An Investigation into the area of Brain Imaging using Machine Learning Techniques

Rahim Hasnani[1] and Muhammad Rafi[2]

SZABIST

Karachi, Pakistan

***Abstract*—This study applies machine learning techniques to brain imaging data. Public EEG data of 122 subjects belonging to both control and alcoholic groups is used. The data contain multiple trials on each subject and each trail uses one of the three stimuli. The data is gathered from 64 different sensors for one second @ 256 Hz. Logistic regression, neural networks and Bayes point machine techniques are applied using Octave and Microsoft Infer.NET for training and prediction. Split sampling method is used for internal validation. Logistic regression and Bayes point machine provided similar results whereas neural networks proved to be slightly better than the other two methods**

.

***Keywords* - Machine Learning, Brain Imaging, Classification, EEG, Logistic Regression, Neural Networks, Bayes Point Machines, Octave, Infer.NET.**

I. Introduction

Brain imaging refers to decoding of brain states using multi-unit arrays, electrocorticography (ECoG), functional Magnetic Resonance Imaging (fMRI), electroencephalography (EEG) and near-infrared spectroscopy (NIRS) [1].

Electroencephalography (EEG) measures the tiny fluctuations in voltage off the neurons of the brain by putting sensors along the scalp that can record electrical activity [2]. This recoding is sampled at a frequency [2].

Decoding of brain imaging is a classification problem [1] and there are a number of machine learning algorithms available that carry out data classification.

In EEG/MEG & fMRI studies, mostly linear methods are used including Linear Discriminant Analysis and its variations [1].

In some cases, regression and correlation is used to carry out pre-processing and dimensionality reduction [3]. In some other studies Genetic algorithms are used to select features [4]. One study has used neuro fuzzy models and rough sets to classify epileptic seizures [5].

A. Brain Imaging Data Set

EEG data for 122 subjects is publicly available. This data is part of a large study to correlate alcoholism with genetics [6]. The full data set is 700 MB and is in the form of text files. Each subject is either in control group or alcohol group.

Each file contains data for a trial for a subject. In each trial a subject is shown a stimulus. The stimulus could be either a single stimulus S1 or two stimuli S1 and S2. When two stimuli are shown, they could be in a matched condition or in a non-matched condition. The stimulus images are pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set [3]. E ach file has data coming from 64 sensors or channels and the data is recorded for 1 seconds @ 256 Hz, i.e., there are 256 samples for each of the 64 sensors. The data value is in micro volts.

II. Research Method

As the objective of the study is to apply machine learning techniques to brain imaging, different learning algorithms will be applied so that a comparison can be made.

A. Machine Learning Methods

Since the data contain the identification of control and alcohol groups, such machine learning will employ supervised learning techniques. Moreover, there are two classes of subjects, this requires a binary classifier. Three different techniques are chosen:

B. Logistics Regression

Logistic regression is used in classification when dependent variable is dichotomous. Logistic regression is represented by the sigmoid equation: [Need Reference]

$$y = 1 / (1 + e^{-z})$$

Where z is defined as:

$$z = \sum x_i w_i$$

C. Neural Networks

Neural networks are an extension to logistic regression where a hidden layer is introduced between the input and output layer to create a more complicated model which can then be compared to logistic regression [Need Reference]. The network is trained by back propagation technique to minimize the cost function over a number of iterations.

D. Bayes Point Machine

Bayes Point Machine (BPM) is a classification algorithm based on Bayesian inference model [7] and has shown to outperform Support Vector Machine (SVM) [7]. The network is trained by assuming that input distribution is Gaussian and finding out the posterior weight factors [7].

E. Tools

The data is migrated to a relational database table so that it can be analysed and summarized correctly. The data is stored in Microsoft SQL Server 2008 Express Edition database.

A windows program is created to read the data from the data files and put them in the database table. Similarly a small program is created to convert the summarized data from the database table to a csv file which can be used by the machine learning programs easily.

Octave is used to carry out logistic regression and neural network classification, whereas Microsoft Infer.NET library is used in a simple .NET program to apply Bayes Point Machine classification.

III. Variable And Measurements

The data represents a single dependent variable which specifies whether the subject is alcoholic or not. All features of the data represent a single measure which is the signal in micro volts. The signal has a number of dimensions which include:

The stimuli identifies which image or pair of images where shown to the subject. The values can be S1, S2 matched or S2 unmatched.

