RESEARCH ARTICLE OPEN ACCESS

# Binary Color Classification For Brain Computer Interface Using Neural Networks And Support Vector Machines

Charmi Sunil Mehta*, Damini Ashok Chopra**, Nidhi Milan Mavani***, Keval Bhavesh Goradia****, Alice N. Cheeran*****
Department of Electrical Engineering, Veermata Jijabai Technological Institute, Mumbai – 400 019, India

## ABSTRACT
As the power of modern computers grows alongside our understanding of the human brain, we move a step closer in transforming some pretty spectacular science fiction into reality. The advent of Brain Computer Interface (BCI) is indeed leading us to a burgeoning era of complete automation empowering our interaction with computer not only with robustness but with also a gift of intelligence. For the fraction of our society suffering from severe motor disabilities BCI has offered a novel solution of overcoming the problems faced in communicating and environment control. Thus the purpose of our current research is to harness the brain's ability to generate Visually Evoked Potentials (VEPs) by capturing the response of the brain to the transitions of color from grey to green and grey to red. Our prime focus is to explore EEG-based signal processing techniques in order to classify two colors; which can be further deployed in future by coupling the actuators so as to perform few basic tasks. The extracted EEG features are classified using Support Vector Machines (SVM) and Artificial Neural Networks (ANN). We recorded 100% accuracy on testing the model after training and validation process. Moreover, we obtained 90% accuracy on re-testing the model with all samples acquired for the task using Quadratic SVM classifier.
***Keywords-*** EEG, Classifier, Artificial Neural Networks (ANN), Support Vector Machines (SVM)

## I. INTRODUCTION

BCI has transformed from a fictional idea to a scientific revelation in neuroprosthetics and alternate methods of communication. BCI is a communication pathway that directly connects signals from the brain to a computer. The human brain is estimated to have around 10 billion neurons each connected on average to 10,000 other neurons. Each neuron receives signals through synapses that control the effects of the signal on the neuron. These synaptic connections are believed to play a key role in the behavior of the brain. EEG signals have a very low frequency range in Hertz as well as they have a low spatial resolution and low Signal to Noise Ratio (SNR). ANN has the unique ability to learn from the examples and to generalize, i.e., to produce the reasonable outputs for new inputs not encountered during a learning process. The neural networks may be regarded as the universal approximators of the measured data in the multidimensional space. They realize two types of approximation: the global and local one. The most representative example of local neural network is the Support Vector Machine (SVM). It is a two layer neural network employing hidden layer of radial units and one output neuron. The procedure of creating this network and learning its parameters is organized in the way in which we deal only with kernel functions instead of direct processing of hidden unit signals.

The main idea is that we get a clear classification of generated evoked potential epochs for our task of color detection. Meanwhile, some pre-processing methods such as band-pass filtering and DC offset removal [Section 3.2] are also performed for signal enhancement in order to increase the computational efficiency.

## II. DATA ACQUISITION

Electroencephalogram (EEG) is the record of the electrical potentials generated by the brain. Specific Harmonic oscillations commonly called rhythms observed in human EEG are classified into five major categories depending on their frequency ranges. These are alpha (α) 8-13 Hz, beta (β) 14-26Hz, gamma (γ) 30 Hz, theta (θ) 4-7.5 Hz, and delta (δ) 0.5-4 Hz [1]. Alpha range is the gamut of the brain signals.

Data for this research is Event Related Potentials (ERPs) generated in the brain in response to stimulus, which can be visual, auditory or somatosensory [2]. Visual Evoked potentials (VEP's), a type of ERPs are elicited due to visual stimulus [1]. VEP's were recorded using Galileo Mizar series (EbNeuro) in EDF format. Dataset consist of 25 recordings acquired from two healthy subjects. Subjects were asked to sit in a relaxed position, while the following task was performed.

**Task**:
1. Eyes open
2. The subject was asked to concentrate on the screen, while the screen showed grey-red (green)-grey
3. Eyes closed

The timings of the transition of the colors were recorded and were used for epoch extraction (section4).

