Emblematical image based pattern recognition paradigm using Multi-Layer Perceptron Neural Network

Md. Abdullah-al-mamun[1], Aminul Islam[2]

[1, 2](Computer Science and Engineering, Rajshahi University of Engineering and Technology, Bangladesh)

Abstract: *The abstract Likewise human brain machine can be signifying diverse pattern sculpt that is proficiently identify an image based object like optical character, hand character image, fingerprint and something like this. To present the model of image based pattern recognition perspective by a machine, different stages are associated like image acquiring from the digitizing image sources, preprocessing image to remove unwanted data by the normalizing and filtering, extract the feature to represent the data as lower dimension space and at last return the decision using Multi-Layer Perceptron neural network that is feed feature vector from got the feature extraction process of a given input image. Performance observation complexity is discussed rest of the description of pattern recognition model. Our goal of this paper is to introduced symbolical image based pattern recognition model using Multi-Layer Perceptron learning algorithm in the field of artificial neural network (like as human-like-brain) with best possible way of utilizing available processes and learning knowledge in a way that performance can be same as human.*

Keywords: *Pattern Recognition; Multi-Layer Perceptron; MLP; Artificial Neural Network (ANN); Backpropagation.*

I. Introduction

"A picture is worth a thousand words." As human beings, we are able to tell a story from a picture based on what we see and our background knowledge [1]. Is it possible to ascertain a semantic concept from an image by a machine that will be performing same as human like brain? In a short, the answer will be "possible". As human brain is capable to easily identify image based object as their learning knowledge is growing from childhood. So if we are propose an artificial neural model that is perform like human-like-brain then it is possible to detect an image object. To able a computer program in semantic understanding by recognition image based pattern, it is need to extract visual feature efficient and effective way and build a model that is capable learning background knowledge like human brain.

Layered neural networks involve a set of input cells connected to a collection of output cells by means of one or more layers of modifiable intermediate connections [3]. The neural network architecture is associating by an input pattern, that representation situation of a pattern, which produced an output class by feed the network. In the context of image processing, the different stages are: acquisition, concern with digitizing the data coming from a sensor such as camera or scanner, localizing, which involves extracting the object from its background, and representation [2]. In the final decision stage, consist of two phases learning and recognition that are associated with each other's. In the learning stage artificial neural network is adapted by the training of known object classes. Network adaptation rate is proportional to the number of training that is performing on the network. To perform recognition phase, object feature vector is feed into the network and according to the previous learning knowledge of the adapted network is return a output feature vector class for the given object feature vector.

II. Model of pattern recognition

An overall organization of the pattern recognition representation can be illustrated with the figure is as following,
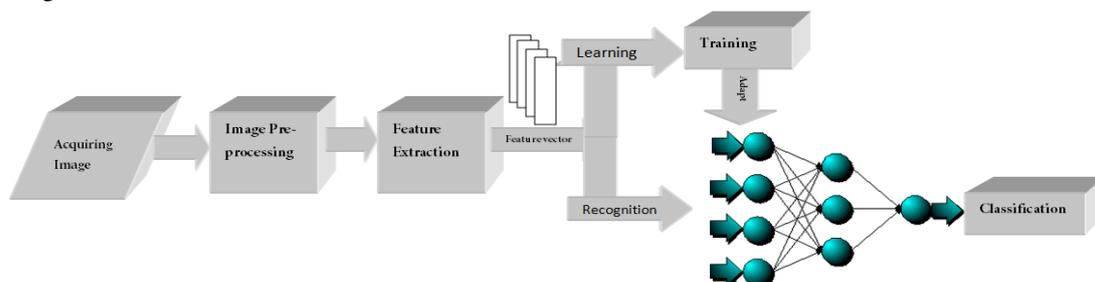

Figure-1: Model of Pattern recognition.

1.1. Image Acquiring

In the acquiring image part input the image document from any hardware device that can be scanned copy image, optical character or something like this. It is the initial stage for the pattern recognition model because without input how to classify?

