Ensemble Based Data Fusion from Parietal Region Event Related Potentials for Early Diagnosis of Alzheimer's Disease

Brian A. Balut, *Student Member IEEE*, Matthew T. Karnick, *Student Member IEEE*, Deborah Green, John Kounios, Christopher M. Clark, and Robi Polikar, *Member IEEE*

***Abstract*—As a natural consequence of steady increase of average population age in developed countries, Alzheimer's disease is becoming an increasingly important public health concern. The financial and emotional toll of the disease is exacerbated with lack of standard diagnostic procedures available at the community clinics and hospitals, where most patients are evaluated. In our recent preliminary results, we have reported that the event related potentials (ERPs) of the electroencephalogram can be used to train an ensemble-based classifier for automated diagnosis of Alzheimer's disease. In this study, we present an updated alternative approach by combining complementary information provided by ERPs obtained from several parietal region electrodes. The results indicate that ERPs obtained from parietal region of the cortex carry substantial complementary diagnostic information. Specifically, the diagnostic ability of such an approach is substantially better, compared to the performance obtained by using data from any of the individual electrodes alone. Furthermore, the diagnostic performance of the proposed approach compares very favorably to that obtained at community clinics and hospitals.**

I. Introduction

Alzheimer's Disease (AD), a neurodegenerative disorder associated with progressive cognitive decline, is the most prevalent form of dementia. The disease causes rapid deterioration of its victims' ability to remember, think, make decisions, and eventually use their motor skills. For decades since its first discovery in 1906, AD was mostly ignored, as it affected primarily the elderly, and most people did not live long enough to experience the symptoms of the disease. However, as the average population age increases – primarily in the developed countries – so does the number of people affected by the disease. Furthermore, the odds look increasingly grim for our most senior citizens: the disease affects less than 1% of those under 60 years of age; but the odds double every five years after 65, reaching an alarming rate of 30-50% of all seniors over the age of 85. The National Institutes of Health and the Alzheimer's Association estimate the current prevalence of the disease at 5 million in the U.S. alone, while the European Union adds another 3 million cases [1]. With an estimated annual cost of over $100 billion for the treatment and care, the disease is no longer ignored.

Furthermore, as if the lack of cure was not devastating enough, there is also no standard procedure for a definitive diagnosis of the disease: the only definitive method of diagnosis is autopsy, where the brain tissue is microscopically examined for the plaques and neurofibrillary tangles that are characteristic hallmarks of AD. Therefore, diagnosis is typically made through longitudinal clinical evaluations that include a series of memory tests and interviews, both with the patient and their caretakers. The diagnostic accuracy of clinical evaluations is estimated to be 90%, when conducted by expert neuropsychologists. Such level of expertise, however, is only available at highly specialized institutions, such as research hospitals. At community hospitals, where most patients are evaluated, the diagnostic accuracy for AD is estimated to be 75%, with a sensitivity of 83%, and specificity of 53%, despite the benefit of longitudinal follow up [2].

Yet, while there is no cure, there is medication that can significantly reduce the progression of the disease. Specifically, if diagnosed early, patients may live 8-20 years beyond initial diagnosis, with improved quality of life. Therefore, the importance of an accurate, non-invasive, and cost-effective diagnostic biomarker that can be made available to community hospitals cannot be overstated.

The analysis of event related potentials (ERPs) of the electroencephalogram (EEG) may provide just such a biomarker. Several studies have shown that certain characteristics of AD, such as cognitive decline, are associated with certain changes in the ERPs [3,4,5]. Most studies, however, could only provide statistical correlations that were simply not strong enough to allow patient specific diagnosis.

Typically, the protocol used to acquire the ERPs for cognitive analysis is the so-called *oddball paradigm*. In this protocol, subjects are instructed to respond to a series of stimuli, by pressing a button, when they hear an occasionally occurring oddball (target) tone of 2 kHz within a series of regular (standard) 1 kHz tones and novel sounds. The ERPs then show a series of peaks, among which the P300 – a positive peak with an approximate latency of 300 ms, seen in response to oddball stimuli only – is of particular interest. Changes in the amplitude and latency of the P300 are known to be altered by neurological disorders affecting the temporal-parietal regions of the brain [3,4,5].

Manuscript received January 31, 2007. This work is supported by the National Science Foundation under Grant No ECS-0239090, and by the National Institute on Aging of the National Institutes of Health under grant number P30 AG10124 - R01 AG022272.

B. Balut, M. Karnick and R. Polikar are with Signal Processing and Pattern Recognition Laboratory of the Electrical and Computer Engineering Department at Rowan University, Glassboro, NJ 08028, USA. (E-mails: {balutb59, karnic08}@students.rowan.edu;

D. Green and J. Kounios are with the Dept. of Psychology, Drexel Univ., Philadelphia, PA 19104 USA(E-mail: {debbie.green; kounios)@drexel.edu).

