



Image Retrieval System using Fuzzy-Softmax MLP Neural Network

KEYWORDS

Image retrieval system; Discrete wavelet transform, Information gain, Neural Network, FS-Multilayer perceptron

G.Ravi Teja

K L University, M. Tech (Computer Science),
Vaddeswaram, Andhra Pradesh, India

M.R. Narasinga Rao

K L University, Professor, Vaddeswaram, Andhra
Pradesh, India

ABSTRACT Many databases contain huge volume of data, mostly in the form of digital images. Digital images such as vector or raster type or medical images such as X-Rays, MRI, and CT are extensively used in research, diagnosis and planning treatment schedule. Large medical institutions produce gigabits of image data every month. For effective utilization of medical images from the archives for diagnosis, research and educational purpose, efficient image retrieval system is essential. Image retrieval systems extract features in the image to a feature vector and use similarity measures for retrieval of images from group of images. Thus the effectiveness of the image retrieval system solely depends upon the feature selection and the way they are classified. The aim of this paper is to implement a novel feature selection mechanism using Discrete Wavelet Transforms (DWT) with Information Gain for feature reduction. Classification results obtained from the proposed method using existing classifiers is compared with the proposed Neural Network model. Results obtained show that the proposed Neural Network classifier outperforms conventional classification algorithms and multi layer perceptron neural network.

1. Introduction

Visual information has been extensively used in the areas of multimedia, medical imaging and other numerous applications. Management of this visual information is challenging as the quantity of data available is very huge and growing exponentially. Digital images play a vital role in diagnosis and treatment schedule planning of a disease. It provides visual information for diagnosis, progress in treatment. Image retrieval of digital medical images from archives is a challenge that is widely researched. Textual annotations of images were the basis on which images were retrieved during the early 80s [1,2]. The images were retrieved using semantic queries. A system which can automatically classify images and retrieve images based on query image is required for efficient use of the archived medical data images. Earlier works in literature include use of visual features with text annotation for image retrieval [3,4]. Modern radiology techniques like CT, PET, MRI, X-Rays, provide essential information required for diagnose and plan treatments to the medical professionals [5]. Thus, efficient storage and image retrieval system for utilization of the images for diagnosis, research and educational purposes are required. Image retrieval based on visual features or image based query wherein the retrieval system responds to a query image by retrieving query similar images from the archive. In this retrieval system, the images in the database are preprocessed automatically to extract features and on the basis of the features, the images are classified. The query image is similarly preprocessed to extract features and based on the similarity measures appropriate images are retrieved from the database. Figure 1 show the block diagram of an image retrieval system.

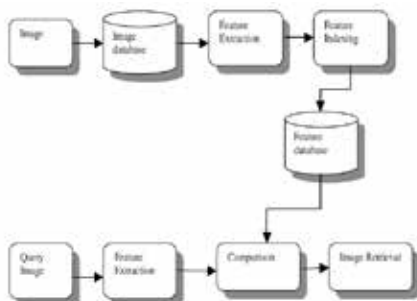


Figure 1: Overview of Image retrieval process

Image retrieval plays a fundamental role in handling large amount of visual information in medical applications [6]. An effectiveness of an image retrieval system depends on:

- Multi magnitude feature vector formed using information extracted from images
- Computing distance metrics
- Identify the images in database with lowest distance metrics from the query image
- How to select features to achieve highest discrimination,
- Combining them effectively,
- Application of proper distance metrics,
- Location of optimal classifier configuration for classification problems,
- scaling/adapting classifier when many classes/features are incrementally introduced and finally,
- Training classifier to maximize classification accuracy.

Generally features such as color, texture, shape, size and spatial relationship are used for classifying images. In medical imaging, color is an effectively used feature; in fields of dermatology [8] color is extensively used as a feature. MRI images, X-Rays are in grey scale, thus color may not be an effective feature for image retrieval. Similarity measures computed from low level image features are mainly used for image retrieval. To automatically categorize medical images, data mining techniques such as decision tree, Bayesian network, Neural networks, Support vector machines are widely used [9].

In this paper it is proposed to extract the feature vector from medical images using Discrete Wavelet Transform (DWT) and feature reduction using Information Gain (IG). The proposed Fuzzy Softmax Multilayer perceptron (FS-MLP) Neural Network is used to classify the obtained feature vectors for the given class.

