

# CERVICAL CANCER DIAGNOSTIC SYSTEM USING ADAPTIVE FUZZY MOVING K-MEANS ALGORITHM AND FUZZY MIN-MAX NEURAL NETWORK

<sup>1</sup>ANAS QUTEISHAT, <sup>2</sup>MOHAMMAD AL-BATAH, <sup>3</sup>ANWAR AL-MOFLEH,  
<sup>4</sup>SHARHABEEL H. ALNABELSI

<sup>1</sup>Asstt Prof., Department of Computer Engineering, Al-Balqa Applied University, Jordan

<sup>2</sup>Asstt. Prof., Department of Computer Science, Jadara University, Jordan

<sup>3</sup>Asstt Prof., Department of Electrical Engineering, Al-Balqa Applied University, Jordan

<sup>4</sup>Asstt Prof., Department of Computer Engineering, Al-Balqa Applied University, Jordan

E-mail: <sup>1</sup>[anas.quteishat@fet.edu.jo](mailto:anas.quteishat@fet.edu.jo), <sup>2</sup>[albatah@jadara.edu.jo](mailto:albatah@jadara.edu.jo), <sup>3</sup> [anwaralmofleh@yahoo.com](mailto:anwaralmofleh@yahoo.com),  
<sup>4</sup>[alnabsh1@gmail.com](mailto:alnabsh1@gmail.com)

## ABSTRACT

Pap smear screening is the most successful attempt of medical science and practice for the early detection of cervical cancer. Manual analysis of the cervical cells is time consuming, laborious and error prone. This paper presents a Neural Network (NN) based system for classifying cervical cells as normal, low-grade squamous intra-epithelial lesion (LSIL) and high-grade squamous intra-epithelial lesion (HSIL). The system consists of three stages. In the first stage, cervical cells are segmented using the Adaptive Fuzzy Moving K-means (AFMKM) clustering algorithm. In the second stage, the feature extraction process is performed. In the third stage, the extracted data is classified using Fuzzy Min-Max (FMM) NN. The empirical results show that the proposed method can achieve acceptable results.

**Keywords:** *Cervical cancer; Pap smear; Adaptive Fuzzy Moving K-means clustering algorithm; Fuzzy Min-Max Neural Network .*

## 1. INTRODUCTION

Early diagnosis of cancer cells plays an important role the curing process. The patient usually undergoes a series of tests and these tests are screened by specialists to diagnose cancer cells. Over the past few decades researchers focused their attention on producing medical decision support systems. Medical diagnosis problems are viewed as pattern recognition problems from the artificial intelligence point of view. Artificial Neural Networks (ANNs) are used to tackle pattern recognition problem. Indeed, ANNs have the ability to integrate data-based analytical techniques such as decision and classification theory, and to use knowledge-based approaches to provide useful information to support the decision making process [1]. From the literature, a lot of successful ANN applications to medical problems can be found. In [1], autonomously learning ANN models were used To classify real coronary care unit (CCU) patient records and trauma patient records into different categories. In [2] a pulmonary disease diagnostic system was proposed. The system used real clinical

data to attempt to treat a whole category of distressed body organs. A model selection method that used the self-organizing map (SOM) for breast cancer diagnosis was demonstrated in [3]. In [4], the performances of a number of ANN models for breast cancer diagnosis were compared and analyzed. A neuromuscular disorder diagnosis system that employed two different ANN models (supervised and unsupervised) for analyzing features selected from electromyography (EMG) was presented in [5].

In this paper an ANN is used as a decision support system for classifying cervical cancer patients. The data collected from the patients are images of their cells. An adaptive Fuzzy Moving K-means (AFMKM) clustering algorithm is used to segment the images. The features are extracted from the segmented images. Finally the extracted features are fed to an ANN in particular the FMM ANN. Figure 1 shows a block diagram for the proposed system.

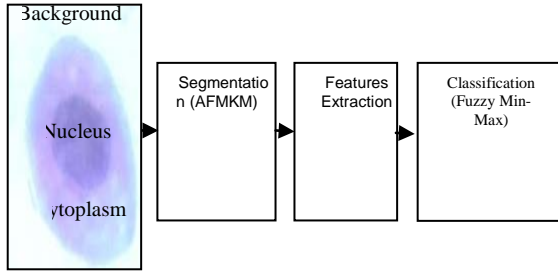


Figure1. Cervical Cancer Computer Vision System Stages

## 2. SEGMENTATION USING AFMKM CLUSTERING ALGORITHM

Data clustering is a method used when creating groups of objects or clusters in such a way that similar objects are grouped together while dissimilar objects are segregated in their distinct clusters[6]. In digital image segmentation applications, the clustering technique is commonly used to segment regions of interest [7-9]and/or detect borders of objects in an image [10-12]. Researchers have implemented numerous clustering algorithms. One of the most important and widely used algorithms is K-means clustering [13]. The K-means method is numerical, unsupervised, non-deterministic, and iterative. It is commonly used in computer vision as a form of image segmentation.