C. M. Clark is with the Department of Neurology, University of Pennsylvania, Philadelphia, PA (e-mail: chris.clark@uphs.upenn.edu).

Contact author: Robi Polikar, phone: (856) 256-5372; fax: (856) 256-5241. E-mail: polikar@rowan.edu .

More recently, EEG analysis combined with signal processing and the use of automated classifiers has enjoyed renewed interest, yielding limited success in some studies, including some of our own [6,7,8]. However, earlier studies have primarily been pilot studies using single channel data and with very few patients (10-25), making statistical generalization very difficult.

Hence our goal in this study is to investigate the feasibility of an automated neural network based diagnosis, that can at least meet (or exceed) the diagnostic accuracy of community hospital physicians, within the constraints mentioned above. We have used a substantially larger patient cohort, specifically recruited for this study, to improve the statistical validity of our diagnostic performance estimates.

In essence, we propose an ensemble of classifiers based algorithm designed to combine complementary information in ERPs obtained from several different electrodes, all located in the parietal region of the brain. This area was chosen due to its reported significance in memory and cognitive skills. Specifically, an ensemble of classifiers is generated, where each classifier is trained on ERP data obtained from separate EEG electrodes, obtained in response to different type of stimuli and analyzed at different frequency bands. The classifiers are then combined through various combination rules to obtain a data fusion based overall classification.

II. Experimental Setup

A. The Oddball Paradigm and the ERP Acquisition

In auditory oddball paradigm, the subject is equipped with a set of headphones, and hears a series of audio tones that occur once every 1 – 1.3 seconds. The majority (65%) of these tones are standard tones at 1 kHz, whereas another 20% are target tones at 2 kHz. In our implementation, we also used novel sounds (15%) obtained from sound clips, as described in [3]. The subject is instructed to respond to target tones only, by pressing a button. It is the rarely occurring target tones that evoke the ERPs, of which the P300 has been the most celebrated component.

The data collection lasted about 30 minutes per subject. Artifactual recordings were rejected by the EEG technician. The remaining recordings were amplified, digitized at 256 samples / second, and partitioned with respect to the nature of the stimulus tone. The ERPs were then segmented to 1 second (256-sample) long segments, including 200 ms of pre- stimulus and 800 ms of post stimulus recording. Segmented recordings were notch filtered at 59-61 Hz, normalized and averaged. The averaging involved 90-250 recordings per patient to obtain robust ERP recordings.

The ERPs obtained from cognitively normal individuals typically exhibit a strong P300 response, with about 300 ms latency after the stimulus. This response is hampered in those individuals whose cognitive skills are deteriorated due to AD. ERPs of such people usually exhibit a weaker P300, with a much extended latency (if any at all). Figure 1 (a) and (b) illustrate two such cases of patients included in this study. This correlation, as mentioned earlier, is a weak one, however: it is not unusual to observe a strong P300 in ERPs of AD patients (particularly during the early stages), and some cognitively normal individuals may have a suppressed P300. Fig. 1 (c) and (d) illustrate two such cases, also included in this study. Therein lies the difficulty of AD diagnosis based on merely visual analysis of the ERPs.

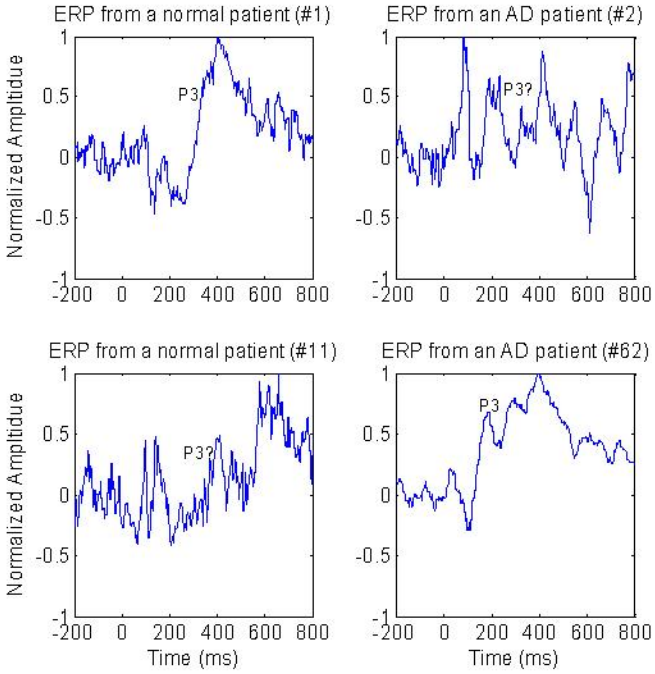

Fig. 1: (a) and (b) show the expected behavior from the P300 from both a normal and an AD patient, whereas (c) and (d) show opposite responses from a different pair of normal and AD patients.

