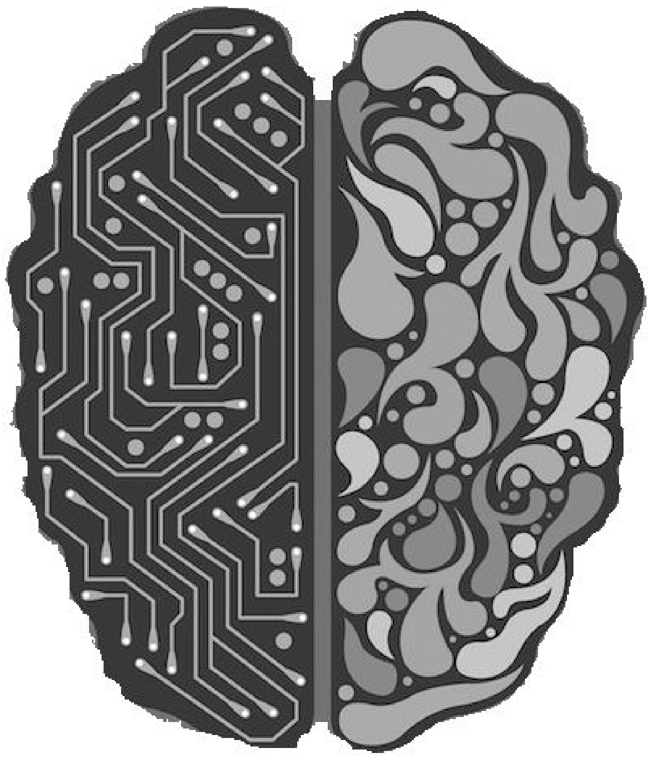
Learning from FMRI Brain Imaging



Aadhavan Sadasivam

Avinash Patil

Manjusha Ravindranath

Satrajit Maitra

Vinit Sheth

Arizona State University

Team Members and Roles

Name	ASU ID	Email	Role	%Contribution
Aadhavan Sadasivam	1215128928	asadasi1@asu.edu	Preprocess the data, Feature Reduction, Implement Logistic regression, Train and Test model. Result analysis.	20%
Avinash Patil	1214167460	apatil20@asu.edu	Literature survey, coding methods for data preprocessing, GNB implementation, training and testing.	20%
Manjusha Ravindranath	1212109143	mravind1@asu.edu	Literature Survey, Data processing & feature reduction, Develop KNN machine learning model using ROI (Region Of Interest), Analyze Results.	20%
Satrajit Maitra	1215097780	smaitra@asu.edu	Data Processing, Feature Reduction, Develop and train Random Forest classifier.	20%
Vinit Girishbhai Sheth	1215128980	vsheth2@asu.edu	Literature Survey, Feature Reduction and finding ROI (Regions of interest), Data manipulation, Implement SVM and Analyze result.	20%

Table 1

All the team members agree on the team members' contributions in terms of both (a) what s/he did and (b) the percentage.

Table of Contents

1. Introduction

Understanding the cognitive state of the brain has been an interesting problem attracting researchers from many domains. With the advent of Functional magnetic Resonance Imaging (fmRI), a 3 dimensional brain imaging method that describes the state of brain by observing the change in blood flow level there was a breakthrough in understanding the state of brain during decision making or any cognitive state. Use of Machine Learning algorithms for predicting the cognitive state became feasible through fMRI data of brain. This is further supported by the evidence that the change of blood level in brain is caused by the neurons. The fact that change in blood flow and neuron activation are correlated gives a strong motivation to use Machine Learning algorithms to detect the cognitive state of the brain. We exercised machine learning methods and used them to train on temporal fMRI patterns which support probabilistic predictions about the cognitive states of the human subject.

1.1. Related Work

Over recent years there has been a growing interest within the computer science community in data processing for fMRI. One of the style of processing involves using Generalized Linear Models (GLM) as in Friston et al. (1995a, 1995b) and Bly (2001). A regression is performed for each voxel, to predict the signal value at that voxel, based on properties of the stimulus. The degree to which voxel activity can be predicted from stimulus features is taken as an indication of the degree to which the voxel's activity is related to the stimulus. This problem is reverse of what we are trying to do. We are trying to determine the stimulus from the voxel's activation.