The moving K-means (MKM) clustering algorithm proposed by [14] has the capability to overcome the three basic problems presented earlier. The algorithm minimizes dead centers and center redundancy problems, as well as indirectly reduces the effect of a trapped center at the local minima. Although the MKM algorithm has been proven to minimize aforementioned problems, it still has the following drawbacks:

1. It is sensitive to noise. The worst case that may happen is that one cluster or center may be assigned as a centroid of a group of noise data.
2. In some cases, clusters or centers are not located in the middle or centroid of a group of data. Thus, this weakness may lead to imprecise results.

These limitations were solved by [15] which has proposed a new clustering algorithm named AFMKM. This algorithm constantly checks the fitness of each center during the clustering process. If the center cannot satisfy a certain criteria, the center will be moved to the group of data with the most active center. The AFMKM algorithm has introduced the concept of fuzzy logic; the decision

to let the pixel join a certain center is based on the membership function.

Consider a problem with N data that needs to be clustered into  $n_c$  centers. Here,  $v_i$  represents the i-th data and  $c_j$  represents the j-th center with predetermined initial value, where  $i = 1, 2, \dots, N$  and  $j = 1, 2, \dots, n_c$ . All data are assigned to the nearest center based upon Euclidean distance. The new position for each center is calculated based upon the Equation as follows:

$$c_j = \frac{1}{n_j} \sum_{i \in c_j} v_i \quad (1)$$

The fitness for each cluster is then calculated using

$$f(c_j) = \sum_{i \in c_j} \left( \|v_i - c_j\| \right)^2; \quad (2)$$

For a satisfactory clustering process, the AFMKM introduced the fuzzy logic to allow each data member to be simultaneously assigned to more than one class with different degrees of membership. The process can be achieved based on the membership function as given by

$$m_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}} \quad (3)$$

where  $d_{ik}$  is the distance from point k to current cluster center i,  $d_{jk}$  is the distance from point k to other cluster centers j, and q is the fuzziness exponent where the typical value is 2. After specifying the membership for each data and applying the fitness calculation process using Equation 2, the original relationship among the centers has to fulfil the equation as follows:

$$f(C_s) < \alpha_a f(C_l) \text{ and } m(C_{sk}) > m(C_{lk}) \quad (4)$$

where  $C_s$  is the cluster with the smallest fitness value,  $C_l$  is the cluster with the largest fitness value,  $\alpha_a$  is a designed small constant with the initial value that equals  $\alpha_0$ , and  $\alpha_0$  is a designed small constant value between  $0 < \alpha_0 < \frac{1}{3}$ ;  $m(C_{sk})$  is the membership value of point k according to the smallest centre, and  $m(C_{lk})$  is the membership value of point k according to the largest centre.

Based on the AFMKM algorithm, if Equation 4 is not fulfilled, the members of  $C_l$  with values lower than  $C_l$  are assigned to the nearest cluster depending on the maximum membership function while the rest remains with  $C_l$ . Then, the new positions of all the existing clusters are recalculated according to Equation 1. Subsequently, the value of  $\alpha_a$  is updated according to the Equation as follows:

$$\alpha_a = \alpha_a - \alpha_a / n_c \tag{5}$$

To improve the clustering process, the whole process has to be repeated until

$$f(C_s) \geq \alpha_b f(C_l) \tag{6}$$

Where  $\alpha_b$  is a designed small constant with an initial value equal to  $\alpha_0$ .

