

# Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease

Robi Polikar<sup>a,\*</sup>, Apostolos Topalis<sup>a</sup>, Deborah Green<sup>b</sup>, John Kounios<sup>b</sup>, Christopher M. Clark<sup>c</sup>

<sup>a</sup>Electrical and Computer Engineering, Rowan University, Glassboro, NJ 08028, USA

<sup>b</sup>Department of Psychology, Drexel University, Philadelphia, PA 19104, USA

<sup>c</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA

## Abstract

Early diagnosis of Alzheimer's disease (AD) is becoming an increasingly important healthcare concern. Prior approaches analyzing event-related potentials (ERPs) had varying degrees of success, primarily due to smaller study cohorts, and the inherent difficulty of the problem. A new effort using multiresolution analysis of ERPs is described. Distinctions of this study include analyzing a larger cohort, comparing different wavelets and different frequency bands, using ensemble-based decisions and, most importantly, aiming the earliest possible diagnosis of the disease. Surprising yet promising outcomes indicate that ERPs in response to novel sounds of oddball paradigm may be more reliable as a biomarker than the more commonly used responses to target sounds.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Alzheimer's disease diagnosis; Wavelets; Event-related potentials; Ensemble classifiers

## 1. Introduction

Neurological disorders that cause gradual loss of cognitive function are collectively known as dementia. Among several forms of dementia, perhaps the most infamous and the most common form is the irreversible and incurable senile dementia of Alzheimer's type, or just Alzheimer's disease (AD), in short. AD, first described by Alois Alzheimer in 1906, was once considered a rare disease, and it was mostly ignored due to elderly people being its primary victim. Today, on the centennial anniversary of the disease's discovery, the situation is much different: as the world's population ages rapidly—primarily in developed countries—so does the number of people affected by the disease. Different estimates vary considerably; however, it is now estimated that there are 18–24 million people suffering from AD worldwide, two-thirds of whom are living in developed or developing countries. This number is expected to reach 34 million by 2025. Up to age 60, AD appears in less than 1% of the population, but its prevalence

increases sharply, doubling every 5 years thereafter: AD affects 5% of 65-year olds, and over 30% of 85-year olds. Beyond age 85, the odds of developing AD approaches a terrifying ratio of 1 in 2 [1,2].

The specific causes of AD are unknown; however, the disease is associated with two abnormal proteins: neurofibrillary tangles clustering inside the neurons, and amyloid plaques that accumulate outside of the neurons of primarily the cerebral cortex, amygdale and the hippocampus. These unusual proteins cause a gradual but irreversible decline in all cognitive (and eventually motor) skills, leaving the victim incapable of caring for him/herself. Furthermore, these proteins can only be identified by examining the brain tissue under a microscope, leaving autopsy as the only method for positive diagnosis. AD not only incapacitates its victim, but it also causes an unbearable grief on the victim's caregiver, and a devastating financial toll on the society with an annual cost of over \$100 billion.

Several biomarkers have been linked to AD, such as the cerebrospinal fluid tau,  $\beta$ -amyloid, urine F2-isoprostane, brain atrophy and volume loss detected by PET or MRI scan [3,4]. However, these methods have either not proven to be conclusive, or remain primarily university or research

\* Corresponding author. Tel.: +1 856 256 5372; fax: +1 856 256 5241.

E-mail address: polikar@rowan.edu (R. Polikar).

hospital-based tools. While clinical and neuropsychological evaluations achieve an average positive predictive value (PPV) of 85–90%, this level of expertise is typically available only at university or research hospitals, and hence remain beyond reach for most patients. Therefore, these patients are evaluated by local community healthcare providers where the expertise and accuracy of AD-specific diagnosis remains uncertain. Our sole metric for community clinics is a recent study that reported 83% sensitivity, 55% specificity and 75% overall accuracy on AD diagnosis by a group of Health Maintenance Organization-based physicians, despite having the advantage of longitudinal followup [5]. Meanwhile, recent development of pathologically targeted medications requires an accurate diagnosis at the earliest stage possible, so that the patient's life expectancy, as well as his/her quality of life, may be significantly improved. Therefore, to have a meaningful impact on healthcare, the diagnostic tool must be inexpensive, non-invasive, accurate, available to community physicians, and be able to diagnose the disease at its earliest stages.

Event related potentials (ERPs) of the electroencephalogram (EEG) may provide such a tool. It is a well-established and reliable procedure, it is non-invasive and readily available to community clinics. However, the ability of EEG signals to resolve AD-specific information is typically masked by changes due to normal aging, coexisting medical illness and levels of anxiety or drowsiness during measurements. Various components of the ERPs, obtained through the oddball paradigm protocol, have previously been linked to cognitive functioning, and are believed to be relatively insensitive to above-mentioned parameters [6–10].

In oddball paradigm, subjects are instructed to respond, typically by pressing a button, to an occasionally occurring target (oddball) tone of 2 kHz, within a series of regular 1 kHz tones. The ERPs then show a series of peaks, among which the P300—a positive peak with an approximate latency of 300 ms that occurs only in response to the oddball stimulus—is of particular interest. Changes in the amplitude and latency of the P300 (P3, for short) are known to be altered by neurological disorders, such as the AD, that affect the temporal–parietal regions of the brain [11]. Polich et al. have shown that increased latency and decreased amplitude of P300 is associated with AD [6,12]. Several other efforts, such as [9,11,13–17], have later confirmed the strong link between AD and P300. More recently, task-irrelevant novel sounds have been included in the protocol that may help distinguish AD from other forms of dementia using the amplitude and latency of the P300 [11]. However, looking at just the P300 component—while provides statistical correlation with AD—does not help in identifying individual patients: cognitively normal people may have delayed or absent P300, and those with AD, in particular in early stages, may still have strong P300, as shown in Fig. 1. The inability of classical statistical approaches in individually identifying specific cases demands sophisticated approaches for such individual identification.

Automated classification algorithms, such as neural networks, can be used for case-by-case identification of individual patient ERPs. The success of such automated classification algorithms strongly depends on the quantity and the quality

of the training data: the available data must adequately sample the feature space, and the features themselves must carry discriminatory information among different classes. Traditionally, the features of the ERPs are obtained either in time (e.g., amplitude and latency of the P300) [11,15–22] or in frequency domain (e.g., power of different spectral bands of the ERP) [23–31]. However, both are suboptimal, since the ERP is a time and frequency-varying non-stationary signal; and a time–frequency-based analysis is more suitable. Despite its now mature history, studies applying time–frequency techniques, such as wavelets, to ERPs have only recently started, and mostly on non-AD-related studies designed specifically for structural analysis of the P300 [32–44].