2. Previous Research

Rigau, et al., [9] proposed a two-step mutual information-based algorithm for medical image segmentation. In the first step, binary space partition splits the image into relatively homogeneous regions. Second step involves clustering around the histogram bins of the partitioned image. The clustering is done by minimizing the mutual information loss of the reserved channel. The proposed algorithm preprocesses the images for multimodal image registration. The multimodal

image registration integrates the information of different images of the same or different subjects. Experimental results using proposed algorithm on different images show that the segmented images perform well in medical image registration using mutual information-based measures.

K. Rajkumar et al., [11] proposed a two step medical image retrieval framework to retrieve similar images. A content based image retrieval framework based on PCA and wavelet was proposed. Wavelet filtering process is used to create a subset of images. The energy efficient wavelet decomposition is used to decompose images and corresponding energies were extracted. The retrieval system uses this subset to search for similar images. Further reduction of dimensions is obtained by applying PCA to the extracted features. Similarity matches of query image and database image was obtained using Euclidean distance. The calculated eigen vectors and the similarity measures were applied to retrieve the medical images. Due to the reduction of searching space efficiency and retrieval accuracy is improved. Experiments conducted using 200 medical images showed that the proposed method has better retrieval accuracy in terms of recall rate and precision.

Kambhatla, et al., 1997 [12] developed local nonlinear extensions of PCA for dimension reduction. The algorithm was applied on both speech and image data. The proposed algorithm is fast to compute and provides accurate representations of the data. PCA and neural network implementations of non-linear PCA were used to compare with the proposed algorithm. Results showed that nonlinear PCA performed better than PCA and the proposed local linear techniques perform better than neural network implementations.

Park, et al., 2003 [13] proposed a method of image classification using neural network. In the preprocessing stage, the object region is extracted using region segmentation techniques. The images are transformed using wavelet transforms. Shape based texture features are extracted from transformed images and are used for classification of the images. The neural network was trained using back propagation learning algorithm. The training of neural network was done using 300 training data composed of 10 images from each of 30 classes. Results showed that the classification rates of 81.7% accuracy were achieved.

Su, et al., 2003 [14] proposed a new feedback approach with progressive learning capability. The proposed approach is based on a Bayesian classifier. The positive and negative feedback are treated with different strategies. The positive examples are used for refining image retrieval results and negative images are used to modify the ranking of the retrieved images. The images are retrieved by estimating Gaussian distribution of the positive examples that represents the desired images for a given query. Bayesian network is used to re-rank the images in the database. PCA is used to update the feature subspace during the feedback process thus reducing sub-space dimensionalities. Thus the feedback process improves the retrieval process. Experimental results show that the proposed method improves the speed, memory and accuracy of the retrieval process.

3. Research Method

This section briefly introduces to Discrete Wavelet Transform (DWT), Information Gain (IG) and the Multi Layer Perceptron (MLP) Neural Network.

3.1 Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. In other words, this transform decomposes the signal into mutually orthogonal set of wavelets, which is the main difference from the continuous wavelet transform (CWT), or its implementation for the discrete time series sometimes called discrete-

time continuous wavelet transform (DT-CWT).

The feature vector from each image was extracted using the discrete wavelet transform. Pixels which are one length away from each other are selected. The algorithm pseudo is given below:

1. Compute Image size MxN
2. For each alternate value 'i' in array M and array size less than M or M+1
3. For each alternate value 'j' in array N and array size less than N or N+1
4. Compute DWT(array[xi,yj])
5. Store computed value in one dimensional array
6. Repeat from step 1 till all images are computed

$$W_{j_0,k} = \frac{1}{\sqrt{M}} \sum_n f[n] \Phi_{i_0,k}[n]. \tag{1}$$

$$W_{j,k} = \frac{1}{\sqrt{M}} \sum_n f[n] \Psi_{i,k}[n] \quad j \geq j_0. \tag{2}$$

Discrete wavelet transform is preferred over Fast Fourier transform due to its simplicity and the reduced time to compute the image coefficients.

