AUTOMATED CLASSIFICATION OF BLASTS IN ACUTE LEUKEMIA BLOOD SAMPLES USING HMLP NETWORK

N. H. Harun[1], M.Y.Mashor[1], A.S. Abdul Nasir[1] and H.Rosline[2]

[1]*Electronic & Biomedical Intelligent Systems (EBItS) Research Group,School of Mechatronic Engineering, University Malaysia Perlis,02600 Jejawi, Arau, Perlis*hazlyna_harun@yahoo.com

[2]*Hematology Department University Hospital, Univesity Science Malaysia, Kubang Kerian,Kelantan,* roslin@usm.kb.my

ABSTRACT. This paper presents a study on classification of blasts in acute leukemia blood samples using artificial neural network. In acute leukemia there are two major forms that are acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL). Six morphological features have been extracted from acute leukemia blood images and used as neural network inputs for the classification. Hybrid Multilayer Perceptron (HMLP) neural network was used to perform the classification task. The Hybrid Multilayer Perceptron(HMLP) neural network is trained using modified RPE(MRPE) training algorithm for 1474 data samples. The Hybrid Multilayer Perceptron (HMLP) neural network produces 97.04% performance accuracy. The result indicates the promising capabilities and abilities of the Hybrid Multilayer Perceptron (HMLP) neural network using modified RPE (MRPE) training algorithm for classifying and distinguishing the blasts from acute leukemia blood samples.

Keywords: Hybrid Multilayer Perceptron, artificial neural network, acute leukemia

INTRODUCTION

Leukemia is a type of cancer in which the body produces large numbers of abnormal cells typically white blood cells (WBC). There are two major kinds of leukemia: chronic and acute (Lim, 2002). The word acute means the diseases grow and progresses rapidly (Lim, 2002). The bone marrow is infiltrated with more than 20% of blast cells and when myeloid cells are affected, the disease is called acute myelogenous leukemia(AML) while the cells affects lymphoid cells, it is called acute lymphocytic leukemia (ALL) (Panovska-Stavridis et.al, 2008).

The early and rapid classification of the acute leukemia diseases, greatly aids in providing the appropriate and effective treatment for that particular type (Khasman et.al, 2010). This is important, as the natural history and reaction to treatment varies according to the type of blast involved in the leukemia process (Panovska-Stavridis et.al, 2008). The original classification scheme proposed by the French-American-British (FAB) Cooperative Group divides AML into 8 subtypes (M0 to M7) and ALL into 3 subtypes (L1 to L3). The FAB classifications of ALL (L1 to L3) which are differentiated based on morphology, including cell size, prominence of nucleoli, and the amount and appearance of cytoplasm (Bennett et.al, 1976 & Panovska-Stavridis et.al, 2008). According to French-American-British (FAB) classification also, the description of cells is small and uniform for ALL-L1. Meanwhile, cells of AML-M1 are large and regular (Bennett et.al, 1976 & Panovska-Stavridis et.al, 2008).

Currently, routine diagnosis and classification of morphological features for acute leukemia blood samples is using microscope evaluation (Sabino et al., 2003). The process is

an exhaustive, burdensome and repetitive work performed by hematologists, technologists or medical expertise (Sabino et al., 2003). To improve the reliability of the diagnosis and decreasing the dependence on human experts, several previous studies developed automated and semi automated diagnosis and classification using artificial neural network for medical images.

Artificial intelligence based on neural network applications has a great impact on the interpretation of medical images (Chenn J.H. et.al, 2004). The researches in neural network implementation for acute leukemia diagnosis have been done by several researchers. By using AML/ALL data sets Toure et al. (2001) proposed Multilayer Perceptron Network (MLP) and the highest performance rate 58% was achieved, Ryu et al. (2002) experimented with Modular Neural Networks as the classifier and the best performance was achieved 75% and finally, Xu et al.(2002) proposed ellipsoid ARTMAP neural network and the best result was 97.1%.

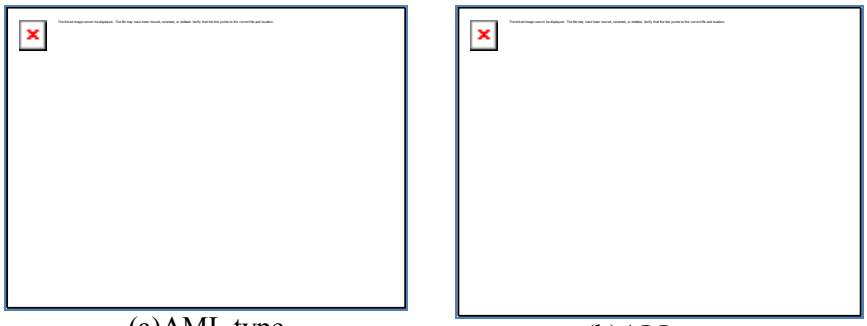
In 2000, Mashor has introduced the Hybrid Multilayered Perceptron Network (HMLP) that has been proven to significantly improve the performance of Multilayered Perceptron Network (MLP). The HMLP network has been tested on other data sets such as on cervical cancer and breast lesions data sets and successfully verified each type of data sets correctly with high percentage in both training and testing phase (Mat-Isa et al.,2004 & Mashor et al., 2007). Besides that, Mashor et al. (2004) had used the HMLP network to classified 3D object using 2D moment and best recognition accuracy of up to 100%.

