ABSTRACT

Background: Visual image classification is a great challenge to the cytopathologist in routine day-to-day work. Artificial neural network (ANN) may be helpful in this matter.

Aims and Objectives: In this study, we have tried to classify digital images of malignant and benign cells in effusion cytology smear with the help of simple histogram data and ANN.

Materials and Methods: A total of 404 digital images consisting of 168 benign cells and 236 malignant cells were selected for this study. The simple histogram data was extracted from these digital images and an ANN was constructed with the help of Neurointelligence software [Alyuda Neurointelligence 2.2 (577), Cupertino, California, USA]. The network architecture was 6-3-1. The images were classified as training set (281), validation set (63), and test set (60). The on-line backpropagation training algorithm was used for this study.

Result: A total of 10,000 iterations were done to train the ANN system with the speed of 609.81/s. After the adequate training of this ANN model, the system was able to identify all 34 malignant cell images and 24 out of 26 benign cells.

Conclusion: The ANN model can be used for the identification of the individual malignant cells with the help of simple histogram data. This study will be helpful in the future to identify malignant cells in unknown situations.

Key words: Artificial neural network (ANN); cytology; effusion; image

Introduction

Artificial neural network (ANN) is a software model that may take an important decision in various medical fields.[1] In cytology, ANN has been used for classification of breast lesions, identification of malignancy in effusion, and in thyroid lesions.[2-4] The identification of benign and malignant cells in cytology is an important task. ANN has rarely been used to identify the individual cells.[5] In this study, for the first time, we used ANN to distinguish benign and the malignant cells with the help of simple histogram data extraction from digital images.

Materials and Methods

This is a retrospective study done on archival slides and no special tests were done in this case. No special ethical clearance was required for this study. Additionally, the identity of the patients was kept as confidential. In this study, we selected digital images of 402 cells from 20 histopathology-proven malignant cases and 20 benign effusion cases. The malignant cells were selected from the
cases of metastatic adenocarcinoma in ascitic fluid. There were 168 benign cells and 236 malignant cells. Figure 1 shows the flowchart of the work. At first, the colored images were taken by a digital camera (Olympus Camera C-4000 zoom) attached with the microscope (Olympus BX51 model) in 40× objective. The nuclear image of each cell was detected by Image J software (NIH, USA) by adjusting the grey threshold value and subsequently converted to 8-bit grey images. A simple histogram was made from each cell [Figure 2] and the histogram data was transferred into an excel sheet and saved as a .csv file. Total and mean count of gray value, standard deviation, minimum and maximum gray value, and the mode of gray value were recorded in each case.

We used Neurointelligence software [Alyuda Neurointelligence 2.2 (577), Cupertino, California, USA] to build the ANN model. We applied the backpropagation neural network for the function of the ANN model. In our previous studies, we noted before that the backpropagation model works best in the ANN model. We did a heuristic search to design the most suitable ANN architecture. We fixed the hidden unit range from 1 to a maximum of 5 and applied “inverse test error” calculation for the best fitness. At least 10,000 iterations were done for each design. Depending on the fitness and error evaluating the r-squared value, the ANN program itself activated the best network design among all other designs.

There were a total of six variables: Total count of pixel, minimum grey, maximum grey, standard deviation of grey, mean grey, and mode of grey value. Therefore, the first layer of ANN model consists of six neurons. The hidden layer neurons are selected by a heuristic process by the software itself as three. The output will be either benign or malignant, so the number of neurons is only one. Therefore, the neural network architecture was 6-3-1. We selected the logistic function for activation of the network. The cells were automatically and randomly partitioned as training, validation, and testing set of images by the program. There were 281 images in training set, 63 images in validation set, and 60 images in test set. The training set was used to train the ANN model for adjustment of network connections and weights among the different nodes. The validation set was used to tune the ANN and retain the best network for the best performance. The “test set” was used to verify the performance of the ANN model on a set of images. This “test set” was used only after the proper training of the ANN on training and validation set. The on-line back propagation method was used for training of the network. About 10,000 iterations at a speed of 609.81/s were done to reduce the network error to 0.0167 [Table 1].