3.2 Haar Wavelet

In mathematics, the Haar wavelet is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal function basis. The Haar sequence is now recognised as the first known wavelet basis and extensively used as a teaching example.

The Haar sequence was proposed in 1909 by Alfréd Haar. Haar used these functions to give an example of a countable orthonormal system for the space of square-integrable functions on the real line. The study of wavelets, and even the term "wavelet", did not come until much later. As a special case of the Daubechies wavelet, the Haar wavelet is also known as D2.

The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines.

The Haar wavelet's mother wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Its scaling function $\phi(t)$ can be described as

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

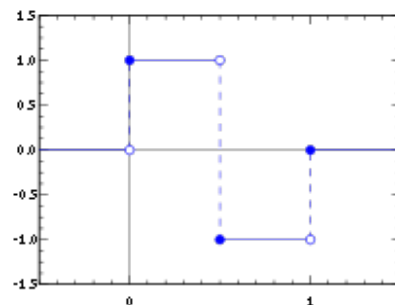


Figure 2: The Haar wavelet

To calculate the Haar transform of an array of n samples:

1. Find the average of each pair of samples. ($n/2$ averages)
2. Find the difference between each average and the samples it was calculated from. ($n/2$ differences)
3. Fill the first half of the array with averages.
4. Fill the second half of the array with differences.
5. Repeat the process on the first half of the array. (The array length should be a power of two)

Two samples, l and r , can be expressed as an average, a , and a difference, d , like in mid-side coding:

$$a = (l + r) / 2 \tag{5}$$

$$d = a - l = r - a \tag{6}$$

This is reversible:

$$l = a - d$$

$$r = a + d$$

3.3 Information Gain

The main aim of information gain criteria is to discover the amount of unique information is added by a feature to the whole feature set. A feature's information gain f can be computed as $F(SUf) - F(S)$, where $F(.)$ is the evaluation criterion and S the selected subset of features. The feature with greater information gain is preferred. Bayes error rate, conditional probability, and information gain are a little information gain criteria. Quinlan suggested a classification algorithm called ID3 that introduced the information gain concept. Information gain is a measure based method, used for selecting best split attributes in decision tree classifiers and indicates the extent to which data's entropy is reduced. It also identifies values of each particular attribute. Each feature basis gets an information gain value, which is used to decide whether a feature is selected or deleted. Hence, a threshold value for feature selection must be established first; a feature is chosen when its information gain value is bigger than the threshold value.

Let a set of s instances be set A and let B be the set of k classes. Let $P(B_i, A)$ be the fraction of the examples in A that have class B_i , then, the expected information for the class membership is given by:

$$Info(A) = - \sum_{i=1}^k P(B_i, A) * \log(P(B_i, A)) \tag{7}$$

If a particular attribute X has y distinct values, anticipated information for the decision tree with X as root is the weighted sum of expected information of subsets of X according to distinctive values. Let A_i be the set of instances whose attribute value of X is X_i .

$$Info_X(A) = - \sum_{i=1}^y \frac{|A_i|}{A} * Info(A_i) \tag{8}$$

Then, difference between $Info(A)$ and $Info_X(A)$ provides information gained by partitioning A according to testing X .

$$Gain(X) = Info(A) - Info_X(A)$$

The higher the information gain, the higher the chances of getting pure classes in a target class if the split is based on the variable with the highest gain.

Information gain selects the feature vectors which are essential for the classification process. On the computed coefficient from DST, the information gain can be computed based on the class attribute. The information gain that has to be computed for an attribute X whose class attribute Y is given by the conditional entropy of Y given X , $H(Y|X)$ is

$$I(Y; X) = H(Y) - H(Y|X) \tag{9}$$

The conditional entropy of Y given X is

$$H(Y|X) = - \sum_{j=1}^i P(X = x_j) H(Y|X = x_j) \tag{10}$$

3.4 Multilayer Perceptron (MLP)

Multilayer perceptron (MLP) is the most favored supervised learning network model. The neural network consists of one input layer, one or more hidden layer and an output layer. The connections between the layers are typically formed by connecting each of the nodes from a given layer to all neurons in the next layer. During the training phase each connection's scalar weight is adjusted. The outputs are got from the output nodes of network. The feature vector x is input at the input layer and the output represents a discriminator between its class and all of the other classes. In training, the training examples are fed and the predicted outputs are computed. The output is compared with the target output and error measured is propagated back through the network and the weights are adjusted.