As HMLP network has the promising potential to classify the various data sets, in the current study HMLP network also has been proposed as blasts classification in acute leukemia blood images. The HMLP network has been trained using modified RPE (MRPE) algorithm.

Blasts classification in Acute Leukemia blood samples using HMLP network

The HMLP network were applied and tested in order to classify the blasts from acute leukemia blood samples. The inputs to the HMLP network are some features of the blasts from acute leukemia images. The images of blasts from acute leukemia blood samples have been captured using *Infinity2* digital camera that is attached to *Leica* microscope. After that, the captured images were revised and approved by hematologists or technologists in order, to verify the correctness of the blasts in acute leukemia cells images. Then, six features have been manually extracted using image processing method from the images of acute leukemia blood samples.

The six features are the area for the blast, nucleus, and cytoplasm, perimeter for the blast and nucleus and ratio of nucleus to cytoplasm. The HMLP neural network receives these six features as the inputs and classifies the blasts in acute leukemia cells into two categories which are AML and ALL types. Samples of acute leukemia cells images are shown in Figure 1.

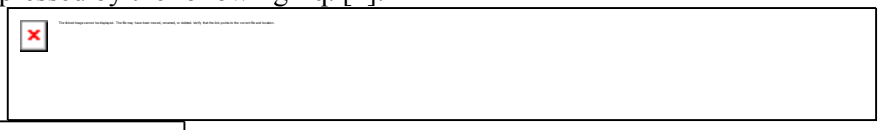


(a)AML type (b)ALL type

Figure 3. Samples of acute leukemia blood images

Hybrid Multilayer Perceptron (HMLP) neural network

Consider HMLP network with one hidden layer as shown in Figure 2. The HMLP network can be expressed by the following Eq. [1]:



where  denote the weights between input and hidden layer, weights between hidden and output layer, and weights between input and output layer respectively.

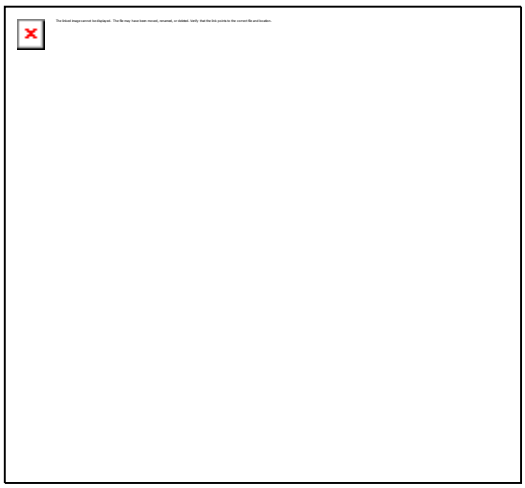
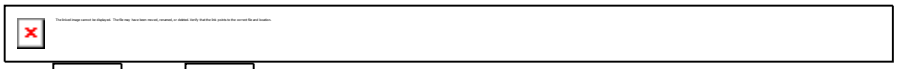


Figure 2. One-hidden layer HMLP network

 and  denote the thresholds in hidden nodes and inputs that are supplied to the input layer respectively;  and  are the number of input nodes, output nodes and hidden nodes respectively.  is an activation function that is selected as sigmoid function. The weights  and  are unknown and should be selected to minimize the prediction error defined as:

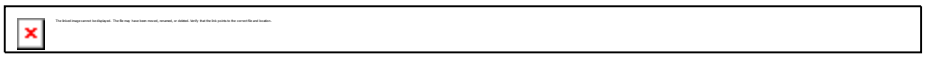


where  and  are the actual outputs and network outputs respectively.

From Eq. (1), the values of  and  must be find out using appropriate algorithm. Chen et.al (1990) proposed recursive prediction error (RPE) to overcome the problems using BP algorithm. Subsequently to improve the convergence rate for RPE algorithm, Mashor (2000) proposed modified recursive prediction error (MRPE) algorithm. The description of MRPE algorithm could be found in Mashor (2000).