Other studies investigated the feasibility of wavelet analysis of EEGs, along with neural networks, but they either did not use ERPs [27,45] or did not specifically target AD diagnosis. For individual AD-specific diagnosis, there have been very few studies that use an appropriate time–frequency analysis, such as discrete wavelet transform (DWT), followed by neural network classification. The results of these primarily pilot studies, such as [46,47], including our previous efforts [48,49] can be characterized as only limited success, due to several reasons: relatively small study cohort with typically 10–30 patients, not targeting diagnosis at the earliest stages, suboptimal selection of classifier model and/or its parameters, as well as the sheer inherent difficulty of the problem. The results therefore remain largely inconclusive.

In this study, we describe a new effort that investigates the feasibility of an automated classification approach that employs multiresolution wavelet analysis; however several factors set this study apart from previous efforts: (i) a very strict and controlled recruitment protocol along with a very detailed and thorough clinical evaluation protocol (see Section 2) is followed to ensure the quality of study cohort; (ii) the study cohort recruited for this study constitutes one of the largest of similar prior efforts; (iii) several different types of wavelets commonly used for analysis of biological signals are compared instead of a single generic wavelet; (iv) single classifier, as well as multiclassifier based ensemble approaches are implemented and compared; (v) analysis is done not only with respect to the general classification performance, but also with respect to commonly used medical diagnostic quantities, such as sensitivity, specificity and PPV, and most importantly; (vi) this study uniquely targets diagnosing the disease at its earliest stage possible, typically before commonly recognized symptoms appear.

In P300 studies, the ERPs are typically obtained from one of the so-called  $P_z$ ,  $C_z$  or  $F_z$  electrodes of the 10–20 EEG electrode placement system shown in Fig. 2, primarily the former one. The common choice of  $P_z$  electrode is well justified, as ERPs are known to be most prominent in the central parietal regions of the cortex [40]. Furthermore, since the P300 is traditionally associated with the oddball tone, only responses to this tone are typically analyzed. In our previous preliminary studies, we have also analyzed the oddball responses from the  $P_z$  electrode. We now investigate the diagnostic information that may reside in data obtained from the other two electrodes,  $C_z$  and  $F_z$ , and obtained in response to the novel tones, as well

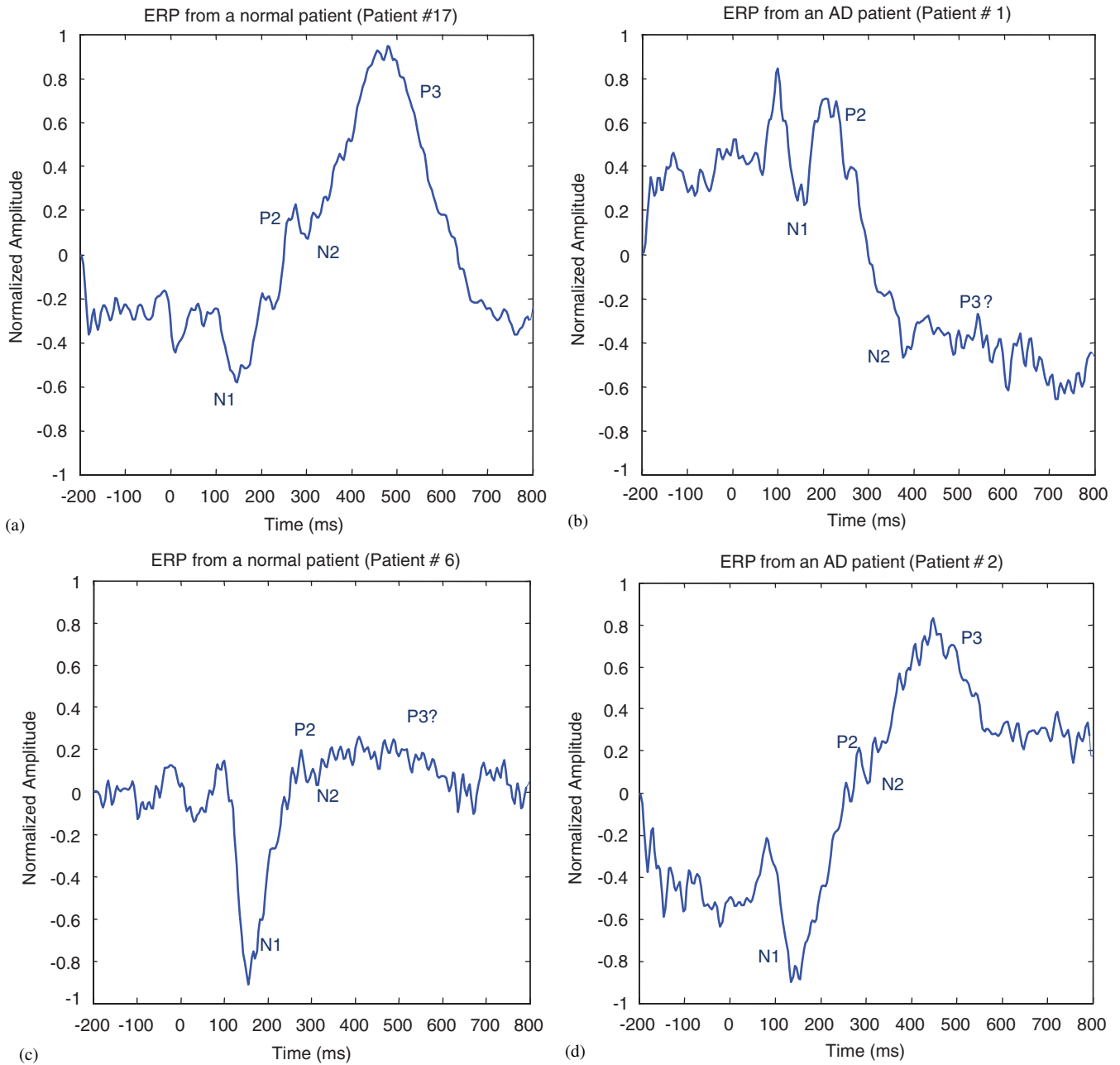


Fig. 1. (a and b) Expected P300 behavior from normal and AD patients, (c and d) not all cases follow this behavior.

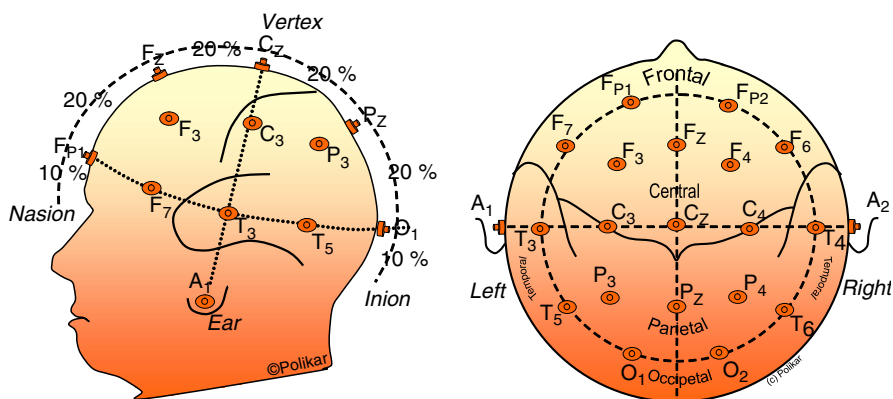


Fig. 2. The 10–20 international EEG electrode placement system.

as the target tones. Our justification for analyzing the remaining two electrodes is the relative and symmetric proximity of  $C_Z$  and  $F_Z$  electrodes to the  $P_Z$  electrode. Our justification for analyzing the responses to the novel tones is the potential information that may be present in other components of the ERP, such as the P3a, that may be more prominent in responses to the novel tones.

