Automatically Early Detection of Skin Cancer:

Study Based on Neural Network Classification

Ho Tak Lau
School of Electrical, Mechanical and Mechatronic Systems, Faculty of Engineering, University of Technology, Sydney, Australia
Ho.T.Lau@student.uts.edu.au

Adel Al-Jumaily
School of Electrical, Mechanical and Mechatronic Systems, Faculty of Engineering, University of Technology, Sydney, Australia
adel@eng.uts.edu.au

***Abstract*— In this paper, an automatically skin cancer classification system is developed and the relationship of skin cancer image across different type of neural network are studied with different types of preprocessing.. The collected images are feed into the system, and across different image processing procedure to enhance the image properties. Then the normal skin is removed from the skin affected area and the cancer cell is left in the image. Useful information can be extracted from these images and pass to the classification system for training and testing. Recognition accuracy of the 3-layers back-propagation neural network classifier is 89.9% and auto-associative neural network is 80.8% in the image database that include dermoscopy photo and digital photo**

Keywords-Skin cancer; classification; neural network; computer based detaction

I. INTRODUCTION

Skin cancer is very common in Australia. According to Cancer Council Australia on survey conducted in 2005, more than 221,000 people are infecting with cancer. There were 34,227 new cases and the death rate was around 12,513 in New South Wales.

The skin cancer malignant melanoma can be recovered if diagnosed and treat it in early stages. Therefore early diagnosing is a crucial issue for patients. However, only experienced doctor is able to classify the skin cancer from other skin diseases. Thus the computer based skin cancer detection is necessary to provide recommendation for non-specialized user. The development of this diagnosis system over 20 years, the accuracy of diagnosis is around 73% to 98%. The variations of diagnosis are sufficiency large and there are lacks of detail of the test methods.

Previously, part of our work published in [1] investigated one type of skin cancer detection based on machine learning. In this paper, a developing an automatically skin cancer detection system to classify the cancer images into either benign or malignant melanoma is presented and discussed through using different approaches. The images in the databases used are contained both digital photo and dermoscopy images. The image databases are collected from Sydney Melanoma Diagnostic Centre in Royal Prince Alfred Hospital and internet website. Total 448 images are collected and separate into four groups to test the accuracy of each group. The grouping policy is determined by the outlook of the image and their properties. Group A contains 93 digital malignant images from internet; group B contains 142 dersmocopy malignant images from Sydney Melanoma Diagnostic Centre; group C and D contains 121 and 92 digital benign melanoma image from internet.

Dermoscopy, also calls Dermatoscopy or Epiluminescence light microscopy (ELM). It was first announced on 1987, it is a kind of imaging technique uses to exanimate lesions with a dermatoscope. The process is done by placing an oil immersion between the skin and the optics. The lighting can magnify the skin that improve on reveal most of the pigmented structure, different color shades that is not visible to naked eye; and allows direct viewing and analysis of the epidermis (the outer layer of the skin) and papillary dermis (the deep vascular inner layer of the skin). Physician uses this technique for diagnosis of skin cancer more efficiency. The historical researches have proved that ELM can improve the diagnostic accuracy by 5 – 30% compare to traditional imaging [2]. Furthermore, ELM has evolution and digitalize that can be used to classification with computer. It given several advantages in diagnosing like considering small suspicious lesion and objective evaluation of parameters: geometry, color and texture; and storage of image and comparable for future development [3]. ELM may improve the accuracy of clinical diagnosing but it requires the experienced dermatology physician to exanimate the image. On Other hand, many research for classification system has been developing over 20 years. Different approach has been announced to improve the result. But most of the published articles usually have not enough information of the implementation. Thus, this paper is interest on compare different approach with their performance and accuracy.

This paper is organized as follows: scheme of classification system is presented in section 2. The different image processing algorithms are explained in section 3. The segmentation methods have clarified in section 4. The feature extraction approach is provided in section 5. Classification results are presented in section 6; and finally the conclusion is made in section 7.

