Diagnostic Utility of EEG Based Biomarkers for Alzheimer's Disease

Charlotte Cecere
Electrical and Computer Engineering
Rowan University
Glassboro, NJ USA
cecere21@students.rowan.edu

Christen Corrado
Electrical and Computer Engineering
Rowan University
Glassboro, NJ USA
corrad68@students.rowan.edu

Robi Polikar
Electrical and Computer Engineering
Rowan University
Glassboro, NJ USA
polikar@rowan.edu

***Abstract*— Alzheimer's disease (AD) is a neurodegenerative disease whose definitive diagnosis is only possible via autopsy. Currently used diagnostic approaches include the traditional neuropsychological tests, and recently more objective biomarkers, such as those obtained from cerebral spinal fluid (CSF), magnetic imaging resonance (MRI), and positron emission tomography (PET). Electroencephalography (EEG), a lower cost and non-invasive alternative, has been previously tried but with mixed success. In this effort, we attempt a more comprehensive analysis and comparison of machine learning approaches using EEG based features to determine diagnostic utility of the EEG. We compared support vector machine (SVM), naïve Bayes, multilayer perceptron (MLP), CART trees, k-nearest neighbor (kNN), and AdaBoost on various sets of features extracted from event related potentials (ERP) of the EEG. Our analysis suggests that there is indeed diagnostically useful information in the ERP of the EEG for early diagnosis of AD.**

***Keywords*—Alzheimer's disease, EEG, event related potentials, automated diagnosis.**

I. INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia, commonly associated with aging. AD is a neuro-degenerative disease that results in – or caused by – neuronal death [1]. Initial symptoms are similar to those associated with normal aging, but then the symptoms deteriorate – typically rapidly – to the point that "those in the final stages of the disease lose their ability to communicate, fail to recognize loved ones, and become bed-bound," [1] with death being the inevitable outcome. While there is no cure for AD, pathologically targeted treatments can slow the progression of the disease, ease the symptoms, and improve the subject's quality of life. As the AD pathology is now suspected to start long before the symptoms begin, the diagnosis of the disease at earlier stages is of paramount importance.

Pathologically, two proteins are implicated with AD: the tau proteins (also known as tangles), and the β-amyloid deposits (also known as plaques) that build up between nerve cells. The specific role of these proteins is not yet known, though they are believed to disrupt the communication among nerve cells, damage the cells, and eventually lead to neuronal death.

Currently, AD diagnosis relies on a battery of cognitive and memory tests, physical and neuropsychological examinations and interviews with the patient and their caregivers over a period of time [1]. The diagnostic accuracy for AD through such clinical evaluations can reach 90% at dementia specialty clinics, however, that accuracy at community hospitals and clinics is estimated as only 75% [2]. The subjectivity as well as the inherent variability associated with clinical evaluations have led researchers to evaluate alternate and more objective biomarkers. These include concentrations of tau and β-amyloid proteins in CSF (obtained through an invasive and unpleasant lumbar puncture that requires an expert physician), or attributes obtained from neuroimaging based biomarkers, such as anatomic volumetric analysis from MRI or metabolic activity based on PET scan – both of which are costly procedures. Neuronal death leads to brain atrophy as well as lack of metabolic activity; the former can be detected with MRI, whereas the latter can be detected with PET imaging [3]. Another modality that has not received as much attention is the EEG, or the event related potentials (ERPs) obtained from the EEG. Unlike MRI and PET, which are anatomical and metabolic biomarkers, EEG is an electrophysiological marker that can provide clues to the neuronal health, or more specifically, to the health of the communications among them.

The ERPs are typically elicited though an oddball paradigm procedure, which requires subjects to identify rare type(s) of "oddball" stimuli embedded in frequent "standard" stimuli. Reduced or delayed P_{300}, a positive peak found approximately 300 milliseconds after stimulus, has been associated with AD [4], however the association is not strong enough to allow individual diagnosis. Several signal processing techniques have also been applied to the ERP/EEG including our prior work on using discrete wavelet transform (DWT) based features to train a multilayer perceptron type neural networks, which also showed mixed success [5].

In this effort, we present our results of a more comprehensive analysis, where optimal subsets of DWT based features are first obtained followed by training a broad spectrum of classifiers to determine the true utility of these features.

II. EXPERIMENTAL SETUP

The 71-subject cohort used in this study was recruited at the University of Pennsylvania, where they also received neuropsy-chological evaluations following the NINCDS-ADRDA criteria for [6] clinical diagnosis. Raw EEG data were collected using an auditory oddball paradigm. The patient, equipped with a pair of headphones and a button, heard a series of tones about one every 1.5 s. There were three types of tones: standard tones at 1 kHz, target tones at 2 kHz, and novel tones – sound clips from Disney movies. The subjects were instructed to push the button every

978-1-4799-3728-8/14/$31.00 ©2014 IEEE

time they heard the target tone, which constituted approximately 20% of all tones (with standard tones and novel sounds constituting 65% and 15%, respectively). The EEG recordings were recorded from 19 electrode locations, placed with respect to the 10-20 electrode placement system, and were sampled at 256 samples/s. The responses, obtained through a 30 minute recording session with frequent brakes, were notched filtered at 60 Hz, segmented into 1-second intervals with respect to stimulus type, time-locked, averaged and amplitude normalized, resulting in the ERPs used in this study. More detailed information can be seen in [4] and [5]. DWT coefficients were obtained from these ERPs using Daubechies 4 wavelets, which were then separated into four frequency bands of 0-1 Hz, 1-2 Hz, 2-4 Hz, and 4-8 Hz, [5], [7]. Data from 11 of the 19 electrodes primarily in the parietal, frontal, and central regions, were used in the final analysis, namely: P_3, P_4, P_Z, P_7, P_8, C_3, C_4, C_Z, F_7, F_8, and F_Z.