The training set of size m can be represented as $T_M = \{(x_1, y_1), \dots, (x_m, y_m)\}$ where $x_i \in R^a$ are the input vectors of dimension a and $y_i \in R^b$ are the output vectors of dimension b and R represents the set of real numbers. Let f_w represent the function with weight w for the neural network. Supervised learning adjusts the weight such that:

$$F_w(X_i) = y_i; \forall (X_i, y_i) \in T_M \tag{11}$$

After the Neural network is trained with all feature vectors, and is tested on new samples its output will be correct to a certain extent.

4. Proposed FS-MLP

The proposed neural network Fuzzy Softmax Multi Layer Perceptron (FS-MLP) Neural Network improves the classification accuracy of traditional MLP Neural Network model by introducing a fuzzy hidden softmax layer. The construction of the proposed model is given in table I.

Input Neurons	36
Output Neurons	1
Number of Hidden Layer	2
Transfer Function of first Hidden Layer	Fuzzy Softmax
Learning Rule of first hidden layer	Back propagation Algorithm
Transfer Function of Second Hidden Layer	Sigmoid
Learning Rule of first hidden layer	Back propagation Algorithm

Table I: Parameters Used In the Proposed Fs-Mlp.

The activation function in a neural network controls the amplitude of the output such that the range of output is between 0 and 1 or -1 to 1. Mathematically the interval activity of the neuron can be shown to be:

$$v_k = \sum_{j=1}^p w_{j,k} x_j \tag{12}$$

Where x_i is the input and w_{jk} is the weights. The output of the neuron, y_k would therefore be the outcome of some activation function on the value of v_k . The most common type of activation used to construct the neural network is the sigmoid function.

A sigmoid activation function uses the sigmoid function to determine its activation. The sigmoid function is given as:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{13}$$

The softmax activation function, (Bridle, 1990), applied to the network outputs ensures that the outputs conform to the mathematical requirements of multivariate classification probabilities [15]. If the classification problem has C classes

or categories, then each category is modeled by one of the network outputs. If Z_i is the weighted sum of products between its weights and input then for the i -th output, i.e.,

$$Z_i = \sum_j w_{ij} y_j \quad (14)$$

Then

$$\text{Softmax}_i = \frac{e^{-Z_i}}{\sum_{j=1}^C e^{-Z_j}} \quad (15)$$

The softmax activation function ensures that all outputs conform to the requirements for multivariate probabilities. That is,

$$0 < \text{softmax}_i < 1, \text{ for all } i=1,2,\dots,C \text{ and}$$

$$\sum_{i=1}^C \text{softmax}_i = 1 \quad (16)$$

3.5 Back Propagation Algorithm

The standard way to train a multi layer perceptron is using a method called back propagation. This is used to solve a basic problem called assignment of credit, which comes up when we try to figure out how to adjust the weights of edges coming from the input layer. Recall that in the single layer perceptron, we could easily know which weights were producing the error because we could directly observe the weights and output from those weighted edges. However, we have a new layer that will pass through another layer of weights. As such, the contribution of the new weights to the error is obscured by the fact that the data will pass through a second set of weights [16].

a. First case: function composition

In the feed-forward step, incoming information into a unit is used as the argument for the evaluation of the node's primitive function and its derivative. In this step the network computes the composition of the functions f and g . Figure 3 shows the state of the network after the feed-forward step. The correct result of the function composition has been produced at the output unit and each unit has stored some information on its left side.

