Int. J. Advance. Soft Comput. Appl., Vol. 2, No. 1, March 2010
ISSN 2074-8523; Copyright © ICSRS Publication, 2010
www.i-csrs.org

Performance Improved PSO based Modified Counter Propagation Neural Network for Abnormal MR Brain Image Classification

D.Jude Hemanth[1], C.Kezi Selva Vijila[2] and J.Anitha[3]

Department of ECE, Karunya University, Coimbatore, India.
jude_hemanth@rediffmail.com
Department of ECE, Karunya University, Coimbatore, India.
vijila_2000@yahoo.co.in
Department of ECE, Karunya University, Coimbatore, India.
rajivee1@rediffmail.com

Abstract

Abnormal Magnetic Resonance (MR) brain image classification is a mandatory but challenging task in the medical field. Accurate identification of the nature of the disease is highly essential for the successful treatment planning. Automated systems are highly preferred for image classification because of its high accuracy. Artificial neural networks are one of the widely used automated techniques. Though they yield high accuracy, most of the neural networks are computationally heavy due to their iterative nature. Low speed neural classifiers are least preferred since they are practically non-feasible. Hence, there is a significant requirement for a neural classifier which is computationally efficient and highly accurate. To satisfy these criterions, a modified Counter Propagation Neural Network (CPN) is proposed in this work which proves to be much faster than the conventional network. For further enhancement of the performance of the classifier, Particle Swarm Optimization (PSO) technique is used in conjunction with the modified CPN. Experiments are conducted on these classifiers using real-time abnormal images collected from the scan centres. These three types of classifiers are analyzed in terms of classification accuracy and convergence time period. Experimental results show promising results for the PSO based modified CPN classifier in terms of the performance measures.

Keywords: *Classification accuracy, Convergence time period, Counter Propagation neural network, Magnetic Resonance and Particle Swarm Optimization.*

1 Introduction

Image classification is a pattern recognition technique in which the unknown input data is assigned to a category based on similarity measure. One of the significant applications of image classification is the medical field in which the abnormal brain tumor images are categorized prior to treatment planning. Accurate identification of the type of the brain abnormality is highly essential since the treatment planning is different for all the brain abnormalities. Any false detection may lead to a wrong treatment which ultimately leads to fatal results. Another significant feature is that the time period taken by the system for recognition should be less. This will lead to quick identification which will increase the success rate of the treatment. But these two criterions are not satisfied by the manual identification techniques which are highly subjective and biased. Hence, quick automated image classification systems with high accuracy are highly essential for the real-time applications.

Several automated image classification systems are available in the literature. Brain tumor classification has been performed using long echo proton MRS signals [1]. The major limitation is the limited number of available spectra for the tumor types which results in inferior classification accuracy [2]. Brain tumor classification has also been implemented using wavelets [3]. But the major drawback is the low convergence rate. Expectation-maximization techniques are also used for brain tumor classification [4]. But the major limitation is the requirement of a spatial probabilistic atlas that contains expert prior knowledge about the brain structures. Statistical classifiers, Probabilistic classifiers, Artificial Neural Networks (ANN) are some of the widely used image classifiers. The major drawback of the statistical classifiers is its inability to classify accurately. On the other hand, probabilistic classifiers suffer from the setback of difficulty in estimating the conditional probabilities. But ANNs outperform the other classifiers because of its flexibility, scalability, tolerance to faults, accuracy, learning [5].

Artificial neural networks based image classification systems are also reported in the literature. Quantitative measurement of various components of normal brain is analyzed using artificial neural networks [6]. But tumor detection is a challenging task which is different from normal brain analysis. A neural network approach for melanoma detection is reported in[7]. But the approach used in this system is computationally heavy. The applications of neural networks is widely analysed by Peterson [4]. Support Vector machines based classification of brain images is implemented and reported in the literature [8]. Unsupervised neural networks are also used for image classification applications [9]. But the accuracy of this approach is very less when compared with the supervised networks.

Several modifications of the conventional neural networks are also implemented for performance enhancement of the classifiers. Vomweg [10] introduced an improved neural network based technique for breast cancer detection. An improved version of back propagation algorithm is used for breast cancer detection [11]. Changhua [12] implemented the multilayer feed forward network with a modified training algorithm. A different version of back propagation neural network is employed for pattern recognition [13]. Modified back propagation algorithms are also developed for computer security applications [14]. Apart from the modifications, optimization algorithms are also useful for improving the efficiency of the classifiers. Genetic Algorithms are widely used to enhance the performance of the k-NN classifiers [15]. Genetic Algorithm in conjunction with the ART neural network improved the performance to a higher extent [16]. Particle Swarm Optimization algorithm is also used for classification applications [17]. Image registration applications have successfully used the PSO technique [18].

In this work, a modification of the conventional CPN is proposed for abnormal brain image classification. PSO technique is used for the concept of feature selection in the classification technique. Abnormal brain tumor images from four groups namely metastase, meningioma, glioma and astrocytoma are used in this work. Experimental analysis show superior results for the PSO based modified CPN technique over the conventional classifiers in terms of the performance measures.

