

Malignancy Detection of Candidate for Basal Cell Carcinoma Using Image Processing and Artificial Neural Network

Armida R. Bayot

Louise Ann M. Samoy

Niño Angelo D. Santiago

Michelle R. Tomelden

Department of Electronics & Communications Engineering
De La Salle University-Manila

and

Edwin Sybingco

Department of Electronics & Communications Engineering
De La Salle University-Manila
Email: sybingcoe@dlsu.edu.ph

Basal Cell Carcinoma (BCC) is a form of skin cancer. Basal cell carcinomas are a result of sun damage to the skin that enlarge slowly and steadily and can invade neighboring tissue. The only way to determine if a skin growth is cancerous is to perform biopsy through removal of a small piece of the skin and having a pathologist look at it under the microscope. This research focuses in the detection of BCC disease by using digital image processing and artificial neural networks. The detection is based on the special microscopic characteristics of basal cell carcinoma such as island formation of multiplying basal cells in the stroma layer, palisading and retraction artifacts that can be seen in low power magnification and the dark ovoid cells that can be observed in high power magnification. The application of the techniques mentioned, the system was able to correctly detect the occurrence of basal cell carcinoma with a percent reliability of 93.33%.

1. INTRODUCTION

Cancer is one of the dreadful diseases that have affected the human race. Up to now, the big C, as what cancer is also popularly known, claims millions of lives and at this point in time nobody have ever found the true and definite cure for it. Most of the time, what was once a speck of cancer cell that eventually causes death was not detected early because of

people's lack of concern to their bodies or maybe because the diagnostic test performed to them was not good or reliable enough. This research aims to help in saving lives by developing a computerized means by which the presence of Basal Cell Carcinoma, the most common type of skin cancer, would be detected. BCC is a result of sun damage to the skin which enlarges slowly and steadily and can invade neighboring tissue as shown in figure 1.

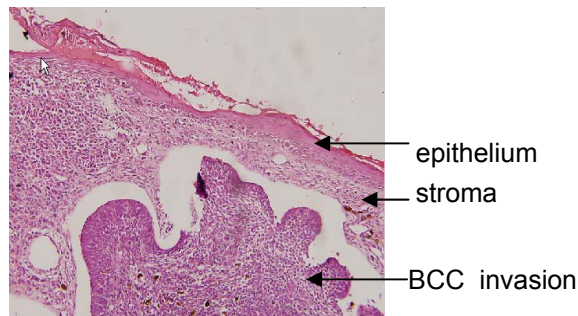


Figure 1. Skin Layer

BCC is best examined under a microscopic view. This process is known as biopsy, where the primary tool is visual inspection. Basal Cell Carcinoma is diagnosed by performing biopsy on suspicious looking lesions. In this procedure, a small part of the infected skin is removed, stained and examined under a microscope [1].

Punch biopsy is considered the primary technique to obtain diagnostic, full-thickness skin specimens. A circular blade or trephine attached to a pencil-like handles is used when performing punch biopsy as shown in figure 2. The instrument is rotated down through the epidermis and dermis, and into the subcutaneous fat. A cylindrical core of tissue is yielded that must be gently handled (usually with a needle) to prevent crushed artifact at the pathologic evaluation. Large punch biopsy sites can be closed with a single suture and generally produce only a minimal scar. Punch biopsy can provide useful information when the differential diagnosis has been narrowed [2].

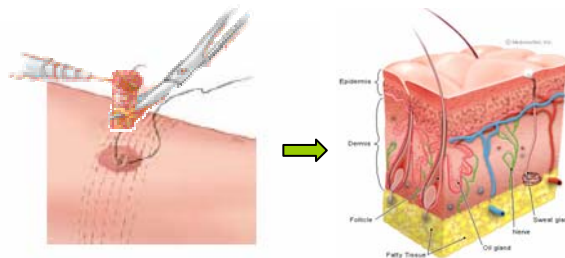


Figure 2. Skin Biopsy

To distinguish skin samples with basal cell carcinoma, BCC arises in the basal cells which are present at the bottom of the epidermis. A person has basal cell carcinoma when the basal cells invade the inner part of the skin. The invasion of basal cells can be distinguished through features that are common among different clinical types. One is the lining of single

layer of basal cells around the periphery with a distinct tendency to palisading. Palisading is the tendency of the cells in the periphery to line up with their nuclei with the long axis perpendicular to the basement membrane and approximately parallel to each other as shown in Figure 3a. Two is the nests of tumor cells are surrounded by stroma. Three a clefting artifact, referred to as retraction artifact is present between the epithelium and the stroma as shown in figure 3b. Another is that the color of the individual cells are more intense than normal and there is little cytoplasm. They are also large, oval or elongated in shape. The nuclei resemble those of basal cells of the epidermis, but basaloma cells differ from basal cells by having a larger ratio of nucleus to cytoplasm and by not showing intercellular bridges.