Hojen-Sorensen, Hansen, and Rasmussen (1999) used Hidden Markov Models (HMM) to learn a model of activity in the visual cortex resulting from a flashing light stimulus. Here the on-off stimulus was recovered as hidden state as HMM. A variety of unsupervised learning methods have also been used for exploratory analysis of fMRI data. Goutte et al. (1998) discussed the use of clustering methods for fMRI data. There have been related work in detecting Cognitive States using machine learning.[8]

The key point of this paper is that it is one of the first attempts to train multiple subjects and contexts, that is, the data from an experiment performed by different subjects can be used as if they were from an experiment performed by a single subject. The classifier then becomes a virtual sensor of cognitive states, thus becoming very useful to brain studies across patients having brain injuries, Alzheimer's disease etc.[8] This [6] work use machine learning models to classify the cognitive state of a human being and to classify whether he or she is reading a sentence or seeing a picture.

2. Problem Statement

The problem we are addressing is to train classifiers to predict whether the subject is perceiving a picture or text. The project will focus on applying machine learning methods to train classifiers on a sparse, noisy and high dimensional data. Appropriate feature selection, feature abstraction and classifier training methods have been applied.

3. Methodology

3.1. Dataset description

The experiment for data collection consists of 54 trials for each subject. Of the 54 trials we consider only 40 trials where the subject is shown a stimuli (Picture / Sentence). Each trial was conducted for a duration of 27 seconds. During the trial, the subject is shown two stimuli (picture / sentence) for 4 seconds each with an interval blank screen for 4 seconds between them. Following the second stimuli there is a blank screen for 15 seconds and at the end of the trial the subject is asked whether the sentence shown describes the picture displayed. During the trial, the state of the brain is recorded for every 500 ms as sequence of 54 images. According to original study the brain retains some activity from last cognition even during the rest therefore, we have incorporated these rest fMRI images along with the stimuli images for classification.

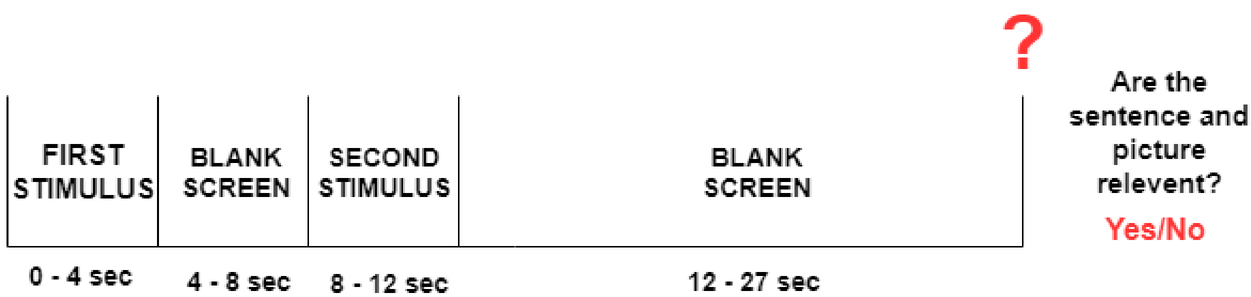


Figure 1: Experimental setup for a trial

For picture Vs sentence classification, we flatten each stimuli into an instance. For each person, we get 40 instances of both sentences and pictures. So, the data for classification has the 80 instances and voxels * 16 attributes.

The FMRI data is for 6 Subjects. 54 trials per subject, and each trial with 54 brain mappings not all the data is useful. To remove noise from data we select only the trails with info.cond > 1.
This gives the trials where the subject was viewing a sentence or a picture. Which gives us 40 good trials. Inside each trial there are 54 brain mappings. We take only first 16 images which represent the data for first 8 seconds (One Fmri image taken every 500ms) from total 27 seconds. In the first 4 seconds (Image number 1 to 8) first stimulus was presented (Sentence or Picture). Then images 17 to 32 are chosen for the next stimulus.
Thecond = 2 represents the trials where the presented sentence was not negated. The value of First stimulus which is either ‘P’ or ‘S’ will tell us which stimulus was presented first.