2 Materials and Methods

The framework for the proposed automated image classification system is shown in Fig. 1. The real time abnormal MR brain images are collected from radiologists. A morphology based pre-processing step is then performed on these images to remove the skull tissues which often interfere with the tumor tissues. Several significant features are extracted from these skull stripped images. Feature selection is then performed to select the optimal feature set. In this work, a modified Counter Propagation neural network is proposed for classification. Initially, the proposed network is trained with the complete feature set. Then the modified CPN is trained with the optimal feature set and the performance measures are calculated. The performance measures used in this work are classification accuracy and convergence time period. Experimental results show promising results for the PSO based Modified CPN in terms of performance measures. Thus the objective of this work is twofold: Proposing a suitable modification in the conventional neural network for practical applications & validating the significance of optimization strategies (PSO) for performance enhancement of the classifiers.

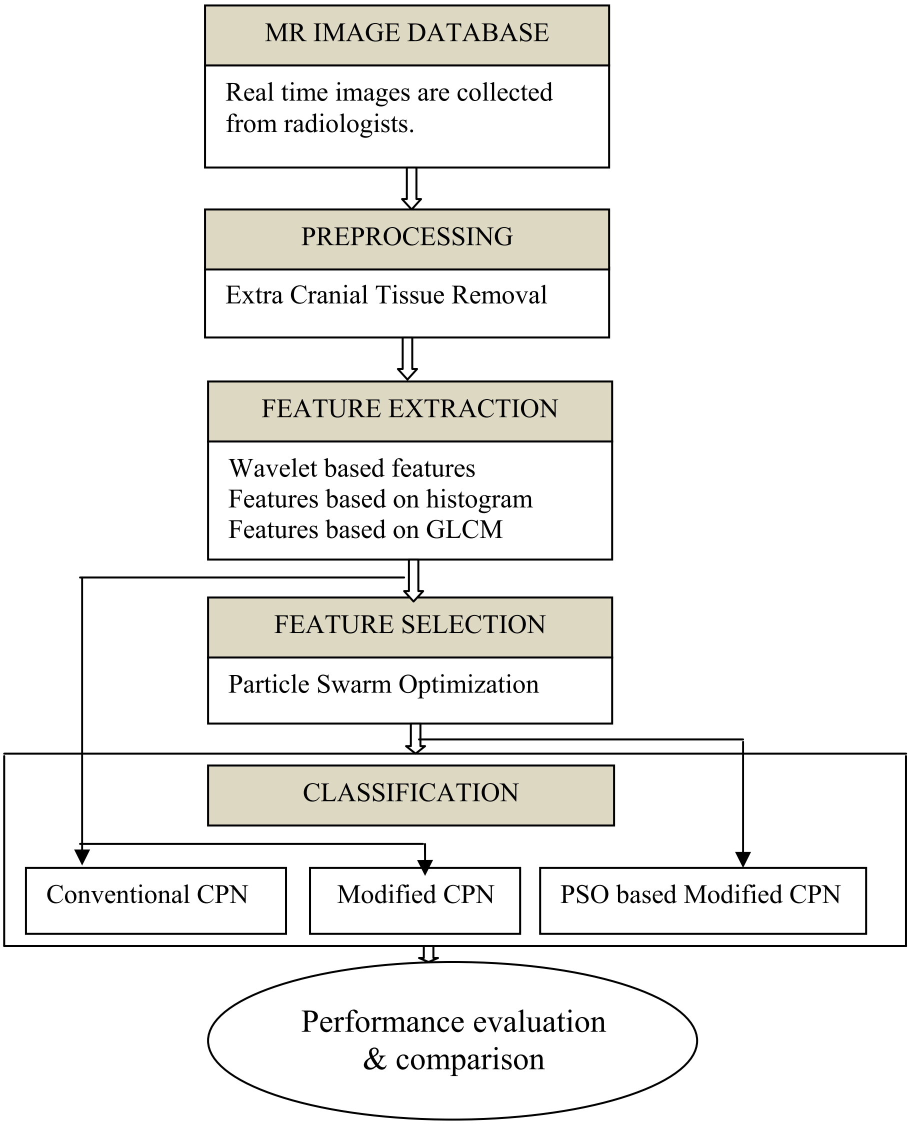
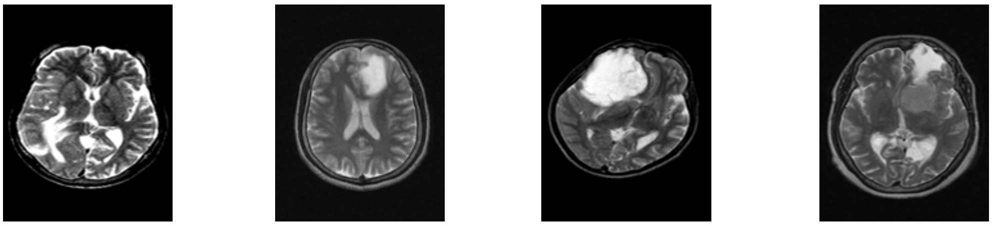


Fig. 1. Flow diagram of the proposed methodology

3 MR Database Collection & Image Pre-Processing

A set of MR brain tumor images comprising of the four tumor types namely meningioma, astrocytoma, glioma and metastase are collected from radiologists. The images used are 256*256 gray level images with intensity value ranges from (0 to 255). Initially, these MRI images are normalized to gray level values from (0 to 1) and the features are extracted from the normalized images. The normalization is performed by dividing each intensity value by the maximum value (255). Since normalization reduces the dynamic range of the intensity values, feature extraction is made much simpler. Some samples of the MRI database have been displayed in Fig. 2.