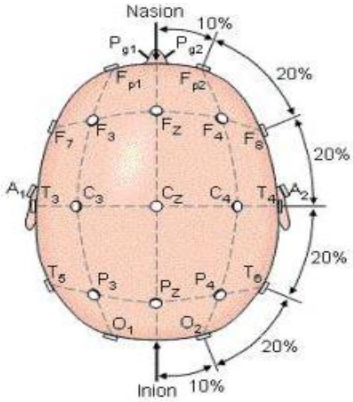


Fig.1 10-20 electrode system

The 10-20 System of Electrode Placement is a method used to describe the location of scalp electrodes. It is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The "10" and "20" refer to the fact that the actual distances between adjacent electrodes are either 10% or 20% of the total front–back or right–left distance of the skull [3].

## III. Data Processing

Signal processing forms an important part of a BCI design, since it is needed in extracting the meaningful information from the brain signal. Processing of the signals is done using EEGLAB, a toolbox in MATLAB [4].

### 3.1. Filtering

The signals consist of noise from unwanted frequency bands. This was removed by [5]

3.1.1 Applying a Notch Filter at 50 Hz to exclude AC Line noise

3.1.2 Applying a FIR filter of order 424 , low pass from 0 to 15 Hz , to get a linear phase delay, and to remove higher frequency components present

3.1.3 The Signal is being sampled at 128 Hz

### 3.2 DC Offset Removal

DC offset is an inherent component in EEG measurement with metallic electrodes components, which adds an arbitrary constant voltage to the original EEG data. DC noise was removed from the recordings subtracting the mean of the signal from each data sample [2].

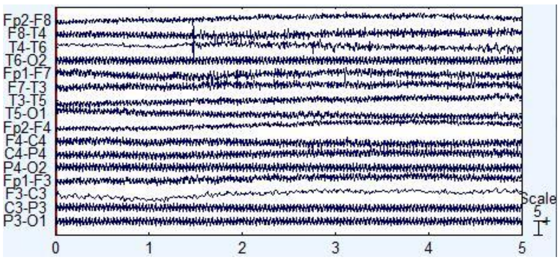


Fig.2 EEG signal before filtering

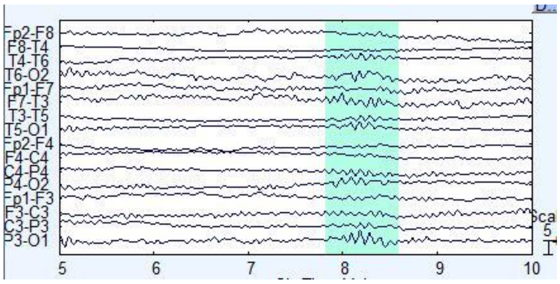


Fig.3 EEG signal after filtering

## IV. Feature Extraction And Selection

After the processing of the signal, the epoch is extracted using EEGLAB, signal of 500ms corresponding to recorded time is chosen.

### 4.1 Channel Selection

According to [6] many of the EEG channels appeared to represent redundant Information. Channels associated with Occipital and parietal lobe of the brain is of the most Importance. Hence Channels P3-O1, P4-O2, T6-O2, and T5-O1 are chosen.

### 4.2 Features

Below mentioned features are calculated for a given epoch, given channel, given sample [7].

4.2.1 Amplitude: It's the maximum value of the potential

4.2.2 Average: It's the Average of the Potential

4.2.3 Average Power

4.2.4 Peak Difference: Difference between Maximum and Minimum value of the Potential

4.2.5 Standard Deviation: Measure of the dispersion from the average

4.2.6 Variance: The average of the squared differences from the mean

Features 1-4 are calculated with the help of GNT Software used along with the acquiring system, while the features 5-6 with EEGLAB.

## V. CLASSIFICATION

### 5.1 Data Division

The data is divided into three subsections: training (60%), validation (15%) and testing (25%). The data is divided randomly into three groups and they are classified using ANN and SVM. The training dataset calculates the gradient (in ANN) and updates the weights and biases in both methods. The error on the validation set which is monitored during the training process decreases during the initial phase of as does the training set error. However, when the network begins to over fit the data, the error on the validation set typically begins to rise. The network weights and biases are saved at the minimum of the validation set error. The test set error is not used during training, but it is used to compare different models [8].