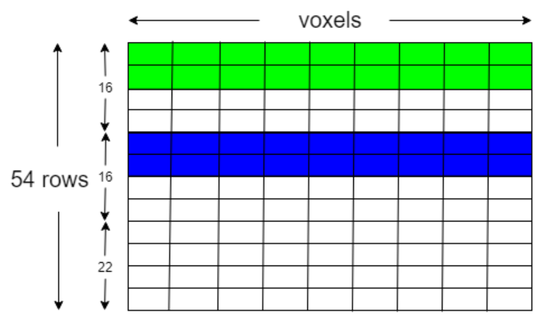


Figure 2: Data for a trial

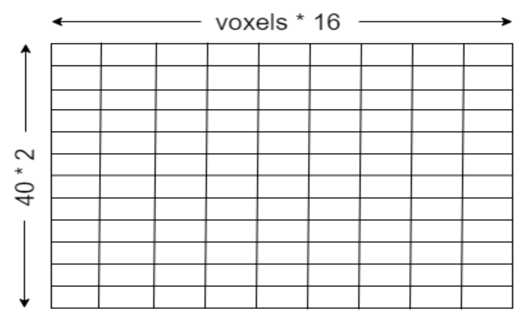


Figure 3: Data Matrix for a subject

3.2. Feature Selection

The data is highly multi dimensional. Data matrix for classification has around 80000 attributes. Plus each part of the brain is associated with unique activities and we are interested in finding the part of the brain that is responsible for the decision making on whether the subject is looking at a sentence or picture. Considering the entire brain data for classification will affect the prediction because of the noise I.e., parts of the brain that are not responsible for decision making. To solve this problem of high dimensionality and noise, we look each into the Region of Interest (ROI) - specific parts of the brain and develop a classifier for each Region of interest. We then combine the voxels from the regions of interest that have high accuracy. The experimental results of ROI based feature selection is briefly described in section 4.2.

We also explored PCA as an alternate method for dimensionality reduction technique along with normalizing the data.

3.3. Machine Learning Algorithms

3.3.1. Gaussian Naïve Bayes

In this experiment we are dealing with continuous data. Typically we assume that continuous values associated with each class are distributed according to a Gaussian distribution. Therefore, for a training data ‘x’ we can first segment the data by class and then compute the mean and variance of ‘x’ in each class. Let μ_k be the mean of values in ‘x’ associated with class C_k ane let σ^2_k be the variance of values in x associated with class C_k. Then the probability distribution of ‘v’, where ‘v’ is some observation, can be computed by:

$$p(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}\, e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

3.3.2. K Nearest Neighbors

KNN or K Nearest Neighbors is a machine modeling technique to find k nearest neighbors of a test sample. The KNN algorithm find the nearest neighbors of the test example amongst all training samples using distance metrics.

We have used Euclidean distance for KNN experiments and is calculated as follows,

$$\begin{aligned} d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}. \end{aligned}$$

3.3.3. Support Vector Machine

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

$$\left[\frac{1}{n} \sum_{i=1}^{n} \max\left(0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b)\right)\right]$$

3.3.4. Logistic Regression

The logistic regression is a discriminative predictive analysis. Logistic regression is used to explain the relationship between one binary variable that is dependent and one or more independent variables [10].

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

3.3.5. Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set. The accuracy of a Random Forest classifier increases with the number of trees. We set the number of trees to be 500 in our classifier.

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(x') \qquad \sigma = \sqrt{\frac{\sum_{b=1}^{B} (f_b(x') - \hat{f})^2}{B - 1}}.$$

4. Results

4.1. Classification Accuracy compared between models

4.1.1. Classification accuracy without feature selection

Algorithm	**Subject1**	**Subject2`**	**Subject3**	**subject4**	**subject5**	**subject6**	**Average**
GNB	60%	69%	90%	65%	78%	82%	74%
5-NN	65%	68.5%	78.75%	77.5%	57.5%	77.5%	71%
LogReg	67%	73%	91%	71%	80%	83%	77%
SVM	67%	74%	91.4%	74%	81.2%	84%	78%
RF	73%	69%	99%	86%	91%	92%	85%

Table 2

4.1.2. Classification accuracy with feature selection

Algorithm	**Subject1**	**Subject2`**	**Subject3**	**subject4**	**subject5**	**subject6**	**Average**
GNB	100%	100%	100%	100%	100%	100%	100%
5-NN	75%	71.25%	92.5%	88.75%	77.5%	87.5%	82%
LogReg	86.25%	75%	96.25%	85%	86.25%	87.5%	86%
SVM	92.5%	75%	96.25%	88.75%	88.75%	92.5%	89%
Random Forest	93.75%	82.5%	98.75%	88%	92%	94.6%	91.6%

Table 3

We trained all the classifiers for all the subjects 25 times with each ROI as input. The subject 3 performed best with the data so we have plotted how each classifier perform with all the ROIs. We have used 5 fold Stratified Cross Validation to measure the accuracy for all the algorithms to avoid overfitting. Logistic Regression is implemented with L2 regularization for logistic Regression and SVM.