(a) (b) (c) (d)

Fig. 2. Sample data set: (a)Metastase (b)Glioma (c)Astrocytoma (d) Meningioma

The soft brain tissues are surrounded by the extra cranial tissues (skull tissues) which is evident from Fig. 2. In the above Fig.s, the outer circular ring corresponds to the extra cranial tissues. These tissues often interfere with the brain tissues which ultimately results in inferior classification accuracy. Hence, the skull tissue removal is a significant pre-proceesing step in the area of brain image analysis. In this work, a series of morphological operations are used to eliminate the skull tissues. Initially, a mask is created using the erosion operation and the connected component operation (image filling operation). This mask is then superimposed on the original image to obtain the skull-stripped images. Features extracted from these pre-processed images are highly significant than the features extracted from the raw images. Thus, the preprocessing step is a mandatory task to guarantee high classification accuracy results. The rest of the image classification system is dealt in detail in individual sections.

4 Feature Extraction

The purpose of feature extraction is to reduce the original data set by measuring certain properties or features that distinguish one input pattern from another pattern . The extracted feature should provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into a feature space. Fourteen features from three categories such as wavelet based features, features based on first order histogram and the features based on gray level co-occurrence matrices (GLCM) are used in this work.

4.1 Wavelet transform based textural features

Mathematical transformations significantly represent the information contained in an image. Among them, wavelet transforms are highly preferred since they provide very sensitive information which is very essential in the texture analysis. The computational complexity is also significantly reduced due to wavelet decomposition. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of wavelets. Any such superimposition decomposes the given function into different levels where each level is further decomposed with a resolution adapted to that level. In this work, a single level decomposition based on Discrete Wavelet Transform (DWT) of the input image is used where the output yields four sub-bands namely approximation sub-band, horizontal sub-band, vertical sub-band and diagonal sub-band. Among these sub-bands, much of the information is present in the approximation sub-band. From the approximation sub-band which is a matrix of reduced size, four textural features namely mean, standard deviation, cluster shade and cluster prominence are extracted and used for classification. The formulas for feature calculation are given below.

Mean:

$$S_1 = \frac{1}{N^2}\sum_{i,j=1}^{N} p(i,j) \tag{1}$$

Standard deviation:

$$S_2 = \left[\frac{1}{N^2}\sum_{i,j=1}^{N}\left(p(i,j) - S_1\right)^2\right]^{\frac{1}{2}} \tag{2}$$

Cluster Shade:

$$S_3 = \sum_{i,j=1}^{N}\left(i - M_x + j - M_y\right)^3 p(i,j) \tag{3}$$

Cluster Prominance

$$S_4 = \sum_{i,j=1}^{N}\left(i - M_x + j - M_y\right)^4 p(i,j) \tag{4}$$

Where p(i,j) = transformed value of the approximation sub-band matrix of size N*N.

$$M_x = \sum_{i,j=1}^{N} ip(i,j)\ ;\ \ M_y = \sum_{i,j=1}^{N} jp(i,j)$$

These features are selected based on the previous works suggested by Arivazhagan [19].

4.2 Features Based On First Order Histogram

The various features such as skewness, kurtosis, energy and entropy based on the first order histogram are computed using the formulae given next.
The first order histogram estimate of p(b) is simply

$$P(b)=\frac{N(b)}{M} \tag{5}$$

where b=a gray level in the image
M=total number of pixels in a neighborhood window centered about an expected pixel.
N(b)=the number of pixels of gray value b in the same window that $0 \le b \le L-1$.
Then the following measures have been extracted by using first order probability distribution.
Skewness:

$$S_S = \frac{1}{\sigma_b^3}\sum_{b=0}^{L-1}\left(b-\bar{b}\right)^3 p(b) \tag{6}$$

Kurtosis:

$$S_K = \frac{1}{\sigma_b^4}\sum_{b=0}^{L-1}\left(b-\bar{b}\right)^4 p(b)-3 \tag{7}$$