For each iteration,  $\alpha_a$  must be reset to  $\alpha_0$ , and

$\alpha_b$  is updated based on

$$\alpha_b = \alpha_b - \alpha_b / n_c \tag{7}$$

In this paper, the AFMKM clustering algorithm is successfully used to segment the ThinPrep® image into three regions; i.e. nucleus, cytoplasm and background. Thus, the background region will be easily eliminated from the image. After the segmentation process was completed, nine features are extracted from both nucleus and cytoplasm. The features are size, grey level, perimeter, red, green, blue, intensity1, intensity2 and saturation [16]. Intensity1, intensity2 and saturation were computed using Equation (8), (9) and (10) respectively [17, 18]. The extracted features will be fed as input data to the intelligent diagnostic part. For this purpose, A Modified Fuzzy Min-Max Neural Network with a Genetic-Algorithm was used.

$$Intensity\ 1 = \frac{1}{3} (Red + Green + Blue) \tag{8}$$

$$Intensity\ 2 = (0.299Red) + (0.587Green) + (0.114Blue) \tag{9}$$

$$Saturation = \sqrt{c_1^2 + c_2^2} \tag{10}$$

Where;

$$c_1 = Red - 0.5Green - 0.5Blue$$

$$c_2 = -\frac{\sqrt{3}}{2} Green + \frac{\sqrt{3}}{2} Blue$$

The next section will describe the FMM ANN used to classify the extracted features from the above described method.

### 3. FUZZY MIN-MAX NEURAL NETWORK

The FMM neural network was first introduced by Simpson (1992) as a supervised classification model. The network structure is built from hyperboxes. A hyperbox is defined by its minimum and maximum points which are created by the input patterns. Figure 1 shows the decision boundary of the FMM network in a two-dimensional space. The FMM learning algorithm consists of three steps: expansion, overlapping test, and contraction [19]. The FMM learning procedure starts by selecting an input pattern, and then finding the closest hyperbox that matches the input pattern. The closest hyperbox in the FMM model is found by using the membership function, which is defined with respect to hyperbox min-max points [19]. The membership value is calculated using equation (1) [19]:

$$b_j(A_h) = \frac{1}{2^n} \sum_{i=1}^n [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - w_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, v_{ji} - a_{hi})))] \tag{11}$$

where B is the membership function, j is the hyperbox,  $A_h = (ah_1, ah_2, \dots, a_j p^h)$  is the hth input pattern,  $V_j = (vj_1, vj_2, \dots, vj_n)$  is the minimum point for  $B_j$ ,  $W_j = (wj_1, wj_2, \dots, wj_n)$  is the maximum point for  $B_j$ , and  $\gamma$  is the sensitivity parameter that regulates how fast the membership values decrease as the distance between  $A_h$  and  $B_j$  increases. The membership function represents the degree to which an input pattern fits in the hyperbox, and the membership value ranges between 0 and 1. If the pattern does not belong to any hyperboxes, even with the expansion process in FMM, a new hyperbox will be created to include the input pattern. In other words, FMM entails a dynamic network structure with online learning capability whereby the number of hyperboxes can be increased when necessary; hence avoiding the problem of re-training as faced by many neural network models with off-line learning capability [20-22]. Further details of FMM can be found in [19]. Figure 2 shows the boundary line for separating two classes using FMM.

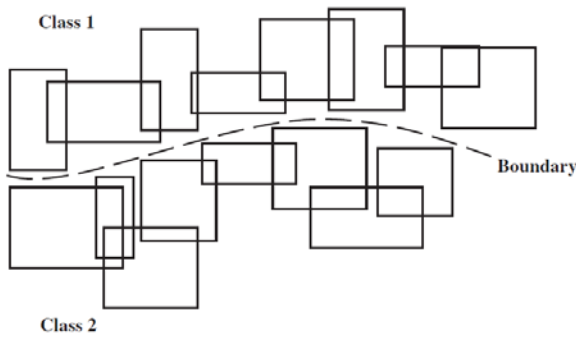


Figure 2. Example Of FMM Hyperboxes Placed Along The Boundary Of A Two-Class Problem

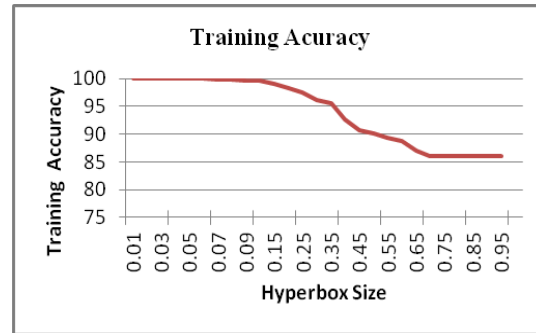


Figure 3. Training Accuracy For The Critical Cancer Dataset.

#### 4. RESULTS AND DISCUSSION

Cervical cancer is a malignant neoplasm arising from cells originating in the cervix uteri. One of the most common symptoms of cervical cancer is abnormal vaginal bleeding, but in some cases there may be no obvious symptoms until the cancer has progressed to an advanced stage.[23] Treatment usually consists of surgery (including local excision) in early stages, and chemotherapy and/or radiotherapy in more advanced stages of the disease.

