# Medical Image Recognition Using Artificial Neural Networks

Dr. Jasim M. Al-Samarai, Dr. Hanan A. Akar and Raied S. Basheer

Received on: 22/ 10/2002

Accepted on: 24 / 2 /2004

**Abstract**

In this paper, an integrated artificial neural system has been designed to recognize and classify images. The system consists of the following three stages: (1) converting images to vectors; (2) feature extraction, i.e., entering these vectors into Self Organizing Feature Map (**SOM**) network and (3) classification for which Learning Vector Quantization (**LVQ**) network is used.

The system has been applied to four classes of gray scale medical images to recognize and classify them. A raw image is entered into the designed system without any preprocessing.

**SOM** network reduces image's dimensions and causes the information not to be influenced by the changes occurred in the image.

**LVQ** network has been compared with conventional methods (Minimum Distance Classification (**MD**) and K-Means Clustering Algorithm) considering recognition rate and recognition time. **LVQ** network gives the best results (highest recognition rate about 94% and least recognition time depends on personal computer in use).

**الخلاصة**

تم في هذا البحث تصميم نظام عصبي اصطناعي متكامل لتمييز الصور وتصنيفها، إذ يتكون هذا النظام من ثلاث مراحل، هي:

١. مرحلة تحويل الصور إلى متجهات.

٢ . مرحلة استخلاص الخواص، أي إدخال هذه المتجهات إلى الشبكة العصبية ذات التنظيم الذاتي Self Organizing Feature Map –SOM-).

٣. مرحلة التصنيف، إذ تم استخدام شبكة (Learning Vector Quantization) لهذا الغرض.

تم تطبيق هذا النظام على أربعة أصناف من الصور الطبية ذات التدرج الرمادي من اجل تمييزها وتصنيفها. وتم إدخال الصورة الخام إلى المنظومة المصممة مباشرة من دون الحاجة إلى أية عمليات للمعالجة الأولية. تعمل الشبكة العصبية ذات التنظيم الذاتي على تقليل أبعاد الصورة وإعطاء خاصية عدم تأثر المعلومات إزاء التغيرات الحاصلة في الصورة.

قورنت شبكة Learning Vector Quantization (LVQ) مع طريقتين تقليديتين وهي الـ Minimum Distance Classification and K- Means Clustering Algorithm من حيث معدل وزمن التمييز.

ولقد أعطت شبكة (LVQ) أعلى معدل ٩٤% واقل زمن للتمييز اعتماداً على سرعة الحاسوب المستخدم.

University of Technology, Electrical and Electronic Engineering Department

## 1. Introduction

Artificial Neural Network technology ANN came from current studies of mammalian brains, particularly the cerebral cortex. Artificial neural networks mimic the way that a human brain copes with incomplete and confusing information set [1]. As a result, in analogy to the brain, an Artificial Neural Network ANN is composed of very simple many calculating units (also called neurons) that are connected to form a network. The structure of the network determines whether one neuron may influence another [2].

Humans can easily perform associative memory tasks, recognize speech and faces, understand situations, and so forth, although it is difficult for recently developed digital computers and computational algorithms to do such things. This is due to the difference in basic principles between digital computers and our brain, i.e., serial processing and parallel distributed processing. **ANNs** have been the focus of active research for the last decade. Many works have been done, which showed the effectiveness of Artificial Neural Networks in pattern recognition, optimization problems, control techniques, and so forth [3].

Pattern recognition may be viewed as a two-part process of feature extraction followed by object classification. First, preliminary mapping from an image to a representation space is made, generally in a significant degree of data reduction. Second mapping then operates on this reduced data to produce a classification or estimation in an interpretation space. Both mapping may be performed using ANN models [4]. In fact, artificial neural systems have been referred to as adaptive pattern recognition systems [5].

D. Y. Tasi described a method for automated classification of ultrasonic heart (echocardiographic) images. The feature of the method is to employ an **ANN** trained by genetic algorithms (**GAs**) instead of the back propagation. With the **GA** the optimal weighting coefficients of the **ANN** are determined. Moreover, the method shows a faster convergence for obtaining the optimal solution in **ANN** training. Experiments on different data sets show the superiority of the **GA**-based method over the back propagation for classification [6].

S. Rachid, N. Niki and H. Nishitani, presented a method for automatic segmentation of sputum cells with color images, to develop an efficient algorithm for lung cancer diagnosis based on a Hopfield neural network. To increase the accuracy in segmentation (the regions of interest), a reclassification technique is used to extract the sputum cell regions within the color image and remove those of the debris cells. The former is then given with the raw image to the input of Hopfield Neural Network to make a crisp segmentation by assigning each pixel to label such as background, cytoplasm, and nucleus [7].

S. Loncaric and D. Kovacevic, presented a novel method for automatic segmentation of computed tomography (**CT**) head images of patients having spontaneous intracerebral brain hemorrhage. The method consists of four major phases. The first phase performs brightness normalization of a **CT** image by applying a k-mean clustering algorithm to the pixel brightness values. Feature extraction based on a special receptive field is done in the second phase. The third phase performs a pixel classification by means of a feed forward error back propagation Neural Network [8].

## 2. Requirements of SOM Learning /Feature Extraction Process

During the **SOM** learning of any image, and also before the starting of feature extraction process of any image, there are many image-processing techniques that must be applied to that image in order to simplify using it by the network. Image normalization, segmentation, and rasterization are these image-processing techniques.