II. AUTOMATICALLY SKIN CANCER DETECTION SYSTEM

The first step in the paper is to establish a general idea and structure of typical image classification system for skin cancer detection. Most of the systems consist of following procedures: image pre-processing to removes the noise and fine hair; Post-processing: to enhance the shape of image; Segmentation: to removes the healthy skin from the image and find the region of interest, usually the cancer cell, remains in the image; Feature extraction extracts the useful information or image properties from the segmented images; Finally, this information is used in the classification system for training and testing purpose. The classification system is usually supported by intelligent classifier, such as neural network and support vector machine. The element of all these ideas will extend in next section.

III. IMAGE PROCESSING ALGORITHMS

The skin cancer images usually contain fine hairs, noise and air bubbles. These feature that is not part of the cancer cell and would reduce the accuracy of the border detection or segmentation. In order to overcome these problems, the first step to do is apply some image processing techniques to the images. In this paper, pre-processing used to referring to remove the unwanted features on the skin and post-processing referring to enhancement to the shape of image. The available methods such as Karhunen-Loève (KL) transform histogram equalization and different kinds of filter are used to achieve these goals. In addition, contrast enhancement can sharpen the image border and improve the accuracy for segmentation.

Since the image database consists of both digital photo and dersmocopy. These images are obtained from different source and the size of the images is non-standard. The first step is to resize the image to have a fixed width 360 pixels but variable size of height. The second step is to remove the background noise from the pictures. The method used here is wavelet de-noise by two-dimensional bior3.3 wavelet. Biorthogonal (bior) is a linear wavelet which advanced used in image reconstruction and decomposition.

Compare Fig. 1 (original image) with Fig. 2 (Image de-noised by wavelet transforms). As can be noticed, the result image given a "blur" image and the detail of the image can be retained.

Next step is image smoothing by median filter. Median filter can be used to preserve edge, remove the noise produced by the image capturing and remove the fine hairs. The size of filter windows is calculated by the method from [4]. The equation refers to a typical size with 768x512 pixels. M and N refer to the dimensions of resized image.

$$n = floor\left(x \sqrt{\left(\frac{M}{768}\right) x \left(\frac{N}{512}\right)}\right)$$

The fine hair is able to mask out by median filter as shown in Fig. 3.

Histogram equalization (HE) algorithm is a smart contrast enhancement technique [5]. The histogram equalization can classify into Global Histogram Equalization (GHE) and Local Histogram Equalization (LHE). GHE's transformation function C(rk) uses the histogram information of the whole input image. The disadvantage of GHE is it

Figure 1. Original image

Figure 2. Image de-noised by wavelet transforms

omits the brightness features of the image; Thus, the gray level with high frequency will nominate the low frequency one. LHE is able to resolve the problem of GHE but also increase the noise. LHS employs small window that scans though every pixel of the image. Only the blocks of pixels that lie between these windows are used for the histogram. Then enhancement by gray level mapping only apply to the centre pixel of that window.

Fig. 4 shows the result of segmentation by threshold with GHE and LHE. GHE (top) has increased the dark background color. LHE (bottom) sharpen the lesion but also reduce the surrounding detail.

IV. AREA OF INTEREST SEGMENTATION

Region of interest (ROI) is extracted by segmentation. The interesting features of melanoma are included within the border since most of the cancer cells are nodule structure. The border structure provides vital information for accurate diagnosis. Many clinical features including asymmetry and border irregularity are calculated from the border. In this paper, threshold and statistical region merging (SRM) are implemented and compare their accuracy with neural network classifier.

The threshold values are calculated by histogram of RGB color band (Red, Green and Blue) of the images. Fig. 5 shows an ideal histogram [6]. The skin and lesions should belong to the two peak values in the histogram since most of the color is belonging to the skin and the lesion. The peak value of skin (Point S) should higher than the lesion (Point M) because skin has lighter color than the lesion. Threshold-low-point (Point V) is located at the lowest point between M and S. The final threshold value T is located between V and S and determined by analysis the pixel distribution of all intensity value. If the pixel in the image is higher than the threshold, then this pixel is set to be 1, else 0. Fig. 6 shows the result of threshold segmentation where the white color represents the area of interest.