## 2. Experimental setup

### 2.1. Research subjects and the gold standard

The current gold standard for AD diagnosis is clinical evaluation through a series of neuropsychological tests, including interviews with the patient and their caregivers. Seventy-two patients have been recruited so far by the Memory Disorders Clinic and Alzheimer's Disease Research Center of University of Pennsylvania, according to the following inclusion and exclusion criteria for each of the two cohorts: probable AD and cognitively normal.

*Inclusion criteria for cognitively normal cohort:* (i) Age > 60; (ii) Clinical Dementia Rating Score = 0; (iii) Mini Mental State Exam (MMSE) Score > 26; (iv) no indication of functional or cognitive decline during the 2 years prior to enrollment based on a detailed interview with the subject's knowledgeable informant.

*Exclusion criteria for cognitively normal cohort:* (i) Evidence of any central nervous system neurological disease (e.g., stroke, multiple sclerosis, Parkinson's disease or other form dementia) by history or exam; (ii) use of sedative, anxiolytic or anti-depressant medications within 48 h of ERP acquisition.

*Inclusion criteria for AD cohort:* (i) Age > 60; (ii) Clinical Dementia Rating Score > 0.50; (iii) MMSE Score  $\leq$  26; (iv) presence of functional and cognitive decline over the previous 12 months based on detailed interview with a knowledgeable informant; (v) satisfaction of National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable AD [50].

*Exclusion criteria for AD cohort:* Same as for the cognitively normal controls.

All subjects received a thorough medical history and neurological exam. Key demographic and medical information, including their current medications (prescription, over-the-counter, or any alternative medications) were noted. The evaluation included standardized assessments for overall impairment, functional impairment, extrapyramidal signs, behavioral changes and depression. The clinical diagnosis was made as a result of these analyses as described by the NINCDS-ADRDA criteria for probable AD [50].

The inclusion criteria for AD cohort were designed to ensure that subjects were at the earliest stages of the disease. One metric is the MMSE, a widely used standardized test for evaluating cognitive mental status. The test assesses orientation, attention, immediate and short-term recall, language and the ability to follow simple verbal and written commands. It also provides a total score placing the individual on a scale of

cognitive function. Cognitive performance as measured by the MMSE shows an inverse relationship between MMSE scores and age/education, ranging from a median of 29 for those 18–24 years of age, to 25 for individuals 80 years of age and older. The median MMSE score is 29 for individuals with at least 9 years of schooling, 26 for those with 5–8 years of schooling and 22 for those with 0–4 years of schooling [51,52]. A grade less than 19 usually indicates cognitive impairment. MMSE is not used for diagnosis alone, but rather for assessing the severity of disease. The AD diagnosis itself is made based on the above-mentioned NINCDS-ADRDA criteria for probable AD.

Of the 72 patients initially recruited, 20 were removed due to various reasons, including those AD patients who—despite satisfying the above requirements—were too demented to be considered at the earliest stage of the disease. Of the 52 remaining patients, 28 were probable Alzheimer's ( $\mu_{AGE} = 79$ ,  $\mu_{MMSE} = 24.7$ ), and 24 were cognitively normal ( $\mu_{AGE} = 76$ ,  $\mu_{MMSE} = 29.6$ ). Note that with an average MMSE score of 25, the AD cohort represents those who are at the earliest stage of the disease, a stage during which the symptoms of the disease may not even be noticeable. While this distinction makes the classification problem all the more challenging, it also sets this study apart from similar earlier efforts. Also, with 52 patients, this effort constitutes one of the larger studies of its kind to date.

### 2.2. ERPs acquisition protocol

The ERPs were obtained using an auditory oddball paradigm while the subjects were comfortably seated in a specially designated room. We used the protocol described by Yamaguchi et al. [11], with slight modifications. Binaural audiometric thresholds were first determined for each subject using a 1 kHz tone. The evoked response stimulus was presented to both ears using stereo earphones at 60 dB above each individual's auditory threshold. The stimulus consists of tone bursts 100 ms in duration, including 5 ms onset and offset envelopes. A total of 1000 stimuli, including frequent 1 kHz normal tones ( $n = 650$ ), infrequent 2 kHz oddball (target) tones ( $n = 200$ ) and novel sounds ( $n = 150$ ) were delivered to each subject with an inter-stimulus interval of 1.0–1.3 s. The subjects were instructed to press a button each time they heard the 2 kHz oddball tone. The subjects were not told about the presence of novel sounds ahead of time, which consisted of unique sound bytes that were not repeated. With frequent breaks (approximately 3 min of rest every 5 min), data collection typically took less than 30 min. The experimental session was preceded by a 1-min practice session without the novel sounds.

ERPs were recorded from 19 electrodes embedded in an elastic cap. The electrode impedances were kept below 20 k $\Omega$ . Artfactual recordings were identified and rejected by the EEG technician. The potentials were finally amplified, digitized at 256 Hz/channel, lowpass filtered, averaged, notched filtered at 59–61 Hz and baselined with respect to the prestimulus interval for a final 257 sample, 1-s long signal. ERPs are often difficult to extract from a single response, due to many variations in

cortical activity. Consecutive successful responses to each tone are therefore synchronized and averaged (after responses with artifacts, responses to missed targets, etc. are removed by the EEG technician) to obtain a robust ERP response. The averaging process, a routine portion of the oddball paradigm protocol, consisted of averaging 90–250 responses per stimulus type each, per patient.

### 3. Methods

#### 3.1. Multiresolution wavelet analysis for feature extraction

Time localizations of spectral components can be obtained by multiresolution wavelet analysis, as this method provides the *time–frequency representation* of the signal. Among many time–frequency representations, the DWT is perhaps the most popular one due to its many desirable properties, and its ability to solve a diverse set of problems, including data compression, biomedical signal analysis, feature extraction, noise suppression, density estimation and function approximation, all with modest computational expense. Considering the audience of this journal, the well-established nature of the wavelet theory, as well as for brevity, we only describe the specific main points of DWT implementation here, and refer the interested readers to many excellent references listed in [53].