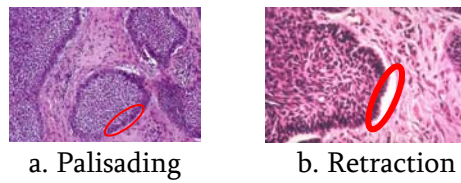


Figure 3. Histology of BCC

These features can be extracted using high and low magnifications. Low magnification was used for the detection of the invasion such as palisading pattern and retraction feature, while high magnification was used for the examination of the color and shape. In this study, digital image processing will be used to effectively extract the features of the skin samples with BCC and non-BCC since digital methods have shown reliable tools for early diagnosis of skin cancer [3]. Biomedical image processing covers biomedical signal gathering, image forming, picture processing, and image display to medical diagnosis based on features extracted from images. Several works done by several researchers used the following image processing techniques: outlining, deblurring, noise cleaning, filtering, search, classical analysis and texture analysis [4]. This study covers some of these techniques which will be discussed in the succeeding sections.

2.0 EXPERIMENTAL SETUP

Shown in figure 4 is the design of the system. The first block is the image acquisition. The pre-diagnosed skin samples are obtained from different hospitals. Afterwards, the skin samples in the slides as shown in figure 5 will be subjected to a light microscope interfaced with a camera, with uniform low and high magnifications. The digital camera serves as the digitizer of the images captured. Image databases were created composed of 150 pairs of images used for the training set and a separate set of 30 pairs of images used as the testing set.

The acquired images are processed as follows: Image enhancement, edge detection, image segmentation and feature extraction. The features extracted are then fed to the artificial Neural Network for training.

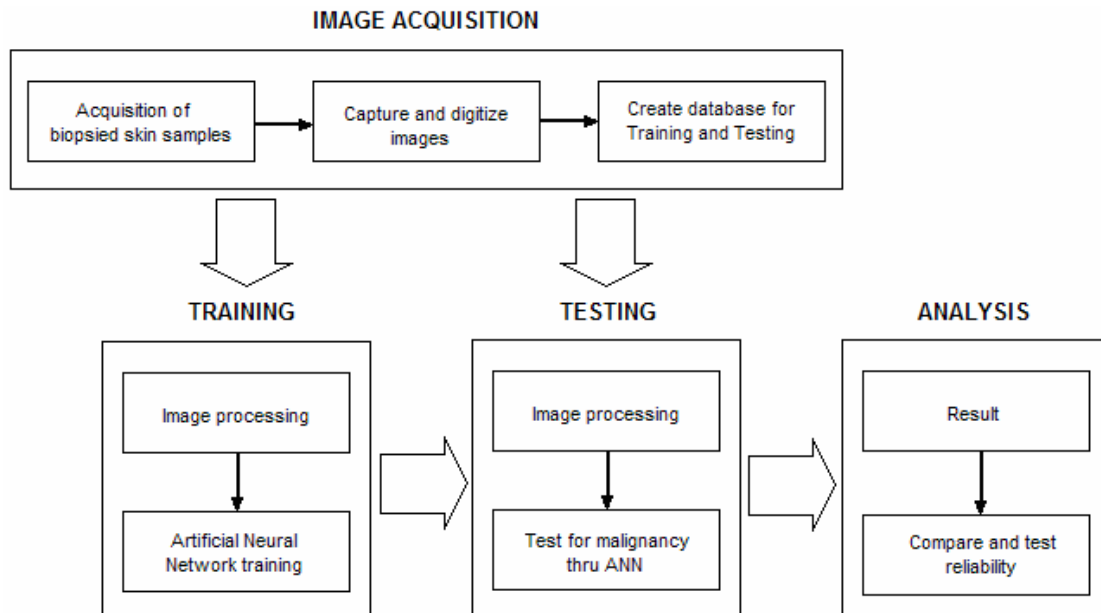


Figure 4. System Block Diagram



Figure 5. Microscopy of slides

The same image processing techniques are applied to the testing set. The image processed test images were applied to ANN. The result of the ANN are compared with the biopsy assessment of the pathologist. The reliability is then computed by getting the ratio of correct results to the total number of test images.

