the hands, and soles of the feet.
- **Connective Tissue:** Connective tissue is one of the four types of biological tissue. It supports connect or separate different types of tissues and organs in the body.
- **Blood Vessels:** The blood vessels are the part of the circulatory system that transports blood throughout the human body. There are three types of blood vessel Arteries-Carry blood away from the heart, Veins-Carry blood to the heart, Capillaries-Found in the muscles and lungs.

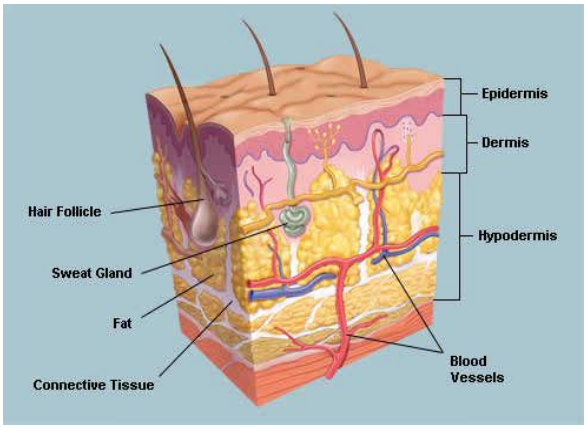

Figure 1: Component of skin

Hypodermis: The hypodermis is the most inner layer of the skin. It invaginates the dermis and is attached to the latter of dermis. It is essentially composed of a type of cells specialized in accumulating and storing fats, known as adipocytes. The hypodermis acts as an energy reserve.

Skin cancer and its type: Skin cancer is abnormal growth of skin cells. There are several types of skin cancer and most

common are melanoma and non-melanoma. Non-melanoma are basal cell skin cancer and squamous cell skin cancer.

- **Melanoma:** Melanoma can occur on the skin. It begins in melanocytes. Melanoma can grow very quickly. It can spread to other parts of the body.
- **Basal cell skin cancer:** Most common skin cancer is basal cell carcinoma. Generally, it is very slow-growing and does not spread to other parts of the body. Because of proper treatment basal cell cancer are completely cured.
- **Squamous cell skin cancer:** Squamous cell skin cancer begins in squamous cell. It may be protected by simple treatment. It also does not spread to other parts of body.

2. RELATED WORK

The correct identification of skin spot based on some certain feature cancer and detects the cancer. This work is focused on extraction of features include contrast, correlation, homogeneity, entropy, radius, standard deviation and perimeter. Back propagation neural network is used as classifier. [1]. Melanoma is the most dangerous skin cancer. Reduction of the cost rather than ordinary system and spend less time to detect cancer. In this frame work, an automated melanoma prescreening system is proposed to diagnose melanoma skin cancer using modified TDLS algorithm and Support Vector Machine (SVM) is used for classification. [2]. Segmentation of skin lesion from the surrounding skin in the dermoscopic images by using Neural Network segmentation algorithm. Different segmentation techniques were applied to the dermoscopic images to segment the skin lesions and evaluated with 3 different metrics, namely sensitivity, accuracy and border error. Segmentation performance shows that Neural Network based lesion segmentation has high sensitivity, accuracy and less border error. [5]. A simple algorithm is used to detect the skin cancer. Skin cancer analysis and detection by Fuzzy C-Means (FCM), Thresholding and Gray Level Co-Occurrence Matrix (GLCM). Feature extraction done by GLCM also calculates the standard deviation. [3]. A technique for early detection skin cancer problem is proposed. The diagnosing methodology use Digital Image Processing Techniques and Artificial Neural Networks for the classification of Malignant Melanoma from other skin diseases. Dermoscopic images were collected and they are processed by various Image processing techniques. The cancerous region is separated from healthy skin by the method of segmentation. [7]. There are several problems like accuracy in the existing works. In this frame work are used to get better accuracy than others.

3. PROPOSED WORK

Figure.2 shows the methodology of this paper which is expanded in the next sections.