Energy:

$$S_N = \sum_{b=0}^{L-1}\left[p(b)\right]^2 \tag{8}$$

Entropy:

$$S_E = -\sum_{b=0}^{L-1} p(b)\log_2\{p(b)\} \tag{9}$$

4.3 Features Based On Gray Level Co-Occurrence Matrices

Spatial gray level co-occurrence estimates image properties related to second-order statistics. Haarlick [20] suggested the use of gray level co-occurrence matrices (GLCM) which have become one of the most well-known and widely used texture features. GLCM $\{P_{(d,\ \theta)}(i,\ j)\}$ represents the probability of occurrence of a pair of gray-levels (i, j) separated by a given distance d at angle θ. The commonly used unit pixel distances and the angles are 0°, 45°, 90° and 135°.
Notations:

$$p_x(i)=\sum_{j=1}^{N_g} p(i,j) \quad ; \quad p_y(j)=\sum_{i=1}^{N_g} p(i,j) \tag{10}$$

$$p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j), k = 2,3.......2N_g; i + j = k \tag{11}$$

$$p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j), k = 0,1........N_g - 1; |i - j| = k \tag{12}$$

$p(i,j) =$ gray level co-occurrence matrix.

The features such as contrast, inverse difference moment, correlation, variance, sum average and difference entropy are calculated using the formulae given below.
Contrast:

$$S_C = \sum_i \sum_j (i-j)^2 P(i,j) \tag{13}$$

Inverse difference moment:

$$S_I = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j) \tag{14}$$

Correlation:

$$S_O = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{15}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and standard deviations of p_x and p_y.
Variance:

$$S_V = \sum_i \sum_j (i-\mu)^2 p(i,j) \tag{16}$$

Sum average:

$$S_A = \sum_{i=2}^{2N_g} i p_{x+y}(i) \tag{17}$$

Difference entropy:

$$S_F = -\sum_{i=0}^{N_{g-1}} p_{x-y}(i) \log\{p_{x-y}(i)\} \tag{18}$$

The features used in this paper are selected based on the previous works [21]. These features work well especially for MRI brain tumor images.

5 Feature Selection

Feature selection refers to the problem of dimensionality reduction of data, which initially consists of large number of features. The objective is to choose optimal subsets of the original features which still contain the information essential for the classification task while reducing the computational burden imposed by using many features. In this work, Particle Swarm Optimization is proposed for feature selection.

5.1 Particle Swarm Optimization (PSO)

The PSO method is a member of wide category of Swarm Intelligence methods for solving the optimization problems. It is a population based search algorithm where each individual is referred to as particle and represents a candidate solution. Each single candidate solution is “an individual bird of the flock”, that is, a particle in the search space. Each particle makes use of its individual memory and knowledge to find the best solution . All the particles have fitness values, which are evaluated by fitness function to be optimized and have velocities which direct the movement of the particles. The particles move through the problem space by following a current of optimum particles.
The initial swarm is generally created in such a way that the population of the particles is distributed randomly over the search space. At every iteration, each particle is updated by following two “best” values, called *pbest* and *gbest*. Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness value). This fitness value is called *pbest*. When a particle takes the whole population as its topological neighbor, the best value is a global best value and is called *gbest*. The detailed algorithm is given below:

Step 1: Set the constants $k_{max,}$ c_1, c_2, r_1, r_2, w.
Randomly initialize particle positions $x_0(i)$ for i=1, 2………p.
Randomly initialize particle velocities $v_0(i)$ for i=1,2……..p.
Step 2: Set k=1.
Step 3: Evaluate function value f_k using design space coordinates $x_k(i)$
If $f_k \geq f_{pbest}$, then pbest(i) = $x_k(i)$
If $f_k \geq f_{gbest}$, then gbest= $x_k(i)$
Step 4: Update particle velocity using the following equation

$$v_{k+1}(i)= w^*(v_k(i))+c_1r_1^*(pbest_k(i) - x_k(i))+c_2r_2(gbest_k - x_k(i)) \quad (19)$$

Update particle position vector using the following equation

$$x_{k+1}(i)=x_k(i)+v_{k+1}(i) \quad (20)$$

Step 5: Increment i. If i > p, then increment k and set i=1.
Step 6: Repeat steps 3 to 5 until k_{max} is reached.
The notations used in this algorithm are:
k_{max} = maximum iteration number
w = inertia weight factor
c_1, c_2 = cognitive and social acceleration factors
r_1, r_2 = random numbers in the range (0, 1).
In this work, each of the twelve features are represented by a chromosome (string of bits) with 14 genes (bits) corresponding to the number of features. An initial random population of 20 chromosomes is formed to initiate the genetic optimization. The initial coding for each particle is randomly generated. The order

of position of the features in each string is mean, standard deviation, cluster shade, cluster prominence, skewness, kurtosis, energy, entropy, contrast, inverse difference moment, correlation , variance, sum average and difference entropy respectively. A suitable fitness function is estimated for each individual. The fittest individuals are selected and the crossover and the mutation operations are performed to generate the new population. This process continues for a particular number of generations and finally the fittest chromosome is calculated based on the fitness function. The features with a bit value "1" are accepted and the features with the bit value of "0" are rejected. The fitness function used in this work is given by

$$Fitness = (\alpha * \gamma) + \beta * \left(\frac{|c| - |r|}{|c|} \right) \tag{21}$$

where γ = classification accuracy

$|c|$ = total number of features

$|r|$ = length of the chromosome (number of '1's)

$\alpha \in [0, 1]$ and β = 1- α

This formula shows that the classification accuracy and the feature subset length have different significance for feature selection. A high value of α assures that the best position is at least a rough set reduct. The goodness of each position is evaluated by this fitness function. The criteria are to maximize the fitness values. An optimal solution is obtained at the end of the maximum iteration. This value is binary coded with fourteen bits. The bit value of "1" represents a selected feature whereas the bit value of "0" represents a rejected feature. Thus an optimal set of features are selected from the PSO technique.