Cancer screening using the Pap smear can identify precancerous and potentially precancerous changes in cervical cells and tissue. Treatment of high-grade changes can prevent the development of cancer in many victims. In developed countries, the widespread use of cervical screening programs has dramatically reduced the incidence of invasive cervical cancer.

A total of 500 reported cases have been used in the classification process (376 normal, 79 LSIL and 45 HSIL), where 60% of the samples have been used for training, 20% of the samples have been used for prediction and 20% of the samples are used for testing. A total of 10 features were extracted using the AFMKM feature extraction method described above. The reported results are the average of 10 runs with each run having a different data sequence in training.

Figure (3) shows the training accuracy for the FMM NN. As the figure indicates the FMM has a perfect 100 % training accuracy over the hyperbox size ranging from 0.01 to 0.06. However as the hyperbox size starts to increase the training accuracy starts to deteriorate. It reaches a minimum of 86% when the hyperbox size is 0.95.

Figure (4) presents the testing accuracy for the cervical cancer dataset. As the figure indicates the maximum testing accuracy was 75% at hyperbox size of 0.04. This accuracy is not as high as the one for the training accuracy because for such a dataset more training samples are needed since it's a hard classification problem. As the hyperbox size starts to increase the testing accuracy will maintain a value of 74%.

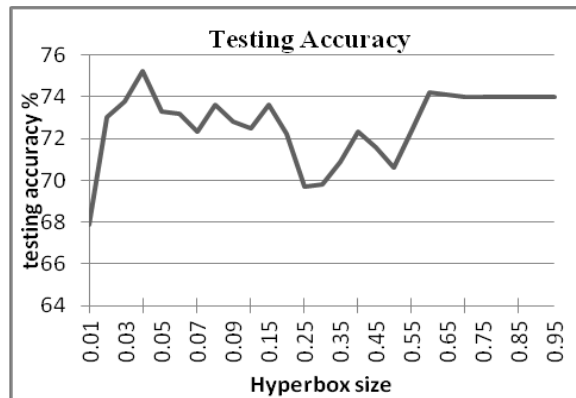


Figure 4. Testing Accuracy For The Cervical Cancer Dataset.

The number of hyperboxes generated is shown in figure 5. As the figure indicates as the hyperbox sizes increase the number of hyperboxes decreases. For instance when the hyperbox size was 0.01 a total of 131 hyperboxes were generated, and when the hyperbox size was 0.95 it was only 3 hyperboxes.

In the testing data set the maximum accuracy (75%) was obtained at hyperbox size of 0.04 using 57 hyperboxes. And the minimum accuracy (74%) was obtained using only 3 hyperboxes. The tradeoff in the testing accuracy is the complexity of the network, for a less complex network the accuracy decreases by 1%, while the network complexity decreases by 94.7%.

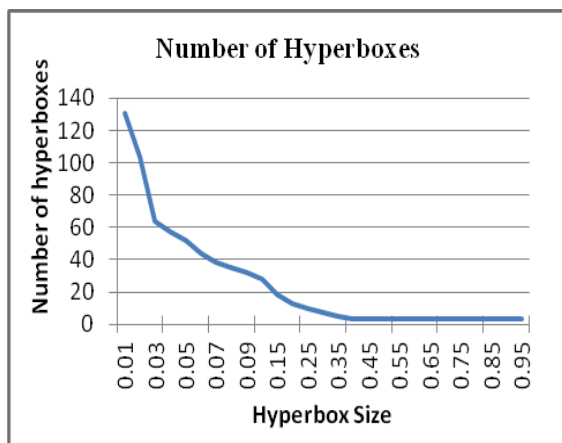


Figure 5. Number Of Hyperboxes Generated.

The accuracy percentage of each class is calculated by dividing the summation of the predicted number of cells over the summation of the original number of samples in each class in all folds.

## 5. CONCLUSION

In this paper a decision support system for cervical data was presented. The classification system consisted of three stages. The first stage was image segmentation using Adaptive Fuzzy Moving K-means. The second stage was feature extracting phase where the feature extraction was used to extract meaningful features from the image data. The extracted features were then fed to the Fuzzy Min-Max neural network. The classification accuracy obtained was 75% using 57 hyperboxes. For a less complex network, 3 hyperboxes were used to obtain a classification accuracy of 74%. The empirical results show that the proposed system can be used to classify cervical cancer datasets.