The trial number identifies a single instance of the trial. In the data, recordings were made for a subject for a stimulus multiple times.

The channel number identifies the sensor. The data is recorded off 64 different channels that were placed on different parts of the head.

The sample number identifies the time of the reading. There were 256 readings taken off the same channel, approx. 3.9 ms apart.

A. *Validation*

Internal validation is done by training the models for 70% of the data and testing it on the full data. The split was done randomly [what is norm in similar studies].

B. *Accuracy*

All three methods carry out the prediction as a probability. In the case of logistic regression and neural network, the sigmoid function returns a value between 0 and 1, whereas in the case of Bayes Point Machine the success probability of Bernoulli distribution also provides a value between 0 and 1. Accuracy is used as a performance measure of binary classifiers [8]. Table 1 shows the classification table which is used with a cut off value of 0.5 to determine positive and negative predicted values.

TABLE I
Classification [8]

		Predicted Values	
		True	***False***
Observed Values	***True***	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

Accuracy is then defined as:

Accuracy = (TP + TN) / (TP + FP + FN + TN) [8]

IV. Data Processing

A. *Initial Data Review*

A program was made to put the data from the files into a relational database table. After disregarding the erroneous files, a total of 10,962 trials were imported, with a total of 179,601,408 records.

There are 45 control subjects and 77 alcoholic subjects in the data. Each subject has been 29 and 119 trials for all three stimuli. There are between 7 and 60 trials per subject for stimulus S1, between 11 and 30 trails per subject for stimulus S1 & S2 matched, and 10 and 30 trials per subject for stimulus S1 and S2 unmatched.

B. *Summarizing Data*

The signals are first summarized over trials for the same stimuli. This reduced the number of data rows from 180 million to 5,996,544 rows or 49,152 features per subject.

This is further reduced by summarizing the 256 samples over 1 second interval. However, to retain the variation (increase/decrease) of signal, variance of the signals over these 256 samples was also summarized. This has further reduced the number of rows to 23,424 with two measures (mean and variance.

This normalized data is now converted to a feature matrix saved in a comma separated file by using a small program. The csv file has 122 rows representing each subject and 384 features representing 64 values of signal means and 64 values of signal variations for each of the three stimuli.

C. *Training Sample*

The training sample comprise of 70% subjects which includes 30 subjects from the control group and 55 subjects from the alcoholic group.

V. Results And Analysis

A. *Logistic regression*

The performance of logistic regression varies with the maximum number of iterations specified to minimize the cost function. The accuracy of logistic regression methods when trained with different values of maximum iterations are given in Table 2 below:

TABLE II
Logistic Regression Results

Max. Iterations	**Accuracy**
50	88.52%
75	90.16%
100	93.44%
125	93.44%
150	92.62%
175	92.62%
200	92.62%

Logistic regression models provided excellent accuracy on a small number of iterations. The performance peak was reached at iteration 80 at 93.44% but settled down to 92.62% (one additional false prediction) from iteration 130 onwards. The performance graph of logistic regression is given in Fig- 1

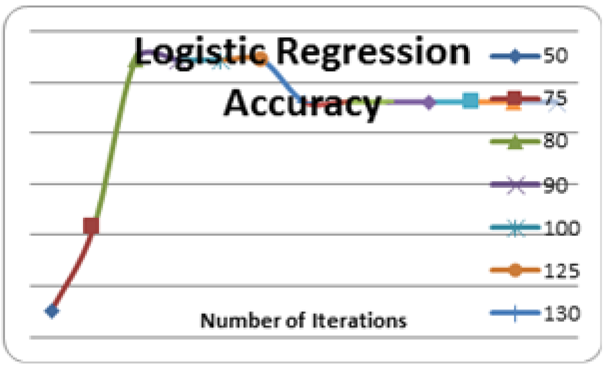

Fig. 1 Logistic Regression Performance

B. Neural Networks

There are a number of parameters that affect the training & performance of neural networks. This includes the number of hidden layers, the number of neurons in the hidden layers, maximum number of iterations specified to minimize the cost function and the learning rate.