1.2. Image Pre-processing

Image preprocessing system preprocesses the image both geometrically and radiometrically by the radiant energy characteristics in such a way that produced the image of the original scenes. By the preprocessing system reduced the noise of the image, normalize the image and filtering the image so that image can't contains the unwanted data.

1.3. Feature Extraction

Feature extraction measured data and builds derived features intended to be informative, non-redundant, facilitating the subsequent learning and generalization steps, in some cases leading to better human interpretations [13]. By the performing operation of feature extraction on the image, it extracts the feature of an image to produce the relevant data vector in a lower dimensional space from the input image.

1.4. Classification

The final part of the classification that is associated with the two parts; one is training process and recognition process. In the training phase, network is adapted from the several training set of data with correct class for a given training set. After the network is trained, classification for new glyphs can be done by extracting features from new glyphs and using the trained network to predict their class [9].

III. Acquiring Image

The opening stage in the workflow progression of any vision system is performing image acquisition in the digital image processing because of without an image there is no query is arise to process it. Image acquisition is the branch of Digital Image Processing can be defined as the action of retrieval of an image from any possible source, generally hardware-based source. After image has been acquired fittingly, there are several preprocessing technique can be applied on the image to execute for the further processing. If the image has not been acquired satisfactorily then the intended tasks may not be achievable, even with the aid of some form of image enhancement [6].

IV. Preprocessing Image

In imaging science, image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image [10]. This image preprocessing system produces a correct image both geometrically and radiometrically by the radiant energy characteristics that is closest as possible of the original scenes.

1.5. Normalization

In the field of image processing normalization refer to the changes the range of pixel intensity values. In the digital image processing it also referred to the dynamic range expansion. Applications include photographs with poor contrast due to glare, for example. Normalization is sometimes called contrast stretching or histogram stretching [10].

Normalization transforms an n-dimensional grayscale image $I : \{\mathbb{X} \subseteq \mathbb{R}^n\} \to \{Min, .., Max\}$ with intensity values in the range (Min, Max), into a new image $I_N : \{\mathbb{X} \subseteq \mathbb{R}^n\} \to \{newMin, .., newMax\}$ with intensity values in the range (newMin, newMax).

The linear normalization of a grayscale digital image is performed according to the formula

$$I_N = (I - Min)\frac{newMax - newMin}{Max - Min} + newMin$$ …………………………………………………...… (1)

According to the histogram analysis, if we see intensity range of the image is [50,180] and that should be converted to the desired range are [0,255]. Now subtracting 50 from each of pixel intensity, that making the range [0,130]. Then each pixel intensity is multiplied by 255/130 to making the range [0,255].

1.6. Filtering

In the field of digital signal processing, filtering is the process to remove the unwanted data, component or feature signal from the original source of data. Filtering is a class of signal processing, removing some frequencies and not others in order to suppress interfering signals and reduce background noise,

Sharpening and Deblurring. Correlations can be removed for certain frequency components and not for others without having to act in the frequency domain [11].

V.   Feature Extraction

In the task of pattern recognition it should be differenced the possible output classes according to the input set of data. So the feature extraction is essentially done after preprocessing of classification phase in pattern recognition and machine learning. In the process of pattern recognition input data is transformed into a reduced set of representation of feature to produced correct output class.

In machine learning, pattern recognition and in image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative, non-redundant, facilitating the subsequent learning and generalization steps, in some cases leading to better human interpretations [13]. The main goal of feature extraction is to extract set of feature to produce the relevant information from the original input set data that can be represent in a lower dimensionality space.

1.7. Approach of Feature Extraction

Generally, an image contains some feature is as - color, texture, shape, edge, shadows, temporal and so on. All features can be classified into two types level feature is as low-level features and high-level features. Low-level features can be extracted directed from the original images, whereas high-level feature extraction must be based on low-level features [8]. Here we should be most concern about color, texture and shape feature that feature extraction details are following.