Statistical region merging (SRM) algorithm is based on region growing and merging. The performance of SRM has a better ROI segmentation result compare to other popular segmentation methods [4]. The theory of region merging starts at a seed point and compare its four neighbor points or pixels. The region is grown when the neighbors pixel share the same homogenous properties. SRM usually work with a statistical test to decide the merging of regions. Fig. 7 shows the result of SRM segmentation where white color represents the area of interest.

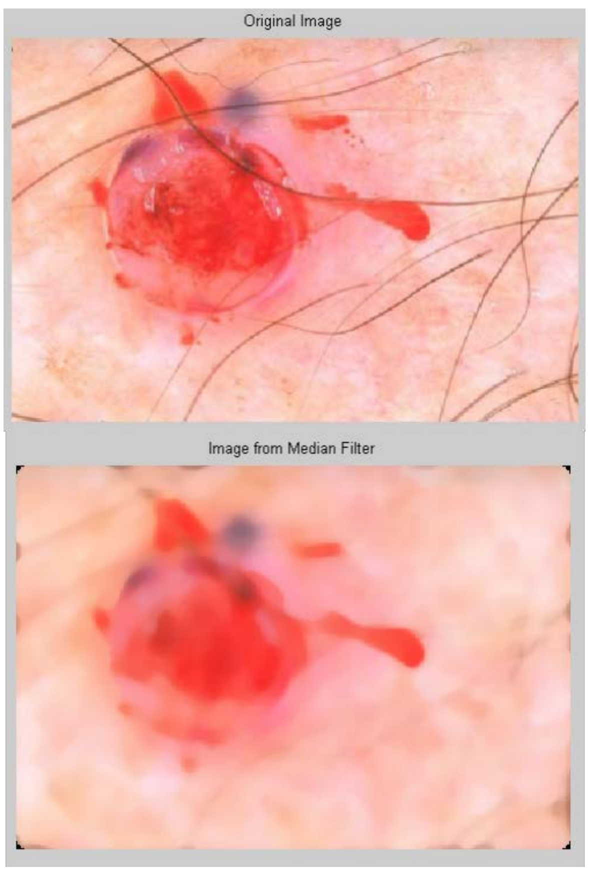

Figure 3. Fig. 3. Original Image (Top) and image after median filtering (Bottom)

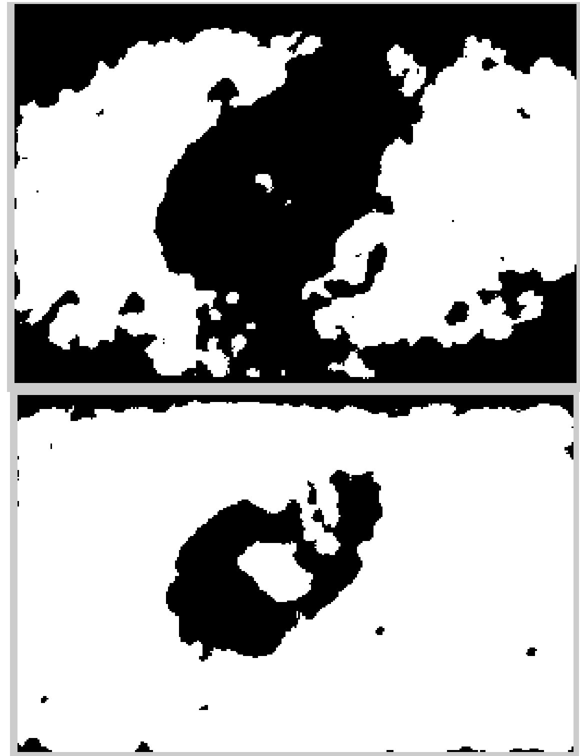

Figure 4. Fig. 4. Result of segmentation by thresholding with GHE (top) and LHE (buttom)

