Neural Network Based Cervical Cancer Classification System Design

Nor Ashidi Mat Isa[1], Tan Kuan Liung[2], Mohd Yusoff Mashor[3], Ahmad Nazri Ali[4], Nor Hayati Othman[5]

[1,2,3,4]Control and Electronic Intelligent System (CELIS) Research Group,
School of Electrical & Electronic Engineering, Universiti Sains Malaysia,
Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia.
[5]Pathology Department, School of Medical Science, Universiti Sains Malaysia,
16150 Kubang Kerian, Kelantan, Malaysia.
[1]Tel: +604-5937788 ext. 6051, Fax: +604-5941023, E-mail: ashidi75@yahoo.co.uk
[2]Tel: +604-5937788 ext. 6020, Fax: +604-5941023, E-mail: kuanliung@hotmail.com
[3]Tel: +604-5937788 ext. 6013, Fax: +604-5941023, E-mail: yusof@eng.usm.my
[4]Tel: +604-5937788 ext. 6020, Fax: +604-5941023, E-mail: nazriali@eng.usm.my
[5]Tel: +609-7602134, Fax: +609-7653370, E-mail: hayati@kb.usm.my

ABSTRACT

Cervical cancer is a lethal cancer that affects women around the world. This study proposes a design of a cervical cancer classification system that can classify cells into normal, LSIL and HSIL categories according to the Bethesda system. The system is designed based on a Hybrid Multilayered Perceptron (HMLP) network, which is integrated to an 8-bit microcontroller embedded device. After testing on 200 sets of input data, the system is found to have accuracy, sensitivity and specificity values of 95.50%, 94.00% and 100% respectively. There are 2.67% of negative false and 5.00% of negative false results. The results show that the system is almost as accurate as a medical expert in classifying cervical cell samples.

Keywords

HMLP neural network, cervical cancer, semi-automated diagnosis system.

1.0 Introduction

The second most common type of cancer that affects women today is cervical cancer. Cervical cancer is a silent cancer. Unlike other cancers that cause pain, noticeable lumps or other early symptoms, cervical cancer has no telltale symptoms until it is so advanced that is usually unresponsive to treatment (WebMD, 2002). Only in its late stage, cervical cancer cause pain in the lower abdominal or back regions. However, most cervical cancer takes many years to develop from normal to dangerous stage. Therefore, the mortality related to cervical cancer can be substantially reduced through early detection and treatment.

Currently, Pap test is the most popular method to detect the occurrence of abnormal cells in and around the cervix. Several previous studies by Breen *et al.* (2001), Framer (2001), Kuie (1996) and Adami *et al.* (1994), showed that the chances for a woman of acquiring cervical cancer is reduced as she has Pap test regularly. However, studies by Othman *et al.*, (1997, 1995), K (1996) and Hislop *et al.* (1994) proved that sometimes the Pap test is not effective. The determination of abnormal cervical cells can sometimes be missed in certain situation. Three major reasons that decrease the accuracy of Pap test diagnosis result are bad Pap smear samples, technical and human errors, and small size of cervical intra-epithelial neoplasia (CIN).

Due to limitation of diagnostic performance by Pap test many supplementary methods have been developed to increase the diagnostic performance of the Pap test. In United States of America, three supplementary diagnosis systems, which are commonly used in medical field and currently approved, by the Food and Drug Administration (FDA) are Papnet, AutoPap and ThinPrep (WebMD, 2002, HTAC, 2002). Papnet is used to rescreen the original smear. A computer system selects over 100 abnormal images from the sample, which are then re-examined using high-resolution video. AutoPap is a computer-assisted primary screening and designed to assist cytotechnologists in reviewing conventionally prepared slides by scoring and ranking each slide according to the likelihood that a slide contains abnormalities. ThinPrep uses the original cervical sample, which is first rinsed in a special solution to thin the mucus and eliminate debris that can obscure the finding. The sample is clearer and cleaner for easier screening process by the experts.

Beside that, an expert system for the classification and diagnosis of squamous lesions in the Pap smear of pre-menopausal women has been developed by Mitlehner *et al.* in 1990. The expert system is called Cytopath. Cytopath questions the user about various diagnostic features taken from Pap smear image. A conclusion is reached when the user's responses match the criteria for a diagnosis.