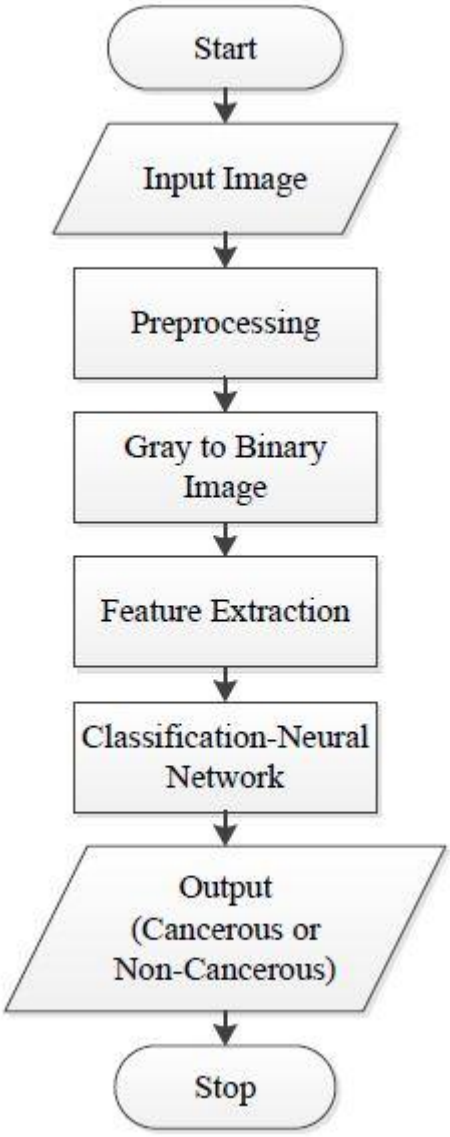

Figure 2: Flow diagram of propossed work

3.1 Preprocessing

Preprocessing is the first step of the frame work. It is important step of image processing. In this step, at first adjusted the intensity of original image for better performance and then adjusted image is filtered by using the 'imfilter' function to remove noise. Finally, the filtered image is converted from RGB image to gray image.

3.2 Gray to binary image

In this step, firstly the gray image is adjusted by using the 'imadjust' function and the adjusted gray image is converted to black and white (BW) image by using 'im2bw' function. Then the affected region is traced that means the white region is traced.

3.3 Feature extraction

Feature extraction is important for the image analysis. By feature extraction many properties of an image can be gained. By the values of feature the system can distinguish the cancerous and non-cancerous skin.

3.3.1 Gray Level Co-occurance Matrix (GLCM)

The GLCM is a powerful tool for image feature extraction by mapping the gray level co-occurrence probabilities based on spatial relations of pixels in different angular directions. The features extracted based on GLCM are: Contrast, Correlation, Energy, and Homogeneity.

- **Contrast:** It returns a measure of the intensity contrast between a pixel and its neighbor over the inter image. Range = [0 (size (GLCM, 1)-1)^2] Contrast is 0 for a constant image.

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j) \quad (1)$$

- **Correlation:** It returns a measure of how correlated a pixel is to its neighbor over the inter image. Range = [-1 1]; If Correlation is 1 positive relation if correlation -1 negative relation if correlation is 0 there is no relation between pixel of image.

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \ \sigma_j} \quad (2)$$

- **Energy:** It returns the sum of squared elements in the GLCM. Range = [0 1] Energy is 1 for a constant image.

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \quad (3)$$

- **Homogeneity:** It returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Range = [0 1]. Homogeneity is 1 for a diagonal GLCM.

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (4)$$

3.3.2 Entropy

Image entropy is the amount of information which must be coded for by a compression algorithm. A perfectly flat image will have zero entropy.

$$\text{Entropy} = -\sum_i p_i (\log_2 p_i) \quad (5)$$

P_i is the probability that the differences between two adjacent pixels are equal to i.