1.8. Color of Feature Extraction

The entire color feature of an image is broadly used to retrieval image feature space for the feature extraction purpose. In general, the color of an image can be specified a color model in terms of 3-D coordinate system. The color model can be RGB (red, green, blue), HSV (hue, saturation, value) and Y,Cb,Cr (luminance and chrominance). The relationship can be defined is as following,

$$H = \cos^{-1}\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\}$$

$$S = 1 - \frac{3\{\min(R,G,B)\}}{V}$$
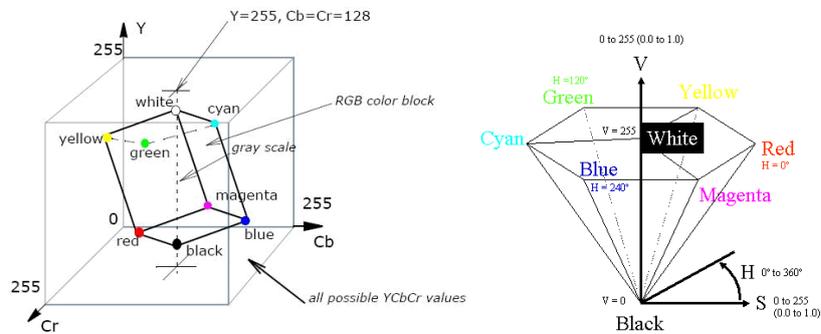
$$V = \frac{1}{3}(R+G+B)$$

Figure-2: RGB and HSV color space.

1.9. Texture of Feature Extraction

Texture is a powerful regional descriptor of an image to retrieval feature extraction process. It is the most important characteristic which has been used to classify and recognize objects and have been used in finding similarities between images in multimedia databases [16]. Consider, the matrix element *p(i,j)* represent the relative frequency of two neighbor pixel at distance *d* with grey levels '*i*' and '*j*' at a given direction. So, Statistical methods based on the grey level co-occurrence matrix can be express,

$$p(i,j) = \frac{P(i,j)}{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}P(i,j)}$$

……………………………………………………………………………………… ……………………………….. (6)

Where, N_g represents the total number of gray level.

1.10. Shape of Feature Extraction

Shape of an image is a primitive feature for description of image content. Shapes are often the archetypes of objects belonging to the same pattern class [4], and it can be as a silhouette (i.e, rotation, scale and translation) of an object. Features calculated from objects contour: circularity, aspect ratio, discontinuity angle irregularity, length irregularity, complexity, right-angleness, sharpness, directedness [16]. Those factors are invariant to the shape descriptors.

VI. Multi-layer Perceptron

Artificial Neural Networks (ANNs) are a family of statistical learning models inspired by biological neural networks (the central nervous systems of animals, in particular the brain) that interconnected to send messages each other and also used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown [7]. In the branch of Artificial Neural Network, A Multi-layer Perceptron(MLP) is the feedforward artificial neural network model that is design in a way that input set are feed from the input layer and that will be produced an output vector set. A MLP consists of two or more layers where each layer contains several number of neuron with nonlinear activation function and each neuron of one layer is fully connected to the next layer as a direct graph (except input layer).
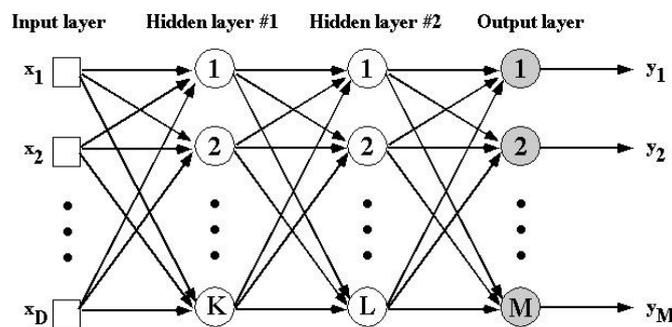

Figure-3: Sample architecture of Multi-Layer Perceptron Neural Network.