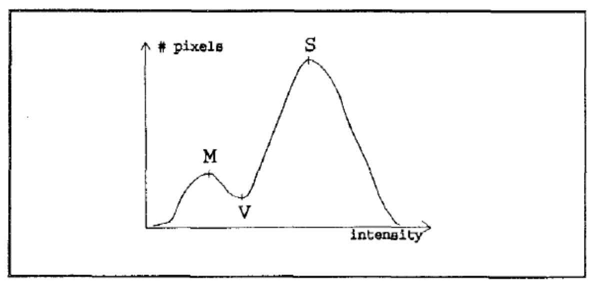

Figure 5. An ideal histogram

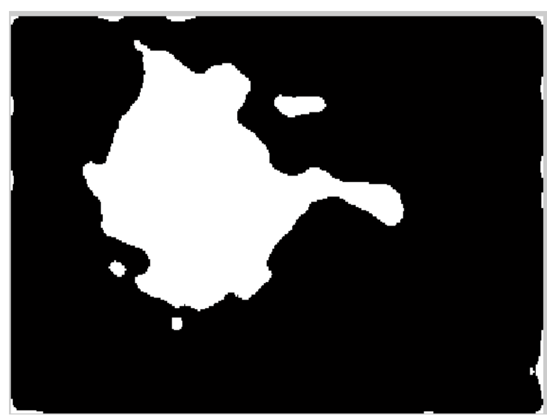

Figure 6. Segmentation by Thresholding

Compare two segmentation method, SRM has a better and more accuracy in segmentation. It includes more useful area and does not put the outer region as part of the area of interest. However, the accuracy for of neural network as classifier could not support that, since by using Back-propagation neural network (BNN), it found that the threshold has 0.5% higher accuracy than SRM in the classification test as illustrate in table 1.

TABLE I. BNN CLASSIFICATION TEST FOR SRM & THRESHOLDING

Method	Training	Validation	Testing	Total
Threshold	98.3%	48.9%	**73.1%**	85.6%
SRM	97%	46.6%	**72.6%**	84.7%

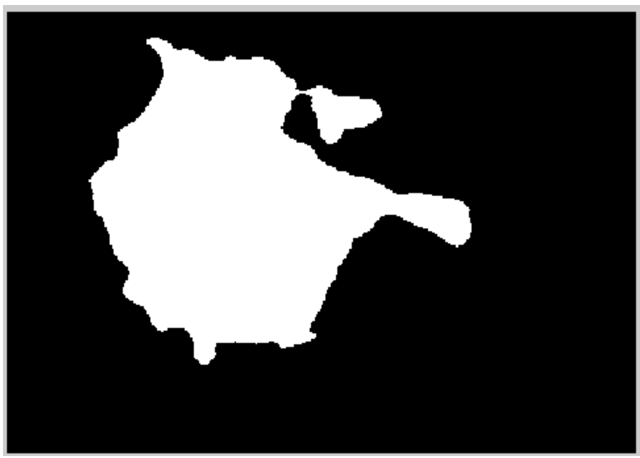

Figure 7. Segmentation by SRM

V. FEATURE EXTRACTION

Wavelet decomposition allows useful cancer's feature to be extracted from the images without clinical knowledge. In this system, 2-D wavelet packet is used and the enhanced image in gray scaled as an input.

