Journal of Engineering Science, Vol. 12, 65–75, 2016

Applying Design of Experiment to Optimise Artificial Neural Network for Classification of Cervical Cancer

Earn Tzeh Tan, Jing Rui Tang and Nor Ashidi Mat Isa*

School of Electrical and Electronic Engineering,
Universiti Sains Malaysia,
14300 Nibong Tebal, Pulau Pinang, Malaysia

*Corresponding author: ashidi@usm.my

Abstract: *Classification of cervical cancer with high accuracy remains as one of the most challenging issues in the automated computer aided diagnosis system. In this study, the significance of four selected features, i.e. nucleus area, cytoplasm area, nucleus mean intensity and nucleus to cytoplasm ratio (N/C ratio) are investigated. These features are extracted from the public image database of Herlev University Hospital, Denmark. Classification of cervical cell is formulated into a two-class problem, dividing the obtained 917 cervical cell images into normal and abnormal cases. T-test analyses show that the four specified features have good potentials as the artificial neural network (ANN) classification elements. The success rates of different ANN framework for classification achieve an average of 86%. The best optimised ANN framework (i.e. eight hidden neurons, "traingd" training algorithm and learning rate of 0.09) achieves an average success rate of 98.50%.*

Keywords: Cervical cancer, design of experiments, factorial design, artificial neural network

1. INTRODUCTION

The 20th century witnessed a significant decline in the incidence and mortality rate of cervical cancer in many developed countries, owing to the widespread of Papanicolaou test (Pap test). Pap test is a screening test in which the pathologists or cytotechnologists examine the samples cell from cervix under light microscope, searching for morphological changes that commonly displayed in the cancerous cells. Cervical cancer is reported according to the worldwide recognised standard, i.e. Bethesda System for Reporting Cervical Cytology.[1] Hyperchromasia and increased nucleus to cytoplasm ratio (i.e. N/C ratio) are among the most commonly used features during screening.[2,3]

Over the past decades, digital data is growing at exponential rate, and analysing all the raw data in manual way is beyond the human limit. The similar issue arises in the medical diagnosis area. Nowadays, the medical information

systems have become larger and rely on the traditional manual data analyses is inefficient for extracting useful information and making decision. This situation stimulates the advancement of big data. Various techniques of statistical analyses, data mining and artificial intelligence (AI) classification are adopted for data segmentation, clustering, classification, etc.[4] Among the conventional classification techniques, artificial neural network (ANN) is widely used to examine and model the relationship between the inputs and outputs of a complex system. Since early 1990s, artificial intelligence (AI)-based system has been available for cervical screening.[5] Various studies reported on the classification of cervical cancer using ANN[6–8] based on different features.

In this study, an ANN-based cervical cancer diagnostic system is proposed to recognise the cervical cell. The significance of four features, i.e. nucleus area, cytoplasm area, nucleus intensity and N/C ratio in classification of normal and abnormal cells are investigated. In general, the development of ANN consists of three stages, i.e. data assembly, training and testing. In fact, performance of ANN highly relies on the training data and training framework. Further analyses on the proposed ANN architecture using two-way factorial design in the classification of cervical cancer are highlighted in this study. Remainder of the paper is organised as follows. Section 2 outlines the methodology and details the experimental setup for statistical design in this case of study. Section 3 presents the experimental results and finally Section 4 concludes the work.

2. EXPERIMENTAL

There are two main stages in the proposed ANN-based cervical cancer diagnostic system. The first stage involves the processing of cervical cell images and feature extraction. The extracted features serve as the input to the ANN. In the second stage, the framework of ANN is developed and further validated by factorial design. The final output of the ANN returns a value of either "0" or "1", indicating whether the presented cervical cell is normal or abnormal.

2.1 Pre-Processing and Feature Extraction of Cervical Cell Images

Cervical cell images from the public image database, the Herlev dataset are used in this study because the ground truth of the images is available. The input colour image is initially converted to gray scale image. The specified four features are then extracted from the ground truth of the images. The rationale of the selection of these features is that they are among the most easily observable characteristics during screening. The features include the following:

1. Nucleus area,
2. Cytoplasm area,
3. Nucleus mean intensity and
4. Nucleus/cytoplasm ratio (N/C ratio).

The nucleus and cytoplasm area are computed in terms of number of pixels. In order to obtain the mean intensity for the nucleus, the ground truth of the nucleus region is employed as the mask and masking is performed on the gray scale image. The average intensity of the nucleus region is then computed. All measurements in this study are normalised. A total of 917 input images are grouped into two categories: normal class consists of 242 images while the other 675 images fall under abnormal group. Examples of cervical cell images with their ground truth are demonstrated in Figure 1.