### 2.1. Normalization

Consider a vector $x = \{x_i | i = 1, 2, 3, n\}$. The normalized vector x is obtained by dividing each component of x by the square root of the sum of squares of all the components:

$$x = \left\{ x_i / \sqrt{L} \,|_{i=1,2,3,\ldots n}, L = \sum_{i=1}^{n} x_i^2 \right\} \quad \ldots\ldots\ldots(1)$$

$\ldots\ldots (3.8)$

Both the weight vector and the input vector are normalized during the operation of the kohonen feature map [9].

### 2.2. Image Segmentation

Image segmentation is the process of dividing the image into subimages, each of which is considered to be a new separate image. Usually, the resulting image segments are assumed to be connected [10]. Fig. 1 shows the process of image segmentation and the process of image rasterization that would be described in the next section.

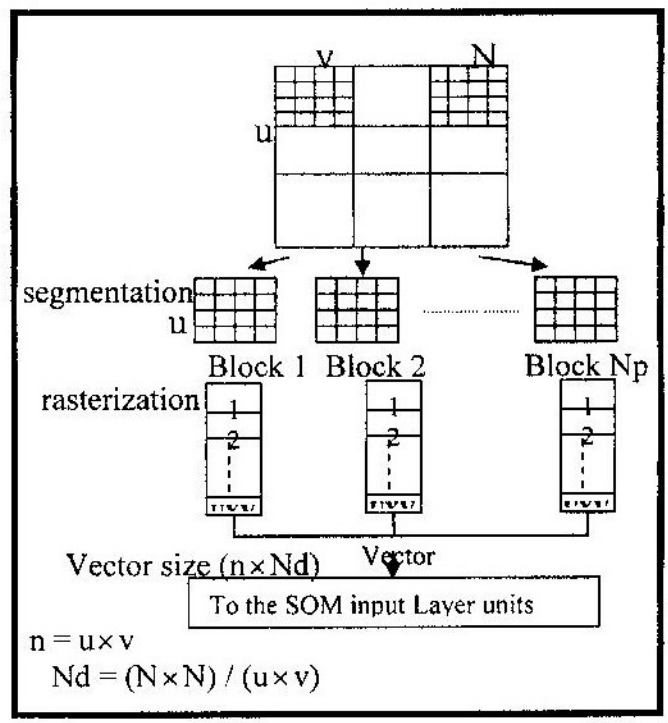


**Fig .1 The Block diagram of the processes of Image Segmentation and Image Rasterization.**

## 2.3. Image Rasterization

Each subimage (segment) is represented, by two-dimensional array, and then each subimage must be converted from a two dimensional block into a one–dimensional vector. This process is called image rasterization. Let us assume that the size of an image is u×v and that g (p, q) is the gray value of the pixel at location (p, q). Then, the image can be represented by a vector:

$$n = uv, (x_1,....,x_i,....,x_n)' \; x = uv \quad .....(2)$$

Where $x_i$=g(p,q). The conversion from the spatial coordinate (p, q) to an index i can be achieved by a transformation [11]:

$$i = (q-1)u + p \qquad .... (3)$$
$$q = 1,.....,v \;\; p = 1,....,u$$

## 3. Kohonen Self Organization Feature Map (SOM)

Self-organizing feature map (SOM) proposed by T. Kohonen is ordered mapping of a high dimensional space of input vector onto a low dimensional discrete topology of neurons. As a simplified definition, we can say that, in a self-organizing feature map, neurons located physically next to each other will respond to classes of input vectors that are likewise next to each other. Large dimensional input vectors are, in a sense, projected down on the two–dimensional maps in a way that maintains the natural order of the input vectors [12].

## 4. Algorithm of SOM

This algorithm includes the following steps [13]:

Step1: Initialize weights.

Initialize weights from input nodes to the output nodes to small random values.

Step 2: Present new input.

Step 3: Compute distance to all nodes.

Compute distance $d_j$ between the input and each output node j using the formula:

$$d_j = \sqrt{\sum_{i=0}^{N}((w_{ij}(t) - x_i(t))^2} \quad .....(4)$$

Where $x_i$ (t) is the input to node i at time t and $w_{ij}$ (t) is the weight from input node i to output node j at time t.

Step 4: Select output node with minimum distance. Select node $j^*$ as that output node with minimum $d_j$.

Step 5: Update weights to node $j^*$ and neighbors.

Weights are updated for node $j^*$ and all nodes in the neighborhood:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)h_{j,j^*}(t)(x_i(t) - w_{ij}(t))$$
$$.....(5)$$

Where: $\eta(t)$ Learning rate parameter.

$h_{j,j^*}(t)$ : Neighborhood function.

Step 6: Repeat by going to step 2.

## 5. Two Phases of the Adaptive Process Ordering and Tuning

Phase 1: ordering phase

This phase lasts for the given number of steps. The neighborhood distance starts as the maximum distance between two neurons, and decreases to the tuning neighborhood distance. The learning rate starts at the ordering phase learning rate and decreases until it reaches the tuning phase.

Phase 2: Tuning phase

This phase lasts for the rest of training or adaptation. The neighborhood distance stays at the tuning neighborhood distance (should include only the winner neuron or its nearest neighbors). The learning rate continues to decrease, but slowly from the tuning phase learning rate. The small neighborhood and slowly decreasing learning rate fine-tune the network, while keeping the ordering learned in the previous phase stable [14].