The ERP data was acquired from 19 electrodes, placed on the scalp according to the 10/20 International System of electrode placement, as shown in Figure 2. During the early stages of this study, ERPs recorded from the Pz electrode, and obtained in response to target tones only were analyzed, as this specific combination was reported to provide the most robust P300 components [4]. Recently, however, we have found that the responses to novel sounds, recorded from other electrodes also provide valuable information [9].

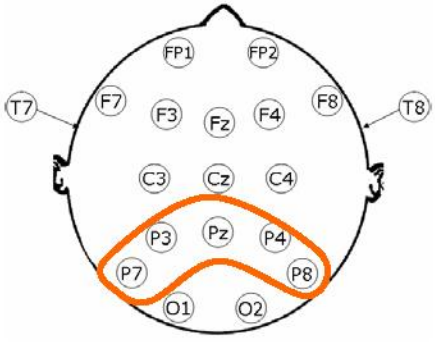

Fig. 2: Scalp Electrode Placement according to the 10/20 International Standard. Parietal electrodes P3, P4, P7, P8 and Pz are used in this study.

B. Patient Cohort

This study started with 28 patients, 14 in each cohort, and gradually grew to 37 to 48 to 51 patients. The results re-

ported in this paper represent our analysis on our final (and complete) cohort size of 71 patients. While relatively small from a computational intelligence perspective, this cohort size constitutes one of the largest studies of its kind.

A strict inclusion and exclusion criteria was followed to ensure data integrity: the inclusion criteria for a cognitively normal patient includes: (a) Age > 60; (b) CDR (clinical dementia rating) score = 0; (c) MMSE (mini mental state exam) score >26; (d) no cognitive decline within two years of testing for the study. The inclusion criteria for a probable AD patient includes: (a) Age > 60; (b) CDR score > 0.5; (c) MMSE score <= 26; (d) cognitive decline over the previous 12 months; (e) meets criteria for probable AD from the National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association [10]. All clinical evaluations were performed by the neuropsychologists at the Memory Disorders Clinics of the University of Pennsylvania.

The final cohorts included 34 AD patients with a mean age of 75 and MMSE score of 24.7, and 37 normal patients with a mean age of 76 and MMSE score of 29.2. Scored on a scale of 0 – 30, the MMSE is used to determine the cognitive level of the individual. A score over 26 is generally considered normal (with level of education taken into consideration), and a score below 19 typically indicates cognitive decline. A score below 12 is considered severe dementia. Note that at an average MMSE score of 24.7, the patients in the AD cohort of this study are at the earliest stages of the disease.

III. Methods

A. Feature Extraction

Traditionally, ERP analysis is performed in time or frequency domain. This analysis reveals only a portion of the information available in the ERP, however, since ERPs are non-stationary signals, whose spectral content vary in time. Other studies have shown that the ERPs, and the P300 component in particular, consist of the superposition of multiple functional components, where these components extend for different, yet overlapping, time intervals in different frequency bands [11]. This makes the discrete wavelet transform (DWT) an appropriate tool for the analysis of ERPs, as also shown in our earlier studies [8].

DWT provides time localizations of spectral components in a signal, thus providing a time-frequency representation. It does so by decomposing a signal into different frequency bands by successive low-pass and high-pass filtering. Each frequency band, organized in consecutive octaves, is analyzed at double the time resolution and half the frequency resolution of the preceding octave, and hence the term multiresolution analysis. The output of each high-pass filter constitutes the DWT coefficients at that level, while the low-pass filter outputs are further decomposed. Daubechies 4 (db4) wavelets were used in this study, chosen due to their good localization properties. DWT is now a well established technique, and detailed information on this topic can be found in many excellent references listed at [12].

The decomposition was carried out for 7 levels for the 256-point long signals. Those frequency bands corresponding to 1-2 Hz (Level 1, L_1), 2-4 Hz (L_2) and 4-8Hz (L_3), are of particular interest in ERP analysis, as the majority of the ERP signal's power lie in these bands. These levels are also known to contain the most cognitive information, as verified by our previous work [9].

Further dimensionality reduction was obtained by removing those DWT coefficients corresponding to pre-stimulus and late post-stimulus regions, and focusing only on 0 to 600 ms regions. The resulting feature vectors were of length 4, 4, and 6 for Levels 1, 2 and 3, respectively.