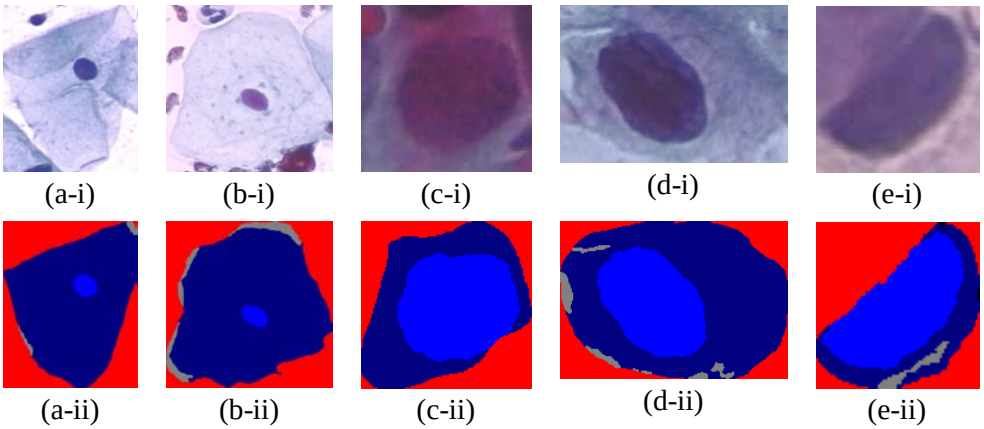


Figure 1: Examples of cervical cell images as in the first row with its ground truth in the second row. For the ground truth images, the light blue region in the centre represents the nucleus and the dark blue region represents the cytoplasm region.

2.2 Classification of Cervical Cell

In this study, a typical three layer feed forward ANN architecture was constructed and trained with back propagation algorithm in Matlab. Inputs to ANN are composed of four extracted features in first stage, and each feature representing a neuron in the input (first) layer. The second layer, which called as hidden layer plays a vital role in modelling the equation to relate the input and the output. The output of the classification is indicated by another neuron assigned in the third layer. In this study, the output of the ANN consists of a neuron, reporting the classification results of the cervical cell, whether the cell is normal or abnormal. A threshold value of 0.5 is introduced to the output neuron,

which means the final value obtained from the neuron is either "0" indicating normal cell or "1" indicating an abnormal cell. The training and the execution of the algorithm is described in detail via a C-style pseudo code outlined in Figure 2.

```
1.  //Xi:vector of input data where i=number of input neuron
2.  //Yk:vector of output data where k=number of output
3.  //Wij:vector of weights between input and hidden layer
4.  //Wjk:vector of weights between hidden and output layer
5.  //Bj & Bk:vector of bias value
6.  //ANN Training
7.  //Training: max epoch:5000; max validation fail:20; goal:1e-2
8.  For (epoch=0; epoch<5000; epoch++) do
9.   if (goal<1e-2 or max validation fail<20) do
10. //Compute outputs of hidden layer neuron
11.   Yj=f(Xi*Wij+Bj);
12. //Compute output of output layer neuron
13.   Yk=f(Yj*Wjk+Bk);
14. //Compute the error Yt:vector of target output data
15.   error=Yt-Yk;
16. //Adjust the weight and bias of each neurons
17.   Execute bp training based on feedback error;
18.   Update Wij,Wjk,Bj,Bk;
19.   Update goal;
20.   end
21.  else do
22.   Stop training;
23.  end
24. end
25. //ANN Testing
26. //Ti:vector of testing data
27. //Compute outputs of hidden layer neuron
28. For (j=0; j<number of hidden neuron; j++) do
29.   Yj=f(Ti*Wij+Bj);
30. //Compute output of output layer neuron
31.  For (k=0; k<number of output neuron; k++) do
32.   Yk=f(Yj*Wjk+Bk);
33.  end
34. end
```

Figure 2: Pseudo code of ANN.

In details, performance of ANN is affected by other parameters' settings in addition to the fixed amount of input and output neuron. Other parameters, which are termed as factors in this study, consisting of the number of hidden neurons, activation function, learning algorithm and learning rate. Several tests involving different factors need to be conducted to resolve the optimised ANN model. Here, an approach called factorial design was conducted to analyse on the influence of each factor to the final ANN results.