4.2. Experimental Analysis

From figure 4.1. and 4.2., we can see that the algorithms perform better with feature selection. This can be explained by the fact that certain regions of the brain are responsible for decision making on whether the subject is looking at a sentence or picture. ROI based feature selection has drastically reduced the noise since we are only considering specific regions of brain associated with the decision making process . From the figures 5.1, 5.2, 5.3, 5.4, 5.5, It can be seen that the Region of Interest 'CALC' has high classification accuracy for all the algorithms. An interesting observation from the above experimental analysis is that GNB does not give out better ROI level accuracy but when top 3 ROI accuracies are combined together, It gives out better accuracy when compared to other algorithms. This result supports the evidence that decision making in brain is handled by parts of the brain combined together.

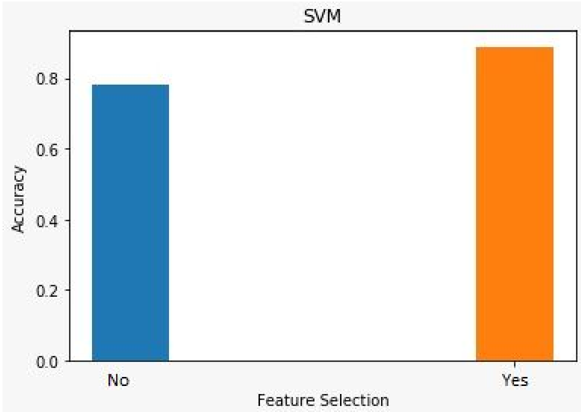


Figure 4.1

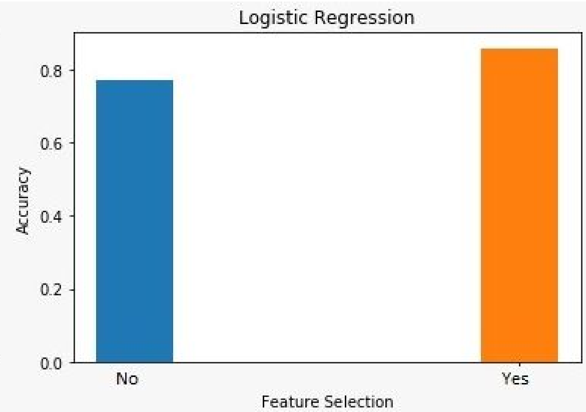


Figure 4.2

From the Table 2 and Table 3, it is seen that the data is linearly separable as Logistic Regression, Linear Support Vector Machine and Gaussian Naive bayes have good prediction accuracy. To deal with this problem, a Receiver Operating Characteristic (ROC) curve is plotted for KNN as shown in Figure 5.6 to further give a confidence score for KNN predictions.

Our comparison of classifiers indicates that Gaussian Naive Bayes (GNB) and linear Support Vector Machine(SVM) classifiers outperform K Nearest Neighbor across all studies. The low prediction accuracy of K Nearest Neighbor can be explained by the fact that KNN gives equal importance to all the features and it’s decision boundary can be easily affected by outliers. In comparing GNB to SVM, GNB is performing well. This can be reasoned by the fact that bias of GNB is more so for small dataset, it gives better accuracy. We have total 80 instances per subject so GNB outperformed all the classifier across all the folds.

