CHAPTER 4

PERFORMANCE COMPARISON OF FEED FORWARD NEURAL NETWORK USING VARIOUS BP ALGORITHMS

4.1 OVERVIEW

The various training algorithms for BPNN is analyzed for obtaining better epileptic seizure detection. In the literature, comparison of the performance of various Back Propagation algorithms are studied in the area of predicting financial stock market (Lahmiri 2011), Haematoma classification of brain images (Bhavana 2014), vertical electrical sounding data inversion application (Srinivas 2012), noise reduction (Badri 2010). The main contribution of the proposed research is to determine the impact of various BP training algorithms in epileptic seizure detection. The method developed has contributed to the achievement of better classification accuracy and minimum execution time in the classification process. The features are extracted using FastICA algorithm and classifier performance obtained for various Back Propagation training algorithms-Gradient Descent, Gradient Descent With Momentum (GDWM), Scaled Conjugate Gradient Algorithm, One Step Secant, Powell-Beale Restarts, Gradient Descent with Adaptive learning rate, Fletcher – Powell Conjugate Gradient and Levenberg Marquardt.
4.2 NEURAL NETWORK TRAINING

Artificial Neural Networks mimic the human brain in its ability to learn from events that happened in the past and apply the same in the future to a similar situation. It consists of a number of processing units along with a node function that determines the output of the node. Back Propagation includes four stages in the training algorithm namely, Initialization of weights, Feed forward, Back Propagation of errors and updating weights and biases.

Training the Back Propagation network reduces the system error to a minimum. The outline of the proposed approach is illustrated in Figure 4.1. The process of epileptic seizure detection approach is composed of the FastICA and Back Propagation algorithm neural network training.

![Figure 4.1 Block Diagram of the Proposed Epileptic Seizure Detection with Performance Comparison](image-url)
The proposed algorithm in the present research is as follows:

1. Loading of the dataset

2. Identification of the components related to epileptic seizure detection using Fast Independent Component Analysis

3. Creation of a network. The feed forward network with the tan-sigmoid transfer function is used in the hidden layer and output layer. 15 neurons (subjective) are used in the hidden layer. The network comprises of two outcome neurons.

4. Random values are assigned to the input weight and bias.

5. Allocation of 80% of the data for training, 10% of the data for validation and remaining 10% of the data for testing.

6. Training and testing of the network, using Back Propagation algorithm.

7. Comparison of the results of training algorithms using accuracy.


The process of epileptic seizure detection approach is composed of the Fast Independent Component Analysis and Back Propagation algorithm Neural Network training. Consequently, trained neural networks are tested and validated with the stored EEG signals for performing the comparative study over all the training functions.
4.2.1  Fast Independent Component Analysis

Independent Component Analysis is a method of recovering the underlying signals from linear mixtures of those signals to regulate a set of components that are maximally independent of each other. When deriving these components, the data are scattered into either spatially or temporally independent components. FastICA can concede interesting information on brain activity by giving access to its independent components.

4.2.2  Gradient Descent Algorithm (GD)

Gradient methods are generally efficient when the function to be optimized is continuous in its first derivative. These methods use information about the slope of the function to follow a search direction of the minimum. The best advantage of GD methods is that convergence can be very fast if the error function shows high gradient towards the global minimum. Starting from this improved initial point, this search quickly converges.

Gradient Descent With Momentum is a predominant training method, concedes a network to retort not only to the confined gradient, but also to the current trends in the error surface. Acting like a low pass filter, momentum avows the network to disregard the small features in the error plane.

4.2.3  Gradient Descent with Adaptive (GDWA) Learning Rate

The Gradient Descent with Adaptive learning rate Back Propagation is mainly dependent on the setting of the learning rate. It also includes the net input/output and transfer functions. The performance of this algorithm is dependent on the given learning training rate.
The training functionalities terminate, when the maximum number of epochs is attained, exceeds the maximum time allocated, minimized performance and the performance gradient drops the minimum gradient value.

### 4.2.4 Scaled Conjugate Gradient (SCG)

Conjugate gradient algorithm craves a line search regarding all iterations. This line exploration is computationally classy. Since it requires that the network retorts to all the training inputs is evaluated several times for each search. The Scaled Conjugate Gradient algorithm was produced to skip the time consuming line search.

The SCG algorithm is based upon a class of optimization approaches well known in numerical investigation as the conjugate gradient Methods. SCG values a speed-up of at least an order of magnitude relative to Back Propagation. The speed-up depends on the convergence principle, that is massive the demand for devaluation in error the bigger the speed-up.