3.3.3 Solidity

It returns a scalar that indicates the proportion of the pixels in the convex hull that are also in the region. It is computed by following.

$$\text{Solidity} = \frac{\text{Area}}{\text{ConvexArea}} \quad (6)$$

3.3.4 Majoraxislength

It returns a scalar that indicates the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

$$\text{Major axis} = a+b \quad (7)$$

Where a, b are the distances from each focus to any point on the ellipse.

3.3.5 Minoraxislength

It returns a scalar that indicates the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

$$\text{Minor axis} = \sqrt{(a+b)^2 + (f)^2} \quad (8)$$

Where f is distance between focus and a, b are the distances from each focus to any point on the ellipse.

3.3.6 Equivdiameter

It returns a scalar that specifies the diameter of a circle with the same area as the region. It is computed as following.

$$\text{ED} = \sqrt{\frac{4*\text{Area}}{\text{pi}}} \quad (9)$$

3.3.7 Perimeter

It returns a scalar that indicates the distance around the boundary of the region. The perimeter calculates the distance between each adjoining pair of pixels around the border of the region. If the image contains discontinuous regions, it returns unexpected results. The formula of distance is given below:

$$\text{Distance} = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} \quad (10)$$

3.3.8 Mean

Mean is the average of sum of all the values in the image matrix.

$$\text{Mean} = \text{sum}(A(i,j))/(r*c) \quad (11)$$

Where A is the area of image and r is row of image matrix and c is the column of image matrix.

3.3.9 *Standard Deviation*

Standard deviation (SD) is measures of how spread out a distribution is. The variance is computed as the average squared deviation of each number from its mean. Standard deviation is the square root of variance.

$$\text{SD} = \sqrt{variance} \quad (12)$$

3.3.10 Convexarea

It returns a scalar that refers the number of pixels in ConvexImage.

3.3.11 Area

Area refers the number of total 'ON' pixel of images.

$$D = \text{nnz}(I(i,j))$$

Where nnz Number of nonzero matrix elements.

3.3.12 Euclidean Distance

The Euclidean distance is commonly used for similarity measurement in image retrieval due to its efficiency. It counts the distance between two vectors of images by computing the square root of the sum of the squared absolute differences. It can be calculated as follows:

$$D_{12} = \sqrt{\sum_{k=1}^{n} (x_{1k} - x_{2k})^2} \quad (13)$$

3.3.13 Manhattan Distance

The Manhattan distance is also known as city block distance. The city block distance metric has robustness to outliers. It is computed by the sum of absolute differences between two feature vectors of images and can be calculated as:

$$D_{12} = \sum_{k=1}^{n} |x_{1k} - x_{2k}| \quad (14)$$

3.3.14 Minkowski Distance

The Minkowski distance is a metric in a normed vector space. It can be considered as a generalization of both the Euclidean distance and the Manhattan distance.

$$D_{12}= \sqrt[p]{\sum_{k=1}^{n} |x_{1k}-x_{2k}|^p} \quad (15)$$

Where p is a variable parameter

When p = 1, is the Manhattan distance

When p = 2, is the Euclidean distance

When p → ∞, the distance is the Chebyshev

For this proposed method p = 3.

3.3.15 Hamming Distance

The Hamming Distance is a number used to denote the difference between two binary strings. Two equal length strings s1 and s2 Hamming distance between the definition will become another one of the minimum needs to be replaced by the number of times.

4. CLASSIFICATION

There are several types of classification algorithm to classify the input skin image. Such as artificial neural network (ANN), support vector machine (SVM), hybrid classifier etc. Among them ANN has been used for the classification. There are various method of ANN like back propagation neural network, feed forward neural network, single layer perceptron and multilayer perceptron (MLP). The feed-forward back-propagation neural network is used for classification. It is like feed forward neural network with back-propagation training algorithm. The back-propagation training algorithm subtracts the training output from the target (desired answer) to obtain the error signal. It then goes BACK to adjust the weights and biases in the input and hidden layers to reduce the error.