B. Classification, Data Fusion and Ensemble Systems

In this work, we employ an ensemble of classifiers approach to data fusion. Ensemble based systems have recently enjoyed an increased attention due to their reported advantages over single classifiers on a variety of applications [13]. Such systems combine several, preferably diverse, classifiers. The diversity is typically achieved by using a different training dataset for each classifier, which then allows each classifier to generate different decision boundaries. Each classifier then makes a different error, and strategically combining these classifiers can reduce the total error. The individual classifiers generated within an ensemble system are combined using one of several combination rules, some of which are reviewed below. Bagging, boosting (AdaBoost), stacking, and mixture of experts are examples of ensemble based classification [13], in all of which the goal is to improve classification performance. Using the ensemble approach for data fusion applications, i.e., combining complementary knowledge from different data sources, has in general been less explored. In this study, we explore using a classifier-fusion type ensemble approach, not just for improving performance on a classification problem, but specifically for combining information from different sources – namely ERPs obtained from different electrodes, in response to different stimuli, analyzed at different frequency bands.

Feed forward neural networks were used to train individual classifiers. Five electrodes and three frequency bands give us 15 combinations of feature vectors for each stimulus type (for a total of 30). The performance of each such combination was first obtained. We have then combined subsets of these 30 classifiers in groups of 3, 5 and 7 using each of the four combination rules discussed below. In all cases, each classifier in the ensemble provides a support for each category, AD or normal. The combination rule then combines these individual supports to calculate an ensemble support for each class. The class with the highest ensemble support is chosen as the final classification.

Sum, Product and Weighted Majority Voting

Let the continuous valued $d_{i,j} \in [0,1]$ represent the degree of support given by classifier T_i to class j; $i=1,\ldots,N$ and $j=1,\ldots,c$, where N is the number of classifiers and c is the number of classes. For each classifier, these supports are normalized to add up to 1 over different classes by the soft-

max transformation, and are often interpreted (albeit sometimes incorrectly) as posterior probabilities. For instance **x**,

$$d_{i,j}(\mathbf{x}) = \exp\left(d_{i,j}(\mathbf{x})\right) \Big/ \sum_{j=1}^{c} \exp\left(d_{i,j}(\mathbf{x})\right) \quad (1)$$

For weighted majority voting (WMV), we only need the labels that are predicted by each classifier, and hence we have binary valued $d_{i,j} \in \{0,1\}$. We also use a voting weight that determines how much weight the ensemble should give to the decision of each classifier:

$$W(T_i) = \log\left(1/\beta_{T_i}\right) \quad (2)$$

where β_T is the normalized training error of classifier T_i. The total support $\mu_j(\mathbf{x})$ given to class j is then

$$\mu_j(\mathbf{x}) = \sum_{j=1}^{N} W(T_i)\, d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in \{0,1\} \quad (3)$$

$$\mu_j(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^{N} d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in [0,1] \quad (4)$$

$$\mu_j(\mathbf{x}) = \frac{1}{N} \prod_{i=1}^{N} d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in [0,1] \quad (5)$$

for WMV, sum and product rules, respectively. Denoting the set of class labels as $\Omega = \{\omega_1, \omega_2, \ldots, \omega_C\}$, and the ensemble decision for instance **x** as $\boldsymbol{\mathcal{E}}(\mathbf{x})$, the ensemble decision is then chosen as ω_m for which the support $\mu_j(\mathbf{x})$, is maximum:

$$\boldsymbol{\mathcal{E}}(\mathbf{x}) = \omega_m \,|\, m = \arg\max_j \left(\mu_j(\mathbf{x})\right),\ j = 1, \cdots, c \quad (6)$$

Dempster-Shafer Combination Rule

Based on Dempster-Shafer (DS) theory of evidence, DS combination rule uses decision profiles to specifically account for the support given to each class when a classifier is trained from instances of different classes. The decision profile DP(**x**) is a TxC matrix of outputs $d_{i,j}$ from all N classifiers for the given **x**, such that the j^{th} column with $d_{1,j}$ through $d_{N,j}$ are the supports from classifiers $T_1 \sim T_N$ to class ω_j, and the i^{th} row with $d_{i,1}$ through $d_{i,c}$ is the support from classifier $T_i(\mathbf{x})$ to classes $\omega_1 \sim \omega_c$. Based on DP(**x**), a decision template DT_j is computed which represents the most typical decision profile for each class ω_j. It is calculated as the average of decision profiles of all training instances of class ω_j

$$DT_j = \left(1/N_j\right) \sum_{\mathbf{x} \in \mathbf{X}_j} DP(\mathbf{x}) \quad (7)$$

where $\mathbf{X}_j$ indicates the set of class ω_j instances, and N_j is the cardinality of this set.

