Machine learning Techniques for Diagnostic Differentiation of Mild Cognitive Impairment and Dementia

Alyssa Weakley¹, Jennifer Williams²

¹Washington State University, Department of Psychology
²Washington State University, School of Electrical Engineering and Computer Science

Introduction

• Accurate and cost efficient classification of mild cognitive impairment (MCI) and Alzheimer’s disease (AD) has benefits of personal and medical importance.
• Based on current literature, it appears that machine learning methods are viable dichotomous classifiers (e.g., Chen et al., 2010; Quintana et al., 2012). However, differentiating between MCI, AD, and neurological healthy older adults using a single model has been problematic.

The purpose of the present research is to use neuropsychological and demographic data to predict:
1. Clinical Dementia Rating (CDR) ratings
2. Clinical diagnoses (cognitively healthy, MCI, AD)

Through the implementation of four machine learning models, and a comparison statistical method:
1. C4.5 Decision tree
2. Naïve Bayes classifier (NB)
3. Artificial Neural Network (ANN)
4. Support Vector Machine (SVM)
5. Multinomial Logistic Regression

A secondary goal of the project is to determine the fewest number of attributes (e.g., neuropsychological tests) required to reliably diagnose individuals.

Method

Participants
• For participant characteristics see Tables 1 and 2. Clinical diagnoses were consistent with Petersen and colleagues criteria, NIA-Alzheimer’s Association workgroup, and NINCDS-ADRDA criteria. CDR ratings were provided by a certified rater.

Table 1: Demographic and feature selected neuropsychological data of participants classified using clinical diagnosis

| Variable or test | Control (n = 161) | MCI (n = 97) | AD (n = 53) | P
|------------------|------------------|--------------|-------------|---|
| Age (years)      | 71.14 8.47       | 71.96 9.43   | 75.70 8.26  | <.005
| Years of Education (years) | 16.28 2.83      | 15.52 2.93   | 16.11 2.83  | <.004
| % Female         | 74%              | 57%          | 45%        | <.001
| TICS             | 34.45 2.27       | 32.48 2.99   | 24.36 5.84  | <.001
| Trails B         | 79.45 27.37      | 123.8 58.4   | 235.25 88.03 | <.001
| Clox I           | 12.92 2.08       | 11.84 2.36   | 8.98 3.47   | <.001
| Design Fluency – Open dots | 10.06 2.87    | 8.25 2.76    | 4.41 2.47   | <.001
| Design Fluency – Switching | 7.63 2.56     | 5.88 2.65    | 2.65 2.18   | <.001
| Functional Status z-score | -3.16 10.05    | 8.10 10.06   | 8.12 1.82   | <.001
| Depression z-score | -2.4 1.83       | 1.84 1.15    | 1.33       | <.001
| Delayed Verbal Memory z-score | 0.40 0.19     | 1.02 1.24    | 1.29        | <.001
| Visual Memory total correct | 60 67       | 38 80       | 103 72     | <.001

Neuropsychological and Demographical Measures
• Demographic variables: age, education, and gender
• Mental status: Telephone Interview of Cognitive Status (TICS)
• Depression: Geriatric Depression Scale (GDS)
• Functional Status
• Neuropsychological Performance:
  - Attention/speeded processing (Symbol Digit Modalities Test, Trails A)
  - Verbal learning and memory (RAVLT, Memory Assessment Scales)
  - Visual learning and memory (7/24, Brief Visual Memory Test)
  - Executive functioning (Trails B, Clox I, Design Fluency of D-KEFS)
  - Working memory (WAIS-III Letter-Number Sequencing)
  - Visual fluency (verbally fluency subtest from the D-KEFS)
  - confrontation naming (BNT)

Table 2: Demographic and feature selected neuropsychological data of participants classified using clinical diagnosis

| Variable or test | Control (n = 154) | MCI (n = 93) | AD (n = 25) | p
|------------------|------------------|--------------|-------------|---|
| Age (years)      | 71.14 8.47       | 71.96 9.43   | 75.70 8.26  | <.005
| Years of Education (years) | 16.28 2.83      | 15.52 2.93   | 16.11 2.83  | <.004
| % Female         | 74%              | 57%          | 45%        | <.001
| TICS             | 34.45 2.27       | 32.48 2.99   | 24.36 5.84  | <.001
| Trails B         | 79.45 27.37      | 123.8 58.4   | 235.25 88.03 | <.001
| Clox I           | 12.92 2.08       | 11.84 2.36   | 8.98 3.47   | <.001
| Design Fluency – Open dots | 10.06 2.87    | 8.25 2.76    | 4.41 2.47   | <.001
| Design Fluency – Switching | 7.63 2.56     | 5.88 2.65    | 2.65 2.18   | <.001
| Functional Status z-score | -3.16 10.05    | 8.10 10.06   | 8.12 1.82   | <.001
| Depression z-score | -2.4 1.83       | 1.84 1.15    | 1.33       | <.001
| Delayed Verbal Memory z-score | 0.40 0.19     | 1.02 1.24    | 1.29        | <.001
| Visual Memory total correct | 60 67       | 38 80       | 103 72     | <.001

Feature Selection
• CDR-FS
  - Age
  - Education
  - Gender
  - Functional Status
  - Visual memory total correct
  - Trails A
  - Letter Fluency
  - Category Fluency
  - Switch Fluency
  - Design Fluency – Solid Dots
  - Short delay verbal memory
  - Long delay verbal memory

Results
• Clinical Diagnosis-FS
  - Age
  - Education
  - Gender
  - Functional Status
  - Visual memory total correct
  - Trails A
  - Letter Fluency
  - Category Fluency
  - Design Fluency – Open Dots
  - Design Fluency – Switching
  - Depression
  - Long delay verbal memory

Table 3: Results from supervised model for original, reduced, and feature selection datasets. This table shows the results of modeling for the clinical diagnosis class with and without CDR total.

Table 4: Results from semi-supervised model for original, reduced, and feature selection datasets, CDR class

Table 5: Results from semi-supervised model for original, reduced, and feature selection datasets.

Conclusions
• We hypothesized that NB and SVM would provide the greatest accuracy. Based on the results, NB achieved the highest accuracy in all cases. However, our prediction that SVM would also provide a high accuracy rate, was not supported.
• Improved accuracy over a traditional statistical method was observed for CDR classification but not for clinical diagnosis.
• We explored the use of feature selection to reduce the number of demographic and neuropsychological data needed to make an accurate classification. We were able to determine which tests were critical when making classification for CDR scores and clinical diagnoses.
• The experiments reported in this paper indicate that artificial intelligence techniques can be used to automate aspects of clinical diagnosis and can provide meaningful insights into which attributes are the most valuable for this diagnosis.