Therefore, this study also focuses on the design of a semi-automated system to classify whether the cervical cells of a patient are normal, low-grade squamous

Table 3: Results for percentage of correct determination of normal, LSIL and HSIL cells.

Normal	100.00%
LSIL	2.00%
HSIL	99.00%

5.0 Conclusion

Generally, the results show that the HRBF network gives very promising results in classifying cervical cells into normal and abnormal cells. The HRBF network produced 100% accuracy, sensitivity and specificity. No false negative and false negative occurred. However, the capability of the HRBF network in further classifying the abnormal cervical cells into LSIL and HSIL cells was poor. The HRBF network could only determine all the normal and the HSIL cells correctly but wrongly classified almost all of the LSIL cells as the HSIL cells.

For the HRBF network, its performance depends on the localisation properties of the network. The HRBF centres are one of the key elements that influence the local properties of the HRBF network (Mashor, 1999). In the case of classifying input data into several classes, the HRBF centres must be able to cluster different types of data separately. In this study, for classifying data into normal and abnormal cells, the HRBF centres successfully cluster both types of data separately, producing high accuracy. However, for classifying the abnormal cells, the HRBF centres could not cluster data for LSIL and HSIL cells separately. Most of the features for both types of cells are redundant. The problem makes the moving k-means clustering algorithm difficult to differentiate between the LSIL and HSIL cells. As a result, the HRBF networks fails to classify abnormal cells into LSIL and HSIL cells. This reduces the percentage of accuracy as well as sensitivity and increases the percentage of false positive.

References

Ashfaq, R., Solares, B. & Saboorian, M. (1996) Detection of Endocervical Component by Papnet System on Negative Cervical Smears. *Diagnosis Cytopathology*. Vol. 15. No. 2. pp. 121-123.

Balasubramaniam, R., Rajan, S., Doraiswami, R. & Stevenson, M. (1998) A Reliable Composite Classification Strategy. *Proc. of IEEE Canadian Conf. on Electrical and Computer Engineering*. Vol. 2. pp. 914-917.

Broomhead, D. S. & Lowe, D. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Complex System*. Vol. 2. pp. 321-355.

Cabaniss, D., Cason, Z., Lemos, L. & Benghuzzi, H. (1997). The Assessment of An Endocervical Component in Cervicovaginal Smears with The Papnet System. *Proc. of 16th Southern Conf. on Biomedical Engineering*. pp. 357-361.

Chen, S., Billings, S. A. & Grant, P. M. (1992). Recursive Hybrid Algorithm for Non-Linear System Identification Using Radial Basis Function Networks. *Int. J. of Control*. Vol. 55. pp. 1051-1070.

Li, Z. & Najarian, K. (2001). Automated Classification of Pap Smear Tests Using Neural Networks. *Proc. of Int. Joint Conf. on Neural Networks*. Vol. 4. pp. 2899-2901.

Ling, F. (1991). Givens Rotation Based on Least Squares Lattice and Related Algorithms. *IEEE Trans. on Signal Processing*. Vol. 39. pp. 1541-1551.

Mashor, M. Y. (2000). Hybrid Multilayered Perceptron Networks. *Int. J. of System Science*. Vol. 31. No. 6. pp. 771-785.

Mashor, M. Y. (1999). Some Properties of RBF Network with Applications to System Identification. *International Journal of The Computer, The Internet and Management*. Vol. 7. NO. 1. pp 34 - 56.

Mat-Isa, N. A., Mashor, M. Y. & Othman, N. H. (2002). Diagnosis of Cervical Cancer Using Hierarchical Radial Basis Function (HiRBF) Network. *Proc. of Int. Conf. on Artificial Intelligence in Engineering and Technology*. Vol. 1. pp. 458-463. 2002.

Mitra, P., Mitra, S. & Pal, S. K. (2000). Staging of Cervical Cancer with Soft Computing. *IEEE Trans. on Biomedical Engineering*. Vol. 47. No. 7. pp. 934-940.