### 5.2 Artificial Neural Networks

ANNs are computational models that are used for machine learning and pattern recognition. Pattern recognition is feed forward networks that can be trained to classify the inputs according to the target classes. Feed forward is a type of ANN in which the information always flows in the forward direction from input to output through intermediate nodes. The classification is done by using nprtool-neural network, pattern recognition tool in MATLAB.

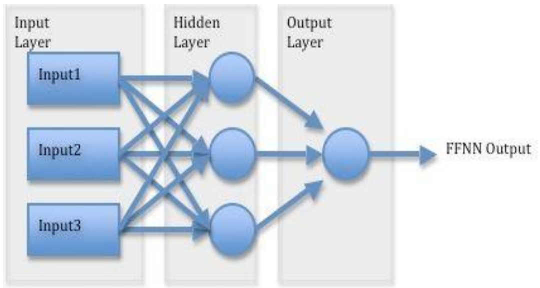


Fig.4 Feed forward Neural Network architecture

When the network is used for classification, the output layer typically has as many nodes as the number of classes and the output layer node with the largest output value gives the class output for a given input.

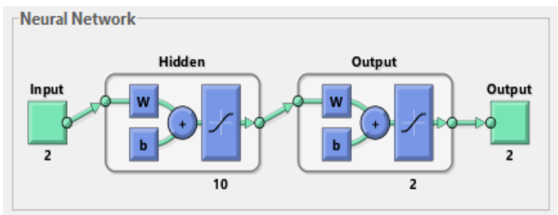


Fig.5 Neural Network architecture for classification

### 5.2.1 Back propagation

Back propagation is a supervised learning method where the activation functions used by artificial neurons have to be differentiable. Its two phases are:

5.2.1.1 Propagation phase

1) Forward propagation of the input through the neural network to get the output activations.
2) Backward propagation of the output activations in order to get deltas of the output and hidden neurons.

5.2.1.2 Weight updating phase

1) Multiply deltas of output with input activations to get gradient of weights.
2) Subtract a ration of gradient from weight, which is learning rate.

The algorithms used are gradient descent and scaled conjugate gradient. For both the algorithms the training stops when any of these conditions occurs:

The maximum number of epochs (repetitions) is reached. The maximum amount of time is exceeded. Performance is minimized to the goal. The performance gradient falls below min_grad. Validation performance has increased more than max_fail times since the last time it decreased (when using validation).

### 5.2.2 Gradient Descent

Gradient descent is a first-order optimization algorithm. The weights and biases are updated in the direction of the negative gradient of the performance function. The larger the learning rate a(k) the bigger the step. If the learning rate is made too large, the algorithm becomes unstable. If the learning rate is set too small, the algorithm takes a long time to converge.

One iteration of this algorithm:

$$y(k+1) = y(k) - a(k)\, g(k) \quad (1)$$

Where, y (k) is a vector of current weights and biases, g (k) is the current gradient, and a (k) is the learning rate. This equation is iterated until the network converges.

### 5.2.3 Scaled Conjugate Gradient

It is a second order technique generally finds a better way to a (local) minimum than a first order technique, but at a higher computational cost. SCG is a batch learning method, so shuffling the patterns has no effect. Indeed one iteration in SCG needs the computation of two gradients, and one call to the error function, while one iteration in standard back propagation needs the computation of one gradient and one call to the error function.[8]

**5.2.4 Performance of the network**

The performance of the network is calculated in terms of mean square error. MSE measures the average of the squares of the "errors." be estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

**5.3 Support Vector Machines**

Support Vector Machines (SVM's) are a relatively new learning method used for binary classification. The basic idea is to find a hyper plane which separates the d-dimensional data perfectly into its two classes. However, since example data is often not linearly separable, SVM's introduce the notion of a kernel induced feature space which casts the data into a higher dimensional space where the data is separable. By introducing the kernel, SVMs gain flexibility in the choice of the form of the threshold separating solvent from insolvent companies [9].

The best hyper plane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyper plane that has no interior data points. The support vectors are the data points that are closest to the separating hyper plane; these points are on the boundary of the slab. The following figure illustrates these definitions, with + indicating data points of type 1 and – indicating data points of type – 2.