METHODOLOGY

In the current study, the HMLP network trained using MRPE algorithm is proposed as acute leukemia cells diagnosis technique. The analysis performance is based on overall performance and accuracy of correct determination of ALL and AML type. In order, to determine the suitability of the HMLP neural network as acute leukemia cells diagnosis technique, the HMLP neural network needs to go through training and testing phases. During the training phase, the weighs and bias of the network are calculated by using MRPE algorithm as mentioned above. The HMLP network was trained using the designing parameter for MRPE as follow;



Based on the values acquired from training phase, the performance of the HMLP network is analyzed to obtain appropriate values for testing phase. In order to find the optimum structure, the HMLP network performance has been analyzed for the optimum number of hidden nodes and epochs. For this situation, the epochs will be set to a certain fixed value. Then, the HMLP network was trained at the appropriate range of hidden nodes. The number of hidden nodes that have given the best performance is then selected as the optimum hidden nodes. After that, by fixing the optimum number of hidden nodes, the epochs will be analyzed in a similar way to obtain the optimum number of epochs that can give the highest or best accuracy.

In this study, 1474 data were used in diagnosing blasts in acute leukemia samples using HMLP network. 1100 data were used as training data while 474 data were used as testing data. From 1100 training data, 555 data for ALL type while 545 were AML type. In addition, for 474 testing data, 253 data for ALL type while 221 data were AML type**.** The distributions of the training and testing data sets are shown in Table 1**.** The data were taken from Hematology Department at Hospital Universiti Sains Malaysia (HUSM).

Table 1 Distribution of training and testing data sets

Types of acute leukemia cells	Training data	Testing data
AML	545	221
ALL	555	253
TOTAL	**1100**	**474**

The inputs of the network are six features which are the area for the blast, nucleus, and cytoplasm, perimeter for the blast and nucleus and ratio of nucleus to cytoplasm. The output nodes of the HMLP network are two which represents the two types of blasts for acute leukemia cells.

RESULTS AND DISCUSSIONS

Table 2 and 3 shows the diagnosis performance of the HMLP neural network using MRPE training algorithm for training and testing phase respectively. The HMLP neural network using MRPE training algorithm produced the highest and best performance at 5 training epochs and 30 hidden nodes.

Table 2.Results for acute leukemia cells size classification in training phase

Classification	True	False	Total	Accuracy (%)
ALL	544	11	555	98.00%
AML	531	14	545	97.43%
Overall	**1075**	**25**	**1100**	**97.72%**

From the Table 2, the results show that the HMLP network classified the ALL type better than AML type in the training phase with 98.00% and 97.43% accuracy respectively. Meanwhile, the overall performance of the HMLP network produced 97.72% of accuracy.

Table 3.Results for acute leukemia cells size classification in testing phase

Classification	True	False	Total	Accuracy (%)
ALL	242	11	253	95.65%
AML	218	3	221	98.64%
Overall	**460**	**14**	**474**	**97.04%**

The results obtained in Table 3 represent that the HMLP network successfully classified 98.64% of AML type as compared to 95.65% of ALL type during testing phase. Besides that, the overall performance of the HMLP network classified 97.04% blasts size from acute leukemia samples correctly.

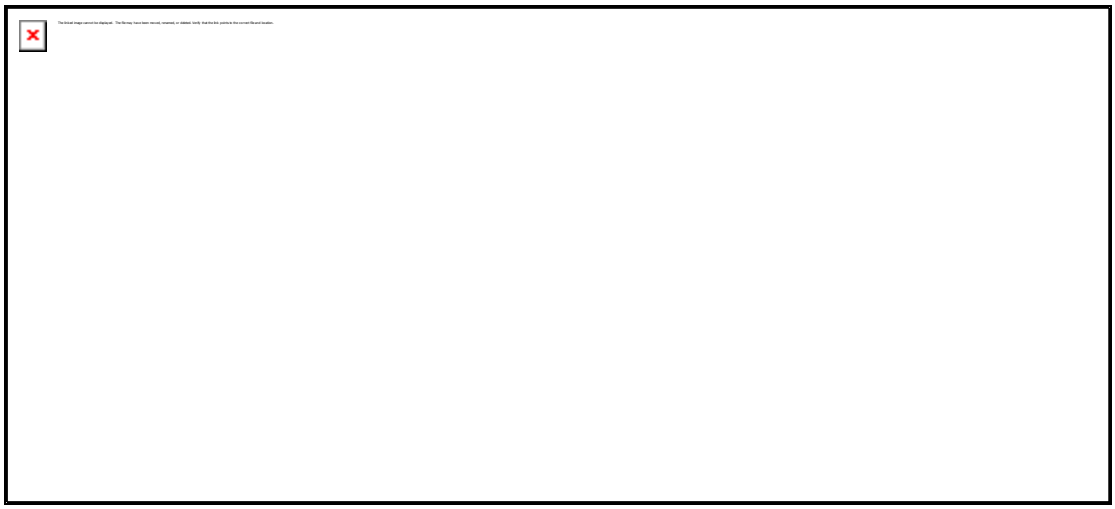


Figure 3. Overall performance accuracy during training and testing phase

Figure 3 indicates the resulted graph for HMLP neural network using MRPE training algorithm was achieved an optimal result at 5 training epochs and 30 hidden nodes. The results represent that the HMLP network has high capability and ability to classify the blasts in acute leukemia samples into two types, namely ALL and AML.