The DWT analyzes the signal at different resolutions (hence, multiresolution analysis) through the decomposition of the signal into several successive frequency bands. The DWT utilizes two sets of functions, a scaling function,  $\phi(t)$ , and a wavelet function,  $\psi(t)$ , each associated with lowpass and highpass filters, respectively. An interesting property of these functions is that they can be obtained as a weighted sum of the scaled (dilated) and shifted versions of the scaling function itself:

$$\phi(t) = \sum_n h[n]\phi(2t - n), \quad (1)$$

$$\psi(t) = \sum_n g[n]\phi(2t - n). \quad (2)$$

Conversely, a scaling function  $\phi_{j,k}(t)$  or wavelet function  $\psi_{j,k}(t)$  that is discretized at scale  $j$  and translation  $k$  can be obtained from the original (prototype) function  $\phi(t) = \phi_{0,0}(t)$  or  $\psi(t) = \psi_{0,0}(t)$  by

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k), \quad (3)$$

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k). \quad (4)$$

Different scale and translations of these functions allow us to obtain different frequency and time localizations of the signal. The coefficients (weights)  $h[n]$  and  $g[n]$  that satisfy (1) and (2) constitute the impulse responses of the lowpass and highpass filters used in the wavelet analysis, and define the type of the wavelet used in the analysis. Decomposition of the signal into

different frequency bands is therefore accomplished by successive highpass and lowpass filtering of the time domain signal. The original time domain signal  $x(t)$  sampled at 256 samples/s formed the discrete time signal  $x[n]$ , which is first passed through a halfband highpass filter  $g[n]$ , and a lowpass filter  $h[n]$ . In terms of normalized angular frequency, the highest frequency in the original signal is  $\pi$ , corresponding to the linear frequency of 128 Hz. According to Nyquist's rule, half the samples can be removed after the filtering, since the bandwidth of the signal is reduced to  $\pi/2$  radians upon filtering. This is accomplished by downsampling with a factor of 2. Filtering followed by subsampling constitutes one level of decomposition, and it can be expressed as follows:

$$d_1[k] = y_{\text{high}}[k] = \sum_n x[n] \cdot g[2k - n], \quad (5)$$

$$a_1[k] = y_{\text{low}}[k] = \sum_n x[n] \cdot h[2k - n], \quad (6)$$

where  $y_{\text{high}}[k]$  and  $y_{\text{low}}[k]$  are the outputs of the highpass and lowpass filters, respectively, after the subsampling. The output of the highpass filter,  $y_{\text{high}}[k]$ , represents *level 1 DWT coefficients*, also called  $d_1$ : *level 1 detail coefficients*. The output of the lowpass filter,  $a_1$ : *the level 1 approximation coefficients*, is further decomposed by passing  $y_{\text{low}}[k]$  through another set of highpass and lowpass filters to obtain level 2 detail coefficients  $d_2$  and level 2 approximation coefficients  $a_2$ , respectively.

This procedure, called subband coding, is repeated for further decomposition as many times desired, or until no more subsampling is possible. At each level, the procedure results in half the time resolution (due to subsampling) and double the frequency resolution (due to filtering), allowing the signal to be analyzed at different frequency ranges with different resolutions. Fig. 3 illustrates this procedure, where the frequency range analyzed with each set of coefficients is marked with “F”. The length of each set of coefficients is also provided, which depends on the specific wavelet used in the analysis. The numbers given in Fig. 3 are for Daubechies wavelets with four vanishing moments, whose corresponding filters  $h[n]$  and  $g[n]$  are of length  $2 \times 4 = 8$ . For example, starting with 257-long signal, the output of each level 1 filter is  $257 + 8 - 1 = 265$  points long, which reduces to 132 after subsampling. A wraparound or truncation can also be used to keep the number of coefficients exactly half of that of the previous level. The total number of all coefficients then adds up to the original signal length ( $\pm 1$  for odd length signals).

An approximation signal  $A_j(t)$  and a detail signal  $D_j(t)$  can be reconstructed from level  $j$  coefficients:

$$A_j(t) = \sum_k a_j[k] \cdot \phi_{j,k}(t), \quad (7)$$

$$D_j(t) = \sum_k d_j[k] \cdot \psi_{j,k}(t). \quad (8)$$

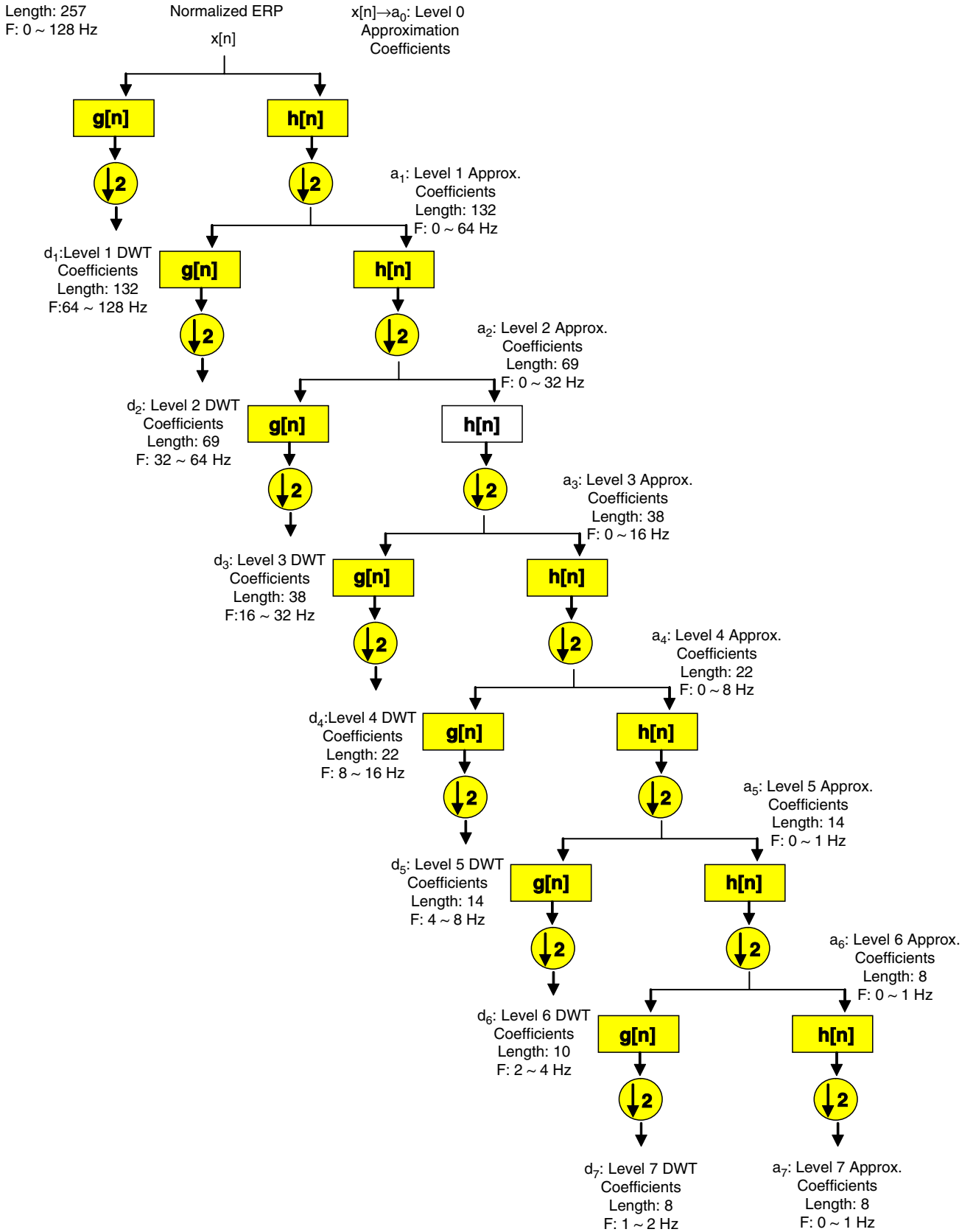


Fig. 3. Seven-level DWT decomposition.

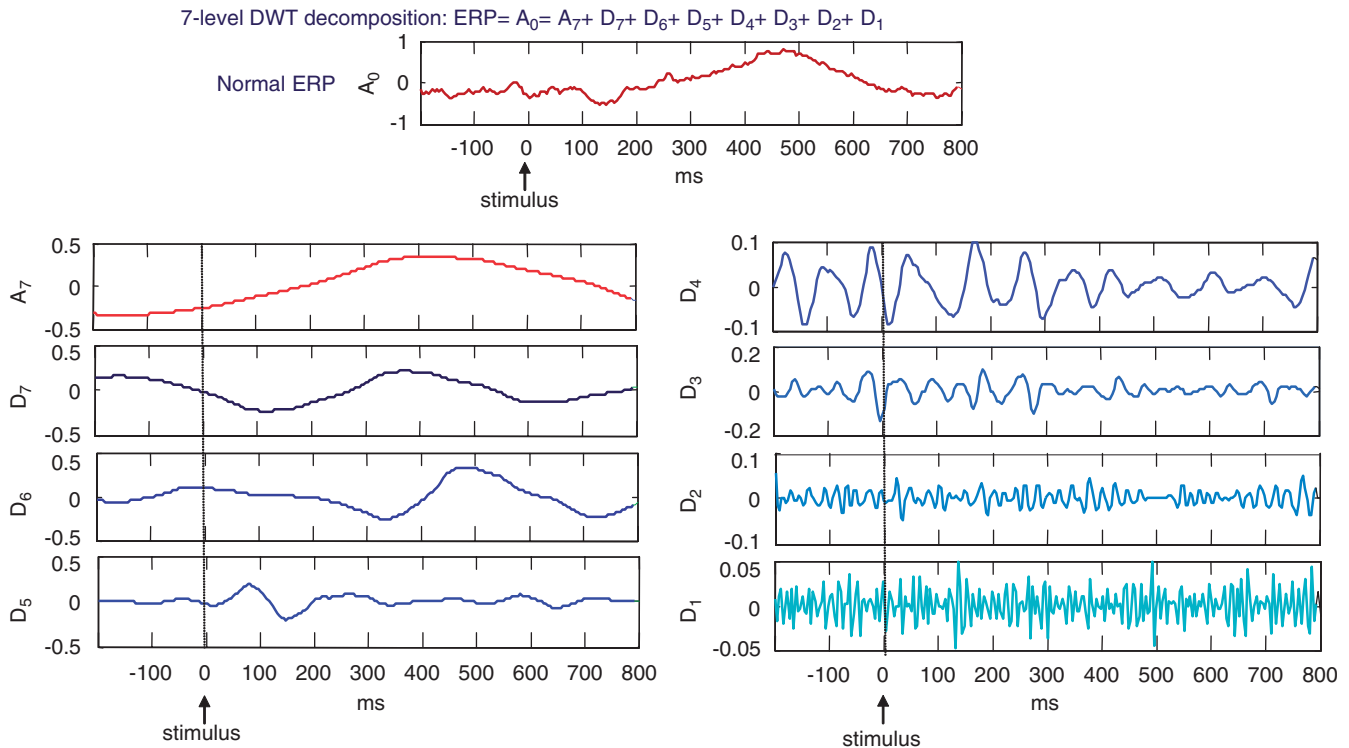


Fig. 4. Reconstructed detail and approximation signals of a cognitively normal person.

The original signal  $x(t)$  can then be reconstructed from the approximation signal  $A_j(t)$  at any level  $j$  and the sum of all detail signals up to and including level  $j$ :

$$\begin{aligned}
 x(t) &= A_j(t) + \sum_{j=-\infty}^j D_j(t) \\
 &= \sum_k a_j[k] \cdot \phi_{j,k}(t) + \sum_{j=-\infty}^j \sum_k d_j[k] \cdot \psi_{j,k}(t). \quad (9)
 \end{aligned}$$

Figs. 4 and 5 illustrate the reconstructed signals obtained at each level of a seven-level decomposition, for the analysis of a cognitively normal and probable-AD patient, respectively. Daubechies wavelets with four vanishing moments were used for these analyses. Whereas signal reconstruction is not required as part of this work, the reconstructed signals in Figs. 4 and 5 indicate that majority of the signals' energy lies in  $D_4$ – $D_7$  and  $A_7$ . The results presented in Section 4 will later confirm this observation.

In this study, we compared four different types of wavelets, including the Daubechies wavelets with four (Db4) and eight (Db8) vanishing moments, symlets with five vanishing moments (Sym5) and the quadratic B-spline wavelets (Qbs). The quadratic B-spline wavelet was chosen due to its reported suitability in analyzing ERP data in several studies [34,35,54–56]. Db4 and Db8 were chosen for their simplicity and general purpose applicability in a variety of time–frequency representation problems, whereas Sym5 was chosen due to its similarity to Daubechies wavelets with additional near symmetry property.

### 3.2. An ensemble of classifiers-based classification

One of the novelties of this work is the investigation of an ensemble of classifiers-based approach for the classification of ERP signals. An ensemble-based system, also known as a multiple classifier system (MCS), combines several, preferably diverse, classifiers. The diversity in the classifiers is typically achieved by using a different training data set for each classifier, which then allows each classifier to generate different decision boundaries. The expectation is that each classifier will make a different error, and strategically combining these classifiers can reduce the total error. Numerous studies have shown that such an approach can often outperform a single classifier system, is usually resistant to overfitting problems, and can often provide more stable results. Since its humble beginnings with such seminal works including, but not limited to [57–65], research in MCSs have expanded rapidly, and become an important research topic [66–68]. A sample of the immense literature on classifier combination can be found in Kuncheva's [66] recent book, the first text devoted to theory and implementation of ensemble-based classifiers, and references therein. The field has been developing so rapidly that an international workshop on MCS has recently been established, and the most current developments can be found in its proceedings [69].