In this study, three manipulating factors are specified namely the number of hidden neurons, learning algorithm and learning rate. The activation function is fixed as sigmoid transfer function (logsig) due to its capability to model non-linear relationship. The response of the factorial design is indicated by the success rate computed in percentage for the ANN to correctly recognise normal and abnormal cells. The null hypothesis of the investigation is each of the three specified factors and their interaction has no significant effect on the response. Based on the requirement of the factorial design, the variables of the three specified factors are reduced to two levels, represented as (+1) and (–1) for the analysis of variance (ANOVA).[9,10] Definition of each factor is shown in Table 1.

Table 1: Factor level defined in two level factorial design.

Factor	Label	Factor level	
		Low (–1)	High (+1)
No. hidden neuron	A	4	8
Learning algorithm	B	"trainlm"	"traingd"
Learning rate	C	0.01	0.09

In order to perform the factorial design, eight tests are carried out in randomised order. The purpose of randomisation is to avoid statistical distortions. In addition, each of the tests is replicate twice to estimate the experimental error for the verification of ANOVA. The design matrices of the factorial design with their respective responses are shown in Table 2. The collected results are further analysed using Minitab. All the features extracted are used to conduct t-test analysis prior development of ANN. A total of 150 sets of normal data and 400 sets of abnormal data are randomly selected from each group to serve as ANN training data. Out of the remaining 367 dataset (92 normal, 275 abnormal), 90 sets of normal data and 200 sets of abnormal data are randomly selected as the performance testing data.

Table 2: Design matrix and responses of the factorial design.

Exp.	Factor			Response	
	A	B	C	Y1	Y2
1	–1	–1	–1	95.75	97.05
2	–1	–1	+1	98.25	96.65
3	–1	+1	–1	50.00	78.65
4	–1	+1	+1	98.25	97.25
5	+1	–1	–1	92.30	90.20
6	+1	–1	+1	95.25	94.35
7	+1	+1	–1	50.00	50.00
8	+1	+1	+1	98.50	98.50

3. RESULTS AND DISCUSSION

Prior to the development of ANN, the significant effect of each individual extracted feature is investigated using *t*-test. In this case of study, 95% confidence interval is set in the *t*-test to investigate the magnitude of the difference. The following hypothesis for all four extracted features (i.e. nucleus area, cytoplasm area, nucleus mean intensity and N/C ratio) are set up for the t-test:

Null Hypothesis, Ho:μd = 0; where μd is the mean difference of each respective feature value between normal and abnormal cell.

Alternate Hypothesis, Ha:μd ≠ 0.

Null hypothesis states that the values of the respected feature are not statistically different between two groups of data whereas the alternate hypothesis states that the difference of the values is statistically significant. The strength of evidence to reject the null hypothesis is indicated by the calculated *p*-value. The null hypothesis is rejected if the resulted *p*-value is less than defined α value, i.e. 0.05. The obtained *t*-test results for each feature are shown in Tables 3 to Table 6 respectively.

Table 3: Two-sample *t*-test for nucleus area.

Condition	N	Mean	Standard deviation
0 : Normal	242	0.087	0.105
1 : Abnormal	675	0.286	0.132

95% CI for mean difference: (–0.215, –0.182)
t-value: –23.52; *p*-value: 0.000

Table 4: Two-sample *t*-test for cytoplasm area.

Condition	N	Mean	Standard deviation
0 : Normal	242	0.504	0.151
1 : Abnormal	675	0.374	0.127

95% CI for mean difference: (0.109, 0.152)
t-value: 12.00; *p*-value: 0.000

Table 5: Two-sample *t*-test for N/C ratio.

Condition	N	Mean	Standard deviation
0 : Normal	242	0.246	0.333
1 : Abnormal	675	1.02	1.00

95% CI for mean difference: (–0.862, –0.690)
t-value: –17.62; *p*-value: 0.000

Table 6: Two-sample *t*-test for nucleus mean intensity.