Assume a digital image sized M x N pixels is transformed by the discrete wavelet as shown in Fig. 8 which produced by the level decomposition map that Fig. 9 indicates, The result of the decomposition L and H stand for low and high frequency components. FL and FH represent low-pass and high-pass filters. Perform discrete wavelet transform to the image. LL(0) is the original image. LH(1), HL(1) and HH(1) are the output of high-pass filter that's represent the horizontal details, vertical details and diagnosing details. LL(1) represents the approximation with the same size of LH(1), HL(1) and HH(1) that's use to perform the second-level decomposition. The images LH(2), HL(2) and HH(2) have finer detail than in LH(1), HL(1) and HH(1). Moreover, the image energy is distributed according to the resolution. Each of these nodes is represent one feature, which then can be used as an input for classification stage. The second level decomposition can generate 16 nodes or features. Two-dimensional wavelet packet returns the coefficients in 2 dimensions matrix. The features are calculated by their mean, maximum, minimum, median, standard deviation and variance. Therefore, 144 features (16 nodes x 9 features) are produced.

Since there is lack of published articles to compare different wavelet, seven different wavelets are used in this paper followed by same classification test to compare them and find out the best results. The 'bior5.5' wavelet found it provided the highest accuracy that achieved 88.5%. On the other hand, the daubechies wavelet ('db1' and 'db10') has the most stable experimental record during the testing.

It proved that Biorthogonal is advance for the image reconstruction and decomposition. The results of comparing different types of wavelets flowed by BNN as a classifier displayed in table 2

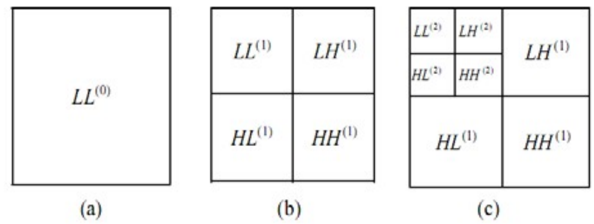

Figure 8. (a) Original image; (b) First-level decomposition; (c) Second-level decomposition

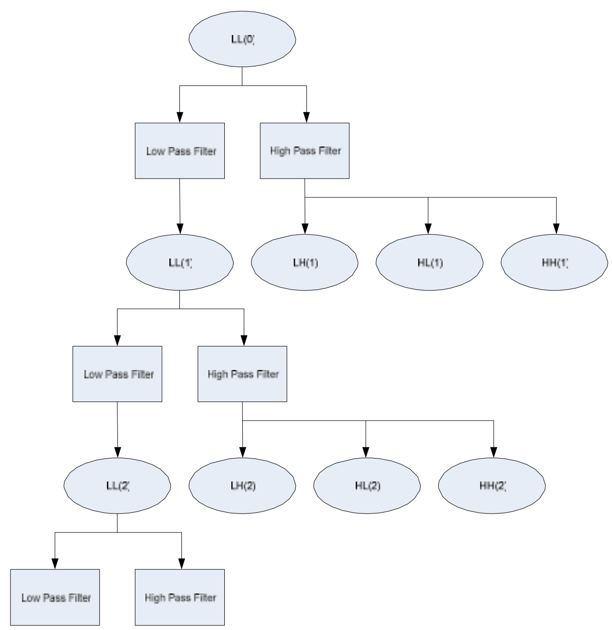

Figure 9. Two level decomposition by two-dimensional wavelet

TABLE II. BNN CLASSIFICATION TEST WITH DIFFERENT WAVELET

Wavelet	Training	Validation	Testing	Total
Db1	98.3%	48.9%	73.1%	85.6%
Db10	100%	48.1%	73.6%	86.9%
Bior1.3	100%	51.1%	68.4%	85.7%
Bior5.5	100%	53.6%	77.4%	88.5%
Coif3	100%	57.1%	69.1%	86.5%
Sym4	99%	44.4%	68.7%	84.5%
Sym7	73.7%	40.8%	53.5%	64.4%

VI. CLASSIFIER RESULTS

Wide ranges of classifiers are available and each one of them has its strengths and weaknesses. Classifier performance depends greatly on the characteristics of the data to be classified and there is no single classifier that works best on all given problems. Various empirical tests have been performed to compare different classifier performance and to find out the relationship between characteristics of data and the classifier performance. Determining a suitable classifier for a given problem is however still more an art than a science. In this paper, two neural networks are used as classifier, Back-propagation neural network (BNN) and Auto-associative neural network (AANN). Fig. 10 shows a typical feed-forward neural network.