A key quantity in DS combination is *proximity*, which represents the (normalized) similarity of the ith row of DT_j to the output of the ith classifier for instance **x**,

$$\Phi_{j,i} = \frac{(1 + \| DT_j^i - D_i(x) \|^2)^{-1}}{\sum_{k=1}^{c} (1 + \| DT_k^i - D_i(x) \|^2)^{-1}} \quad (8)$$

where $\| . \|$ is Euclidean norm. Finally, we compute belief b_j

$$b_j(D_i(x)) = \frac{\Phi_{j,i}(x) \prod_{k \neq j} (1 - \Phi_{k,i}(x))}{1 - \Phi_{j,i}(x)[1 - \prod_{k \neq j} (1 - \Phi_{k,i}(x))]} \quad (9)$$

which indicates the ensemble's *degree of belief* that classifier T_i correctly identifies instances from class ω_j

The final support given to class ω_j by the ensemble can then be obtained as

$$\mu_j = K \prod_{i=1}^{T} b_j(D_i(x)), \qquad j = 1, \ldots, C \quad (10)$$

where K is a normalizing constant such that $\sum \mu_j(\mathbf{x}) = 1$

IV. Results

Previously, we have analyzed responses from the Cz, Fz, and Pz electrodes, as these were reported to contain the most relevant diagnostic information. When analyzed individually, the best diagnostic performance of 76% was obtained from the Pz electrode, in response to novel sounds at Level 2, on 52 patients [9]. In this study, we investigate whether there is complementary information in ERPs obtained from the surrounding parietal region electrodes of P3, P4, P7, and P8. ERPs from all five electrodes, in response to both types of stimuli (target and novel), and at each of the three frequency bands were first analyzed individually for the now expanded 71 patient cohort. This analysis provided us with the baseline individual performances of each of the 30 electrode – frequency – stimulus combinations. The diagnostic classification performances and the corresponding confidence intervals listed in Table 1 are obtained as averages of 10 independent leave-one-out trials, where in each trial the MLP type classifiers were randomly initialized. The best individual performances are highlighted in bold.

We observe from Table 1 that typical individual diagnostic performances are in the mid 50 to low 70% range. The best performing classifiers (at 72% and 70%) are those ob-

TABLE 1
SINGLE ELECTRODE CLASSIFIER PERFORMANCE.

Electrode	Mean (%)	CI (%)
Target Level 1: 1-2 Hz		
P3	**63.2**	**2.4**
P4	60.1	2.3
P7	54.8	2.9
P8	54.4	2.9
PZ	52.5	2.4
Target Level 2: 2-4 Hz		
P3	60.6	3.0
P4	60.6	2.8
P7	51.8	3.9
P8	55.6	3.3
PZ	58.5	2.8
Target Level 3: 4-8 Hz		
P3	50.6	3.1
P4	41.0	4.5
P7	59.7	2.6
P8	35.9	3.3
PZ	44.7	3.2

Electrode	Mean (%)	CI (%)
Novel Level 1: 1-2 Hz		
P3	60.7	2.4
P4	61.7	2.8
P7	51.6	2.2
P8	60.7	2.4
PZ	**67.5**	**2.8**
Novel Level 2: 2-4 Hz		
P3	**67.8**	**2.3**
P4	62.7	1.7
P7	61.0	3.1
P8	**66.3**	**2.2**
PZ	**72.3**	**1.8**
Novel Level 3: 4-8 Hz		
P3	64.9	1.9
P4	53.9	2.8
P7	56.1	1.8
P8	51.6	1.9
PZ	**70.1**	**2.8**

tained with data from the Pz electrode at level 2 and 3, in response to novel sounds. The best performing single electrode/frequency band out of the target responses was P3 at level 1 with 63.24% classification performance. To ensure a performance boost from the ensemble, the ensemble would have to perform significantly better than these benchmarks.

To determine whether complementary information is provided by ERPs obtained from different electrode - frequency – stimulus combinations, they were combined in groups of 3, 5 and 7 classifiers. An exhaustive search through all possible ensemble combinations of 3, 5 and 7 classifiers for both target tone and novel sound responses were performed. This process results in 455 combinations of 3 classifiers, (15 choose 3), 3003 combinations of 5, and 6435 combinations of 7 classifiers for *each* stimulus type.

First, we combined classifiers trained with different stimulus types separately. The diagnostic performance of such 3, 5 and 7 classifier ensembles are shown in Table 2 for the top five performing combinations (where frequency levels are indicated as subscripts, e.g. Pz_3 indicates electrode Pz at level 3 (4-8 Hz)). The combination rule that provided the listed performances is also shown in Table 2.