The neural network was designed with one hidden layer. For all the other parameters, the neural network method was executed with different values to look at the accuracy which is provided next. Also, because neural network are initially assigned random values for its weights, five readings are taken for each set of configuration values and the average as well as the best ones are provided below in Table 3. Only the promising results are provided here:

TABLE III

NEURAL NETWORK RESULTS

Neurons in Hidden Layer	Max. Iterations	Learning Rate	Average Accuracy	Best Accuracy
15	100	1	86.72%	88.52%
15	200	3	88.69%	91.80%
15	300	0.1	91.47%	95.08%
15	400	1	91.80%	93.44%
25	100	0.3	88.52%	90.98%
25	200	0.3	91.14%	93.445
25	300	0.3	93.11%	95.08%
25	400	0.3	94.59%	95.08%
30	100	0.3	88.11%	90.98%
30	200	0.1	92.46%	95.08%
30	300	0.1	93.44%	94.26%
30	400	0.1	94.11%	95.98%
30	400	0.3	93.11%	96.72%

The best result provided by neural networks is 96.72% with 30 neurons in the hidden layer, trained with 400 iterations at a learning rate of 0.3.

As for the average result over 5 runs, the best performance is 94.59% accuracy for the network with 25 neurons in the hidden layer, trained with 400 iterations at a learning rate of 0.3

The effect of neurons in the hidden layer, number of iterations and learning rate are presented in graphs of Figure 2, Figure 3, and Figure 4 respectively. It can be seen that there is not much difference in performance as number of hidden neurons are increased from 25 to 30. Also, performance improves as number of iterations is increased.

There is little effect of different learning rates on the performance of the neural networks, however, it can be seen that 0.3 has slightly better performance over others.

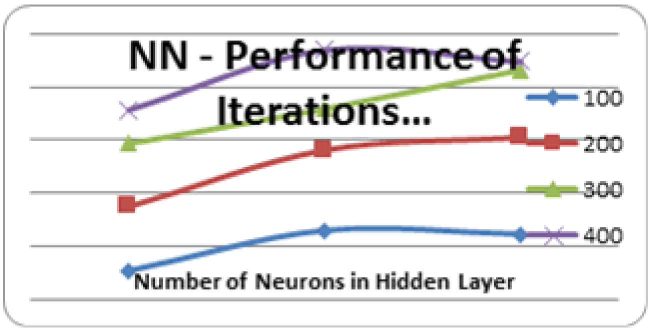

Fig. 1 Neural Network Performance - # of Hidden layers

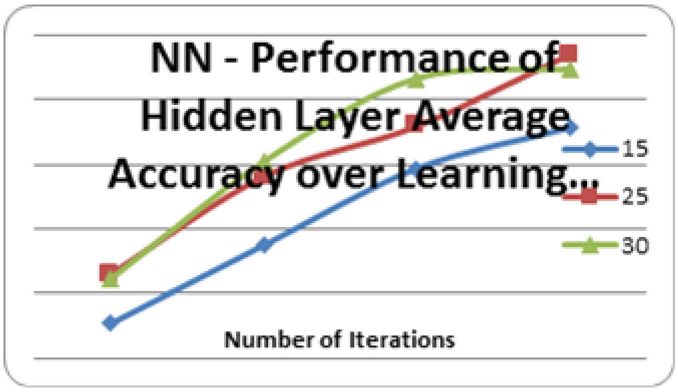

Fig – 3 Neural Network Performance - # of Iterations

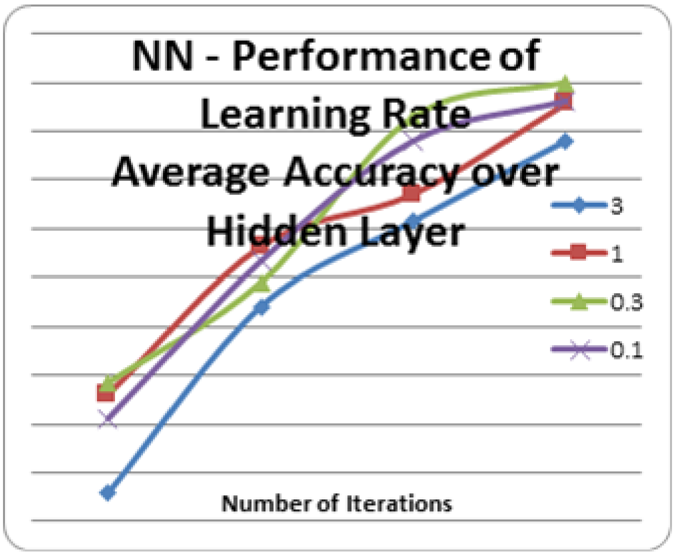

Fig – 4 Neural Network Performance – Learning Rate

C. Bayes Point Machine

Bayes point machine took considerably more time to execute. This may be because of the general performance of Bayesian classification algorithms or because of the way tool was used.