Powell, M. J. D. (1988). Radial Basis Functions for Multivariable Interpolation: A Review. *Proc. of IMA Conf. on Algorithms for The Approximation of Functions and Data*. pp. 143-167.

Tumer, K., Ramanujam, N., Ghosh, J. & Richards-Kortum, R. (1998). Ensembles of Radial Basis Function Networks for Spectroscopic Detection of Cervical Precancer. *IEEE Trans. on Biomedical Engineering*. Vol. 45 No. 8. pp. 953-961.

intraepithelial lesions (LSIL) or high-grade squamous intraepithelial lesions (HSIL) cell. LSIL and HSIL denote cases where there are abnormal growths in the cervical cells, with the latter being more severe and may lead to the development of cancer.

2.0 Methods

The main component of the designed system is the 8051 8-bit microcontroller. Four parameters will be input into the system, namely nucleus size, cytoplasm size, nucleus greylevel and cytoplasm greylevel. These parameters are extracted from cervical cell samples of patients. The Hybrid Multilayered Perceptron (HMLP) network, a form of neural network, will be used to process these inputs. The network is trained using the Modified Recursive Prediction Error (MRPE) algorithm. This network is integrated to the system before processing the above-mentioned inputs. After processing, the result: normal, LSIL or HSIL will be displayed as the output.

2.1 Hybrid Multilayered Perceptron Neural Network

A hybrid multilayered perceptron (HMLP) is an enhanced version of the multilayered perceptron (MLP) network. The proposed network allows network inputs to be connected directly to the input nodes via some weighted connections to form a linear model in parallel with the nonlinear, original MLP model (Mashor, 2000).

A HMLP network with one hidden layer is shown in Figure 1. HMLP network with one hidden layer can be expressed by the following equation (Mashor, 2000):

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F\left(\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t) + b_j^1\right) + \sum_{i=0}^{n_i} w_{ik}^2 v_i^0(t) \quad \text{for } 1 \le k \le m \qquad (1)$$

where w_{ij}^1, w_{jk}^2, w_{ik}^2 denote the weights between input and hidden layer, weights between hidden and output layer, and weights between input and output layer respectively. b_j^1 and v_i^0 denote the thresholds in hidden nodes and inputs that are supplied to the input layer respectively. F(•) is an activation function. A sigmoid function is normally selected as the activation function.

2.2 Modified Recursive Prediction Error

The Modified Recursive Prediction Error (MRPE) algorithm is used to train the HMLP neural network. It is based on structured learning error correction and is a modified version of the Recursive Prediction Error Algorithm (RPE). This study proposed the MRPE algorithm based on its ability to converge at a small value and has a faster convergence rate compared to its predecessor. More information of this algorithm is found in reference (Mashor, 1999).