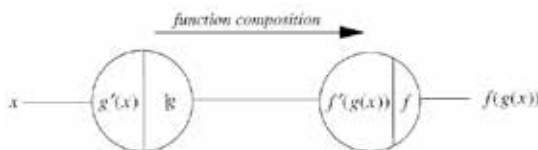


Figure 3: Result of the feed-forward step

In the backpropagation step the input from the right of the network is the constant 1. Incoming information to a node is multiplied by the value stored in its left side. The result of the multiplication is transmitted to the next unit to the left. We call the result at each node the traversing value at this node. Figure 4 shows the final result of the backpropagation step, which is $f'(g(x))g'(x)$, i.e., the derivative of the function composition $f(g(x))$

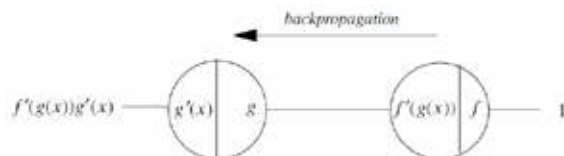


Figure 4: Result of the backpropagation step

Implemented by this network. The backpropagation step provides an implementation of the chain rule. Any sequence of function compositions can be evaluated in this way and its derivative can be obtained in the backpropagation step. We

can think of the network as being used backwards with the input 1, whereby at each node the product with the value stored in the left side is computed.

b. Second case: function addition

The next case to consider is the addition of two primitive functions. Figure 5 shows a network for the computation of the addition of the functions f_1 and f_2 . The additional node has been included to handle the addition of the two functions. The partial derivative of the addition functions with respect to any one of the two inputs is 1. In the feed-forward step the network computes the result $f_1(x) + f_2(x)$. In the backpropagation step the constant 1 is fed from the left side into the network. All incoming edges to a unit fan out the traversing value at this node and distribute it to the connected units to the left. Where two right-to-left paths meet, the computed traversing values are added. Figure 6 shows the result $f_1'(x) + f_2'(x)$ of the backpropagation step, which is the derivative of the function addition $f_1 + f_2$ evaluated at x . A simple proof by induction shows that the derivative of the addition of any number of functions can be handled in the same way.

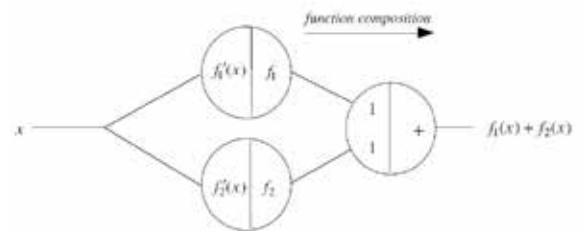


Figure 5: Addition of functions

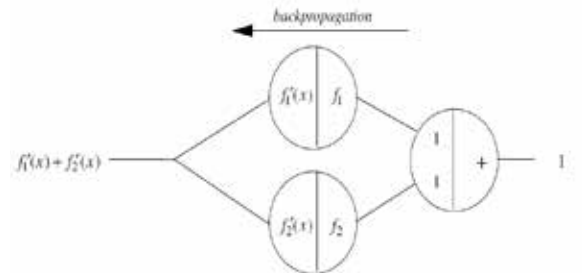


Figure 6: Result of the backpropagation step

c. Third case: weighted edges

Weighted edges could be handled in the same manner as function compositions, but there is an easier way to deal with them. In the feed-forward step the incoming information x is multiplied by the edge's weight w . The result is wx . In the backpropagation step the traversing value 1 is multiplied by the weight of the edge. The result is w , which is the derivative of wx with respect to x . From this we conclude that weighted edges are used in exactly the same way in both steps: they modulate the information transmitted in each direction by multiplying it by the edges' weight.

After choosing the weights of the network randomly, the backpropagation algorithm is used to compute the necessary corrections. The algorithm can be decomposed in the following four steps:

1. Feed-forward computation
2. Backpropagation to the output layer
3. Backpropagation to the hidden layer
4. Weight updates

The algorithm is stopped when the value of the error function has become sufficiently small.

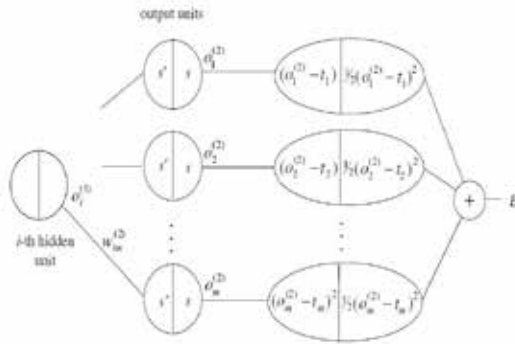


Figure 7: Extended multilayer network for the computation of E

1. Feed-forward computation

The evaluated derivatives of the activation functions are calculated and stored at each unit.