## 6. Learning Vector Quantization (LVQ)

Learning Vector Quantization (**LVQ**) is a pattern classification method in which each output unit represents a particular class or category (several output units should be used for class). The weight vector for an output unit is often referred to as a reference (or codebook) vector for the class that the unit represents. It is assumed that a set of training patterns with known classifications is provided, along with an initial distribution of reference vectors (each of which represents a known classification). After training, an **LVQ** net classifies an input vector by assigning it to the same class as the output unit that has its weight vector (reference vector) closest to the input vector [15].

## 7. Algorithm of LVQ

This algorithm includes the following steps [15]:

Step 1: Initialize reference vector; weight vectors could be initialized by using one of the following methods:

1- Assign initial weight vectors and classification randomly.

2- Take the first m training vectors (from different m classes) and use them as initial weight vectors; the remaining vectors are then used for training.

3- Use k. means clustering, or the self–originating map to place the initial weight vectors. Each weight vector is then calibrated by determining the input patterns that are closest to it, finding the class that the largest number of these input patterns belongs to, and assigning that class to the weight vector.

Step 2: While stopping condition is false, do steps 3 – 7.

Step 3: For each training input vector x, do steps 4 – 5.

Step 4: Find j so that $\|x - w_j\|$ is a minimum. .... (3.20)

.... (3.21)

Step 5: Update $w_j$ as follows:

If $T=C_j$ then

$$w_j(t+1) = w_j(t) + \alpha_n \left[x - w_j(t)\right]$$

If $T \neq C_j$, then

$$w_j(t+1) = w_j(t) - \alpha_n \left[x - w_j(t)\right]$$

Step 6: Reduce learning rate.

Step 7: Test stopping condition.

The condition may specify a fixed number of iterations (i.e., executions of step 2) or the learning rate reaching a sufficiently small value.

where:

x training vector $(x_1 \ldots, x_i, \ldots x_n)$

T correct category or class for the training vector

$w_j$ weight vector for jth output unit $(w_{1j}, \ldots, w_{ij}, \ldots, w_{nj})$.

$C_j$ category or class represented by jth output unit.

$\|x - w_j\|$ Euclidean distance between input vector and weight vector for jth output unit.

## 8. Implementation of Medical Image Recognition Process on Artificial Neural Network

The flow chart of the medical image recognition process is shown in Fig. 2. Medical image recognition process on **ANN** involves two phases:

1. Training Phase:

Training phase can be summarized by the following steps: .... (6)

Step 1: Read the images pixels from the file, and then divide them to non overlapping blocks (image segmentation).

Step 2: Rasterizing the images by converting each two-dimensional block into one-dimensional vector using Eq.2. .... (7)

Step3: Normalize the images by converting them from the range [0-255] into binary image [0,1].

Step 4: Apply the rasterized vectors of the training image into input layer of **SOM**.

Step 5: Train the **SOM** net with the rasterized vectors of training image.

Step 6: While there are more rasterized vectors of training images go to step 4.

Step 7: Calibrate each weight vector by determining the rasterized vectors of training images that are closest to it, finding the class that the largest number of these rasterized vectors of training images belongs to, and assigning that class to the weight vector.

Step 8: Save the feature weight vectors of **SOM** net after completing its training.

Step 9: Apply the rasterized

vectors of the training image into input layer of **LVQ**.

Step 10: Train the **LVQ** net with the rasterized vectors of training image and feature weight vectors of **SOM** net.

Step 11: While there are more rasterized vectors of training images go to step 9.

Step 12: Save the weight (reference) vectors of **LVQ** net after completing its training.

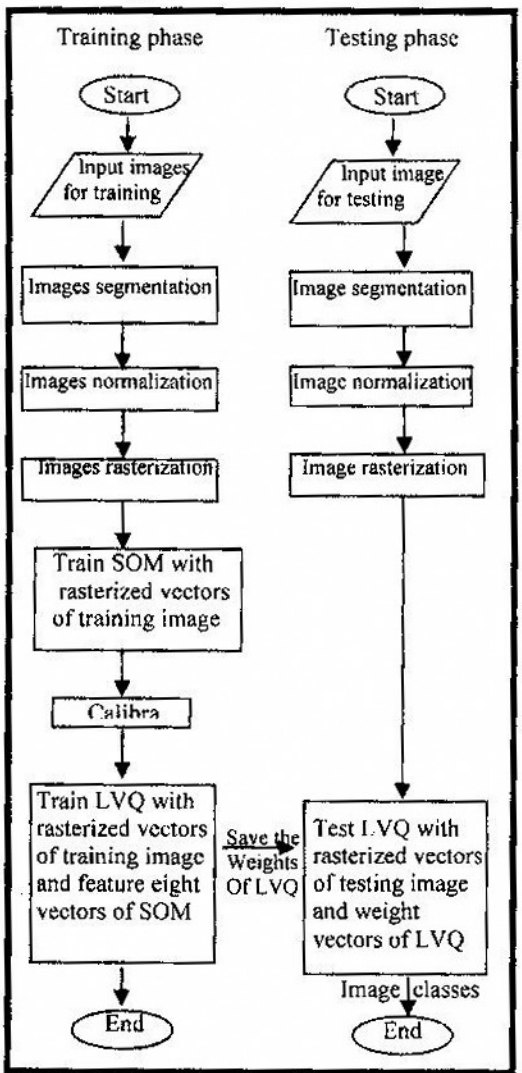


**Fig.2 The flow chart of medical image**

2. Testing Phase:

For each testing image, do steps 1-5:

Step 1: Read the test image pixels from the file, and then divide it to non-overlapping blocks (image segmentation).