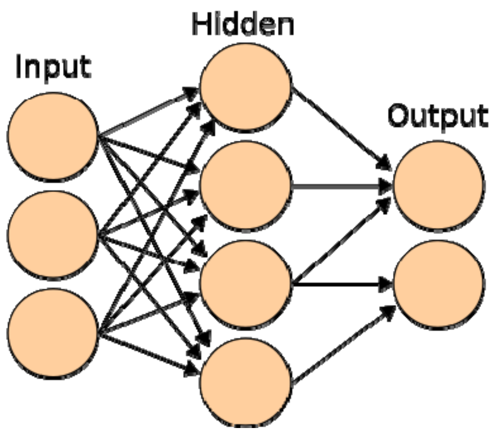

Figure 10. A typical feed-forward neural network

A. *Back-Propagation Neural Network (BNN)*

Back-propagation neural network is one of the most common neural network structures, as it is simple and effective. Back-propagation is the generalization of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions [7]. The hidden and output layer nodes adjust the weights value depend on the error in classification. The modification of the weights is according to the gradient of the error curve, which points in the direction to the local minimum. BNN is benefit on prediction and classification but the processing speed is slower compared to other learning algorithms.

From the results in table 3, the best result with highest overall accuracy is 89.9%. The best BNN is three hidden layer with 40, 25 and 10 neurons for each hidden layer. The accuracy is increase with number of neuron in hidden layer. However, number of hidden layer cannot improve the result but it could reduce the probability of over-fitting.

B. *Auto-Associative Neural Network*

Auto-associative neural network is a network made of equal size of the input layer and the output layer, and smaller

TABLE III. BNN CLASSIFICATION RESULT WITH DIFFERENT LAYERS

No of Layer	No of Neuron	Training	Testing	Validation	Total
1	10	82.6%	55.6%	63.9%	74.3%
1	20	99.3%	60.7%	66.7%	85.7%
1	30	100.0%	52.6%	77.4%	88.5%
1	40	98.8%	50.4%	77.6%	87.6%
2	10, 5	79.4%	47.4%	56.7%	69.4%
2	20, 10	98.1%	51.9%	74.4%	86.4%
2	30, 20	99.8%	55.6%	69.7%	86.3%
2	40, 20	99.9%	51.8%	77.4%	88.3%
3	10, 8, 6	95%	60.8%	69.9%	84.1%
3	20, 12, 8	97.6%	62.9%	76.6%	87.8%
3	30, 20, 10	99.1%	62.2%	71.4%	87.1%
3	40, 25, 10	99.4%	61.5%	80.3%	89.9%

size in hidden layer. The structure of AANN is shown on Fig. 11. The input layer and mapping layer form a compression network and the De-mapping and output layer act as a decoder network. Unlike the others neural network, the desired output of AANN is identity from the input.

The best AANN testing result found is 20 neurons in the first and third layer with overall accuracy 80.8% as table 4 illustrated. Unlike BNN, ANN provides a stable classification result in different number of neuron. However, when the layer 1 and layer 3 have different size of neuron, the classifier result has a significant low accuracy diagnosing result.