2. Backpropagation to the output layer

The backpropagation path from the output of the network is propagated starting from output layer.

3. Backpropagation to the hidden layer

Each unit in the hidden layer is connected to each unit in the output layer with an edge of weight W . The backpropagated error up to unit i in the hidden layer must be computed taking into account all possible backward paths. The backpropagated error can be computed in the same way for any number of hidden layers.

4. Weight updates

It is very important to make the corrections to the weights only after the backpropagated error has been computed for all units in the network. Otherwise the corrections become intertwined with the backpropagation of the error and the computed corrections do not correspond any more to the negative gradient direction.

5. Training Dataset

Nearly 100 images were used in the experimental setup containing five class labels. The top 40 relevant attributes were selected using information gain. Figure 8 shows some of the images used in this work.

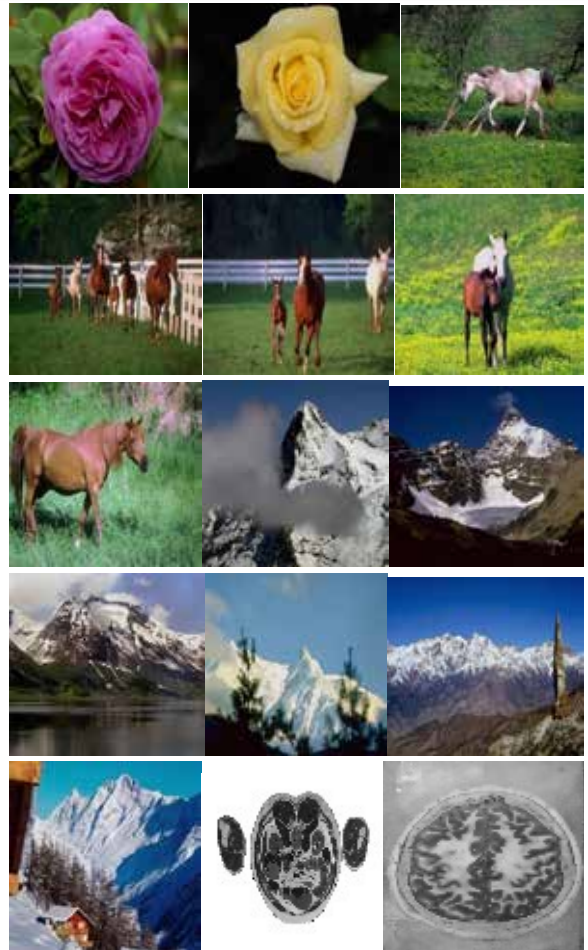


Figure 8: Some of the images used in this work.

6. Testing Results

The results obtained from the proposed FS-MLP Neural Network are shown in figure 9.

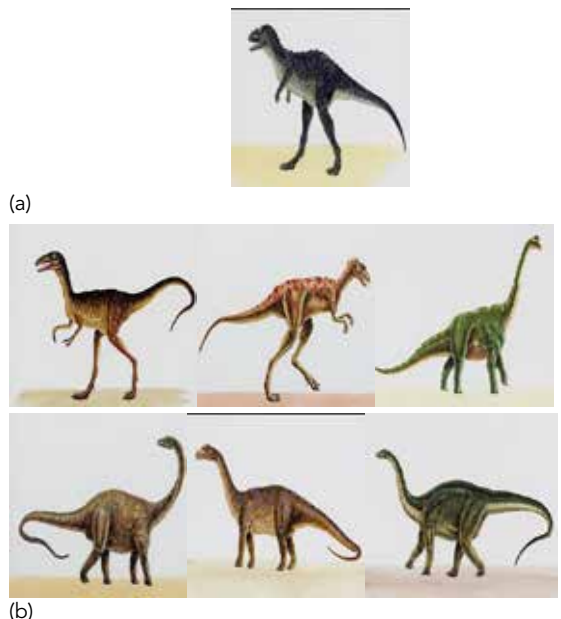


Figure 9: (a) Query image from database. (b) Query results.

The results obtained from normal MLP Neural Network and the proposed FS-MLP Neural Network is shown in figure 10.