Step 2: Rasterizing the test image by converting each two-dimensional block into one-dimensional vector.

Step3: Normalize the test image by converting it from the range [0-255] into the range [0-1].

Step 4: Apply the rasterized vectors of the testing image into input layer of **LVQ**.

Step 5: Find the network node (winner node) which has the minimum distance using Euclidean distance between rasterized input vectors of testing image and each weight vectors of **LVQ** net by using Eq. 7.

## 9. Minimum Distance Classifier (MD)

Classification by minimum distance is one of the earliest and most fundamental supervised algorithms in automatic classification. Let us assume that the minimum distance classification is required to classify patterns into one of the R categories. Each of the R classes is represented by prototype points $P_1$, $P_2$,..., $P_R$ being vectors $x_1$, $x_2$,...., $x_R$, respectively. The Euclidean distance between input pattern x and the prototype pattern vector $z_i$ is expressed by the norm of the vector $x-z_i$ as follows:

$$\|x - z_i\| = \sqrt{(x - z_i)'(x - z_i)} \qquad \ldots. (8)$$

A minimum-distance classifier computes the distance from pattern x of unknown classification to each prototype. Then, the category number of that closest, or smallest distance, prototype is assigned to the unknown pattern. Calculating the squared distance from equation (2.) yields:

$$\|x - z_i\|^2 = x'x - 2z_i'x + z_i'z_i, \quad \ldots\ldots\ldots(9)$$

for i=1,2,….,R

Obviously, the term $x'x$ is independent of i and shows up in each of the R distances under evaluation in equation (9). Thus, it will suffice to compute only R terms, $2z_i'x - z_i'z_i$ for i =1,…., R, in (9), and to determine for which the term $z_i$ takes the largest of all R values. It can also be seen that choosing the largest of the terms $z_i'x - 0.5z_i'z_i$ is equivalent to choosing the smallest of the distances $\|x - z_i\|$. This property can now be used to equate the term with a discriminate function $g_i(x)$:

$$g_i(x) = z_i'x - \frac{1}{2}z_i'z_i, \text{ for i=1,2,….,R} \quad ..(10)$$

It becomes clear now that the discriminate function (10) is of the general linear form, which can be expressed as:

$$g_i(x) = w_i'x + w_{i,n+1}, \text{ for i=1,2,….,R} \quad \ldots. (11)$$

The discriminate function coefficients that are weights $w_i$ can be determined by comparing (10) and (11) as follows [16]:

$$w_i = z_i \qquad \ldots. (12)$$

$$w_{i,n+1} = -\frac{1}{2}z_i'z_i,$$

## 10. K-Means Clustering Algorithm

The most flexible-clustering algorithm used for unsupervised image classification is named K-means clustering. K-means makes the assumption that the number of cluster centers that will be required to adequately represent the sample space is known as priori. The K-means algorithm is implemented in the following steps:

Step 1: Initialize

Choose the number of cluster (K). For each of these K clusters choose an initial cluster center:

$$\{z_1(1), z_2(1), \ldots, z_k(1)\} \quad \ldots\ldots\ldots(13)$$

Where 1 refers to iteration number 1. The starting values can be arbitrary but are generally taken to be the value of the first K of the sample vectors.

Step 2: Distribute samples

Distribute all sample vectors. It is meant that each sample vector $x^{(p)}$ is attached to one of the K clusters according to the following criteria: $x^{(p)} \in S_j(t)$ if

$$\left\|x^{(p)} - z_j(t)\right\| < \left\|x^{(p)} - z_i(t)\right\| \quad \ldots\ldots(14)$$

For all i=1,2...K, $i \neq j$

$S_j(t)$ represents the population of cluster j at iteration t. Ties in equation (14) may be resolved arbitrarily.

Step 3: Calculate new cluster centers:

Using the new cluster membership sets established in step 2, recalculate the position of each cluster center such that the sum of the distances from each member vector to the new cluster center is minimized, specifically we wish to minimize $J_j$ where:

$$J_j = \sum_{x^{(p)} \in S_j(t)} \left\|x^{(p)} - z_j(t+1)\right\|^2 \quad \ldots(15)$$

j=1,2...K

The value of $z_j$ (t+1) which minimizes equation (15) is simply the mean taken over the samples of $S_j$ (t). Therefore, the new cluster center is calculated using equation (16) as follows:

$$z_j(t+1) = \frac{1}{N_j} \sum_{x^{(p)} \in S_j(t)} x^{(p)} \quad (16)$$

Where $N_j$ is the number of sample vectors attached to $S_j$ during step 2.

Step 4: Check for convergence

The condition for convergence is that no cluster center has changed its position during step 3. This condition can be expressed mathematically as follows:

$$z_j(t+1) = z_j(t) \quad \text{where} \quad j=1,2,\ldots,K \quad \ldots(17)$$

If equation (17) is satisfied, then convergence has occurred. Otherwise, iterate by going to step 2 [17].