Condition	N	Mean	Standard deviation
0 : Normal	242	0.3053	0.0979
1 : Abnormal	675	0.3734	0.0720

95% CI for mean difference: (–0.082, –0.055)
t-value: –9.92; *p-value*: 0.000

From *t*-test, since the resulted *p*-values for all the extracted features are close to 0.000, thus the null hypothesis for all four cases is rejected. In other words, there is sufficient statistical evidence to conclude that the values among the four extracted features between the normal and abnormal cell are not equal. Hence, these features show potential in the case of cervical cells as the inputs of a classification system. Some other details obtained from *t*-test are listed as follows:

1. For 95% confidence interval, the mean value of nucleus area for abnormal case is between 0.182 and 0.215 higher than the normal case.
2. For 95% confidence interval, the mean value of cytoplasm area for abnormal case is between 0.109 and 0.152 lower than the normal case.
3. For 95% confidence interval, the mean value of N/C ratio for abnormal case is between 0.690 and 0.862 higher than the normal case.
4. For 95% confidence interval, the mean value of nucleus mean intensity for abnormal case is between 0.055 and 0.082 higher than the normal case.

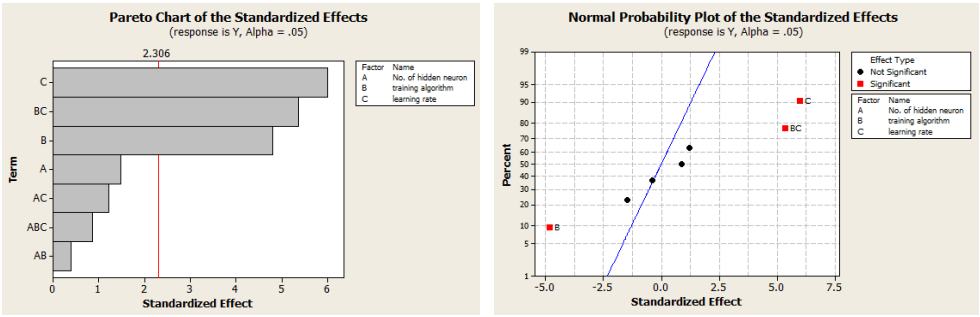
For the analyses of ANN framework using factorial design, the significant effects of the manipulating factor are obtained from the ANOVA analysis. From the ANOVA results (shown in Table 7) generated in Minitab, the factor B, C and the interaction between factor B and C have a significant effect on the response with the *p*-value less than chosen α value. The smaller number of the *p*-value shows the greater strength of effect. Among the specified first order factor, the learning rate has the most significant effect followed by the training algorithm, whereas the number of hidden neuron is not statistically significant. For the second order factor, besides of the interaction between learning rate and training algorithm, other interactions are not significant. The better visualisations of the effect for each factor are shown in Pareto chart and normal probability plot as demonstrated in Figure 3a and 3b respectively.

In order to further analyses on the effects, the interaction plot as shown in Figure 4 between learning rate and training algorithm is examined instead of their main effect plot. From the interaction plot, it can be concluded that the performance between both training algorithms (i.e. "traingd" and "trainlm") do not show much difference when the learning rate is set at 0.09. However, the results drastically drop if "traingd" and learning rate of 0.01 are implemented at the same time. Thus, in this case of study "traingd" cannot be used at low learning rate since it will diminish the final ANN performance.

Table 7: ANOVA result.

Term	Effect	*t*-value	*p*-value
A	–5.344	–1.48	0.176
B	–17.331	–4.81	0.001
C	21.631	6.00	0.000
A*B	–1.444	–0.40	0.699
A*C	4.394	1.22	0.258
B*C	19.331	5.36	0.001
A*B*C	3.144	0.87	0.408

S = 7.208; R–Sq = 92.04%; R–Sq(adj) = 85.08%



(a) (b)

Figure 3: (a) Pareto chart and (b) Normal probability plot.

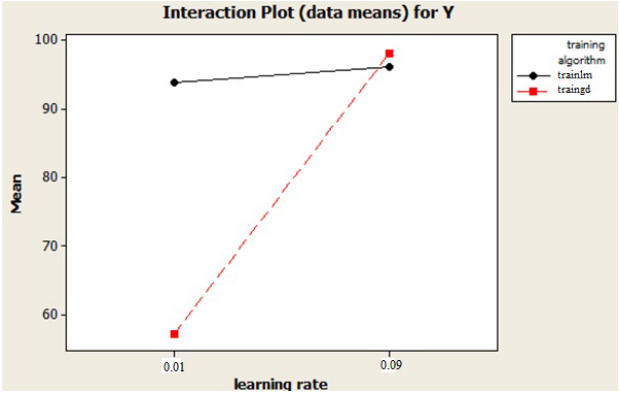


Figure 4: Interaction plot between learning rate and training algorithm.