3.0 SYSTEM DESIGN

The samples consists of 30 slides were taken from the Philippine General Hospital and De La Salle University Medical Center. 15 of which were diagnosed as BCC and the other 15 are differential diagnoses of BCC. From the 15 slides that are considered BCC, 75 images were taken that contains BCC invasion. The other 75 images were taken from the 15 slides that are considered non BCC samples.

The BX51 Olympus Microscope was used to acquire images from the sample slides. The BX51 is equipped with 20x and 100x magnification enough to view the necessary

characteristics of the acquired images for the detection of BCC. The images are saved as JPEG files with a uniform resolution of 680 x 512.

The complete histological characteristics of BCC are characterized using low magnification and high magnification. In the low magnification images were resized from its original image for faster processing. Histogram equalization is applied to the resized image resulting to a better contrast to identify the lighter pixels from a darker pixel as shown in figure 6.

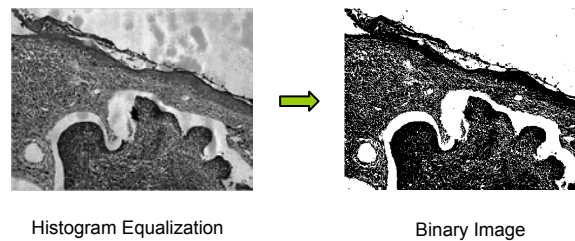


Figure 6. White Background Identification

The technique to separate the background from foreground in gray scale images based on the object area was very effective[5]. After identifying the background and the area of interest as shown in the binary image, grids will be now made to represent each neighboring pixels in a larger scale for each grid the color of each pixel are average as shown in figure 7. The resulting average-colored grid image was converted to grayscale. Only the luminance is considered since based from the images sampled, the cancer cell island formation has a darker color compared to its background or stroma. And since pathologist and hospitals have different staining, relying on luminance would be a safe way to eliminate the “staining problem”, and will do in order to detect each cancer formation. Based from the converted luminance of each gray pixel, average threshold for the whole image was obtained in order to detect the bcc formation and eliminate the stroma. Each gray grid pixels are examined on how they relate on each other, if the value of their luminance is near or completely contrasting. Lighter pixels usually represent unwanted areas and darker pixels were retained, which represents the bcc island formation. Through this, based from the average of the gray grids, lighter grids will be eliminated, and the formation of each bcc will remain. After selecting islands, it will then undergo checking if it has the same luminance value from other islands and still check if it is near the luminance of the background.

After undergoing these processes, feature extraction such as palisading and retraction are evaluated. Shown in figure 7 is an image which represents palisading, which is found on the island’s outline, and retraction, which is a gap formed near the bcc island formation. These features are essential for the preliminary investigation for bcc detection.

Palisading value was obtained through the number of pixels present in the outline compared to the “actual” outline of the image. Retraction on the other hand was measured through its area, as present near to the bcc island formation.

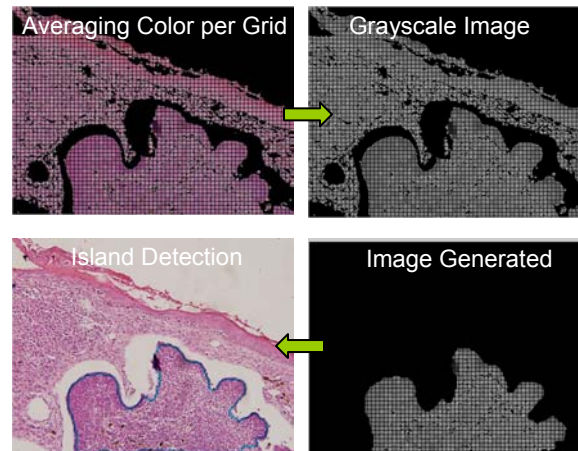


Figure 7. Region of Interest Extraction

Now after low magnification detection, high-magnification detection is used to verify if the selected island is really bcc or not. This is done to double check and to really examine each cell. In detecting the cells in high magnification, preliminary image processing was done first as just as done in low magnification. White background and area to be sampled for threshold should be first identified. Based from the binary image, darker areas are the area to be sampled for threshold as shown in figure 8.