The ensemble classification algorithm of choice for this study was Learn<sup>++</sup>, originally developed for efficient learning of novel information [70,71]. Inspired in part by AdaBoost [64], Learn<sup>++</sup> generates an ensemble of diverse classifiers, where

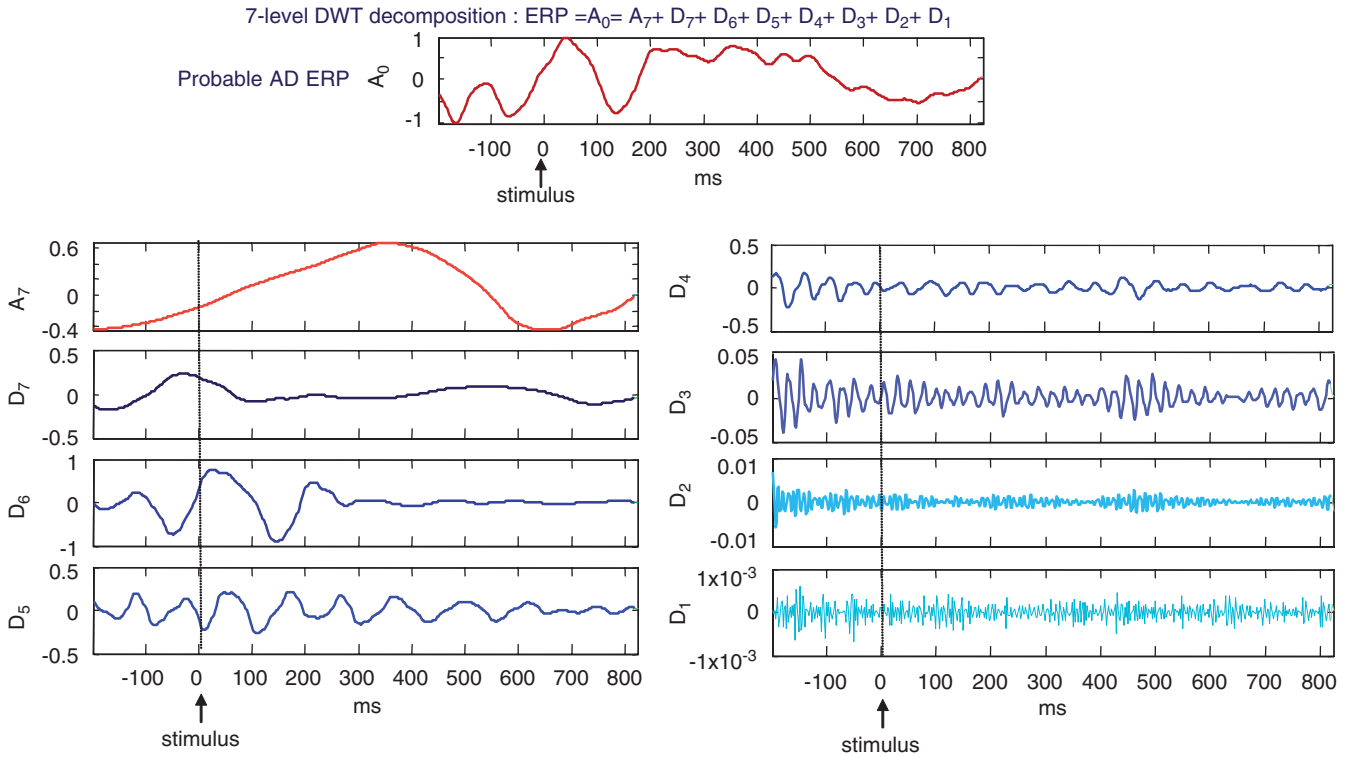


Fig. 5. Reconstructed detail and approximation signals of a probable-AD patient.

each classifier is trained on a strategically updated distribution of the training data that focuses on instances previously not seen or learned. The inputs to Learn<sup>++</sup> algorithm are (i) the training data  $S$  comprised of  $m$  instances  $x_i$  along with their correct labels  $y_i \in \Omega = \{\omega_1, \dots, \omega_C\}$ ,  $i = 1, 2, \dots, m$ , for  $C$  number of classes; (ii) a supervised classification algorithm *BaseClassifier*, generating individual classifiers (henceforth, hypotheses); and (iii) an integer  $T$ , the number of classifiers to be generated. The pseudocode of the algorithm and its block diagram are provided in Figs. 6 and 7, respectively, and described below in detail.

The BaseClassifier can be any supervised classifier, whose *instability* can be adjusted to ensure adequate diversity, so that sufficiently different decision boundaries can be generated each time the classifier is trained on a different training data set. This instability can be controlled by adjusting training parameters, such as the size or error goal of a neural network, with respect to the complexity of the problem. However, a meaningful minimum performance is enforced: the probability of any classifier to produce the correct labels on a given training data set, weighted proportionally to individual instances' probability of appearance, must be at least  $\frac{1}{2}$ . If classifiers' outputs are class-conditionally independent, then the overall error monotonically decreases as new classifiers are added. Originally known as the Condorcet jury theorem (1786) [72–74], this condition is necessary and sufficient for a two-class problem ( $C = 2$ ), and it is sufficient, but not necessary, for  $C > 2$ .

An iterative process sequentially generates each classifier of the ensemble: during the  $t$ th iteration, Learn<sup>++</sup> trains the Base-

**Input:**

- Training data  $S = \{(x_i, y_i)\}$ ,  $i = 1, \dots, m$  with labels  $y_i \in \Omega = \{\omega_1, \dots, \omega_C\}$ .
- Supervised algorithm BaseClassifier.
- Integer  $T$  specifying the number of classifiers.

Initialize  $w_1(i) = D_1(i) = 1/m$ ,  $\forall i, i = 1, 2, \dots, m$

**Do for**  $t = 1, 2, \dots, T$ :

1. Set  $D_t = w_t / \sum_{i=1}^m w_t(i)$  so that  $D_t$  is a distribution.
2. Draw training  $TR_t$  subset from  $D_t$ .
3. Obtain  $h_t$  by training with data  $TR_t$ .
4. Calculate the error of  $h_t$ ,  $\epsilon_t = \sum_{i: h_t(x_i) \neq y_i} D_t(i)$  on  $S$

If  $\epsilon_t > 1/2$ , discard  $h_t$  and **Go to** Step 2.