6 Classifiers

In this work, three types of neural network classifiers are used. The counter propagation neural network along with two of its modifications is implemented for abnormal brain tumor image classification.

6.1 Conventional CPN

Counter propagation neural network is a hybrid neural network employing both supervised and unsupervised training methodologies. A CPN consists of three layers: the input layer, the competition layer and the output layer. Given the input training set $(\overline{X_i}, \overline{Y_i})$ i = 1, 2……N, where $\overline{X}_i = (x_{i1}, x_{i2}......x_{in})$ and $\overline{Y_i} = (y_{i1}, y_{i2}.....y_{in})$, the configuration of the network is as follows: number of neurons in the input layer = n, number of neurons in the competition layer=N. Fig. 3 shows the architecture of CPN used in this work.

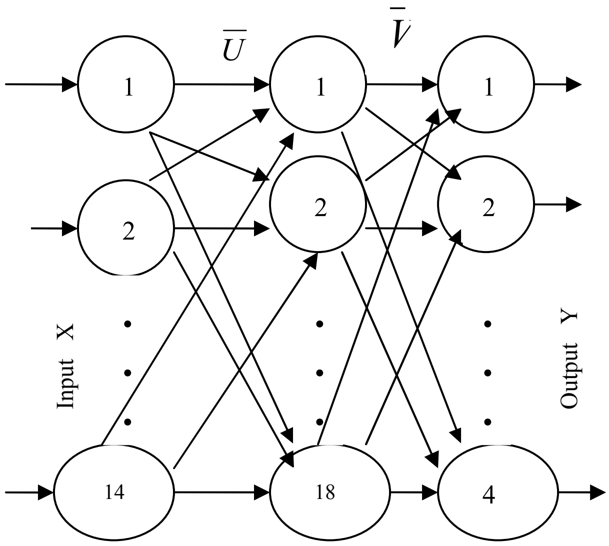


Fig. 3. Topology of CPN

6.1.1 Algorithm

The learning algorithm proceeds in two steps. In the first step, each component of the training instance $\overline{X_i}$ is presented to the input layer. Let U_{ij} be the arbitrary initial weights vector assigned to the links connecting input node i with the competition node j. The transfer function of the competition layer is defined by the Euclidean distance d_j between the weight vector $\overline{U_j}$ and the input vector $\overline{X_k}$ as

$$d_j = \left\|\overline{U_j} - \overline{X_k}\right\| = 1/2\left(\sum_i \left(U_{ij} - X_{ki}\right)^2\right)^{1/2} ....j = 1,2\ldots\text{N} \quad (22)$$

For each $\overline{X_k}$, each node in the competition layer competes with the other nodes, and the node with the shortest Euclidean distance wins. The output of the winning node is set to 1 and the rest to 0.thus, the output of the jth node in the competition layer is

$$Z_j = 1.0 \text{ if } d_j < d_i, \quad (23)$$
$$Z_j = 0.0.........\text{Otherwise}$$

The weight updation between the input layer and the competition layer is given by

$$U_{ij}^{t+1} = U_{ij}^{t} + \alpha\left(x_{ki} - U_{ij}^{t}\right)Z_j \quad (24)$$

where t is the iteration number and α , the learning coefficient is such that 0< $\alpha \leq 0.8$, as suggested by Heicht- Neilsen. After the weight vectors $\overline{U_{ij}}$ have stabilized, as the second step, the output layer begins learning the desired output. The weight adjustments for the output layer is given by

$$V_{jl}^{t+1} = V_{jl}^{t} + \beta\left(y_{kl} - V_{jl}^{t}\right)Z_j \qquad (25)$$

where the learning coefficient lies in the range $0 < \beta \leq 1.0$.

A stabilized set of weights are obtained by training the neural network with the input samples from the four abnormal categories. After training, the network is tested with the unknown data set. The major drawback of the CPN is the high convergence time period which leads to computational complexity. A suitable modification is made in the conventional CPN to improve the convergence rate besides yielding sufficient classification accuracy.

6.2 Modified CPN

In order to enhance the convergence rate of the conventional CPN, a suitable modification is performed in the training methodology. Usually, CPN employs the unsupervised training methodology for weight adjustment between the input and hidden layer and supervised training methodology between the hidden and output layer. In this work, the stabilized weights between the hidden and output layer are obtained without any training procedure. The architecture of the modified CPN is same as that of the conventional CPN.