III. Methods

A nested dual cross-validation scheme, each 5-fold, was used to determine feature sets as well as classifier parameters, with a separate test dataset used for model evaluation. Each potential feature set contained the DWT coefficients of a different combination of 11 electrodes, two stimulus tones (oddball and novel) and the aforementioned 4 frequency bands, resulting in 88 possible feature sets. Six classification algorithms were used: the SVM, naïve Bayes, the MLP, kNN, CART, and AdaBoost using CART as the base weak classifier.

IV. Results

Our preliminary results are summarized in Table I, where for each feature set tested, Acc is the averaged classification accuracy obtained based on 5-fold cross validation, Tone is the stimulus type (T: Target, N: Novel), Elec. is the electrode, Fr. is the frequency band (in Hz) in which the best performance was obtained, Er is the average error, SE is the sensitivity, PPV is positive predictive value, and SP is the specificity for the best performing feature sets.

TABLE I. SUMMARY CLASSIFICATION RESULTS

	Acc %	Tone	Elec	Fr	PPV%	SE %	SP%
KNN	85.7	N	P_7, P_8	2-4	66.7	100	80.0
NB	85.7	N/T	C_3, P_7, F_7, C_Z	0-4	100	66.7	100
SVM	78.6	N/T	P_4, P_Z	2-4	75	85.8	71.4
MLP	82.1	N/T	P_3, P_Z	0-2	80.8	80.8	83.3
CART	85.7	N	C_4, P_4	0-4	88.9	88.9	80.0
ADA	85.7	N/T	F_7	0-3	77.8	100	71.4

The classification algorithms that had the highest accuracy were kNN, CART, AdaBoost, and naïve Bayes at 85.7% classification accuracy (12 out of 14 correct in each test subset of 14 subjects). Based on a parameter sweep using an internal cross validation (with a separate test subset set aside), we observed that kNN performed best with 7-9 neighbors and a distance metric of either city block or Minkowski (surprisingly, these two metrics perform better than the Euclidean distance). The feature sets that had the highest accuracy typically used novel tones and electrodes in the frontal and parietal regions. This is a satisfying outcome, as these are the areas that are known to be first affected by AD.

The naïve Bayes classification algorithm classified best using electrodes C_3, P_7, F_7, and C_Z. Based on a parameter sweep we observed that naïve Bayes performed best with a kernel distribution with prior obtained empirically from the training set.

The SVM had a highest accuracy of 78.5%, classifying best using mostly parietal region electrodes and some central region electrodes. SVM has its best performance when trained with frequencies in the 2-4 Hz range. A parameter sweep for SVM was also used to determine the optimal kernel parameter sigma of 10 (for an RBF kernel), and the box parameter of infinity.

The MLP neural network had an average accuracy of 82.1%. It also classified best with parietal electrodes P_3 and P_Z and lower frequencies (at or below 4 Hz) using 15 hidden layers.

The CART classification algorithm performed best using 5 maximum divisions per node, while merging leaves and pruning, with the prune criterion based on 'error'. The algorithm did better with larger numbers of minimum leaves, namely 8 and a prior probability obtained empirically from the training dataset. Novel tones and electrodes C_4 and P_4 provided the best accuracy.

The AdaBoost algorithm had better performance using just 5 classifiers in the ensemble. The algorithm had the highest accuracy at 85.7% and performed best with the F_7 electrode at low frequencies.

V. Discussion and Conclusions

These preliminary results indicate that there is indeed diagnostically useful information in the event related potentials, typically with ERPs obtained in response to novel tones and in the 0 – 4 Hz range. These results are in the same general range of earlier efforts, both by us and others, and reside in the low to mid 80% range. Of course, these results were obtained when the feature sets were evaluated individually, and we believe that when used in a strategic combination, a data fusion based approach that combines the feature sets that include complimentary information may significantly improve the diagnostic accuracy. Of course, the diagnostic accuracy may be further improved if features from other modalities (such as MRI, PET, etc.) are added. Our future work will include additional classification models, and, more importantly, data fusion algorithms to determine optimal combination of features from any given modality as well as from different modalities.

References

[1] A. Association, "Alzheimer ' s disease facts and figures," *Alzheimer's Dement.*, vol. 9, no. 2, 2013.

[2] A. Lim, D. Tsuang, W. Kukull, D. Nochlin, J. Leverenz, W. McCormic, J. Bowen, L. Teri, J. Thompson, E. R. Peskind, M. Raskind, and E. B. Larson, "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series," *J. Am. Geriatr. Soc.*, vol. 47, no. 5, pp. 564–569.

[3] K. Blennow, M. J. de Leon, and H. Zetterberg, "Alzheimer's disease.," *Lancet*, vol. 368, no. 9533, pp. 387–403, Jul. 2006.

[4] J. Polich, C. L. Ehlers, S. Otis, a J. Mandell, and F. E. Bloom, "P300 latency reflects the degree of cognitive decline in dementing illness.," *Electroencephalogr. Clin. Neurophysiol.*, vol. 63, no. 2, pp. 138–44, Feb. 1986.

[5] R. Polikar, A. Topalis, D. Green, J. Kounios, and C. M. Clark, "Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease.," *Comput. Biol. Med.*, vol. 37, no. 4, pp. 542–58, Apr. 2007.

[6] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, vol. 34, no. 7, pp. 939–939, Jul. 1984.

[7] B. Balut, M. Karnick, D. Green, J. Kounios, C. Clark, and R. Polikar, "Ensemble Based Data Fusion from Parietal Region Event Related Potentials for Early Diagnosis of Alzheimer's Disease," *IJCNN*, 2007.