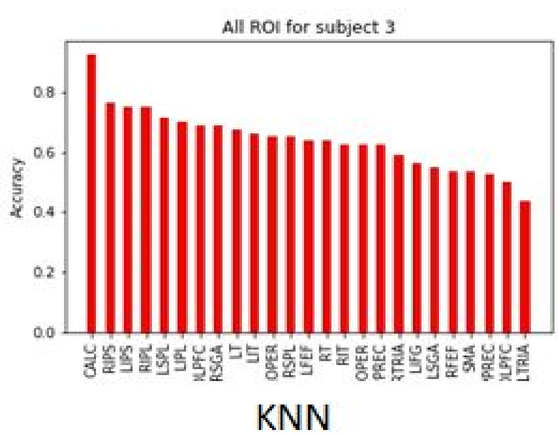


Figure 5.1

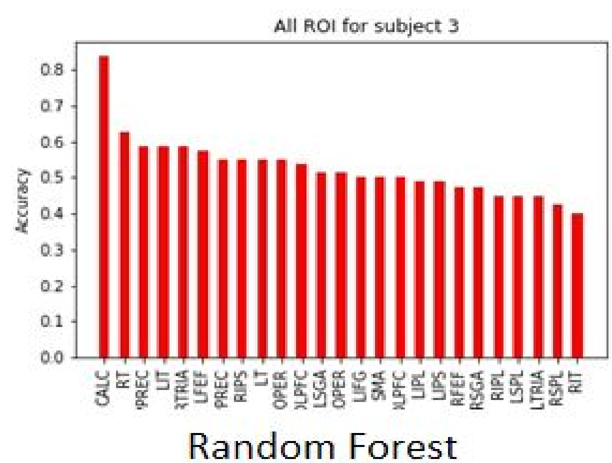
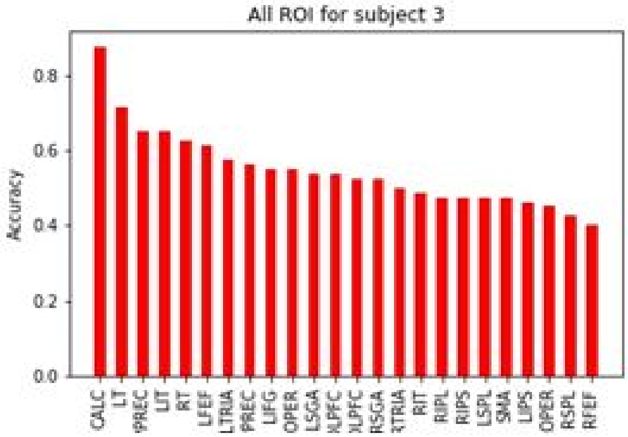
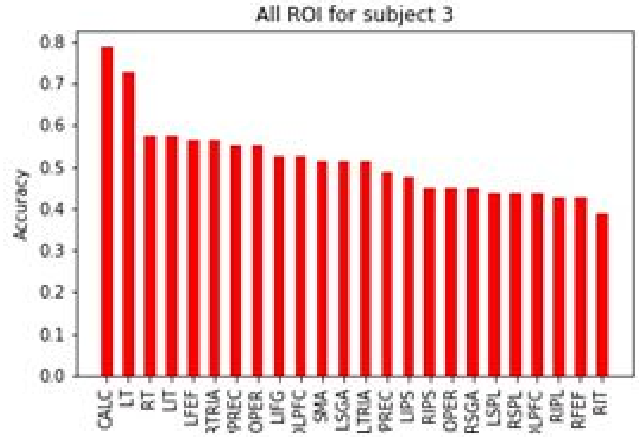


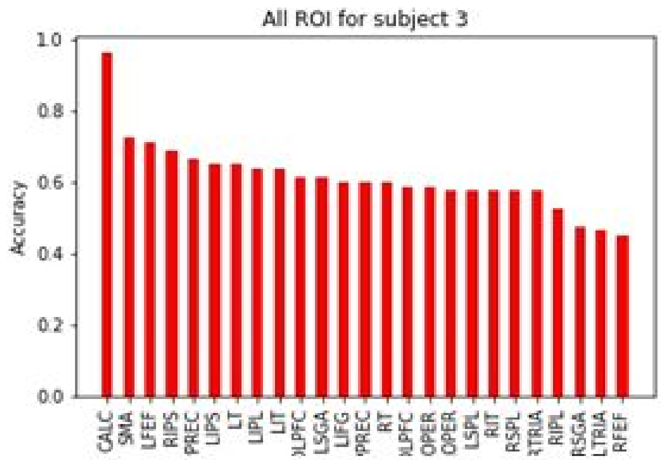
Figure 5.2



Logistic Regression



GNB



SVM

Figure 5.3 (LR), Figure 5.4 (GNB), Figure 5.5 (SVM)

Receiver Operating curve plots the true positive rate (Sensitivity along Y-axis) against the false positive rate (Specificity along X-axis) for different threshold values. ROC curve was used in conjunction with KNN experimentation to confirm prediction by giving a confidence score to prediction.