## 11. Data Base

The samples of training and testing images are collected by taking images for four diseases (Bronchogenic Carcinoma, Meningioma, Renal Cell Carcinoma and Ankle Fracture). 36 images are collected with different sizes. The images change their sizes to (64×64) pixel. Then, the colors information removes from the

images. The resultant is gray images with 256 gray scales.

In training process, 5 images (for each of Bronchogenic Carcinoma and Renal cell Carcinoma) and 4 images (for each of Meningioma and Ankle Fracture) are used for training the network. Namely, 18 images are used to train the network. Another 5 images (for each of Bronchogenic Carcinoma and Renal Cell Carcinoma) and another 4 images (for each of Meningioma and Angle Fracture) are used for testing the network after training. Namely, 18 images are used to test the network (Fig. 3).

## 12. Experimental Results

Table (1) shows the recognition rate of medical image recognition system as the size of **SOM** (m) is varied. The size of **SOM** is chosen by trail and error. Each application has its own size of **SOM**. It is clear from the table (4.1) that the low value of m leads to mixing between the classes and decreases the recognition rate of the system. The high value of (m) leads to increase specialization of the network and decrease the generalization of the network. Medium value (m=81) among these values will give the trade-off between them.

Table (2) shows the generalization of the system as the number of classes increases using the same size of SOM (m=25, m1=m2=5, m=m1 × m2). It is clear from the table that the recognition rate of the system decreases using the same size of SOM network. Increasing the number of classes leads to increase the different features that they carry. This requires increasing the size of network and number of training images.

Table (3) shows the recognition rate of the system as the dimensions of **SOM** vary. With **SOM** of two-dimensions, the size of **SOM** is 9 nodes per dimension. The initial width of neighborhood topological function $\sigma_o$ is 4. For **SOM** of one-dimension, the number of nodes is 81 and the initial width is 40.

Table (4) shows the recognition rate of the system as the initial value of learning rate parameter $\eta_o$ varies. The test values are between 0 and 1. The best performance of the network for initial value of learning rate parameter is 0.1. Large values of $\eta_o$ lead to instability in the network and unsatisfactory learning. Small values of $\eta_o$ lead to excessively slow learning. Fig.4a shows exponential decay of learning rate parameter $\eta(t)$ during ordering phase (first 1000 iterations). During that phase, the learning rate parameter $\eta(t)$ starts with an initial value $\eta_o$=0.1 and then decreases to 0.037. Fig.4b shows linear decay of learning rate parameter in convergence phase (from 1000 to 3000 iterations). During this phase, the learning rate parameter de$\eta(t)$creases linearly to 0.005. Table (5) shows recognition rate of the

system as the initial width of topological neighborhood function varies. In Table (5a), the nodes of **SOM** arrange in two-dimensions. Number of nodes per dimension is 9 nodes (m=81,m1=m2=9). In Table (5b), the nodes in SOM arrange in one dimension. The number of nodes is 81 nodes (m=81). It is clear from the tables that when $\sigma_0$ decreases, the topological neighborhood function decreases and the numbers of excited neuron decrease. The features do not organize in geometrically relevant way within the output layer. As a result, the recognition rate of system will decrease. Fig.5 shows exponential decay for width $\sigma(t)$ during ordering phase (first 1000 iterations). The width starts with an initial value $\sigma_0$=4 and then decrease to 1. In convergence phase, the value of width is zero.

During ordering phase, the topological ordering of the weight vectors takes place; therefore, the initial width $\sigma_0$ was 4.

The second phase (convergence phase) needs to fine-tune the feature map and; therefore, provides an accurate statistical quantification of the input space. During this phase, the neighborhood function $h_{j,j^*}$ contains only the winner neuron or its neighbors; therefore, the initial width $\sigma_0$ was zero.

Table (6) shows recognition rate of training images when LVQ and conventional methods (minimum distance classifier (MD) and K-means clustering algorithm) are used for classification process. In each case the SOM is used for feature extraction process.

Table (7) shows recognition rate of testing images when **LVQ** and conventional methods MD and K-means clustering algorithm are used for classification process. In each case, **SOM** is used for feature extraction process. From the table it is clear that the **LVQ** gives the best recognition rate for testing images. The recognition rate for testing images using **LVQ** is 94%, using **MD** is 89% and using K-means clustering is 63.15%. K-means clustering algorithm is influenced by the number of cluster centers specified, the choice of initial cluster centers, the order in which the samples are taken and the geometrical properties of the data. The minimum distance classification is influenced by the order in which the samples are taken and the geometrical properties of the data.

Table (8) shows recognition time of medical image recognition system when LVQ and conventional methods (MD and K-means clustering algorithm) are used for classification process. In each case, SOM is used for feature extraction process. From the table it is clear that the best recognition time is obtained when LVQ is used. The worse recognition time is obtained when K-means clustering is used. In each case, the number of classes is 4, the **SOM** has two dimensions of nodes and the structure of **LVQ** is 256

nodes in input layer and 14 nodes in output layer.

## 13. Conclusions

The **SOM** network used for feature extraction maintains recognition and the generalization of medical image recognition system.

The recognition rate of medical image recognition system can be improved by using an appropriate size and suitable parameters of **SOM** network.

Trail and error method used in determining the size and parameters for **SOM** network gives acceptable recognition rate after completing the design of the network.