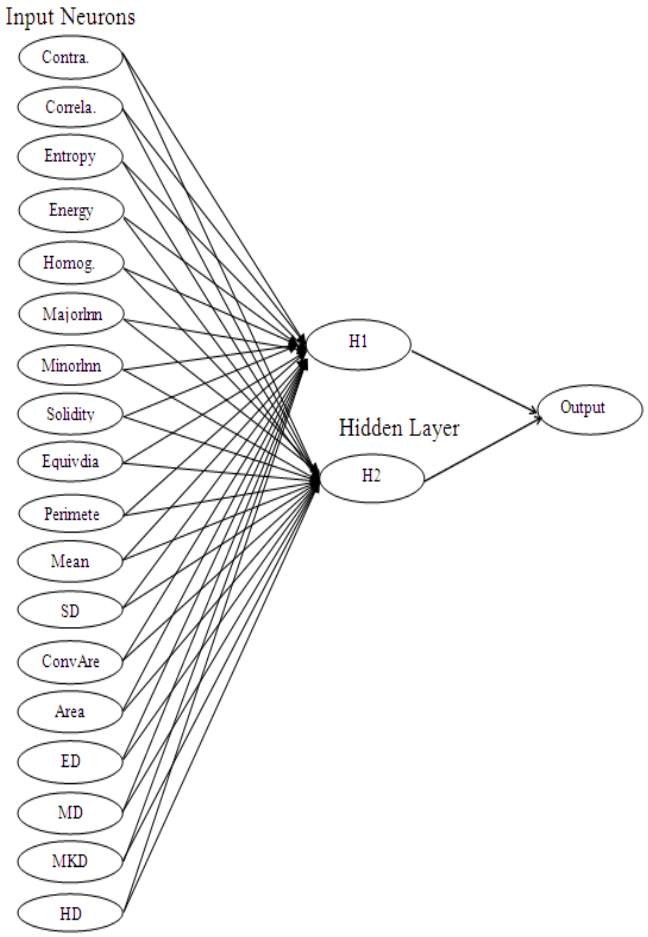

Figure 3: Feed-forward back-propagation neural network

Parameters:

No. of Input Neurons = 18

Training Class = 2

Targets: 0.4 for Class-1: Cancerous Skin

Targets: 0.9 for Class-2: Non-cancerous Skin

5. RESULT AND DESCISSION

The MATHLAB simulation tool is used for this algorithm. Most of all sampled image are collected from National Cancer Institute and American Cancer Society. By using the sample image, the system extracts value of images and detects the cancerous and non-cancerous skin.

Table 1: Feature extracted values of skin images

Feature	Cancer 1	Non-cancer1	Cancer 2	Non-cancer2	Cancer 3	Non-cancer3	Cancer 4	Non-cancer4	Cancer 5	Non-cancer5
Contrast	0.0037	0.012	0.005	0.0112	0.0034	0.0111	0.0072	0.0111	0.0077	0.0076
Correlation	0.9879	0.9696	0.9882	0.9775	0.9931	0.9613	0.9855	0.9613	0.9842	0.9753
Entropy	0.6998	0.837	0.8836	0.9986	0.9965	0.662	0.998	0.662	0.977	0.6968

Energy	0.6886	0.5944	0.5728	0.4896	0.4992	0.7019	0.4944	0.7019	0.508	0.6853
Homogeneity	0.9981	0.994	0.9975	0.9944	0.9983	0.9944	0.9964	0.9944	0.9962	0.9962
MajorAxisLength	152.5791	73.0381	181.9354	103.3157	326.6582	45.6638	237.7496	45.6638	243.5545	65.4218
MinorAxisLength	125.8061	59.4517	131.609	53.3744	145.6085	37.6507	179.8393	37.6507	157.8586	46.711
Solidity	0.9578	0.9588	0.9763	0.9665	0.9688	0.9676	0.9415	0.9676	0.7642	0.986
EquivDiameter	136.5525	65.1242	154.1178	76.0213	216.3136	41.3671	203.5528	41.3671	175.6941	55.233
Perimeter	479.1026	225.1371	538.4996	280.4508	825.9554	141.4386	765.4529	141.4386	959.2346	184.7523
Mean	0.1892	0.2669	0.3019	0.4803	0.5349	0.1719	0.5266	0.1719	0.411	0.1878
SD	0.3917	0.4424	0.4591	0.4996	0.4988	0.3773	0.4993	0.3773	0.492	0.3905
Convex Area	15291	3474	19108	4700	37934	1389	34563	1389	31726	2430
Area	14645	3331	18655	4247	36750	1344	32542	3100	24244	1355
ED	6.1708	4.6012	8.0393	6.3176	10.9219	3.5745	8.379	4.6014	9.5527	2.9555
MD	58.5473	27.7819	78.8661	47.6937	141.6869	18.6011	86.6367	29.333	105.3205	16.746
MKD	2.974	2.583	3.8499	3.3074	4.7489	2.1014	3.9971	2.5433	4.3946	1.6819
HD	0.2269	0.2671	0.2629	0.3785	0.4723	0.2022	0.2888	0.1956	0.4196	0.1675