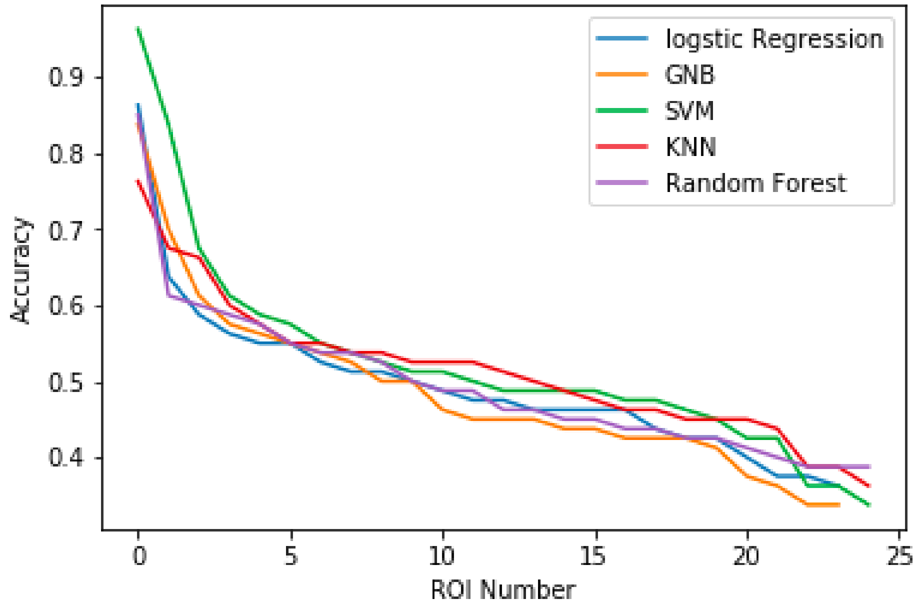


Figure 6: Comparison across all models

5. Conclusion and Future Work

We have trained several classifiers to predict whether the subject is perceiving a picture or text. And as expected the classification algorithms perform better when we look at certain parts of the brain. This supports the evidence that a certain part of the brain is responsible for decision making in humans. Though the prediction on single subject gives out good accuracy, prediction on multiple subjects does not give promising results. The small size of the dataset possesses another limitation on training and across multiple subjects. This work understands the cognitive state of the brain under predefined scenarios. Learning the cognitive state of the thorough undefined scenarios will lead to a better understanding of the human decision making process and it might bring in potential breakthrough in many current world problems like treating alzheimer's disease.

6. References

1. Mitchell et al, 2004: http://www.cs.cmu.edu/~tom/mlj04-final-published.pdf.
2. Zhou, Jiayu, et al. "Modeling disease progression via fused sparse group lasso." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.
3. Xang, Xuerui, Tom M. Mitchell, and Rebecca Hutchinson. "Using machine learning to detect cognitive states across multiple subjects." CALD KDD project paper (2003).
4. Mitchell, Tom M., et al. "Classifying instantaneous cognitive states from fMRI data." AMIA annual symposium proceedings. Vol. 2003. American Medical Informatics Association, 2003. Classifying Instantaneous Cognitive States from fMRI Data
5. Wang, Xuerui, Rebecca Hutchinson, and Tom M. Mitchell. "Training fMRI classifiers to detect cognitive states across multiple human subjects." Advances in neural information processing systems. 2004.
6. T.M. Mitchell, R. Hutchinson, R.S. Niculescu, F.Pereira, X. Wang, M. Just, and S. Newman, Machine Learning, Vol. 57, Issue 1-2, pp. 145-175. October 2004, "Learning to Decode Cognitive States from Brain Images,"
7. R. Hutchinson, T.M. Mitchell, I. Rustandi, submitted to HBM 2005, "Learning to Identify Overlapping and Hidden Cognitive States from fMRI Data"
8. https://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-73/www/papers/XueruiReport-10-2002.pdf
9. https://www.medcalc.org/manual/roc-curves.php