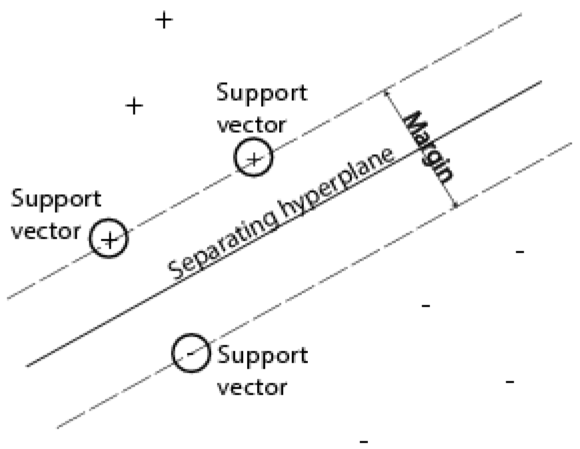


Fig.6 Nonlinear transformation with kernels

Some binary classification problems do not have a simple hyper plane as a useful separating criterion. For those problems, there is a variant of the mathematical approach that retains nearly all the simplicity of an SVM separating hyper plane.

This approach uses these results from the theory of reproducing kernels. There is a class of functions K(x, y) with the following property. There is a linear space *S* and a function *φ* mapping *x* to *S* such that:

$$K(x, y) = \langle \varphi(x), \varphi(y) \rangle. \quad (2)$$

The dot product takes place in the space S. This class of functions includes:

Polynomials: For some positive integer d:

$$K(x, y) = (1 + \langle x, y \rangle)\, d. \quad (3)$$

Radial basis function (Gaussian): For some positive number *σ*:

$$K(x, y) = \exp(-\langle (x-y), (x-y) \rangle / (2\sigma^2)) \quad (4)$$

Multilayer perceptron (neural network): For a positive number $p_1$ and a negative number $p_2$:

$$K(x, y) = \tanh(p_1 \langle x, y \rangle + p_2). \quad (5)$$

In k-fold cross-validation, we first divide the training set into k subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining k-1 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified [9].

In this case the samples are divided in the same way as discussed in section 5. Here the model is trained and validated while testing is on the best model chosen out of many. Finally all the samples are being tested on the best model with the help of MATLAB [10].

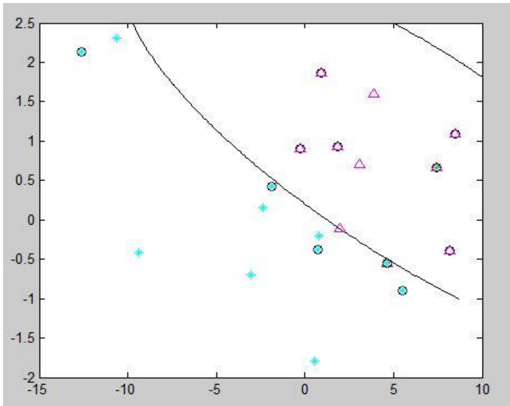


Fig.7 Quadratic SVM

## VI. Results And Analysis

Nos. 1, 6 are features amplitude, average, average power, peak difference, standard deviation, variance, (a) is channel P3-O1 and (b) is channel T5-O1. The last row indicates all features from 4 channels that are taken into consideration. 24

dimension data was reduced to 2 dimensions and by trying various combinations and the ones with optimum result are displayed.

Table 1

| Features | SCG | | GD | |
|---|---|---|---|---|
| | Test Accuracy | Final Accuracy | Test Accuracy | Final Accuracy |
| 1,2(a) | **100** | **85** | 80 | 70 |
| 1,3(a) | 80 | 75 | 60 | 65 |
| 2,4(a) | **100** | 80 | 60 | 50 |
| 4,5(b) | 80 | **85** | 40 | 65 |
| 4,6(b) | 60 | 75 | 60 | 60 |
| All | 80 | **85** | 80 | 75 |

Table2

| Features | Linear | | Quadratic | |
|---|---|---|---|---|
| | Test Accuracy | Final Accuracy | Test Accuracy | Final Accuracy |
| 1,2(a) | **100** | **85** | 80 | **85** |
| 1,3(a) | 60 | 75 | 60 | 75 |
| 2,4(a) | 80 | 65 | **100** | **85** |
| 4,5(b) | 40 | 65 | 80 | 80 |
| 4,6(b) | 60 | 70 | 80 | 80 |
| All | 80 | 85 | 80 | **90** |