Consider, in the MLP has number of hidden layer is one then it can be express as the function $f:R^D => R^M$, where D is the input vector size and M is the output vector size $y=f(x)$. Here, the output layer matrix notation can be express as,

$f(x)=G(b_j+W_j(S(b_i+W_ix))$

Where b_i, b_j are the bias vector for each neuron; W_i,W_j are the weight of the connected link and G,S are the activation function.

The vector of the hidden layer can be express as $h(x) = \emptyset(x) = s(b_i + W_ix)$ so that weighted matrix $W_i \in R^{D \times D_h}$ is connected the input vector to hidden layer. Here, sigmoid function is used to activation function that can be express as $\varphi(v) = \frac{1}{1+e^{-v}}$

1.11. Learning

Learning is the process of a system that it enables the system to do that's types of task efficiently for the next time. It is essential for unknown environments, i.e., when lack of output control to make decision. The problem of learning is to take in prior knowledge and data (e.g., about the experiences of the agent) and to create an internal representation (the knowledge base) that is used by the agent as it acts [18]. Appling the learning process algorithm within the multilayer network architecture, the synaptic weights and threshold are update in a way that the classification/recognition task can be perform efficiently. A *class* of functions   training process means fined $f^* \in F$ using a set of *observations* which solves the task in some *optimal* sense [19]. Here to perform the learning process for the multilayer perceptron neural network we have used error backpropagation learning rule.

Consider a multilayer neuron network that has *d* dimensional input feature with *M* numbers of neuron in the hidden layer which is produced *k* numbers of output vector. And sigmoid function is used as activation function is

$\varphi(v) = \frac{1}{1+e^{-v}}$

Now calculated the activation for Hidden layer b_j is following,

$b_j = \sum_{i=0}^{d} W_{ji}x_i$ Where, j=1,2,....,M

W_{ji} is the weight for the network connection from i[th] level input layer to j[th] level hidden layer where x_i is the feature input. Now transferring each activation by the non-linear activation function,

$$Z_j = \varphi(b_j) = \frac{1}{1+e^{-b_j}}$$

In the context of neural network, Z_j interpreted the output of the hidden layer unit.

Now for the hidden layer, the Zj is linearly combined to the activation of the Kth output layer.

$$a_k = \sum_{j=0}^{M} W_{kj} z_j \quad \text{Where, k=1,2,...,K}$$

Transferring each activation by non-linear activation function is following,

$$y_k = \varphi(a_k) = \frac{1}{1+e^{-a_k}}$$

These equations may be combined to give the overall equation that describes the forward propagation through the network, and describes how an output vector is computed from an input vector, given the weight matrices:

Now to get the output vector from the input vector, we have combined the equation is as following,

$$y_k = g\left(\sum_{j=0}^{M} W_{kj}\, h\left(\sum_{i=0}^{d} W_{ji} x_i\right)\right) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(7)}$$

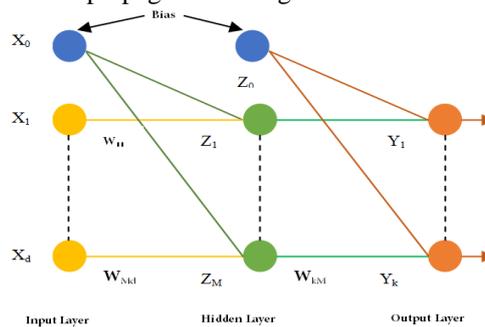That can be describing as a forward propagation throughout the network.

Figure-4: 3-Layer MLP with bias connection.

1.12. Learning through Backpropagation

Back-propagation learning rule perform of two sweeps called forward sweep and backward sweep throughout different layer in the network. In forward sweep or forward pass, an activity pattern (input vector) is applied to the sensory nodes of the network, and its effect propagates through the network layer by layer [17]. In the time of forward pass, the synaptic (network link) weight are fixed and after pass the values throughout the layer an output is produced as the actual response of the network. If the produced output is not equal to the desired output then an error signal is produced and to reduce the error, it needs to perform backward pass throughout the network. During the backward pass, the synaptic weights are adjusted in accordance with an error-correction rule [17]. By the performing sweeps, the produced output error is propagated backward through the network.