**Result**

The ANN model was applied in the test set images. All 34 images of malignant cells were correctly recognized by the system. Out of 26 cases of benign cells, 24 were identified correctly by the ANN model. Table 2 shows the confusion matrix of the training, validation, and test set image results. The relative importance of different variables was additionally highlighted in Table 3. The greatest value was given to the count of gray value of the histogram (58.86) followed by the mean gray value (21) and maximum gray value (6).
Discussion

ANN has been previously used for the classification of different lesions of breast, thyroid, etc., based on the acquired data from the images of the lesions. A digital image contains a large number of pixels with a good amount of information. Therefore, using a digital image data directly in the network will have few thousands to lakhs of neuronal data depending on the size of the image. It may not be possible for ANN to handle such an enormous number of neuronal inputs. Zheng et al. used pixel intensity data to classify the images of the various blood cells.[7] Herein, we have used histogram of simple 8-bit gray images of different cells and prepared an ANN. Unlike other studies, we had only few neuronal inputs from the histogram. Surprisingly, we noted that these simple histogram values of the images can be used to classify the character of the cells in most of the cases. The present ANN system was able to diagnose all the malignant cells and 24 out of 26 (92%) benign cell images in the test set. The whole procedure is very simple and minimal time-consuming. In future, the total procedure can be made automated. This present work highlights the enormous potential of image histogram and ANN to recognize automated detection of malignant cells.

Financial support and sponsorship
No grant has been taken for this study.

Conflicts of interest
There is no conflict of interest in this study.

References

1. Dey P, Dey R. Artificial neural network — Mechanism and application in pathology. Indian J Pathol Microbiol 2002;45:371-4.
2. Isa NA, Esugasini S, Mashor MY, Othman NH. Fine needle aspiration cytology evaluation for classifying breast cancer using artificial neural network. Am J Appl Sci 2007;4:999-1008.
3. Dey P, Logasundaram R, Joshi K. Artificial neural network in diagnosis of lobular carcinoma of breast in fine-needle aspiration cytology. Diagn Cytopathol 2013;41:102-6.
4. Subbaiah RM, Dey P, Nijhawan R. Artificial neural network in breast lesions from fine-needle aspiration smear. Diagn Cytopathol 2014;42:218-24.
5. Barwad A, Dey P, Susheilia S. Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology. Cytometry B Clin Cytom 2012;82:107-11.
6. Karakitsos P, Cochand-Priollet B, Guillausseau PJ, Pouliakis A. Potential of the back propagation neural network in the morphologic examination of thyroid lesions. Anal Quant Cytol Histol 1996;18:494-500.
7. Zheng Q, Milthorpe BK, Jones AS. Direct neural network application for automated cell recognition. Cytometry A 2004;57:1-9.
8. Available from: http://rsbweb.nih.gov/ij. [Last accessed on 29th February 2016].

Table 2: Confusion matrix of training, validation, and test set

| Target Output diagnosis | Benign | Carcinoma |
|-------------------------|--------|-----------|
| Training set Benign     | 115    | 26        |
| Carcinoma               | 2      | 138       |
| Validation set Benign   | 24     | 6         |
| Carcinoma               | 3      | 30        |
| Test set Benign         | 24     | 2         |
| Carcinoma               | 0      | 34        |

Table 3: Relative importance of input data

| Input                              | Importance (%) |
|------------------------------------|----------------|
| Count of gray value                | 58.886549      |
| Mean of gray value                 | 21.49769       |
| Standard deviation of gray value   | 5.252215       |
| Minimum of gray value              | 4.208878       |
| Maximum of gray value              | 6.176414       |
| Mode of gray value                 | 3.978254       |

Table 2: Confusion matrix of training, validation, and test set

Table 3: Relative importance of input data

Announcement

“QUICK RESPONSE CODE” LINK FOR FULL TEXT ARTICLES

The journal issue has a unique new feature for reaching to the journal’s website without typing a single letter. Each article on its first page has a “Quick Response Code”. Using any mobile or other hand-held device with camera and GPRS/other internet source, one can reach to the full text of that particular article on the journal’s website. Start a QR-code reading software (see list of free applications from http://tinyurl.com/yzh2tc) and point the camera to the QR-code printed in the journal. It will automatically take you to the HTML full text of that article. One can also use a desktop or laptop with web camera for similar functionality. See http://tinyurl.com/2bw7fn3 or http://tinyurl.com/3ysr3me for the free applications.