The accuracy of BPM with different iterations is given in Table 4 below:

TABLE IV
BAYES POINT MACHINE RESULTS

# of Iterations	Accuracy
1	79.51%
2	85.25%

# of Iterations	Accuracy
3	92.62%
5	92.62%
10	92.62%

D. Comparison of Methods

It is possible to compare the models because all three algorithms were given the same training and testing data for prediction. Bayes Point Machine was able to stabilize to its best performance after only 3 iterations, whereas neural networks took up to 400 iterations to perform. However, Bayes point machine took more time to compute the weights but this could be because of the library and how it is being used.

Given that the number of features is much larger than the number of training samples, a better comparison can be made when features are reduced and/or other internal validation methods are used for evaluation of performance.

VI. Conclusions

Three different learning algorithms were trained and tested on the same data set. The performance of logistic regression and bayes point machine stabilized at 92.62% whereas neural networks have shown better performance.

It is concluded that all three algorithms are capable of performing over a large set of features

However, this must be mentioned that bayes point machine implementation has not been carried out in detail and there may be additional parameters and settings that may have effects on its performance.

VII. Future Directions

An attempt has been made to apply different machine learning techniques to brain imaging data. This research can be further extended in a number of directions, including:

- Other internal validation techniques can be used like k-fold cross validation and/or bootstrapping.
- The data was summarized over trials and over time using central measures of mean and variance. Other summary measures can be used and comparisons can be made to identify what summarization functions are more effective in brain imaging data. Similarly, models with summarized features can be compared with non-summarized features to see the information loss.
- The models can be reviewed to identify the brain regions that correlate strongly with the classification.
- In addition to summarization, no attempt was made to carry out feature selection. This can be useful extension to the study.
- In addition to accuracy computations, other methods of comparisons like specificity [9], sensitivity [9], Area under Receiver Operating Curve (AUC) [10] can be evaluated.
- Neural networks with more than one hidden layer can be constructed and compared with the existing models [why was it not done for this study?].

References

[1] S. Lemm, B. Blankertz, T. Dickhaus, K. Muller, "Introduction to machine learning for brain imaging",

[2] NeuroImage, No 56 (2011), November 2010, pp 387-399.

[3] Wikipedia, Electroencephalography – Wikipedia, the free encyclopedia, "Electroencephalography", [online], Available: http://en.wikipedia.org/wiki/EEG [Accessed: August 25, 2011]

[4] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of Alzheimer's disease from EEG signals: where are we standing?", Current Alzheimer Research, Vol 7 (6), Sep 2010, pp. 487-505.

[5] M. Schroder, M. Bogdan, W. Rosenstiel, T. Hinterberger, and N. Birbaumer, "Automated EEG Feature Selection for Brain Computer Interfaces", Proceedings of 1st International IEEE EMBS Conference on Neural Engineering, Capri Island, Italy, March 2003, pp. 626-629.

[6] P. Jahankhani, K. Revett, and V. Kodogiannis, "A rule based approach to classification of EEG datasets: a comparison between ANFIS and rough sets", Proceedings of Ninth Symposium of Neural Network Applications in Electrical Engineering NEUREL 2008, Sep 2008, pp. 157-160

[7] UCI, UCI Machine Learning Repository: EEG Database Data Set, EEG Database Data Set, [online], Available: http://archive.ics.uci.edu/ml/datasets/EEG+Database [Accessed: August 25, 2011]

[8] R. Herbrich, T. Graepel, and C. Campbell, "Bayes Point Machines", Journal of Machine Learning Research, Vol 1, 2001, pp. 245-279.

[9] Wikipedia, Accuracy & precision – Wikipedia, the free encyclopedia, Accuracy & precision, [Online], Available: http://en.wikipedia.org/wiki/Accuracy_and_precision, [Accessed: 12 November 2011]

[10] Wikipedia, Sensitivity & specificity – Wikipedia, the free encyclopedia, Sensitivity & specificity, [online], Available: http://en.wikipedia.org/wiki/Sensitivity_and_specificity, [Accessed: 12 November 2011]

[11] C. X. Ling, J. Huang, H. Zhang, "AUC: a Statistically Consistent and more Discriminating Measure than Accuracy", Proceedings of 18th International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003, pp. 519-526.