## REFERENCES:

- [1] C. P. Lim, R. F. Harrison, and R. L. Kennedy, "Application of autonomous neural network systems to medical pattern classification tasks," *Artificial Intelligence in Medicine*, vol. 11, pp. 215-239, 1997.
- [2] G. P. K. Economou, C. Spiropoulos, N. M. Economopoulos, N. Charokopos, D. Lymberopoulos, M. Spiliopoulou, et al., "Medical diagnosis and artificial neural networks: a medical expert system applied to pulmonary diseases," in *Proceedings of the 1994 IEEE Workshop Neural Networks for Signal Processing 1994*, pp. 482-489.
- [3] D. West and V. West, "Model Selection for a Medical Diagnosis Decision Support System: A Breast Cancer Detection Case," *Artificial Intelligence in Medicine*, vol. 20, pp. 183-204, 2000.
- [4] T. Kıyan and T. Yıldırım, "Breast Cancer Diagnosis Using Statistical Neural Networks," in *International XII. Turkish Symposium on Artificial Intelligence and Neural Networks*, 2003, pp. 754-760.
- [5] C. S. Pattichis, C. N. Schizas, and L. T. Middleton, "Neural network models in EMG diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 42, pp. 486-496, 1995.
- [6] Gan G, Ma C, and W. J, *Data clustering: theory, algorithms, and applications*: SIAM, 2007.
- [7] S. Hojjatoleslami and J. Kittler, "Region growing: A new approach," *IEEE Trans. Image Process*, vol. 7, pp. 1079-1084, 1998.
- [8] Y. Chang and X. Li, "Adaptive image region-growing," *IEEE Trans. Image Process.*, vol. 3, pp. 868-872, 1994.
- [9] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Patt. Anal. Mach. Intell.*, pp. 641-647, 1994.
- [10] N. Mat-Isa, "Automated Edge Detection Technique for Pap Smear Images Using Moving K Means Clustering and Modified Seed Based Region Growing Algorithm," *Int. J. Comput. The Internet Manage*, vol. 13, pp. 45-59, 2005.
- [11] W. Zhi and H. Sai-Xian, "An adaptive edge-detection method based on Canny algorithm " *Image Graphics*, p. 8, 2004.
- [12] Z. Yuqian, G. Weihua, and C. Zhencheng, "Edge Detection Based on Multi-Structure Elements Morphology," in *The Sixth World Congress on Intelligent Control and Automation*, 2006, pp. 9795-9798.
- [13] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 24, pp. 881-892, 2002.
- [14] M. Mashor, "Hybrid training algorithm for RBF network," *International Journal of The Computer, The Internet and Manage.*, vol. 8, pp. 50-65, 2000.
- [15] N. Mat-Isa, S. Salamah, and U. Ngah, "Adaptive fuzzy moving K-means clustering algorithm for image segmentation," *IEEE*





- Trans Consum. Electron.* , vol. 55, pp. 2145-2153, 2009.
- [16] N. Mustafa, N. Mat-Isa, M. Mashor, and N. Othman, "Capability Of New Features Of Cervical Cells For Cervical Cancer Diagnostic System Using Hierarchical Neural Network," *International Journal of Simulation, Systems, Science and Technology*, vol. 9, pp. 56-64, 2008.
- [17] A. Weeks, *Fundamentals of Electronics Image Processing*. Washington: Spie Optical Engineering Press, 1996.
- [18] C. Zhang and P. Wang, "A New Method of Color Image Segmentation Based on Intensity and Hue Clustering," presented at the Proc. of 15th Int. Conf. on Pattern Recognition, 2000.
- [19] P. K. Simpson, "Fuzzy min-max neural networks. I. Classification," *Neural Networks, IEEE Transactions on*, vol. 3, pp. 776-786, 1992.
- [20] T. Nakashima, T. Uenishi, and Y. Narimoto, "Off-line learning of soccer formations from game logs," in *World Automation Congress (WAC), 2010*, 2010, pp. 1-6.
- [21] M. Benaim and M. Samuelides, "A rigorous result about the off-line learning approximation," in *Neural Networks, 1991., IJCNN-91-Seattle International Joint Conference on*, 1991, p. 979 vol.2.
- [22] S. M. Odeh and M. Khalil, "Off-line signature verification and recognition: Neural network approach," in *Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on*, 2011, pp. 34-38.
- [23] V. Kumar, A. Abbas, N. Fausto, and R. Mitchell, *Robbins Basic Pathology*, 8 ed.: Saunders Elsevier, 2007.