Table3

| Features | Cubic | | MLP | |
|---|---|---|---|---|
| | Test Accuracy | Final Accuracy | Test Accuracy | Final Accuracy |
| 1,2(a) | 60 | 70 | 60 | 65 |
| 1,3(a) | 60 | 70 | **100** | 75 |
| 2,4(a) | 80 | 80 | 80 | 75 |
| 4,5(b) | 40 | 65 | 40 | 40 |
| 4,6(b) | 60 | 70 | 40 | 55 |
| All | 60 | **90** | 60 | 60 |

Table 4

| Features | Quadratic SVM | | Neural Network (SCG) | |
|---|---|---|---|---|
| | Test Accuracy | Final Accuracy | Test Accuracy | Final Accuracy |
| 4,8 | 80 | 85 | 100 | 85 |
| 8,16 | 100 | 85 | 100 | 80 |
| All | 80 | **90** | 80 | 85 |

On comparing the results from the table 1 SCG gave better performance, from table 2-3 quadratic gives better performance, from table 4 SVM proves to be better than NN.

## VII. CONCLUSION

After proposing the above mentioned models for classification of color, this work can be further extended in terms of many colors and online analysis e.g. simulating a scenario of traffic light signals in virtual environment or to identify and explore any possibility of analyzing the EEG signals and developing BCI applications for color blind and/or blind people. Since such applications are quite novel in their fields of BCI therefore requires enormous research work in interdisciplinary fields.

## VIII. ACKNOWLEDGEMENT

This work was undertaken as the realm of research for final year project. The authors would like to thank Prof. Rizwan Ahmed for imparting us knowledge on 'Advanced Digital Signal Processing' course and for his support and motivation, and Sir Jamshedjee Jeejebhoy Group of Hospitals (J. J. Marg, Mumbai, India) for providing facilities to use EEG acquisition system.

## REFERENCES

[1] Jorge Baztarrica Ochoa, "*EEG Signal Classification for Brain Computer Interface Applications*", ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE, March 28th, 2002.

[2] Md Z. Ali,"*EEG-BASED ASSESSMENT OF DRIVER'S COGNITIVE RESPONSE IN VIRTUAL TRAFFIC LIGHT ENVIRONMENT*",The Faculty of the College of Graduate Studies Lamar University,May 2012

[3] ABDUL-BARY RAOUF SULEIMAN, TOKA ABDUL-HAMEED FATEHI,"*FEATURES EXTRACTION TECHNIQES OF EEG SIGNAL FOR BCI APPLICATIONS",*Computer and Information Engineering Department College Of Electronics Engineering, University of Mosul Mosul, Iraq

[4] Ms. Antara Bhattacharya, Dr. N. G. Bawane, Ms. S. M. Nirkhi,"*Brain Computer Interface Using EEG Signals*",G.H.R.C.E, Nagpur

[5] Proakis, *Digital Signal Processing: Principles, Algorithms, And Applications, 4/E*, (Pearson Education India, 01-Sep-2007

[6] William. O. Tatum, Aatif. M. Husain, *"Handbook of EEG Interpretation"(Demos publication)*

[7] Onur Varol,Assoc. Prof. Müştak Erhan Yalçın,"*RAW EEG DATA CLASSFICATION AND APPLICATIONS USING SVM",*Istanbul Technical University Electrical -Electronics Engineering Faculty,May 2010

[8] Nandish.M, Stafford Michahial, Hemanth Kumar P, Faizan Ahmed,"Feature Extraction and Classification of EEG Signal Using Neural Network Based Techniques",*International Journal of*

*Engineering and Innovative Technology (IJEIT)* Volume 2, Issue 4, October 2012

[9] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin,"*A Practical Guide to Support Vector Classification*",Department of Computer Science National Taiwan University, Taipei 106, Taiwan

[10] Deon Garrett, David A. Peterson, Charles W. Anderson, Michael H. Thaut,"*Comparison of Linear and Nonlinear Methods for EEG Signal Classification*"

[11] D. J. Krusienski, D. J. McFarland, and J. R. Wolpaw, An evaluation of autoregressive spectral estimation model order for brain-computer interface applications, IEEE EMBS Ann. Int. Conf. New York, 1323-1326, 2006.