6.2.1 Training Algorithm

The training algorithm involves the following two phases:

(i) Weight adjustment between the input layer and the hidden layer

The weight adjustment procedure for the hidden layer weights is same as that of the conventional CPN. It follows the unsupervised methodology to obtain the stabilized weights. Eqns (22) - (24) summarises this procedure. This process is repeated for a suitable number of iterations and the stabilized set of weights are obtained. After convergence, the weights between the hidden layer and the output layer are calculated.

(ii) Weight adjustment between the hidden layer and the output layer

The weight adjustment procedure employed in this work is significantly different from the conventional CPN. The weights are calculated in the reverse direction without any iterative procedures. Normally, the weights are calculated based on the criteria of minimizing the error. But in this work, a minimum error value is specified initially and the weights are estimated based on the error value. The detailed steps of the modified algorithm are given below.

Step 1: The stabilized weight values are obtained when the error value (target-output) is equal to zero (or) a predefined minimum value. The error value used for convergence in this work is 0.01. The following procedure uses this concept for weight matrices calculation.

Step2: Supply the target vectors t_l to the output layer neurons

Step 3: Since $(t_l - y_l)$ = 0.01 for convergence, the output of the output layer neurons is set equal to the target values as

$$y_l = t_l - 0.01 \tag{26}$$

Step 4: Once the output value is calculated, the sum of the weighted input signals (y_in_l) can be estimated. Since the sigmoid activation function is used, the following equation yields the value for y_in_l.

$$y\_in_l = \ln\left[\frac{y_l}{|1-y_l|}\right] \tag{27}$$

Step 5: Based on the values of y_in_l, the weight matrix v_{jl} is calculated using the following expression. The matrix equation is solved using Math toolbox in MATLAB.

$$y\_in_l = \sum h_j \cdot v_{jl} \tag{28}$$

Where h_j is the output value of the hidden layer and this value is obtained at the completion of phase 1.

Thus without any training methodology, the weight values are estimated. This technique accounts for higher convergence rate since one set of weights are estimated directly. A further improvement can be made in this network by considering the weight adjustment procedure between the input layer and the hidden layer. Even though it cannot be made iterative-free, the number of input training samples can be reduced using optimization algorithms which further reduce the time taken for the training procedure.

6.3 PSO Based Modified CPN

The third classifier used in this work is a hybrid PSO based CPN classifier. The objective for using the optimization algorithm is twofold: (i) dimensionality reduction which improves the convergence rate and (ii) elimination of insignificant features which improves the classification accuracy. In this work Particle Swarm Optimization algorithm is used for optimal feature selection. The extracted features are subjected to this optimization technique which finally yields the optimal feature set. The number of neurons used in the input layer for this PSO based CPN classifier is reduced since the number of optimal features is lesser than the complete feature set. Also, the mathematical calculations are minimized because of the reduced size of the weight matrix. Hence, a significant reduction in the time period is achieved for the weight adjustment of the hidden

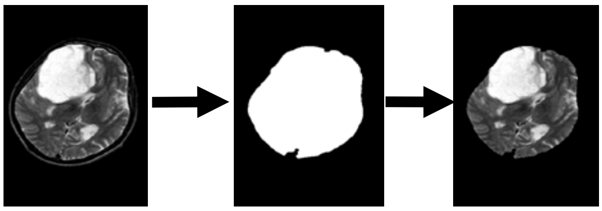
layer neurons. The training methodology for output layer neurons is same as that of modified CPN. Thus, modified CPN is better than conventional CPN and PSO based CPN is superior to modified CPN.
The experiments are carried out on an IBM PC Pentium with processor speed 700 MHz and 256 MB RAM. The software used for the implementation is MATLAB (version 7.0) [22] developed by Math works Laboratory.

7 Results and Discussions

Experiments are conducted on 478 real time brain images from four abnormal classes. These images are tested with all the three classifiers. The performance measures used in this work are classification accuracy and convergence time period. Classification accuracy is the ratio of the number of correctly classified images to the total number of images. Convergence time period is the time taken for training process and testing process. Initially, the results of the pre-processing technique are reported followed by the feature selection and the classification accuracy results.

7.1 Image Pre-processing results

Fig. 4 shows the output of the pre-processing techniques. In this section, the result is displayed only for a sample image.



(a) (b) (c)

Fig. 4. Pre-processing System Overview

Fig. 4(a) shows the MR input image. Fig. 4(b) depicts the mask which has been created to remove the extra cranial tissues. Fig. 4(c) shows the skull extracted image which is obtained by the superimposition of Fig. 4(a) and Fig. 4 (b). This procedure is repeated for all the images to make them extra cranial tissue free images. Thus the output images are free from the skull tissues (outer tissues) which guarantee high accuracy for the subsequent stages.

7.2 Feature selection results

All the features extracted from the input images does not account for high accuracy. Hence, a feature selection process is highly essential for eliminating the insignificant features. In this work, PSO is used as an optimization algorithm for feature selection. Table 1 shows the features selected by the optimization algorithm.