**SOM** network of one dimension gives the same recognition rate of **SOM** of two dimensions provided that the initial width of topological neighborhood function is very high.

Medical image recognition system needs to learn more than one image per disease in order to increase the ability of recognition.

The use of **ANN** is proven to be outstandingly faster than the conventional methods, with an acceptable degradation in the recognition rate.

The selection of training area for **ANN** should be made under considerations different from those of the conventional methods. In the **ANN**, the samples are preferred to be selected from the class border, as this will provide more flexibility and accuracy for the **ANN**.

## References

[1] S. P. Chitra, "Use Neural Network for Problem Solving", Measurements and Control, April, 1993, pp.43-52.

[2] J. Branke, "Evolutionary Algorithms for Neural Network Design and Training", Technical Report, No.322, University of Karlsruhe, Institute AIFB, 1995, pp.1-21.

[3] M. Fukumi, S. Omatu and Y. Nishikawa, "Rotation-Invariant Neural Pattern Recognition System Estimating a Rotation Angle", IEEE Transactions on Neural Networks, Vol. 8, No. 3, May, 1997, pp. 566-579.

[4] M. B. Reid, L. Spirkovska and E. Ochoa, "Simultaneous Position, Scale, and Rotation Invariant Pattern Classification Using Third Order Neural Networks", The International Journal of Neural Networks and Applications, Vol.1, No. 3185, July, 1989, pp. 154-158.

[5] L. I. Burke, "Introduction to Artificial Neural Systems for Pattern Recognition", Computers Ops Res, Vol. 18, No. 2, April, 1991, pp. 211-220.

[6] D. Y. Tsai, "Classification of Heart Diseases in Ultrasonic Images Using Neural Networks Trained by Genetic Algorithms", International Conference on Signal Processing Proceedings, Beijing, China, October, 1998, pp. 959-962.

[7] S. Rachid, N. Niki, H. Nishitani, S. Nakamura and S. Mori, "Segmentation of Sputum Color Image for Lung Cancer Diagnosis Based on Neural Networks", Image Analysis and Processing, 9th International Conference, Florence, Italy, September, 1997, pp. 461-468.

[8] S. Loncaric and D. Kovacevic, "A Method for Segmentation of CT Head Images", Image Analysis and Processing, 9th International Conference, Florence, Italy, September, 1997, pp. 388-395.

[9] W. B. Rudiger and R. Michael, "Noise Suppressing Sensor Encoding and Neural Signal Orthonormalization", IEEE Transactions on Neural Network, Vol.9, No.4, July, 1998, pp. 613-618.

[10] J. Teuber, "Digital Image Processing", Prentice Hall, New York, 1993.

[11] K. J. Cio and I. Shin, "Image Recognition Neural Network: IRNN", Neurocomputing, Vol. 7, No. 2185, March, 1995, pp. 159-185.

[12] R. I. Chang and P. Y. Hsiao, "Fast Self Organization by Query Based Algorithm and its Applications", International Symposium on Speech, Image Processing and Neural Networks, Hong Kong 88, 13-16 April, 1994, pp. 85-88.

[13] R. P. Lippmann, "An Introduction to Computing with Neural Nets", IEEE Assp Magazine, April, 1987, pp. 4-22.

[14] H. Demuth and M. Beale, "Neural Network Toolbox for Use with Matlab, the Math Works Inc, MA, USA, 1998.

[15] L. Fausett, "Fundamentals of Neural Network Architectures, Algorithms, and Applications", Prentice Hall International, Inc, 1994.

[16] Pandya and Macy, "Pattern Recognition with Neural Network in C++", CRC. Press, 1996.

[17] J. M. Zurada, "Introduction to Artificial Neural Systems", Jaico Publishing House, 1996.