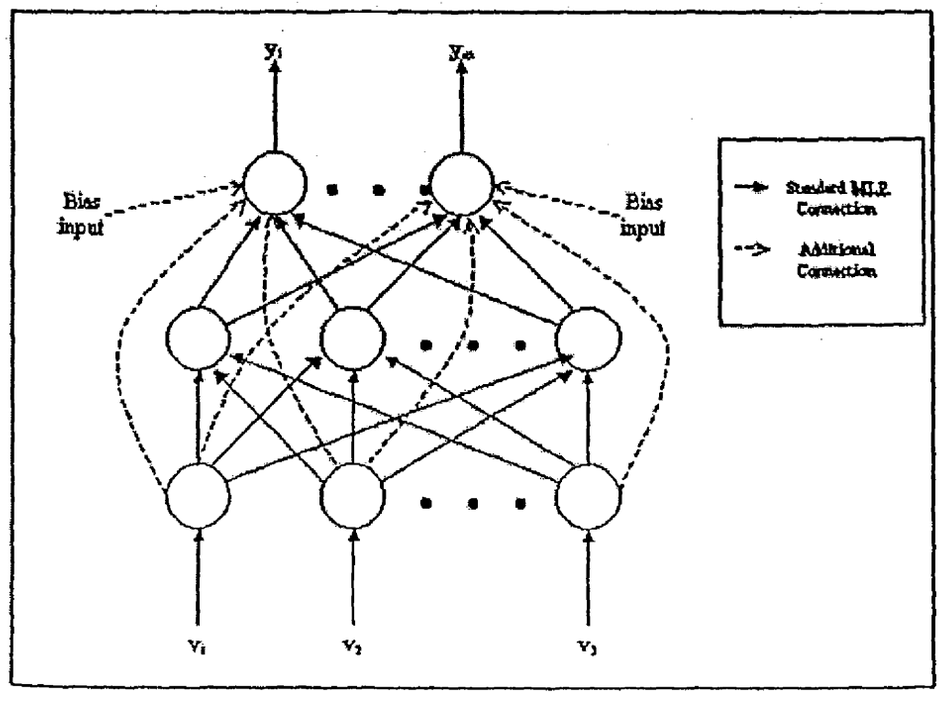

Figure 1: Hybrid multilayered perceptron network

2.3 System Design

For this study, an 8-bit microcontroller, the Atmel 8051 is used as the main component for mathematical computations and interface purposes. It is chosen because it is low cost and easily available. Other components include memory (8K EEPROM and 8K RAM), the 8255 I/O port, RS232 port, 4 x 3 keypad, LCD display. The block diagram for the overall system is shown in Figure 2. The HMLP network is integrated into the 8051's internal memory. The training weights are kept inside the EEPROM. The 4 x 3 keypad functions as an input device to key in the parameters. The Liquid Crystal Display (LCD) functions as an output to show whether the results are Normal, LSIL or HSIL.

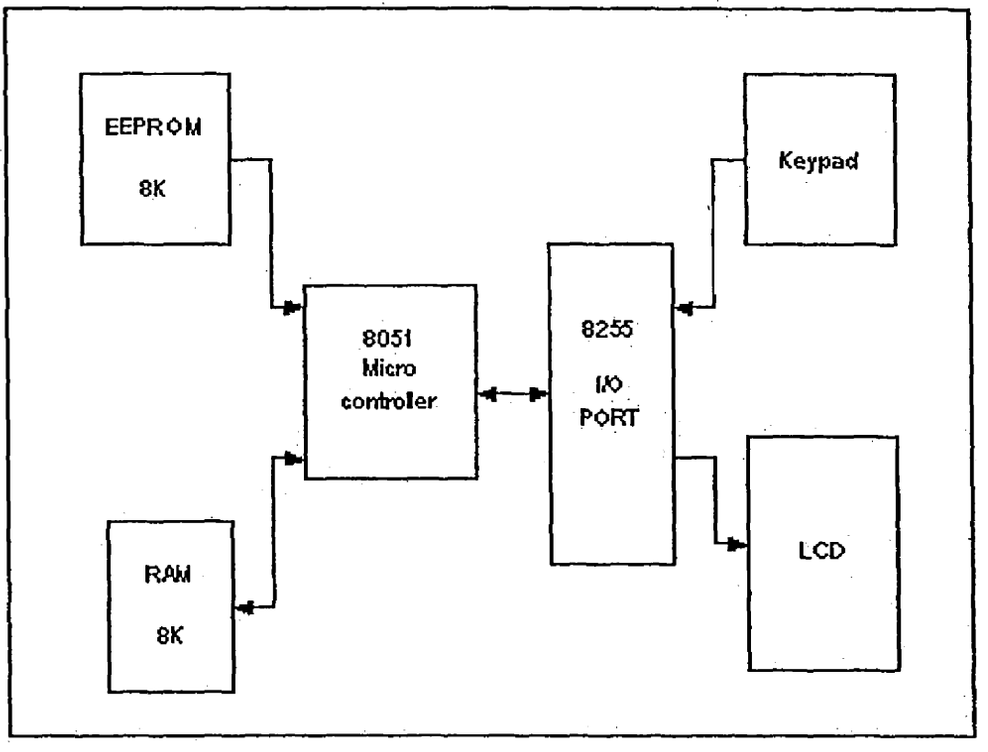

Figure 2: Block Diagram of the Cervical Cell Classification System

3.0 Results

In an experiment to determine the suitability of the device, a total of 200 sets of input parameters are obtained from patients. From the data, 50 sets are normal cells, 50 sets are LSIL cells and another 100 sets are HSIL cells. Prior to the testing, medical experts have classified the data, based on their own diagnosis, into normal, LSIL and HSIL categories. The experts' diagnosis will be used as the de-facto standard to grade the suitability of the system. The suitability itself is determined by five criteria, namely accuracy, sensitivity, specificity, negative false and false positive. After inputting the parameters into the system and observing its outputs, it is discovered that the system managed to achieve diagnosis performance as shown in Table 1. Table 2 shows the percentage of correct determination of normal, LSIL and HSIL using the proposed system.