5. Set  $\beta_t = \epsilon_t / (1 - \epsilon_t)$ . Obtain the composite hypothesis through weighted majority voting
 
$$H_t = \arg \max_{y \in \Omega} \sum_{i: h_t(x_i) = y} \log(1/\beta_t)$$
6. Compute error of  $H_t$ :  $E_t = \sum_{i: H_t(x_i) \neq y_i} D_t(i)$
7. Set  $B_t = E_t / (1 - E_t)$ , and update the weights:
 
$$w_{t+1}(i) = w_t(i) \times \begin{cases} B_t, & \text{if } H_t(x_i) = y_i \\ 1, & \text{otherwise} \end{cases}$$

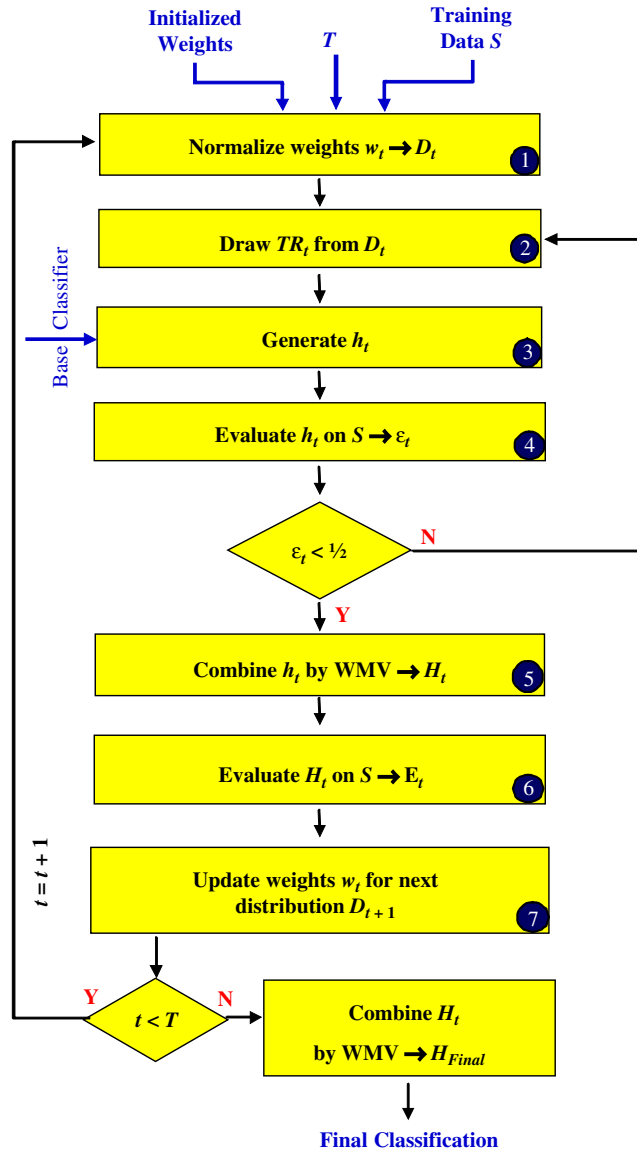
**Output** the final hypothesis:

$$H_{Final}(x) = \arg \max_{y \in \Omega} \sum_{t: h_t(x) = y} \log\left(\frac{1}{\beta_t}\right)$$

Fig. 6. Learn<sup>++</sup> pseudocode.

Classifier on a judiciously selected subset  $TR_t$  (about  $\frac{2}{3}$ ) of the current training data to generate hypothesis  $h_t$ . The training subset  $TR_t$  is drawn from the training data according to a dis-



Fig. 7. Learn<sup>++</sup> block diagram.

tribution  $D_t$ , which is obtained by normalizing a set of weights  $w_t$  maintained on the entire training data  $S$ . The distribution  $D_t$  determines which instances of the training data are more likely to be selected into the training subset  $TR_t$ . Unless a priori information indicates otherwise, this distribution is initially set to be uniform, giving equal probability to each instance to be selected into  $TR_1$ . At each subsequent iteration loop  $t$ , the weights previously adjusted at iteration  $t - 1$  are normalized (in step 1 of the loop),

$$D_t = w_t / \sum_{i=1}^m w_t(i) \quad (10)$$

to ensure a proper distribution. Training subset  $TR_t$  is drawn according to  $D_t$  (step 2), and the BaseClassifier is trained on  $TR_t$  (step 3). A hypothesis  $h_t$  is generated by the  $t$ th classifier, whose error  $\epsilon_t$  is computed on the current data set  $S$  as the

sum of the distribution weights of the misclassified instances (step 4),

$$\epsilon_t = \sum_{i:h_t(\mathbf{x}_i) \neq y_i} D_t(i) = \sum_{i=1}^m D_t(i) [h_t(\mathbf{x}_i) \neq y_i], \quad (11)$$

where  $[ \cdot ]$  evaluates to 1, if the predicate holds true, and 0 otherwise. As mentioned above, we insist that  $\epsilon_t$  be less than  $\frac{1}{2}$ . If this is the case, the hypothesis  $h_t$  is accepted, and its error is normalized to obtain

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}, \quad 0 < \beta_t < 1. \quad (12)$$

If  $\epsilon_t > \frac{1}{2}$ , the current hypothesis is discarded, and a new training subset is selected by returning to step 2. All hypotheses generated thus far are then combined using weighted majority voting to obtain the *composite hypothesis*  $H_t$  (step 5), for which each hypothesis  $h_t$  is assigned a weight inversely proportional to its normalized error. Those hypotheses with smaller training error are awarded a higher voting weight, and thus have more say in the decision of  $H_t$ , which then represents the current ensemble decision:

$$H_t = \arg \max_{y \in \Omega} \sum_{t:h_t(\mathbf{x})=y} \log(1/\beta_t). \quad (13)$$

It can be shown that the weight selection of  $\log(1/\beta_t)$  is optimum for weighted majority voting [66]. The error of the composite hypothesis  $H_t$  is computed as the sum of the distribution weights of the instances that are misclassified by the ensemble decision  $H_t$  (step 6),

$$E_t = \sum_{i:H_t(\mathbf{x}_i) \neq y_i} D_t(i) = \sum_{i=1}^m D_t(i) [H_t(\mathbf{x}_i) \neq y_i]. \quad (14)$$

Since individual hypotheses that make up the composite hypothesis all have individual errors less than  $\frac{1}{2}$ , so too will the composite error, i.e.,  $0 \leq E_t < \frac{1}{2}$ . We normalize the composite error  $E_t$  to obtain

$$B_t = \frac{E_t}{1 - E_t}, \quad 0 < B_t < 1, \quad (15)$$

which is then used for updating the distribution weights assigned to individual instances,

$$\begin{aligned} w_{t+1}(i) &= w_t(i) \times B_t^{1 - [H_t(\mathbf{x}_i) \neq y_i]} \\ &= w_t(i) \times \begin{cases} B_t & \text{if } H_t(\mathbf{x}_i) = y_i, \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \quad (16)$$