Now we can evaluate the training network using gradient descent function. If error function can be define as *E* then derivation form is $\frac{\partial E}{\partial W_{ji}}$ and $\frac{\partial E}{\partial W_{kj}}$

1.12.1. Error Function

To train an MLP we need to define an error function. Again we use the sum-of-squares error function, obtained by summing over a training set of N examples:

For the training set *N*, the error function *E* can be defined as sum-of-square error function that is summing over training sample *N*.

$$E = \sum_{n=1}^{N} E_n = \frac{1}{2} \sum_{k=1}^{K} (y_{nk} - t_{nk})^2$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(8)}$$

Where, y_{nk} is the output vector that is get from the forward propagation and t_{nk} is the targeted output for the output layer.

So the gradient descent for the output layer is sum of overall training set,

$$\frac{\partial E}{\partial W_{kj}} = \sum_{n=1}^{N} \frac{\partial E_n}{\partial W_{kj}}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(9)}$$

Similarly, Gradient descent for the hidden layer,

$$\frac{\partial E}{\partial W_{ji}} = \sum_{n=1}^{N} \frac{\partial E_n}{\partial W_{ji}}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(10)}$$

1.12.2. Hidden-to-Output Weight

The error gradients for the hidden-to-output weights from equation (9) is,

$$\frac{\partial E_n}{\partial W_{kj}} = \frac{\partial E_n}{\partial a_{nk}}\frac{\partial a_{nk}}{\partial W_{kj}} = \delta_{nk}z_{nj}$$

……………………………………………………………………………………… (11)

Where, the error signal δ_{nk} can be defined as derivation form of the gradient error E_n with respect to the activation a_{nk}

$$\delta_{nk} = \frac{\partial E_n}{\partial a_{nk}} = \frac{\partial E_n}{\partial y_{nk}}\frac{\partial y_{nk}}{\partial a_{nk}} = (y_{nk} - t_{nk})$$ …………………………………………………(12)

And, $z_{nj} = \frac{\partial a_{nk}}{\partial W_{kj}}$

1.12.3. Input-to-Hidden Weight

The error gradients for the input-to-hidden weights can be express as, $\frac{\partial E}{\partial W_{ji}}$. The error signal for hidden unit j

$$\delta_{nj} = \frac{\partial E_n}{\partial b_{nj}} = \sum_{k=1}^{k}\frac{\partial E_n}{\partial a_{nk}}\frac{\partial a_{nk}}{\partial b_{nj}} = \sum_{k=1}^{k}\delta_{nk}\frac{\partial a_{nk}}{\partial b_{nj}}$$

……………………………………………………………………………(13)

Here, $\frac{\partial a_{nk}}{\partial b_{nj}} = \frac{\partial a_{nk}}{\partial z_{nj}}\frac{\partial z_{nj}}{\partial b_{nj}} = w_{kj}h'(b_{nj})$ because of a_{nk} is the weighted sum of W_{kj}. And z_{nj} is the first derivation of b_{nj}. So from the equation (13) we get,

$$\delta_{nj} = h'(b_{nj})\sum_{k=1}^{k}\delta_{nk}w_{kj}$$ …………………………………………………………………….………(14)

This equation (14) is described as back-propagation error formula. So the derivation of input-to-hidden weight can be express as,

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial b_{nj}}\frac{\partial b_{nj}}{\partial w_{ji}} = \delta_{nj}x_{ni}$$ ……………………………………………………………………….……..(15)

Equation (15) will be applied recursively for the next level hidden layer unit.