Table 1: Results for classification of cervical cells into normal, LSIL and HSIL cells.

Accuracy	95.50
Sensitivity	94.00
Specificity	100.00
Negative False	2.67
False Positive	5.00

Table 2: Results for percentage of correct determination of normal, LSIL and HSIL cells.

Normal	100.00
LSIL	90.00
HSIL	96.00

4.0 Conclusion

The result showed the proposed system is 100% accurate in discerning between normal and abnormal (LSIL and HSIL) cervical cancer cells. The system functions almost as accurately as a medical expert in differentiating between LSIL and HSIL cells, with 90.00% and 96.00% accuracy respectively. From the experiment, it can be deduced that the classification given by the system matches almost as same as to that of the medical experts. On condition that the input parameters are extracted correctly, the system will be as accurate as an expert in classifying between normal and abnormal cervical cell samples, and almost as accurate when classifying between LSIL and HSIL cells.

As a conclusion, the proposed system is very accurate in classifying cervical cell samples. However, the system is only tested for 200 sets of input data. Further work should be done in obtaining more sets of data to really test the suitability of the system for practical cases.

References

Adami, H. O., Ponten, J., Sparen, P., Bergstrom, R., Gustafsson, L. & Friberg, L. G. (1994). Survival Trend After Invasive Cervical Cancer Diagnosis in Sweden Before and After Cytologic Screening. *Cancer*. 73(1). 140-147.

Breen, N., Wagener, D. K., Brown, M. L., Davis, W. W. & Barbash, R. B. (2001). Progress in Cancer Screening Over a Decade: Results of Cancer Screening From the 1987, 1992 and 1998 National Health Interview Surveys. *J. of the National Cancer Institute*. 93(22). 1704-1713.

Framer, P. S. (2001). Screening for Cancer: Progress but More Can Be Done. *J. of the National Cancer Institute*. 93(22). 1676-1677.

Hislop, T. G., Band, P. R., Deschamps, M., Clarke, H. F., Smith, J. M. & Ng, V. T. Y. (1994). Cervical Cancer Screening in Canadian Native Women: Adequacy of The Papnicolaou Smear. *The J. of Clinical Cytology and Cytopathology*. 38(1). 29-32

HTAC. (2002). Pap Smears and Prevention of Cervical Cancer. Citing from internet source URL *http://www.health.state.mn.us/htac/papq&a.html*.

Kuie, T. S. (1996). *Cervical Cancer: Its Causes and Prevention*. Singapura: Times Book Int..

Mashor, M.Y. (1999). Nonlinear System Identification using HMLP Networks, *Journal- Institution of Engineers, Malaysia*, Vol 6, No 2, pp 33-39.

Mashor, M.Y. (2000). Hybrid Multilayered Perceptron Networks, *International Journal of Systems Science*, Vol 31, No 6, pp 771-785

Mitlehner, W., Cronin, E., Bronzino, J. & Veranes, R. (1990). Cytópath: An Expert System for The Classification and Diagnosis of Squamous Lesions in The Pap Smears of Pre-Menopausal Women. *Proc. of The 12th Annual Int. Conf. on The IEEE Engineering in Medicine and Biology Society*. 12(3). 1291-1292.

Othman, N. H., Ayub, M. C., Azîz, W. A. A., Muda, M., Wahid, R. & Selvarajan, S. (1997). Pap Smears – Is It An Effective Screening Methods for Cervical Cancer Neoplasia? – An Experience with 2289 Cases. *The Malaysian J. of Medical Sciences*. 4(1). 45-50.

WebMD. (2002). How Can Cervical Cancer Be Prevented?. Citing from internet source URL *http://www.webmd.com/content/dmk/dmk_article_3961643*.