Table 1. Optimal Feature Set

Feature type	Whole feature set	Features selected by PSO
Features based on wavelet transform	Mean, standard deviation, cluster shade & cluster prominence	Cluster shade & cluster prominence
Features based on first order histogram	Skewness, Kurtosis, Energy, Entropy	Skewness, Kurtosis & Entropy
Features based on GLCM	Contrast, inverse difference moment, correlation, variance, sum average, difference entropy	Contrast, inverse difference moment, correlation & difference entropy

From the above table, it is evident that the number of features is reduced through the optimization algorithm. Two features are selected from the first category; three features are selected from the second category and four features are selected from the third category. Since mean and standard deviation are purely intensity based features, they are sidelined by the optimization algorithm. Energy is also directly dependent on the statistical property of the input data and hence it is least preferred over skewness, kurtosis and entropy. Other features such as correlation and difference entropy represent the similarity measure and average information respectively and hence these features carry more importance than the other features. Thus, the total number of features is 14 and the features selected by PSO are 9. Thus, a significant dimensionality reduction is achieved through the optimization algorithm. This optimization technique not only improves the convergence rate but also enhances the classification accuracy results to a higher extent.

7.3 Classification accuracy results

Three classifiers are tested in this work and two of the classifiers use the complete feature set and one of the classifier use the optimal feature set. Among 478 images, 30 images from each class are used for training and the remaining images are used for testing. The classification accuracy for the conventional CPN network is shown in Table 2.

Table 2. Performance measure of CPN network for the tumor types

Type	No.of training images	No.of testing images	Number of correctly classified images	Classification accuracy (%)
Metastase	30	87	78	89.6
Meningioma	30	95	86	90.5
Astrocytoma	30	75	62	82.6
Glioma	30	101	91	90.0

From the above table, it is evident that meningioma images have been classified to a higher extent and the level of misclassification in the astrocytoma images is comparatively higher than the other types. Then, the classification is performed with the modified CPN and the classification accuracy is noted. Table 3 lists the classification accuracy results of the GA optimized classifier.

Table 3. Performance measure of modified CPN for the tumor types

Type	No.of training images	No.of testing images	Number of correctly classified images	Classification accuracy (%)
Metastase	30	87	79	90.8
Meningioma	30	95	87	91.6
Astrocytoma	30	75	62	82.6
Glioma	30	101	92	91.0

The classification accuracy of the modified CPN is almost same as that of the conventional CPN. Since both the networks use the same features, there is no significant change in the accuracy. But the advantage of the modified CPN is the superior convergence rate since the network is free from iterations. Table 4 lists the classification accuracy results of the PSO based classifier.

Table 4. Performance measure of PSO based CPN for the tumor types

Type	No.of training images	No.of testing images	Number of correctly classified images	Classification accuracy (%)
Metastase	30	87	83	95.4
Meningioma	30	95	91	95.7
Astrocytoma	30	75	69	92.0
Glioma	30	101	98	97.0

A significant improvement in classification accuracy is obtained for the PSO based CPN over the other classifiers which is evident from the above results. Since the insignificant features are eliminated, the accuracy has been substantially increased. Table 5 lists the overall classification accuracy of the classifiers.

Table 5. Overall performance of the classifiers

Classifier	Overall classification accuracy (%)
Conventional CPN	88.175
Modified CPN	89.0
PSO based CPN	95.02

From the above table it is evident that the PSO based CPN is superior over the other classifiers in terms of classification accuracy.

7.4 Computational complexity and convergence rate

In terms of computational complexity, the PSO based CPN is highly efficient over the other classifiers. Computational complexity is indirectly measured by the number of neurons used and the size of the weight matrix. Table 6 shows the parameters used by the classifiers.

Table 6. Parameters used by the classifiers

Classifier	Number of input neurons used	Number of features used	Size of the weight matrix
Conventional CPN	14	14	14×18
Modified CPN	14	14	14×18
PSO based CPN	9	9	9×18

From the above table, it is clearly understood that the architecture of the PSO based CPN is highly simplified over the other classifiers. It also reveals the less number of mathematical computational operations involved in PSO based CPN. This simplicity is highly essential for real-time environment which makes the PSO based CPN ideal for practical applications. The three classifiers are further analyzed in terms of the time period required for convergence. Table 7 depicts the convergence rate of the classifiers.

Table 7. Convergence rate of the classifiers

Classifier	Average training time period (CPU secs)	Average testing time period (CPU secs)
Conventional CPN	45.2	0.09
Modified CPN	10.02	0.09
PSO based CPN	6.57	0.09

The average convergence time period of the PSO based CPN is very much less when compared with the other classifiers. The PSO based CPN also yields a significant classification accuracy which reveals the potential of this network. Thus, the proposed network is superior over the other conventional classifiers in terms of the performance measures.