VII. Multi-layer Perceptron Learning Algorithm

1. Initializes all weights and thresholds with minimum random value.
2. Present input $X_p = x_0,x_1,x_2,…,x_{n-1}$ and target output $T_p = t_o,t_l,t_2,. . .,t_{m-l}$ where n is the number of input nodes and m is the number of output nodes. For classification, T_p is set to zero except for one element set to 1 that corresponds to the class that X_p is in [20].
3. Calculate the output for each layer

$$y_{pj} = f\left[\sum_{i=0}^{n-1} w_i x_i\right]$$

   This output will be input for the next layer. Where, layer has contains n numbers of nodes and w is the weight for the each link.
4. If the targeted output is not equal the desire outputs then it need to apply backward sweep to reduce the error level by updating the weight and threshold value.
5. Now adjust weight for the output layer can be calculated by,

$$W_{jk} = W_{jk} + n_2K_2\delta_{ak}O_{aj}O_{ak}(1 - O_{ak})$$

   Adjust threshold for the output layer can be calculated by,

$$U_{ok} = U_{ok} + n_2K_2\delta_{ak}O_{ak}(1 - O_{ak})$$

6. Adjust weight for the hidden layer can be calculated by,

$$W_{ij} = W_{ij} + n_1K_1O_{ai}O_{aj}(1 - O_{aj})\sum\delta_{ak}W_{jk}$$

   Adjust threshold for the hidden layer can be calculated by,

$$uh_j = uh_j + n_1K_1O_{aj}(1 - O_{aj})\sum\delta_{ak}W_{jk}$$

   The backward sweep repeated until error difference between actual and desire value reduced within the minimum level.

VIII. Complexity of Multi-layer Perceptron Learning Algorithm

The performance of pattern recognition can be influence by the Multi-Layer perceptorn neural network for the following reasons is describe below,

---

1.13. Parameters of MLP
1.13.1. Numbers of Input
Increasing the number of input to feed the input layer can be increased the complexity of network topological structure and may be influencing the overall network (like number of hidden layer) structure. That may be caused the increased the error rate. This can be clear by an example is that if the number of input is 100 then to get an optimal topology network it need to about 300 neurons for the hidden layer!

1.13.2. Numbers of Hidden Layer
Significant change can be shown if the numbers of Hidden layer are increased in the network. For example, to solve the XOR problem by 2-Layer and 3-Layer it can be classify is as following,
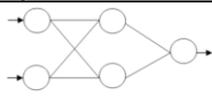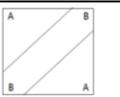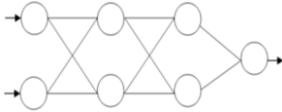
Layer	Figure	Classification
2-Layer		
3-Layer		

Table-1: Comparison between Network and Classification.

Local minima can be considered to occur when two or more disjoint classes are categorized as the same. These amounts to a poor internal representation within the hidden units, and so adding more units to this layer will allow a better recoding of the inputs and lessen the occurrence of these minima [20].

1.13.3. Number of Epochs
Increasing the number of epochs of the network can provide the better performance but many times it may be return the poor result for the recognition system. This partially can be attributed to the high value of learning rate parameter as the network approaches its optimal limits and further weight updates result in bypassing the optimal state [19]. This process is known as over-learning.

1.13.4. Learning rate parameter
The variation of the learning rate parameter can be influence the network performance. If the learning rate parameter value is high then network neurons and link values are updating more quickly comparing if the learning rate parameter value is less to reach the optimal state.

1.14. Feature Extraction
Feature extraction plays a big role for the total operational process for any pattern recognition system in the field of machine learning system. Because of the feature extraction process build a source raw data into convert in a way that it should be informative, non-redundant and successive for the learning or recognition process. So it should be apply better feature extraction process produced set of data that should be obtain a linear or non-linear transformation of the original set of data.

1.15. Data Collection
Input source data is important each stages for the pattern recognition because this data will be used for the further processing like feature extraction, classification and recognition of a pattern. So the data collection method should be design in a way that input data is garbage or rush free and it is analyzable for the complex multivariate techniques. To perform inception operation on the data it needs to check the quality of the data, calculating the summary statistics of the data an sketch the plots of the data to get the structure of the data [19].