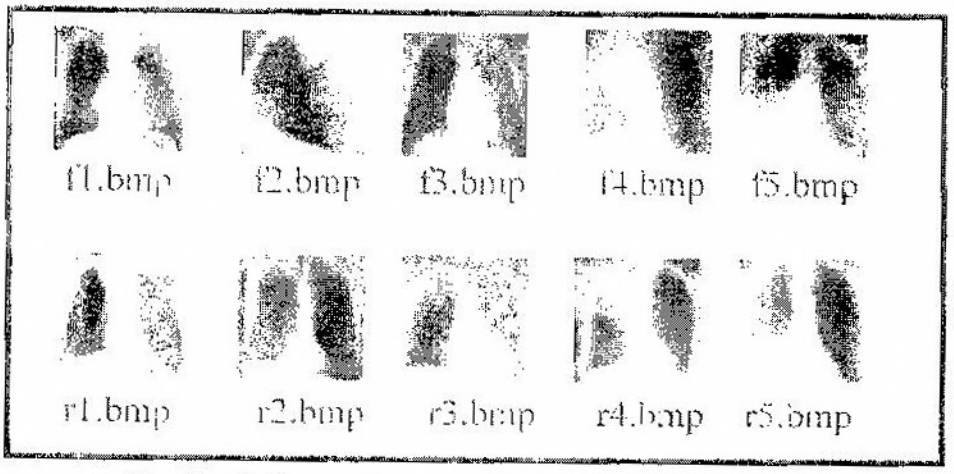


**Fig.3-a1 Bronchogenic Carcinoma (class 1).**

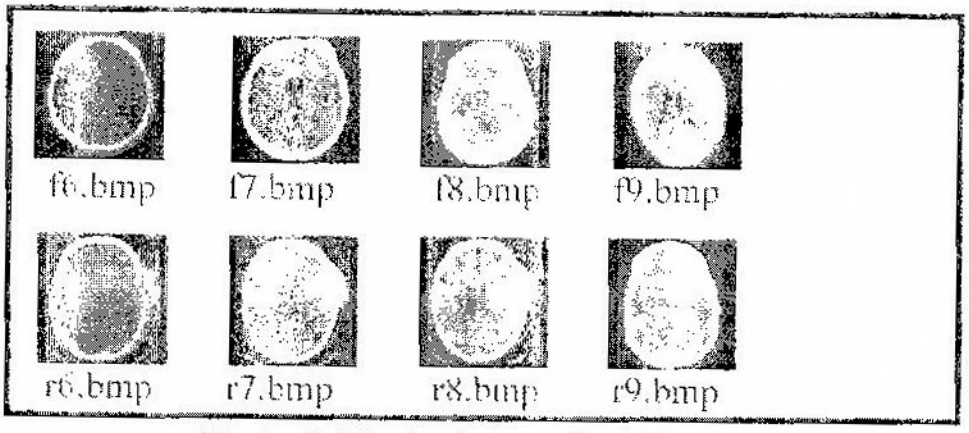


**Fig 3-a2 Meningioma (class 2).**

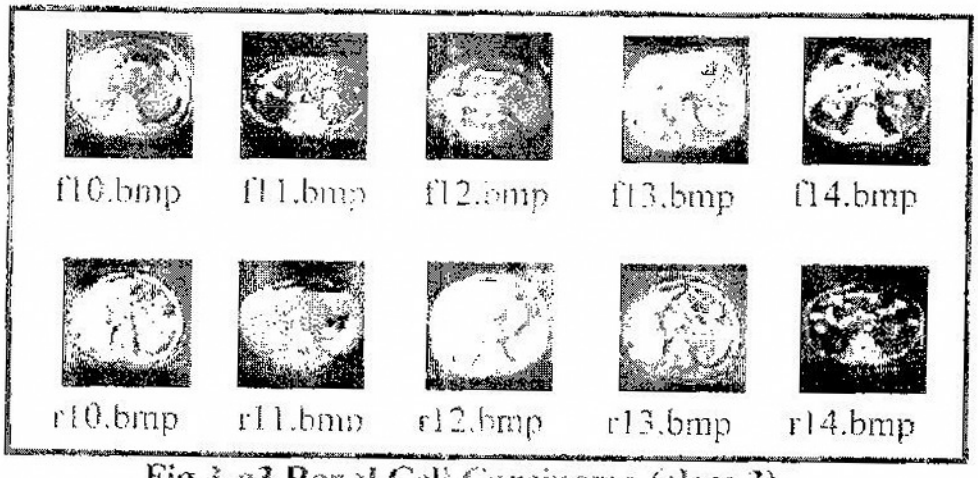


**Fig.3-a3 Renal Cell Carcinoma (class 3).**

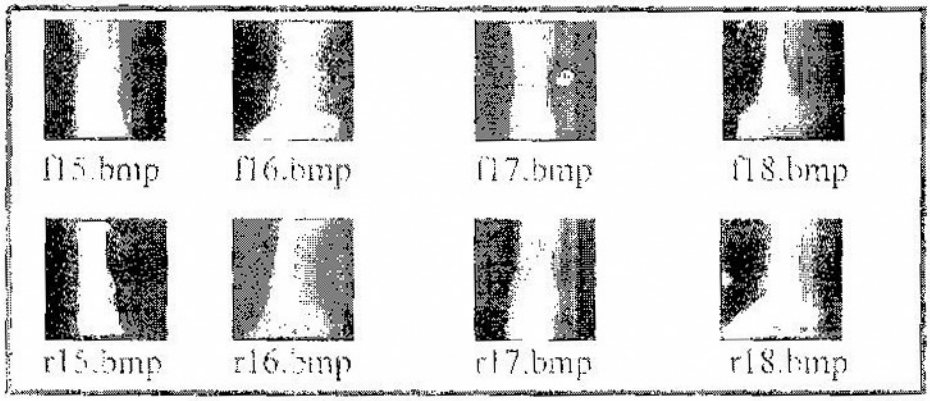


**Fig 3-a4 Ankle Fracture (class 4).**

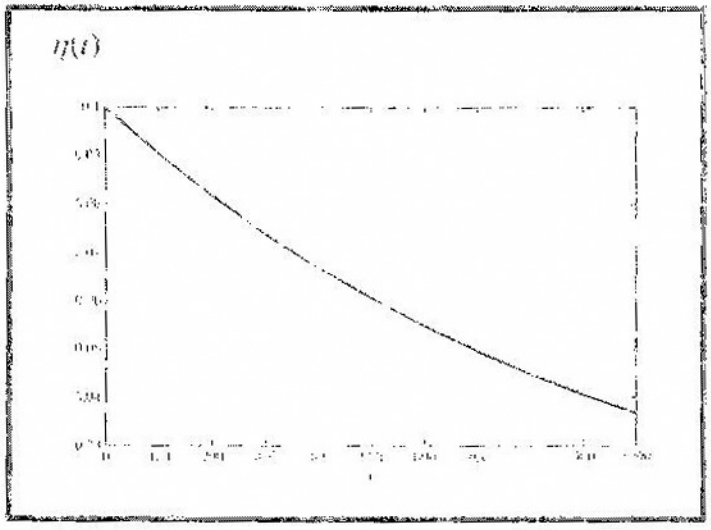


**Fig.4a Exponential decay of learning rate parameter η(t) during ordering phase.**

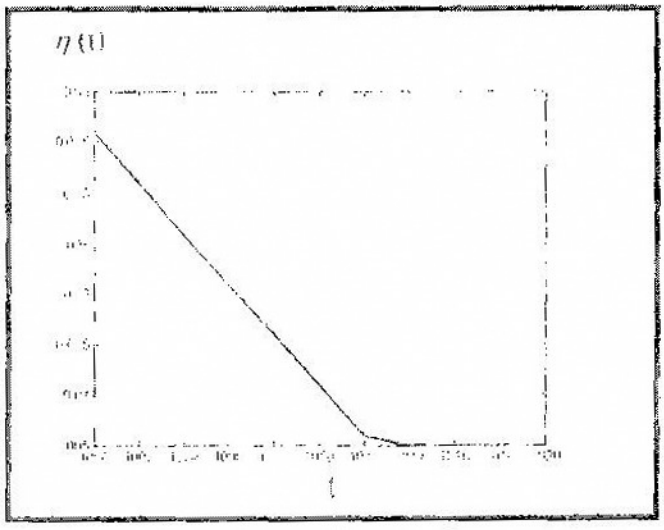


**Fig.4b Linear decay of learning rate parameter η(t) during convergence phase.**

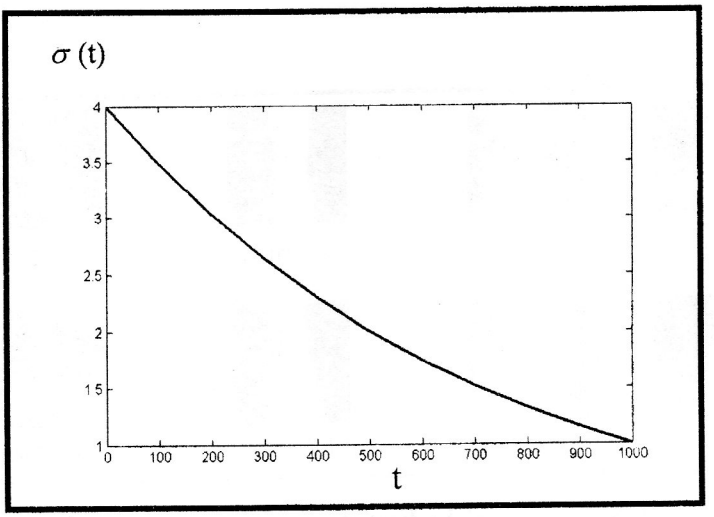


**Fig.5 Exponential decay for width of topological neighborhood function during ordering phase.**

**Table (1) Recognition rate of medical image recognition system with varying the size of SOM.**

| SOM Size ( m ) m=m1 × m2 | m=49 m1=m 2=7 | m=81 m1=m2=9 | m=121 m1=m2=11 | m=225 m1=m2=15 |
|---|---|---|---|---|
| Recognition Rate ( R% ) | 89 | 94 | 94 | 84.2 |

**Table (2) Recognition rate of medical image recognition system as number of classes are increases using the same size of SOM network (m=25, m1=m2=5, m=m1 × m2).**

| Number Of classes | 2 | 3 | 4 |
|---|---|---|---|
| R% | 100 | 100 | 89 |
| | | | |

**Table (3) Recognition rate of the medical image system as the dimensions of SOM vary.**

| SOM Dimensions | 1 | 2 |