Interesting observations can be made from Table 2. First, the ensemble of classifiers trained with different electrode–frequency combinations provide statistically significant performance improvement (reaching 79%) over single classifiers listed in Table 1. This indicates that, given any stimulus type, ERPs obtained from different electrodes and limited to different frequency bands do carry complementary information. Second, as we incorporate an increasing number of ERPs from different combinations, the diagnostic performance increases up to combinations of 7 classifiers (with no further improvement when more than 7 classifiers were fused). Third, the electrodes whose ERPs performed well individually (such as Pz) were included in all top performing ensembles: an expected, yet a satisfying outcome. Finally, the sum rule, in general, appears to work better than others. However, those ensembles in top 10, but not in top 5, were obtained through different combination rules, and the difference in performances were not statistically significant. This indicates that the specific choice of the combination rule may not make as much of a difference.

Our final analysis within this effort was to determine whether ERPs obtained in response to target and novel stimuli provided complementary information with respect to each other. To do so, we have picked the 6 most frequently occurring electrode - frequency band combinations from the 25 best performing ensemble combinations for each of the target and novel stimuli, for a total of 12 electrode - frequency band – stimulus combinations. Using one classifier for each such combination, 3, 5 and 7-classifier ensembles were then exhaustively formed and evaluated using all four combination rules. The results are shown in Table 3, where ERPs in responses to target tones are indicated with a "t" appended to their electrode name.

Even more intriguing observations can be made from Table 3. Most importantly, comparing the diagnostic performances in Table 3 to those in Table 2, it is clear that there is additional and complementary information in ERPs obtained in response to different types of stimuli. In each of the 3, 5,

TABLE 2
BEST PERFORMING ENSEMBLES FOR TARGET AND NOVEL RESPONSES FOR ENSEMBLES OF 3, 5 AND 7 CLASSIFIERS.

Target - Combinations of 3				Novel - Combinations of 3			
Electrodes/ Levels	Mean (%)	CI (%)	Comb Rule	Electrodes/ Levels	Mean (%)	CI (%)	Comb Rule
$P3_1$,$P3_2$,$P7_3$	70.0	2.8	Sum	Pz_2,Pz_3,$P8_2$	75.4	2.4	Sum
$P3_1$,$P4_1$,$P4_2$	69.6	3.0	Sum	Pz_1,Pz_3,$P8_2$	74.8	2.2	Sum
$P3_1$,$P4_2$,$P7_3$	69.3	2.3	Sum	Pz_2,Pz_3,$P3_2$	74.5	2.4	Sum
Pz_3,$P3_1$,$P4_2$	69.2	3.2	Sum	Pz_2,Pz_3,$P4_3$	74.2	2.2	Sum
$P3_1$,$P3_2$,$P4_2$	69.2	2.4	Sum	Pz_2,Pz_3,$P3_3$	74.2	2.2	Sum
Target - Combinations of 5				**Novel – Combinations of 5**			
Pz_3,$P3_1$,$P3_2$, $P4_2$,$P7_3$	70.1	3.2	WMV	Pz_1,Pz_3,$P3_2$, $P3_3$,$P8_1$	77.3	2.0	Sum
$P3_1$,$P3_2$,$P4_1$, $P4_2$,$P7_3$	70.1	2.2	Sum	Pz_2,Pz_3,$P3_1$, $P7_2$,$P8_2$	77.3	1.5	Sum
Pz_2,Pz_3,$P3_2$, $P3_3$,$P7_3$	69.9	2.9	Sum	Pz_1,Pz_2,Pz_3, $P3_3$,$P8_2$	77.0	1.1	Sum
Pz_2,$P3_1$,$P3_2$, $P4_1$,$P7_3$	69.7	2.6	Sum	Pz_1,Pz_2,Pz_3, $P3_2$,$P3_3$	76.9	1.7	Sum
Pz_2,Pz_3,$P3_1$, $P4_1$,$P4_2$	69.2	2.4	Sum	Pz_2,Pz_3,$P3_3$, $P7_3$,$P8_2$	76.5	2.5	Sum
Target - Combinations of 7				**Novel - Combinations of 7**			
Pz_2,$P3_1$,$P3_2$, $P4_2$,$P7_2$,$P7_3$, $P8_1$	70.0	2.1	Prod	Pz_1,Pz_2,Pz_3,$P3_2$, $P3_3$,$P8_1$,$P8_2$	79.0	2.2	Sum
Pz_1,Pz_2,$P3_1$, $P3_2$,$P3_3$,$P7_3$, $P8_1$	69.4	3.6	Prod	Pz_1,Pz_2,Pz_3, $P3_3$,$P7_2$,$P8_1$,$P8_2$	78.7	2.0	Sum
Pz_3,$P3_1$,$P3_2$, $P4_1$,$P4_2$,$P7_3$, $P8_2$	69.4	2.6	Sum	Pz_1,Pz_2,Pz_3, $P3_3$,$P7_1$,$P7_2$,$P8_1$	78.7	1.6	Sum
Pz_2,Pz_3,$P3_1$, $P3_2$,$P4_2$,$P7_3$, $P8_1$	69.2	2.5	Prod	Pz_1,Pz_2,Pz_3, $P3_1$,$P3_3$,$P7_2$,$P8_1$	78.6	1.5	Sum
Pz_3,$P3_1$,$P3_2$, $P4_1$,$P4_2$,$P7_3$, $P8_1$	69.0	2.6	Sum	Pz_1,Pz_2,Pz_3, $P3_2$,$P3_3$,$P7_2$,$P8_1$	78.5	2.2	Sum