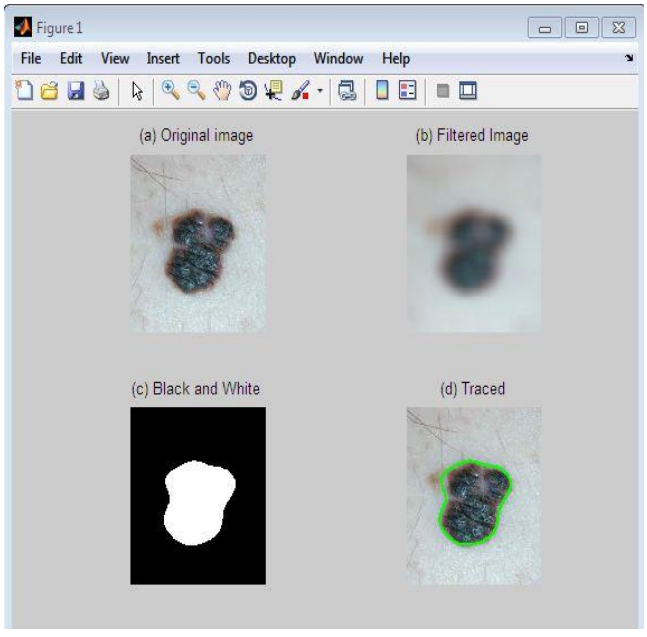

Figure 4: Cancerous image

(a) Original image, (b) Filtered image, (c) Black and White image, (d) Traced image

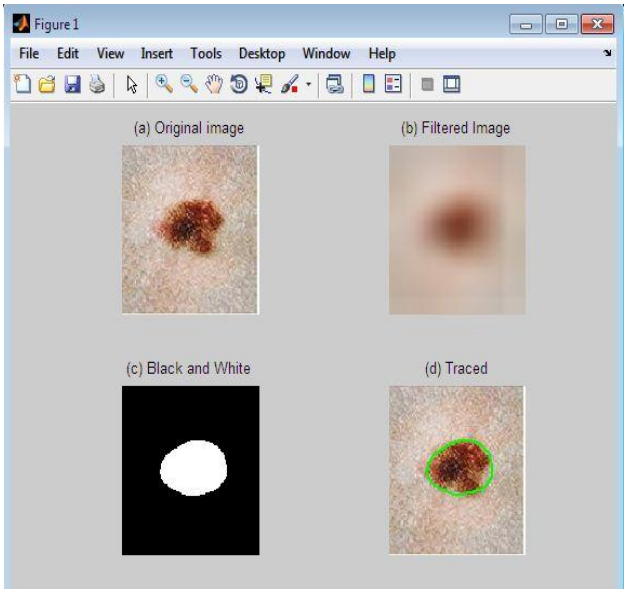

Figure 5: Non-cancerous image

(a) Original image, (b) Filtered image, (c) Black and White image, (d) Traced image.