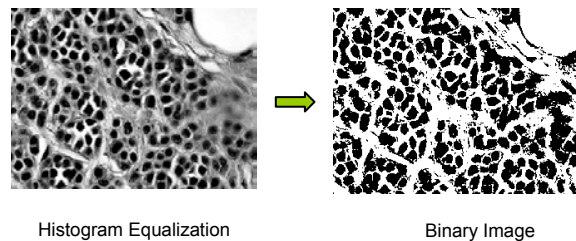


Figure 8. White Background Identification

Based from the binary image shown, color per grid was also averaged, as what have done in low magnification, and it was then converted to grayscale for averaging the luminance per grid. The derived average grayscale is used as a basis for threshold in converting the image into black and white. Detected cells based from the threshold were then subjected to checking if it has the same luminance compared to others through getting their average grayscale value. Noise and specs were eliminated to give the binary image a cleaner look as shown in figure 9.

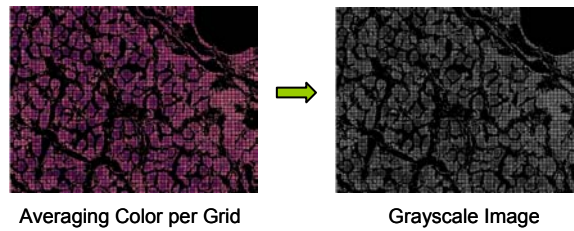


Figure 9. Region of Interest

The binary image is then subjected to feature extraction namely: eccentricity and solidity that determines whether they are really bcc or not. As what have mentioned earlier, eccentricity was a measure how cells are elongated, while solidity measures how pixel filled the area of cells.

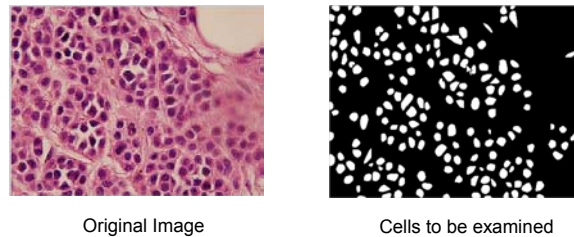


Figure 10. Feature Extraction

After detecting features for both low and high magnification, palisading, retraction, eccentricity and solidity were fed to neural networks. Neural networks have been used by [6] to identify benign or malignant signal characteristics.

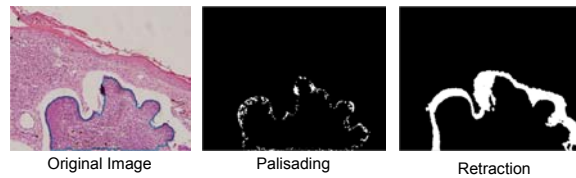
The quantified features were computed using image processing and the array of these quantitative values were inputted to the artificial neural network. The artificial neural network used was a feed forward back propagation type of network with a 4x7x1 structure. The 4 corresponds to the number of features inputted at the ANN namely the palisading, retraction, solidity and eccentricity. The 7 corresponds to the number of hidden nodes and the 1 denotes the final output wherein 1 means BCC and a 0 denote non-BCC.

4.0 DATA AND RESULT

After extracting the features or characteristics of each image, these are then translated into quantitative values. The quantitative values of these features are based on the following parameters of the images: area, major axis length, minor axis length, solidity, bounding box, centroid, orientation and eccentricity. These quantitative values are inputted to the ANN as the training set. Shown in figure 11 is an example of an image whose features in low power magnification are extracted. The first indication of BCC is the presence of island formation. If there is any island found, its palisading characteristic and the retraction artifact around it are quantified. The valued for palisading and retraction artifact are their computed areas. On the other hand, if there isn't any island formation detected, the values for palisading and

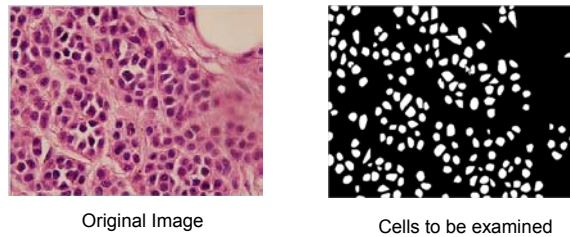
retraction artifact are automatically set to zero. Shown in figure 11 the palisading is the area where the color in the periphery of the island is darker due to the lining up of the basal cells. The retraction artifact is the noticeable white area around the island that takes its shape due to continuous pushing.