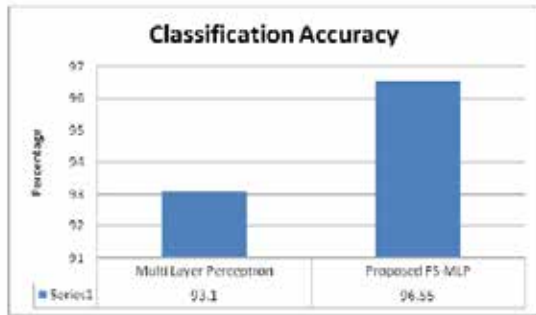


Figure 10: Classification accuracy measured in percentage.

7. Summary and Conclusion

In this paper it was proposed to extract features using Discrete Wavelet Transform (DWT) and select the top attributes based on class attribute using information gain. The extracted features were trained with the existing MLP Neural network classifier and compared with the proposed FS-MLP neural network. The classification accuracy of the proposed method improved by a percentage of 3.45. Using less number of features in the proposed method decreases the overall processing time for a given query.

REFERENCE

- [1] N. S. Chang, and K. S. Fu, "Query by pictorial example," IEEE Trans. on Software Engineering, Vol.6, No.6, pp. 519-524, Nov.1980. | [2] S. K. Chang, C. W. Yan, D. C. Dimitroff, and T. Arndt, "An intelligent image database system," IEEE Trans. on Software Engineering, Vol.14, No.5, pp. 681-688, May 1988. | [3] J. Dowe, "Content-based retrieval in multimedia imaging," In Proc. SPIE Storage and Retrieval for Image and Video Database, 1993. | [4] C. Faloutsos et al, "Efficient and effective querying by image content," Journal of intelligent information systems, Vol.3, pp.231-262, 1994. | [5] Sonali Bhadoria and C.G. Dethé, "Study of Medical Image Retrieval System" 2010 International Conference on Data Storage and Data Engineering, pg 192-6. | [6] H. Müller, N. Michoux, D. Bandon, and A. Geissbühler. A review of contentbased image retrieval systems in medical applications-clinical benefits and future directions. In International Journal of Medical Informatics, volume 73, pages 1–23, 2004. | [7] T. Lehmann, M. G'uld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. Wehn. Content-based image retrieval in medical applications. | [8] Y. Cheng, R. Swamisai, S. E. Umbaugh, R. H. Moss, W.V. W. V. Stoecker, S. Teegala, and S. K. Srinivasan. Skin lesion classification using relative color features. In Skin | [9] C.-Y. Lin, J.-X.Yin, X. Gao, J.-Y.Chen, and P.Qin, "Asemantic modelling approach for medical image semantic retrieval using hybrid Bayesian networks," in Proc. 6th Int. Conf. Intell. Syst. Des. Appl. (ISDA), 2006, pp. 482–487. | [10] Rigau J, Feixas M, Sbert M, et al. Medical image segmentation based on mutual information maximization. In Proc of MICCAI 2004 (Saint-Malo). Berlin: Springer-Verlag, 2004. 135–142. | [11] K. Rajakumar, "An Integrated Approach for Medical Image Retrieval Using PCA and Energy Efficient Wavelet Transform", European Journal of Scientific Research ISSN 1450-216X Vol.51 No.3 (2011), pp.340-348. | [12] Nandakishore Kambhatla, Todd K. Leen, "Dimension Reduction by Local Principal Component Analysis". Neural computation, October 1, 1997, Vol. 9, No. 7, Pages 1493-1516. | [13] Soo Beom Park, Jae Won Lee, Sang Kyoan Kim, "Content-based image classification using a neural network." Pattern Recognition Letter, Vol 25, Issue 3, Feb 2004, Pages 287-300. | [14] Zhong Su, Hongjiang Zhang, Stan Li, and Shaoping Ma, "Relevance Feedback in Content- Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning." IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 12, NO. 8, AUGUST 2003, pp: 924-937. | [15] Bridle, J.S "Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition" In: F.Fogelman Soulie and J.Herault (eds.), Neurocomputing: Algorithms, Architectures and Applications, Berlin: Springer-Verlag, pp. 227-236. | [16] R. Rojas: Neural Networks, Springer-Verlag, Berlin, 1996