8 Conclusion and Future Work

This work explores the necessity for optimization algorithms to enhance the performance of the classifiers. In this paper, PSO is used as the optimization algorithm and it is used along with the modified CPN classifier. Experimental results suggest a 5-6% improvement in the classification accuracy for the PSO based CPN over the other classifiers. Approximately 8 times increase in the convergence rate is also achieved by the PSO based CPN classifier which is highly essential for real-time applications. Hence, an optimization technique is highly essential irrespective of the classifiers used. Thus, the application of PSO algorithm for performance improvement of the neural classifier is explored in the context of abnormal brain tumor image classification.
As an extension of this work, a different optimization algorithm can be used to estimate the performance of the classifiers. A different set of features also can be used to improve the classification accuracy. Emphasis may be given for selection of the classifiers to enhance the performance measures. These experiments can be carried out on a different set of database in order to generalize the technique. Irrespective of the modifications and the systems used, this work highlights the significance of optimization algorithm for accurate and quick image classification systems.

Acknowledgment

The authors sincerely thank M/S. Devaki Scan Centre, Madurai, India for their support in providing the image database and expert validation.

References

[1] Lukas, L., Devos, A. and Suykens, A., “Brain tumor classification based on long echo proton MRS signals”, *Artificial Intelligence in Medicine*, vol. 31, (2004), pp.73-89.
[2] Sandeep,C., Patnaik, L. and Jaganathan, N., “Classification of MR brain images using wavelets as input to SVM and neural network”, *Biomedical signal processing and control*, vol.1, (2006), pp.86-92.

[3] Prastawa,M., Bullit,E. And Moon,N., "Automatic brain tumor segmentation by subject specific modification of atlas priors", *Medical image computing*, vol.10, (2003), pp.1341-1348.
[4] Egmont-Petersena,M et.al., "Image processing with Neural Networks-a review", *Pattern Recognition*, (2002), pp. 2279-2300.
[5] Fausett, L., *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Englewood Cliffs, NJ: Prentice Hall, (2004).
[6] Raff, U etal., "Quantitation of grey matter, white matter, and cerebrospinal fluid from spin-echo MR images using an artificial neural network technique", *Journal of Medical Physics,* IET digital library, vol.21, (1994), pp. 1927-1933.
[7] Tomatis, S etal., "Automated melanoma detection: Multispectral imaging and neural network approach for classification", *Journal of Medical Physics*, IET digital library, vol. 30, (2003), pp. 207-212.
[8] Luts, J etal., "A MRI and MRSI based brain tumor recognition using LS-SVM with probabilities", *Artificial Intelligence in Medicine*, vol.40, (2007), pp. 87-102.
[9] Mellsen,W., Wehrens, R. and Buydens, L., "Supervised Kohonen networks for classification problem", *Chemometrics and Intelligent Laborotary Systems*, vol.83, (2006), pp. 99-113.
[10] Vomweg, T. et al., "Improved artificial neural networks in prediction of malignancy of lesions in contrast enhanced mammography", *Journal of Medical Physics*, vol.30, (2003), pp.23541-2350.
[11] Knowles, A., Gibbs, P. and Turnbull, L.W., "Improved classification of breast DCE-MRI using a neural network ensemble", *Proc. Int. Soc. Magn. Reson. Med*., vol.8, (2000), pp. 2157-2163.
[12] Changua.Y etal., "An efficient hidden layer training method for the multilayer perceptron", *Journal of Neurocomputing*, vol.70, (2006), pp.525-535.
[13] Shamsudin.S.M, Darus.M and Sulaiman.M.N., "Classification of reduction invariants with improved BPN", *International journal of mathematics and mathematical sciences*, vol.30, (2002), pp. 239-247.
[14] Khalil.S., "A back propagation neural network for computer network security", *Journal of Computer Science*, vol.2, (2006), pp. 710-715.
[15] Peterson.R, Doom.T and Raymer.M., "GA facilitated KNN classifier optimization with varying similarity measures", *Proceedings of Conference on Genetic and evolutionary computation*, (2005), pp.1549-1550.
[16]Palaniappan.R and Eswaran.C., "Using genetic algorithm to select the presentation order of training patterns that improves ARTMAP classification performance", *Applied Soft Computing*, vol.9, (2009), pp.100-106.
[17] Sousa,T. Silva, A. And Neves,A., "PSO based data mining algorithms for classification tasks", *Parallel Computing*, vol.30, (2004), pp.767–783.
[18] Bhattacharya, M and Das, A., "Registration of multimodality of medical imaging of brain using PSO", *Proceedings of the First International Conference on Intelligent Human Computer Interaction*, (2009), pp.131-139.
[19] Arivazhagan.S and Ganesan.L., "Texture classification using wavelet transform", *Pattern Recognition Letters*, vol.24, (2003), pp. 1513-1521.

[20] Haarlick. R.M., "Statistical and structural approaches to texture", *IEEE Transactions on Systems, Man and Cybernatics*, vol.67, (1979), pp.786-804.
[21] Xuan, X. Etal., "Statistical structure analysis in MRI brain tumor segmentation", *International conference on image and graphics*, vol.22, (2007), pp. 421-426.
[22]MATLAB, User's Guide, The Math Works, Inc., Natick, MA 01760, (2002).