For the high power magnification as shown in figure 12, the characteristic of the cells are examined in detail. The cells in BCC are characterized as dark and ovoid, the solidity and eccentricity are thus analyzed. The % eccentricity tells the percentage of the cells examined that met the threshold value over the total number of cells. On the other hand, % solidity tells the percentage of the cells examined that met the threshold value over the total number of cells. As can be noticed in table 1, BCC samples possess high values than those of non-BCC samples. While in some cases that the non-BCC samples may also have a comparably high value in one feature, the other features offset this by their very low values. For example, image 86 has a very high % solidity, but its % eccentricity is very low as well as its palisading and retraction values. For image 100, both its % eccentricity and % solidity are somewhat high; however its palisading and retraction have low values. After getting all these features in low and high power magnification, these four features are simultaneously fed and analyzed by the ANN. The ANN is responsible for the detection of BCC and non BCC based on the extracted features.



Presence of island formation:	Yes
	58.3193
Retraction artifact:	171.04

Figure 11. Low Magnification



% Eccentricity:	0.2772
	0.9837

Figure 12. High Magnification

Table 1. Training Set Result

Training Sample	Palisading	Retraction	Eccentricity (ovoidness)	Solidity (darkness)
<i>Basal Cell Carcinoma</i>				
1	42.5661	332.140	0.7276	0.8889
4	42.7701	16.5600	0.8384	0.7475
7	24.7321	96.8800	0.7570	0.8500
11	37.0832	13.1900	0.6402	0.6296
27	22.0603	15.5300	0.7059	0.6928
<i>Non-Basal Cell Carcinoma</i>				
76	0	0	0.2772	0.9837
86	0	0	0.4722	0.9444
88	19.9002	42.9600	0.6942	0.4515
89	16.6201	1.6800	0.5789	0.7632
100	3.8697	0.0001	0.5000	0.8333

To test the reliability of the technique, 30 pre-biopsied samples – 15 BCC and 15 non-BCC – whose features have been extracted using image processing are fed into the trained artificial neural network. The ANN was able to correctly detected 14 out of 15 BCC (malignant) and 14 out of 15 non BCC resulting to 93.33% reliability.

5.0 CONCLUSION

The designed processed was able to accomplish a percent reliability of 93.33% based on the samples gathered from the two hospitals mentioned. The technique was developed from the consultation with the pathologist and medical practitioners and it was proven that a certain pattern can be acquired from the histological characteristics of BCC. Comparing against the available non-BCC samples used for this research, it can be said that the pattern is only inherent to BCC. The qualitative evaluation of a pathologist was successfully translated into quantitative values through the use of image processing, which was communicated to the computer trained to recognize the pattern inherent to BCC. Using the proper threshold values, the algorithm was able to acquire the numerical values of different features such as invasion, palisading characteristics of the cells, retraction artifact, and the shape and solidity of the individual cells. A supervised neural network has been proven to be a fast and reliable tool for identification although the learning process could take quite a number of epochs in order to arrive with an acceptable value of accuracy. The use of grayscale images in

manipulating the data images has been proven to be sufficient for the recognition of necessary patterns since the features worked upon are color-independent. Due to confidentiality of samples, a limited number of slides were used in study and for this reason the authors recommends further evaluation of the technique by considering more samples.

6.0 REFERENCES

- [1] University Health Systems of Eastern Carolina [Online]. Available: <http://www.uhseast.com/132773.cfm>
- [2] American Academy of Family Physicians [Online]. Available: <http://www.aafp.org/afp/20020315/1155.html>
- [3] Green, A., Martin, N., Pfitzner, J., O'Rourke, M., Knight, N., 1994. "Computer image analysis in the diagnosis of melanoma", *J. Am. Acad. Dermatol.*, vol. 31, pp. 958-964, 1994.
- [4] Huang, H.K. "Biomedical image processing", *CRC Crit. Rev. Bioeng.*, vol. 5 no. 3, pp. 185-271, 1981.
- [5] Kazhiyur-Mannar R., Smiraglia, D.J., Plass, C., Wenger, R., "Contour Area Filtering of two-dimensional electrophoresis images", *Medical Image Analysis.*, vol. 10, pp. 353-365, 2006.
- [6] Nattkemper, T.W., "Tumor feature visualization with unsupervised learning", *Medical Image Analysis.*, vol. 9, pp. 344-351, 2005.

ABOUT THE AUTHORS

Armida R. Bayot, Louise Ann M. Samoy, Niño Angelo D. Santiago and Michelle R. Tomelden earned their BS-ECE degrees from De La Salle University-Manila in October 2005.



Edwin Sybingco is an assistant professor of the ECE Department of De La Salle University. His research interests include application of Digital Signal Processing in communication, multimedia and biomedical applications.