**Machine-learning Techniques for Building a Diagnostic Model for Mild Cognitive Impairments**

# **Method**

**Participants**

120 people participated in this study. Sixteen participants were excluded from the study based on exclusion criteria. 49 healthy adults with no objective cognitive problems were in the control group, 36 adults with mild cognitive problems were in the MCI group, and 19 patients with Alzheimer's were in the Alzimer disease group. The total number of participants was divided into two groups: group 1 (MCI + control = 68) and group 2 (AD + control = 85). Participants were aged 55 to 95 years.

Table 1. Subject information.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MCI-AD  n = 85 | MCI-HC  n=68 | AD  (n =19; xF/yM ) | | | MCI  (n =36; xF/yM ) | | | HC  (n=49; xF/yM) | | | characteristics |
|  |  | Range | SD | Mean | Range | SD | Mean | Range | SD | Mean |  |
|  |  |  |  |  |  |  |  |  |  |  | Age (mean sd) |
|  |  |  |  |  |  |  |  |  |  |  | education(mean sd) |
|  |  |  |  |  |  |  |  |  |  |  | CDR |

The numbers refer to baseline data. AD=Alzheimer's disease, MCI=Mild Cognitive Impairment, HC=Healthy Control, CDR= Clinical Dementia Rating.

**Eye-tracking experiments**

We used Simon, Pen and pencil, n-back and eye-tracking methods. The eye-tracking experiments were carried out in a quiet lab environment. 1000 Desktop Mount with monocular eye-tracking. The eye tracking system is completely safe and does not need to restrain or control the subject's head, and this tool only records eye movements. Each person is placed at a distance of 20 cm from the screen. The system follows both eyes of the subject and its accuracy is 0.5 degrees of eye angle. The sampling rate was set to 1000 Hz. Spatial resolution is 0.2 degrees and error rate is 0.3 degrees. Eye movements, such as saccades and fixations, are then calculated with respect to the predefined Areas of interest (AOIs). The eye-tracker was calibrated for each participant using a 9-point calibration procedure, and drift-corrected between Trial 1 and Trial 2

**Data preparing**

Three machine learning algorithms were implemented in Python using Orange (Curk et al., 2005): Logistic Regression (LR), Gaussian Naive Bayes (GNB), and Support Vector Machine (SVM) with radial basis function kernel type. Our approach was to adapt the techniques from machine learning-based classification methodology from Computer Science. The specifics of the training and prediction stages of the classification algorithms are described below. We explored a number of popular and state-of-the-art classification algorithms from computer science, ultimately focusing on three representative methods that resulted in the best performances on our preliminary experiments Given the small size of our data set, we forego parameter optimization and use the default parameters; i.e., for LR we use a ridge regression parameter of 10−8, and for SVM we use a first degree polynomial kernel and a complexity parameter of 1.0.

**Logistic Regression (LR)**

The logistic regression model is used for prediction by fitting the training data to the parameterized logistic regression function. At test (prediction) stage, the optimal parameters are used to compute the most likely class for each example based on the feature (parameter) values.

**Gaussian Naive Bayes (GNB)**

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes’ theorem with the “naive” assumption of conditional independence between every pair of features given the value of the class variable. Bayes’ theorem states the following relationship, given class variable y and dependent feature vector X1 through Xn.

**Support Vector Machine (SVM)**

Before introducing the kernel combination method, we first briefly review the standard single-kernel SVM algorithm. The main idea of SVM is summarized as follows. First, the linearly no separable samples are mapped from their original space to a higher or even infinite dimensional feature space, where they are more likely to be linearly separable than in the original lower-dimensional space, through a kernel-induced implicit mapping function. Then, a maximum margin hyper plane is sought in the higher dimensional space

**Validation**

We evaluate the classifier using leave-one-out cross validation, in which at every iteration one data point is held out as a test point, and all remaining points are used for feature selection and classifier training.

Accuracy: The fraction of correctly classified subjects out of all the subjects in the test set.

Sensitivity: The ratio of correctly classified impaired subject’s to the total number of impaired subjects in the test set.

Specificity: The ratio of correctly classified normal control subjects to the total number of the control subjects in the test set.

Area under the ROC curve (AUC): The area under the Receiver Operating Characteristic (ROC) curve, which is a common