Finally, the optimum architecture in this case of study is easily observed from the cube plot as shown in Figure 5. In overall, the results show a good potential of using the four extracted features (i.e. nucleus area, cytoplasm area, nucleus mean intensity and N/C ratio) as the ANN classification elements. The success rates of different ANN framework investigated in this study achieves an average of 86%. The best performance ANN is achieved when eight hidden neurons, "traingd" and learning rate of 0.09 are implemented together, for an average success rate of 98.50%. On the other hand, since the number of hidden neuron is not a significant factor, the combination of 4 hidden neurons, "traingd" and learning rate of 0.09 also yields a good result (97.75%).

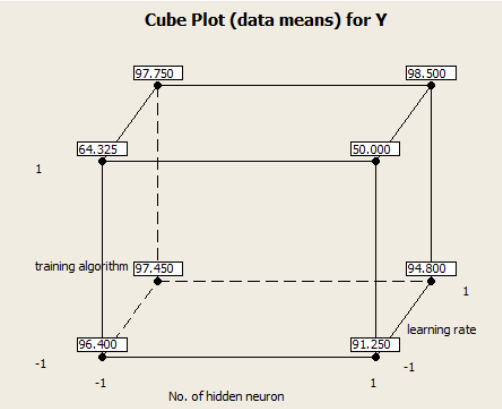


Figure 5: Cube plot for ANN.

For the evaluation of optimised ANN model, total of 290 sets of data which different from the ANN training data are tested. The obtained results are summarised in Table 8. Out of the 200 abnormal data, only six data are predicted as normal (failed case), giving a success rate of 97%, whereas the others data are all correctly classified as normal group.

Table 8: Results of performance for optimised ANN model.

Subject	Predicted as		Average accuracy (%)
	Normal [0]	Abnormal [1]	
Normal [0]	90	0	100
Abnormal [1]	6	194	97.0

4. CONCLUSION

In this study, an ANN-based cervical cancer diagnostic system is proposed to categorise the cervical cell into normal and abnormal case. Significance of four commonly used features in Pap test, i.e. nucleus area, cytoplasm area, nucleus mean intensity and nucleus to cytoplasm ratio (N/C ratio) is studied. Data for testing and training are extracted from the Herlev dataset. Two-sample t-test results reported that all the features have good potentials for ANN classification. By employing factorial design, the best optimised ANN framework (i.e. eight hidden neurons, "traingd" training algorithm and learning rate of 0.09) achieves an average success rate of 98.50%.

5. ACKNOWLEDGEMENTS

This study was partially supported by the Ministry of Higher Education (MOHE), Malaysia under MyPhD Scholarship and Majlis Kanser Nasional (MAKNA), Malaysia under research grant entitled "Development of an Intelligent Screening System for Cervical Cancer".

6. REFERENCES

1. Nayar, R. & Wilbur, D. C. (2015). *The Bethesda system for reporting cervical cytology*. 3rd ed. New York: Springer.
2. Shield, P. W., Daunter, B. & Wright, R. G. (1992). Post-irradiation cytology of cervical cancer patients. *Cytopathology*, 3(3), 167–182.
3. Kreider, J. W. et al. (1985). Morphological transformation in vivo of human uterine cervix with papillomavirus from condylomata acuminata. *Nature*, 317(6038), 639–641.
4. Negnevitsky, M. (2011). *Artificial intelligence: A guide to intelligent systems*. 3rd ed. Ontario, Canada: Pearson Education Canada, 165–216.
5. Dybowski, R. & Gant, V. (2007). *Clinical applications of artificial neural networks*. Cambridge: Cambridge University Press, 380.
6. Romeo, M. et al. (1998). Infrared microspectroscopy and artificial neural networks in the diagnosis of cervical cancer. *Cell Mol. Biol. (Noisy-le-grand)*, 44(1), 179–187 .
7. Singh, S. et al. (2015). Neural network based automated system for diagnosis of cervical cancer. *Int. J. Biomed. and Clin. Eng.*, 4(2), 26–39.
8. Sulaiman, S. N. et al. (2015). Improvement of features extraction process and classification of cervical cancer for the NeuralPap system. *Procedia Comp. Sci.*, 60, 750–759.

9. Wong, W. H. et al. (2015). Two level half factorial design for the extraction of phenolics, flavonoids and antioxidants recovery from palm kernel by-product. *Ind. Crops Prod.*, 63, 238–248
10. Montgomery, D. C. (2009). *Design and analysis of experiments*. 7th ed. Hoboken, NJ: John Wiley & Sons Ltd.