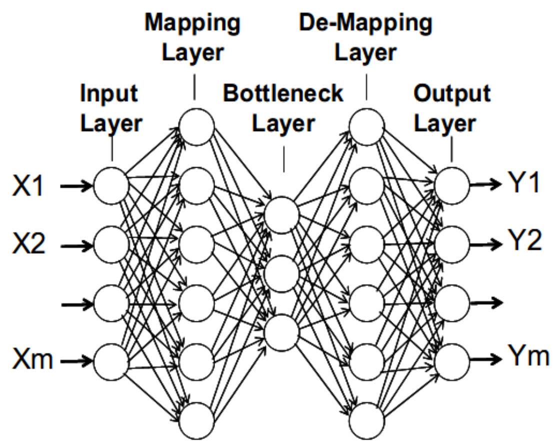

Figure 11. Structure of AANN

TABLE IV. TABLE OF AANN CLASSIFICATION RESULT WITH SIZE OF NEURONS

Layer 1 to 4				Training	Validation	Testing	Total
10	4	10	4	86.9%	59.3%	69.3%	77.8%
10	5	10	4	83.1%	56.3%	68.9%	73.5%
20	4	20	4	89.4%	58.5%	68.9%	79.1%
20	10	20	10	91.4%	61.5%	68.5%	80.8%
30	4	30	4	89.9%	59.3%	73%	80.4%
30	10	30	4	88.4%	55.2%	64.1%	76.9%
40	4	40	4	89.2	63.3%	68.5%	79.9%
40	20	40	4	89.8%	54.1%	63.3%	77.3%
40	10	30	4	41.4%	39.7%	39.3%	40.3%

VII. CONCLUSION

The rate of patient with skin cancer will be increasing if pollution still damaging the ozone layer. The risk of ultra violet light is a hidden damage to our body skin. It is hard to prevent and the effect can be accumulated. Early detection is important to the patients of the skin cancer.

Building an image classification system for early skin cancer detection include different stages, pre and post processing, feature extraction and classifier. The pre-processing resizes the image that improves the speed performance and removes the superfluous feature such as the noise and fine hair. Post-processing enhances the image quality and sharpens the outline of the cancer cell. Feature extraction decomposes the useful feature without prior clinical knowledge. The classifier uses neural network proved advance on predict new image.

The paper presented a study which can be concluded that there are some possible factors of low classification result. The image database is not feasible and too small; the variation between dermoscopy and digital image is large. Since dermoscopy and digital image are used both in testing. The imaging processing methods are not unique and their variation is large. The result has shown clearly in the group A. Group A has a relatively low classification result to group B, C and D. A possible improvement and generalization is to deal with the effect of large variation and overcome of it by using larger database.

However for our study with both types of image and for two types of skin cancer, overall result is 89.9% for back-propagation neural network and 80.8% for auto-associative neural network.

.

REFERENCES

[1] A. Chiem, A. Al-Jumaily, R. N. Khushaba 2007, A Novel Hybrid System for Skin Lesion Detection, *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference*, pp. 567 - 572

[2] P. Rubergni, G. Cevenini, M. Burroni, R. Perotti, G. Dell Eva, P. Sbano, C. Miracco, P. Luzi, P. Tosi, P. Barbini, L. Andreassi 2002, Automated diagnosis of pigmented skin lesions, *International Journal of Cancer*, Volume 101, Issue 6, pp. 576-580

[3] U. Wollina, M. Burroni, R. Torricelli, S. Gilardi, G. Dell'Eva, C. Helm, W. Bardey 2007, Digital dermoscopy in clinical practice: a three-centre analysis, *Skin Research and Technology Journal*, Volume 13, pp. 133-142

[4] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss 2007, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, S. W. Menzies 2008, Border detection in dermoscopy images using statistical region merging, Skin Research and Technology 14, pp. 347-353

[5] M. Abdullah-Al-Wadud, Md. H. Kabir, M. A. A. Dewan, O. Chae 2007, A Dynamic Histogram Equalization for Image Contrast Enhancement, *IEEE Transactions on Consumer Electronics*, Vol. 53 Issue 2, pp. 593-600

[6] T. Lee, V. Ng, D. McLean, A. Coldman, R. Gallagher, J. Sale 1995, A Multi-Stage Segmentation Method for Images of Skin Lesions, *IEEE Pacific Rim Conference on Communication, Computer and Signal Processing*, pp. 602-605.

[7] H. Demuth, M. Beale, M. Hagan 2008, Neural Network Toolbox User's Guide, *The Mathworks*, Version 6.