CONCLUSION

The HMLP network trained using MRPE training algorithm has been proposed to analyze and classify blasts into two types which are acute lymphocytic leukemia (ALL) and acute myelogenous leukemia (AML). Six features from acute leukemia blood samples, which are area for the blast, nucleus, and cytoplasm, perimeter for the blast and nucleus and ratio of nucleus to cytoplasm, have been used as HMLP network inputs data. From the results, it has been proved that HMLP neural network has successfully classified the blasts in acute leukemia blood samples with high percentage in training and testing phase. Furthermore, the HMLP network successfully classified 242(51%) of acute lymphocytic leukemia (ALL) as well as 218(46%) acute myelogenous leukemia (AML) cells correctly, from 474 total data during the testing phase. Only 14 (3%) data is miss - classified. It is concluded that, artificial intelligence using HMLP neural network can contribute efficient and accurate diagnosis for acute leukemia blood samples.

ACKNOWLEDGMENTS

We would like to express our thanks to UniMAP and Malaysian Government for supporting this research in term of research grant. This research is funded under Fundamental Research Grant Scheme (Grant No. 9003 00129).

REFERENCES

Bennett,J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A.,Gralnick H.R & Sultan,C:(1976). *Proposals for the classification of the acute leukaemias. French–American–British (FAB) Cooperative Group.Br J Haematol*.33,451–458.

Chen, S., Cowan, C. F. N., Billings, S. A., & Grant, P. M.:(1990).*A Parallel Recursive Prediction Error Algorithm for Training Layered Neural Networks. International Journal of Control*.6,1215-1228.

Chenn, J.H. & Wei, C.L. : (2004).Application of Probabilistic Neural Networks to the Class Prediction of Leukemia and Embryonal Tumor of Central Nervous System. *Neural Processing Letters* .19, 211–226.

Khasman,A. & Al-Zgoul,E. :(2010) .Image Segmentation of Blood Cells in Leukemia Patients. *Recent Advances in Computer Engineering and Applications.* 104-109.

Lim, G. C. C. :Overview of Cancer in Malaysia. (2002).*Japanese Journal of Clinical Oncology, Department of Radiotherapy and Oncology, Hospital Kuala Lumpur*.

Mashor M.Y.:(2000).Hybrid Multilayered Perceptron Networks.*International Journal. of System and Science*.6,771-185.

Mashor,M. Y., Osman,M. K., & Arshad,M. R.:(2004).3D Object Recognition Using 2D Moments and HMLP Network. *IEEE Proceeding of the International Conference on Computer Graphics, Imaging and,Visualization*. 26-130.

Mat-Isa, N. A.,Mashor,M. Y. , & Othman,N. H.:(2004).Classification for Cervical Cancer Cells Using HMLP Network with Confidence Percentage and Confidence Level Analysis.*International Journal of The Computer, The Internet and Management*.1,17-29.

Panovska-Stavridis, I., Cevreska, L., Trajkova, S., Hadzi-Pecova, L., Trajkov, D., Petlichkovski, A., Efinska-Mladenovska, O., Sibinovska, O., Matevska, N., Dimovski, A. and Spiroski, M. : (2008). Preliminary Results of Introducing the Method Multiparameter Flow Cytometry in Patients with Acute Leukemia in the Republic of Macedonia. *Maced J Med Sci.*36-43.

Ryu, J. & Cho, S. -B.: (2002).Gene expression classification using optimal feature/classifier ensemble with negative correlation. *Proceedings of the 2002 International Joint Conference on Neural Network*. Vol 1,198–203.

Sabino,D.M.U.,Costa, L.F.,Martins,S.L.R.,Calado,R.T, & Zago,M.A..(2003).Automatic Leukemia Disease.*Article Acta Microspica*.12,1-6.

Toure,A. & Basu, M. : (2001) .Application of neural network to gene expression data for cancer classification. *Proceedings of the 2001 International Joint Conference on Neural Networks*. Vol 1,583–587.

Xu, R., Anagnostopoulos, G. & Wunsch, D.(2002). Tissue classification through analysis of gene expression data using A new family of ART architectures. *Proceedings of the 2002 International Joint Conference on Neural Networks*.Vol 1,300–304.