SCG is entirely automated incorporating no user dependent parameters and obviates a time consuming line search, which Conjugate Gradient uses in each iteration to complete a convenient step size. Integrating issue reliant structural information in the architecture of a Neural Network often lowers the overall complexity. This algorithm combines the exemplary trust region approach used in the Levenberg Marquardt algorithm, with the Conjugate gradient method.

The SCG schedule may require more iteration to congregate than the other conjugate gradient algorithms, but the number of estimations in all the iterations is significantly reduced as no line search is performed.
4.2.5 One Step Secant (OSS)

Since the Broyden-Fletcher-Goldfarb-Shanno (BFGS) methods consume more storage and computation in all the iterations than the Conjugate gradient algorithm, there is need for a secant approximation with lesser storage and computation prerequisite. The One Step Secant approach is an effort to bridge the gap between the Conjugate gradient algorithms and the Quasi-Newton (secant) algorithms.

There is no contingency for storing the complete Hessian matrix. It deduces that at all iterations; the prior Hessian was the identity matrix. This has the supplementary advantage that the new search route can be premeditated without computing a matrix inverse. This algorithm obligates less storage and computation per epoch than the BFGS algorithm and needs lightly more storage and computation per epoch than the Conjugate gradient algorithms. It can combine features of full quasi-Newton algorithms and conjugate gradient algorithms.

4.2.6 Powell-Beale Restarts (PBR)

In Powell-Beale Restarts, the search direction will be systematically reset to the negative of the gradient, for all conjugate gradient algorithms. The reset point will be fixed where the number of iterations is equivalent to the amount of network parameters (weight and biases), but there are other reset approaches that can progress the competence of training.

For this practice, the present work resumes if there is very little orthogonality left between the present gradient $g_k$ and the past gradient $g_{k-1}$. This is experienced with the following discrimination.

$$ |g^T_{k-1}g_k| \geq 0.2 \|g_k\|^2 $$

(4.1)
If this state is satisfied, the search direction is reset to the adverse of the gradient.

4.2.7 Fletcher-Powell Conjugate Gradient (FPCG)

The Fletcher-Powell Conjugate Gradient function updates the weight and bias values including Fletcher-Powell renovate. The input set comprises of details about the Neural Network, delayed input vectors, target vectors, initial input delay conditions, batch size and time steps. The algorithm needs the smallest storage requirements among all the other compared algorithms. Moreover, the training function enormously affects the performance with respect to weight and bias values.

4.2.8 Levenberg Marquardt Back Propagation (LMBP)

Levenberg Marquardt Back Propagation (LMBP) is another optimal way of training the Neural Networks resulting in minimum errors. Training patches weight and bias values in accordance with Levenberg Marquardt optimization. LM is the fastest Back Propagation algorithm in image processing modules and is highly suggested as the first choice of supervised algorithm even though it requires more memory than the other algorithms.

4.3 RESULTS AND DISCUSSION

The EEG dataset used in the present research is the same as the previous chapter with four sets each containing 100 single channel EEG data (Andrzejak et al 2001). The experimental simulation has been carried out in MATLAB. Various network training functions used in updating the weight and bias values are traingd for Gradient Descent, traingdm for Gradient Descent With Momentum, trainscg for Scaled Conjugate Gradient Algorithm,
trainoss for One Step Secant, traincgb for Powell-Beale Restarts, traingda for Gradient Descent with Adaptive learning rate, traincgf for Fletcher – Powell Conjugate Gradient and trainlm for Levenberg Marquardt. Figure 4.2 shows the accordance between the 100 epochs and the mean squared error rate with the Gradient descent algorithm without any adaptive and momentum combinations in that. From the above training, it is found that the Regression R-value of the algorithm is 0.9246 and the best validation performance is 0.093783. Accuracy is obtained from the confusion matrix. GD algorithm takes the maximum number of iterations that is 100 for producing an accuracy of 90.75% in 9.6099 seconds.

![Figure 4.2 Gradient Descent Algorithm](image)

The Neural Network is also inspected in gradient descent with an adaptive algorithm in spite of the examination of further competence. The Figure 4.3 demonstrates the training performance of the dataset affords the Regression R-value of 0.93527 and gives the best validation performance rate of 0.082295. It contributes to 98.5% of accuracy in 9.5387 seconds.
Figure 4.3 Gradient Descent with Adaptive Learning Rate

The next result analysis is made with the training approach called Gradient Descent with Momentum. The Figure 4.4 shows the performance rate of the gradient descent with momentum algorithm. The training approach contributes the Regression R-value 0.9739. It is glassy that, with the complete accommodation of 100 epochs, it contributes the best validation performance rate of 0.057991, but with 100% accuracy in 9.8426 seconds.