Table 2: Test with cancerous skin image

Type of Image	No. of Image	Successful Detection	Accuracy (%)
Image group1	10	10	100
Image group2	10	9	90
Image group3	10	10	100
Image group4	10	10	100
Total	40	39	97.50

Table 3: Test with Non- cancerous skin image

Type of Image	No. of Image	Successful Detection	Accuracy (%)
Image group1	10	9	90
Image group2	5	5	100
Image group3	4	4	100
Total	19	18	96.67

The implemented system has tested with many data. Some of the tested histories are given the above tables. In the above tables, most of the experimented results give 100% accuracy it indicates that expected results exactly. On the other hand experimented result approximate 100% indicate slight distance from the expected result.

All the values of cancer and non-cancer images are trained at first in neural network. Then the trained network has been tested with the some test values. 40 and 19 images are used for test. Among them 40 images are cancerous image and in Image group2 one image is not correctly detected and rest all types are correctly detect that shown in Table 2.

Among them 19 images are non-cancerous. In Non-cancerous Image group1, one image is not correctly detected and Non-cancerous Image group2 and Image grup3 is successfully detected that shown in Table 3.

6. CONCLUSION

In this paper, the proposed method is developed for the skin cancer. Skin cancer is most common cancer of human. There are various types of skin cancer some of most dangerous and relatively some are less dangerous. It is total preventable with proper conscious. In the preprocessing images are adjusted to obtain accurate feature extraction. Neural network is to classify processing images. This study shows the combination between co-occurrence matrix and neural network is providing technique for detecting the cancerous and non-cancerous cell of skin.

In future, the future work will be based on developing algorithm to identify the various skin cancers with its type and to improve the overall efficiency and to reduce the computational time. Take more features in future to get more accuracy and other additional steps**.**

7. REFERENCES

[1] Rashi G, Saranjeet S, "Skin Cancer Detection using GLCM Matrix Analysis and Back Propagation Neural Network Classifier" International Journal of Computer Apllication IJCA , Volume 112, 9 February 2015.

[2] I. S. Akila, V. Sumathi, "Detection of Melanoma Skin Cancer using Segmentation and Classification Algorithm " National Conference on Information and Communication Technologies (NCICT 2015)

[3] Sonali J, D. K. Kamat, "Analysis and Detection of Skin Cancer" IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Volume 9, Issue 4, PP 50-54, Jul-Aug 2015.

[4] Mariam A. S, Mia S. M, Amr S "Automatic Detection of Melanoma Skin Cancer using Texture Analysis" International Journal of Computer Application IJCA, Volume 42,-No. 20 9 March 2015.

[5] Rani N, Nalam M and Mohan A, "Detection of Skin Cancer Using Artificial Neural Network" International Journal of Innovations & Advancement in Computer Science IJIACS, Volume 2, Issue 1 January 2014.

[6] Rajam P.J.J, Thasneem H.A.A, "Detection of Skin Lesions in Dermoscopic Images" International Journal of Recent Development in Engineering and Technology, Volume 2, Special Issue 3,pp 193-198, February 2014.

[7] Md. Amran H. B, Ibrahim A, Md. Kamal U "Image Processing for Skin Cancer Features Extraction",. International Journal of Scientific & Engineering Research Volume 4, Issue 2, February-2013

[8] Sarika. C, Seema. B, "Artificial Neural Network for Skin Cancer Detection", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 5, September- October 2014

[9] Sigurdur. S, Peter. A. Philipsen, Lars. K. Hansen, " Detection of skin cancer by classification of Raman spectra", IEEE TRANSACTION ON BIOMEDICAL ENGINEERING

[10] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins "Digital Image Processing Using MATLAB", Second Edition

[11] Ramteke S.N and Shweta V.J, "Analysis of Skin Cancer Using Fuzzy and Wavelet Technique – Review & Proposed New Algorithm," International Journal of Engineering Trends and Technology (IJETT), Volume 4, Issue 6, june 2013

[12] Maurya R, Surya K .S, Maurya K. A and Ajeet, "GLCM and Multi Class Support Vector Machine based Automated Skin Cancer Classification," IEEE journal, vol 12, pp 444-447, 2014.