|---|---|---|
| R% | 94 | 94 |

**Table (4) Recognition rate of medical recognition system as the initial value of learning rate parameter $\eta_0$ varies.**

| Initial Value of Learning rate parameter $\eta_0$ | 0.05 | 0.08 | 0.1 | 0.5 | 0.9 |
|---|---|---|---|---|---|
| R% | 78.94 | 89.47 | 94 | 78.94 | 78.94 |

**Table (5) Recognition rate of the system as the initial width of topological neighborhood function $\sigma_0$ varies (a) for two dimensions. (b) for one dimensions.**

| Initial width of SOM ($\sigma_0$) | 0 | 2 | 4 |
|---|---|---|---|
| R% | 26.32 | 84.21 | 94 |

**Table (5a)**

| Initial width of SOM ($\sigma_0$) | 5 | 20 | 40 |
|---|---|---|---|
| R% | 63.15 | 78.94 | 94 |

**Table (5b)**

**Table (6) Recognition rate of training medical images as classification method varies.**

| Classification Method | LVQ | MD | K-means |
|---|---|---|---|
| R% | 100 | 100 | 78.94 |

**Table (7) Recognition rate of testing medical images as classification method varies.**

| Classification Method | LVQ | MD | K-means |
|---|---|---|---|
| R% | 94 | 89 | 63.15 |

**Table (8) Recognition time of testing medical images as classification method varies.**

| Classification Method | LVQ | MD | K-means |
|---|---|---|---|
| Recognition Time (sec) | 0.28 | 0.33 | 4.23 |