IX. Conclusion
Presenting the pattern recognition sculpt to identify image based object, we are present the multi-layer perception learning algorithm for learning and classification method for decision making scheme with pre-processing technique after getting the input image. The significant feature of the Multi-Layer Perceptron is that the learning rates are dynamically computed each epoch by an interpolation map and error function is transformed into a lower dimensional error space and the reduced error function is employed to identify the variable learning rates [9]. In the field of artificial neural network, it is really difficult to recognize an image based object by the multi-layer perceptron learning algorithm due to real time complexity of the pattern whereas pattern can be linear or non-linear. Multi-Layer Perceptrons can be return 100% accurate result only for the

linearly separable problems. Whereas for the non-linear separable pattern is difficult to recognition. Because the learning process of a multilayer perceptron requires the optimization of an error function $E(y,t)$ comparing the predicted output y, and the observed target t [15]. So a lot of efforts are associated to improving higher recognition accuracy.

References

[1]   Dong ping, Tian, "A Review on Image Feature Extraction and Representation Techniques", International Journal of Multimedia and Ubiquitous Engineering, Vol. 8, No. 4, July, 2013, pp. 385-396.
[2]   Sarvda Chauhan, Shalini Dhingra, "Pattern Recognition System using MLP Neural Networks", International Journal of Engineering Research and Development, e-ISSN: 2278-067X, Volume 4, Issue 9 (November 2012), PP. 43-46.
[3]   "A Versatile Machine Vision System for Complex Industrial Parts" IEEE Trans. On Computers, Sep. 1977, No. 9, Vol. C-27.
[4]   L.daF. Costa, J.R.M. Cesar, "Shape Analysis and Classification: Theory and Practice", CRC Press, Florida, USA, 2001
[5]   Eugene P, Angela B, "What Is Image Acquisition in Image Processing", wisegeek.com, 2003-2015 Conjecture Corporation.
[6]   A D Marshall, R R Martin, "Computer Vision, Models and Inspection", World Scientific Series in Robotics and Intelligent Systems: Volume 4, ISBN: 978-981-02-0772-4 (hardcover), March 1992.
[7]   En Wikipedia, "Artificial neural network", 17 July 2015.
[8]   E. Saber, A.M. Tekalp, "Integration of color, edge and texture features for automatic region-based image annotation and retrieval," Electronic Imaging, 7, pp. 684–700, 1998.
[9]   Dharm Singh, Neha Mehta, Pallavi Purohit,"Text Based Image Recognition using Multilayer Perceptron", IP Multimedia Communications, A Special Issue from IJCA - www.ijcaonline.org, pp. 143-146.
[10]   Rafael C. Gonzalez; Richard E. Woods (2008). "Digital Image Processing", Prentice Hall. pp. 1–3. ISBN 978-0-13-168728-8.
[11]   en Wikipedia, "Filter_(signal_processing)", 4 July 2015.
[12]   Gaurav Kumar,Pradeep Kumar Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems",2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 5-12, Febuary 2014.
[13]   En Wikipedia, "Feature_extraction", 7 July 2015.
[14]   Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, "Introduction to Information Retrieval", Cambridge University Press. 2008, ISBN. 0521865719.
[15]   Luís M. Silva, J. Marques de Sá, Luís A. Alexandre, "Data classification with multilayer perceptrons using a generalized error function", ScienceDirect, elsevier.com, Neural Networks 21 (2008) 1302-1310, pp. 1302-1310.
[16]   Ryszard S. Chora´s, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", International journal of biology and biomedical engineering, pp. 6-16 ,Issue 1, Vol. 1, 2007.
[17]   Amrita Hirwani, Sandeep Gonnade, "Character Recognition Using Multilayer Perceptron", Amrita Hirwani et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, pp. 558-661.
[18]   David Poole, Alan Mackworth, "Artificial Intelligence", Cambridge University Press, 2010.
[19]   Md. Abdullah-al-mamun, Mustak Ahmed, "Hypothetical Pattern Recognition design using Multi-Layer Perceptorn Neural Network for supervised learning", Proc. in International Journal of Advanced Research, 2015.
[20]   R Beale, T Jackson, "Neural Computing: An Introduction", Bristol, Philadelphia and New York.