TABLE 3
COMBINATIONS OF TARGET AND NOVEL RESPONSES IN ENSEMBLES OF 3, 5, AND 7 CLASSIFIERS

Target/Novel - Combinations of 3						
Elec-trodes/Levels	Mean (%)	CI (%)	Sens (%)	Spec (%)	PPV (%)	Comb Rule
Pz_3, $P3_2$, $P3t_1$	78.2	2.2	77.1	78.4	70.8	Sum
Pz_3, $P3_2$, $P3_3$	78.0	3.1	82.1	75.6	68.1	Sum
Pz_2, Pz_3,$P3_3$	77.5	3.1	76.8	77.3	66.9	Sum
Pz_1, Pz_3, $P3_2$	77.3	2.0	77.9	71.9	66.6	Sum
Pz_3, $P3_3$, $P3t_1$	76.5	3.3	76.5	79.2	70.3	Sum
Target/Novel - Combinations of 5						
Pz_2, Pz_3,$P3_2$, $P3_3$,$P3t_1$	80.9	2.3	78.2	80.3	73.9	Sum
Pz_1, Pz_3,$P3_2$, $P3_3$,$P3t_2$	80.9	1.9	79.7	79.2	66.1	Sum
Pz_1, Pz_2,Pz_3, $P3_2$,$P7t_3$	80.8	1.2	80.3	81.4	71.8	DS
Pz_3,$P3_2$,$P3_3$, $P3t_1$,$P7t_3$	80.4	2.2	80.0	80.8	68.0	Sum
Pz_2, Pz_3,$P3_3$, $P3t_1$,$P7t_3$	80.1	2.6	76.7	79.5	68.7	Sum
Target/Novel - Combinations of 7						
Pz_1, Pz_3,$P3_2$,$P3_3$, $P3t_1$, $P3t_2$, $P7t_3$	**81.7**	**3.8**	**79.7**	**84.1**	**71.3**	**Sum**
Pz_3,P3-2,$P3_3$,$P3t_1$, Pzt_3, $P3t_2$, $P7t_3$	81.5	3.4	76.8	82.7	74.6	DS
Pz_3,P3-2,$P3_3$,$P3t_1$, $P3t_2$, $P4t_2$, $P7t_3$	81.4	6.1	78.8	80.3	72.4	DS
Pz_3,$P3_2$,P8-2,$P3_3$, $P3t_1$,$P4t_2$, $P7t_3$	81.2	2.6	80.6	81.9	74.1	DS
Pz_3,$P3_2$,$P3_3$,Pz_2t, $P3t_1$, $P3t_2$,$P7t_3$	81.1	3.6	81.5	80.8	71.0	DS

or 7 classifier ensembles, the performance consistently increases compared to corresponding ensembles using single stimulus type: 75% to 78% with 3 classifiers, 77% to 80.9% with 5 and 79% to 81.7% with 7 classifiers.

Second, as before, individually best performing electrode-frequency-stimulus combinations are included in the best performing ensembles, and including additional classifiers always improves diagnostic performance up to 7 classifiers. Third, the sum rule is still the best combiner; however the differences between the performance of the sum and other rules were not statistically significant.

It is also customary in similar medical studies to provide additional diagnostic indicators, other than just the performance of the proposed technique. Three most commonly used indicators are sensitivity, specificity and positive predictive value (PPV). Sensitivity is the ability of a medical test to correctly identify the target group. In this application, it is the proportion of true AD patients correctly identified as AD patients by the classification system. Specificity is the ability of a test to correctly identify the control group: the proportion of cognitively normal patients correctly identified as normal. Finally, PPV is the probability that the patient has the disease, given that the test result is positive. In this study, PPV is the proportion of those patients identified as AD patients by the classifier, who actually have AD. These diagnostic indicators are also provided in Table 3, and are particularly promising when compared to those reported for community hospitals.