Figure 4.4 Gradient Descent with Momentum
The approach for automatic detection of epileptic seizure is then inspected with Scaled Conjugate Gradient method where the training of the network is described with trainscg. The process reaches the best validation performance in 47 iterations. The Scaled conjugate training method affords 100% accuracy in 8.9952 seconds.

The Figure 4.5 shows the illustration of the results of Scaled Conjugate Gradient training which contributes the Regression R- value of 0.99749. The best validation performance rate of this training is determined as 0.010643. Besides the above techniques, the Scaled conjugate gradient network method affords good performance rate accommodating less epochs.

![Figure 4.5 Scaled Conjugate Gradient](image)

The SCG approach of the network training begets 100% accurate results in 8.9952 seconds. The examination proceeds with the conviction of the efficacy rate of One Step Secant method of training. The Figure 4.6 shows the correlation between the number of epochs desired for better accurate prediction and the Mean squared error rate of OSS. The best validation performance is 0.0092 at epoch 53.
Figure 4.6 One Step Secant

The Regression R-value of this technique is decided as 0.98877. It affords 100% accurate results in 11.6329 seconds time duration. The Figure 4.7 indicates the performance rate of Powell-Beale Restarts. The Regression R-value of this training algorithm is 0.9821. The training procedure based on the network parameters such as weights and biases and follows a search direction concept for competent results.

Figure 4.7 Powell-Beale Restarts
It attends the best performance rate in the accomplishment of 40 iterations or epochs. The accuracy obtained through this training is 100% in 8.3196 seconds. The best validation performance is calculated as 0.010777.

Figure 4.8 Fletcher- Powell Conjugate Gradient

Figure 4.8 illustrates the performance of FPCG. The best validation performance is attained at 36th iteration and produces 0.012244 performance rate. It produces 100% accurate results in 9.8426 seconds. The training function used is traincfgf. The technique demonstrates its adaptability by exploiting fewer numbers of iterations. The Regression R-value of this conceit is determined as 0.98227.

Fletcher-Powell Conjugate gradient attains the target performance rate with the accomplishment of lower epoch value and less duration. The Figure 4.9 demonstrates the statistical training analysis of Levenberg Marquardt Back Propagation that is exhibited by trainlm.
Though trainlm is generally the fastest Back Propagation algorithm in the toolbox, and is eminently implied as a first choice supervised algorithm, and requires more memory than the other algorithms, it is not applicable for the network training with respect to the constraints. Hence, it produces Not a Number (NaN) results which could be studied with just four iterations. LM algorithm results in 55.5% accuracy with an execution time of 29.5872 seconds. The efficiency discrimination of the above explained methodology is made in terms of accuracy and the time needed for generating the results. The Figures 4.10 and 4.11 show the comparison chart for the scrutiny of the above specified network training classification.
Figure 4.10  Comparison of Accuracy for the Various Training Function Algorithms

Figure 4.11  Comparison of Execution Time (sec) for Various Training Function Algorithms
On correlating the performance of these aforementioned algorithms, highest accuracy with less execution time has been obtained for mostly the Conjugate Gradient based Back Propagation algorithm.

4.4 SUMMARY

The imperious antecedent of the present research is to inspect the performance evaluation of distinct training functions in epileptic seizure detection from the recorded EEG brain signals. Independent sub-components are segregated from the recorded brain signals for the Fast Independent Component Analysis, subsequently the signals are trained with some Neural Network techniques with disparate training functions such as Gradient Descent Algorithm, Gradient Descent With Momentum, One Step Secant, Scaled Conjugate Gradient, Powell-Beale Restarts, Gradient Descent with Adaptive learning rate, Fletcher-Powell Conjugate Gradient (FPCA) and Levenberg Marquardt Back Propagation.

From the results of all the above mentioned training functions, the performance evaluation of the trained network in accordance with the recorded EEG brain signals is determined. By the experimental results, it is evident that mostly the Conjugate Gradient based Back Propagation method is the best training function which contributes 100% result accuracy in epileptic seizure detection with a limited number of iterations and diminished time consumption. Attempts have then been made to improve the classification accuracy for more number of classification problem by making use of FastICA as a preprocessing step, STFT for denoising, and feature extraction.