V. Conclusion

In this study, we investigated the diagnostic accuracy and performance of an ensemble-of-classifiers-based data fusion approach for early diagnosis of Alzheimer's disease. The classifiers were trained on wavelet coefficients of ERPs obtained at different electrode locations, in response to different stimulus types and band-limited to different frequency bands. Specifically, ERPs obtained from all parietal region electrodes, in response to both target (oddball) and novel stimuli were first individually analyzed and evaluated, and then exhaustively combined with each other to obtain an ensemble of classifiers. The ensemble classifiers were then combined using four different combination rules.

On a cohort of 71 patients, the results indicate that there is indeed complementary information in ERPs obtained from different electrodes and in response to different stimuli, and the ensemble combination of classifiers trained on each provides a statistically significant performance improvement. Of the five electrodes, Pz, P3 and P7 appear to provide the most informative diagnostic specific discriminatory information, as they appear most often in all top performing combinations. However, since the performance differences among the five top performing combinations were not statistically significant, it seems that all parietal region electrodes carry some relevant information.

Most importantly, we note that a recent study estimates the community clinic-based physicians' diagnostic performances at 83% sensitivity, 53% specificity and 75% overall diagnostic performance [2]. While the results of this study are quite satisfactory in their own right, from a computational intelligence perspective, they are particularly meaningful within the context of this application. This is because the ensemble generalization performance in 80% range exceeds the 75% diagnostic performance of trained physicians at community-based healthcare providers—despite the physicians' benefit of a longitudinal study, and despite the early diagnosis emphasis of our study. Furthermore, with sensitivity and specificity also reaching 80% ranges, these results are particularly promising. Overall, the proposed ensemble based data fusion approach for early diagnosis of AD exceeds the diagnostic performance of the community hospitals, with clinically and statistically significant margins.

References

[1] Alzheimer's Association, "Alzheimer's Disease Statistics," Available at: http://www.alz.org/AboutAD/statistics.asp.

[2] A. Lim, W. Kukull, D. Nochlin, J. Leverenz, and others, "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series," *J. of the American Geriatrics Soc.*, 47, (1999), 564-569.

[3] S. Yamaguchi, H. Tsuchiya, S. Yamagata, G. Toyoda, S.Kobayashi, "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, vol.112, no. 2, pp. 195-203, 2002.

[4] D. Linden, "The P300: Where in the brain is it produced and what does it tell us?" *The Neuroscientist*, vol. 11, no. 6, 563-576, 2005.

[5] J. Polich, "P300 in clinical applications," In *Electroencephalography*, E. Niedermeyer, F. Da Silva, Ed. Philadelphia: Williams & Wilkins, pp.1073-1091,1999.

[6] A. Petrosian, D. Prokhorov, W. Nanson and R. Schiffer, "Recurrent neural network based approach for early recognition of Alzheimer's disease in EEG," *Clinical Neurophysiology*, vol. 112, no. 8, pp. 1378-1387, 2001.

[7] N. Stepenosky, A. Topalis, H. Syed, D. Green, J. Kounios, C. Clark, R. Polikar, "Boosting based classification of event related potentials for early diagnosis of Alzheimer's disease," *IEEE Eng. in Medicine and Biology Society Conf.*, Shanghai, China, September 2005.

[8] G. Jacques, J.L. Frymiare., J. Kounios, C. Clark, and R. Polikar, "Multiresolution wavelet analysis and ensemble of classifiers for early diagnosis of Alzheimer's disease," *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, vol. 5, pp. 389-392, Philadelphia, PA, 2005.

[9] R. Polikar, A. Topalis, D. Green, J. Kounios, C.M. Clark, "Comparative multiresolution analysis and ensemble of classifiers approach for early diagnosis of Alzheimer's Disease," *Computers in Biology and Medicine*, Volume 37, Issue 4, April 2007, Pages 542-558.

[10] G. McKhann, D. Drachman, M. Folstein, R. Katzman, and others, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Dept. of Health and Human Services Task Force on Alzheimer's Disease," *Neurology* vol.34, pp.939–944, 1984

[11] T. Demiralp, A. Ademoglu, "Decomposition of event related brain potentials into multiple functional components using wavelet transform," *Clinical Electroencephalography*, vol. 32, no. 3, pp. 122-138, 2001.

[12] M. Unser, editor, Gallery at www.wavelet.org, Available at: http://www.wavelet.org/phpBB2/gallery.php

[13] L. Kuncheva, *Combining Pattern Classifiers*, New York: Wiley, 2004.