Eq. (16) indicates that the distribution weights of the instances correctly classified by the composite hypothesis  $H_t$  are reduced by a factor of  $B_t$ . Effectively, this increases the weights of the misclassified instances, making them more likely to be selected into the training subset of the next iteration. Readers familiar with the AdaBoost algorithm have undoubtedly noticed the overall similarities, but also the key difference between the two algorithms: the weight update rule of Learn<sup>++</sup> specifically targets learning novel information from data by

focusing on those instances that are not yet learned by the ensemble, whereas AdaBoost focuses on instances that have been misclassified by the previous classifier. This is because, the weight distribution of AdaBoost is updated based on the decision of a single previously generated *hypothesis*  $h_t$  [64], whereas Learn<sup>++</sup> updates its distribution based on the decision of the current *ensemble* through the use of the composite hypothesis  $H_t$ . This procedure forces Learn<sup>++</sup> to focus on instances that have not been properly learned by the *ensemble*. It can be argued that AdaBoost also looks, albeit indirectly, at the ensemble decision since, while based on a single hypothesis, the distribution update is cumulative. However, the update in Learn<sup>++</sup> is directly tied to the ensemble decision, and hence been found to be more efficient in learning new information in our previous trials on benchmark data sets.

The final hypothesis is obtained by combining all hypotheses that have been generated thus far:

$$H_{final}(\mathbf{x}) = \arg \max_{y \in \Omega} \sum_{t: h_t(\mathbf{x})=y} \log \left( \frac{1}{\beta_t} \right), \quad t = 1, \dots, T. \quad (17)$$

For any given data instance  $\mathbf{x}$ ,  $H_{final}$  chooses the label  $y \in \Omega : \{\omega_1, \dots, \omega_C\}$  that receives the largest total vote from all classifiers  $h_t$ , where the vote of  $h_t$  is weighted by its normalized performance  $\log(1/\beta_t)$ .

### 3.3. Implementation details

As mentioned in the Introduction, several factors set this study apart from previous efforts: a large cohort; analysis using three different electrodes  $P_Z$ ,  $F_Z$  and  $C_Z$ , instead of just the standard  $P_Z$ , and two different stimuli target and novel tones; analysis with four different wavelets and several different frequency bands; analysis with an ensemble of classifiers approach, as well as a single classifier; and the effect of the above-mentioned variables on commonly used medical diagnostic quantities, including sensitivity, specificity and PPV, instead of just overall generalization performance (OGP).

#### 3.3.1. Feature extraction

For each patient, six sets of ERPs were extracted and averaged: responses to novel and target tones, from each of the  $P_Z$ ,  $C_Z$  and  $F_Z$  electrodes. All averaged ERPs were decomposed into seven levels using one of the four types of wavelets: Daubechies with four and eight vanishing moments (Db4, Db8), symlets with five vanishing moments (Sym5) and quadratic B-splines (Qbs). Of the eight frequency bands created by the decomposition, the following bands were used for further analysis: approximation at 0–1 Hz, and details at 1–2, 2–4, 4–8 and 8–16 Hz. Detail coefficients at 16–32, 32–64 and 64–128 Hz were not considered for analysis, as the ERPs are known not to include any relevant frequency components in these intervals. In fact, the P300 is known to reside in 0–4 Hz interval, primarily around 3 Hz.

The number of coefficients created at each level depends on the analysis wavelet, more specifically the number of its filter tabs. Since each recording started 200 ms before the stimulus

and lasted for exactly 1 s, the middle coefficients were extracted in each case to remove those DWT coefficients corresponding to prestimulus baseline as well as large latency poststimulus baseline. The extracted coefficients corresponded to approximately 50–600 ms duration after the stimulus.

#### 3.3.2. Classification

We have first tried a single classifier system, using the multilayer perceptron as the base model. We have experimented with several architectural parameters, such as the number of hidden layer nodes in the 5–50 range, and the error goal in the 0.005–0.1 range. As a result of thousands of independent trials, an MLP architecture of 10 hidden layer nodes and 0.01 error goal was decided as the common architecture for all experiments. We have then implemented a Learn<sup>++</sup>-based ensemble system, where we have tried several different numbers of classifiers in the ensemble, from 3 to 25. In general, a five-classifier ensemble provided good results. While Learn<sup>++</sup> is independent of the base classifier, and can work with any supervised classification algorithm, MLP with the same architecture mentioned above was chosen as the base classifier for a fair comparison.

#### 3.3.3. Validation process: leave-one-out

In all cases listed below, generalization performance was obtained through leave-one-out cross-validation. According to this procedure, a classifier is trained on all but one of the available training data instances, and tested on the remaining instance. Its performance on this instance, 0% or 100%, is noted. The classifier is then discarded and a new one—with identical architecture—is trained again on all but one training data instance, this time leaving a different data instance out. Assuming that there are  $m$  training data points, the entire training and testing procedure is repeated  $m$  times, leaving a different instance as a test instance in each case. The mean of  $m$  individual performances is then accepted as the estimate of the performance of the system.

The leave-one-out process is considered as the most conservative—and, of course, computationally most costly—estimate of the true performance of the system, as it removes the bias of choosing particularly easy or difficult instances into training or test data. Due to the delicate nature of the application, and in order to obtain a reliable estimate of the true performance of this approach, we decided to use the leave-one-out procedure (instead of two-way splitting of the data into training and test data sets, or a  $k$ -fold cross-validation) despite its computational complexity. In order to further confirm the validity of the results, all leave-one-out validations were repeated three times.

#### 3.3.4. Diagnostic performance figures

While generalization performance is the traditional figure of merit in evaluating machine learning algorithms, more descriptive quantities are often used to evaluate medical tests and procedures. Sensitivity, specificity, PPV and negative predictive value (NPV) are four such quantities commonly used in medical diagnostics. Table 1 summarizes the concepts defined below.

Table 1  
Category labels for defining diagnostic quantities

Number of patients	True condition	
	Probable AD	Cognitively normal
<i>Classification decision</i>		
Probable AD	A	B
Cognitively normal	C	D

*OGP*: In pattern recognition, this is the average leave-one-out validation performance of the classifiers, or average generalization performance on test data. OGP represents the average probability of correct decision. Within the medical community, OGP is also known as the accuracy of the test: the ratio of patients the classification system is expected to correctly identify.

*Sensitivity*: Formally defined as the probability of a positive diagnosis given that the patient does in fact have the condition, sensitivity is the ability of a medical test to correctly identify the target group. In the context of this application, sensitivity is the proportion of true AD patients correctly identified as AD patients by the classification system.

*Specificity*: Formally defined as the probability of a negative diagnosis given that the patient does not have the disease, specificity is the ability of a test to correctly identify the control group. In this study, specificity is the proportion of cognitively normal patients, correctly identified as normal.

*PPV*: PPV is defined as the probability that the patient has the disease, given that the test result is positive. It is calculated as the proportion of the sample population that is correctly identified by the test as the target group, among all those identified as target, correctly or otherwise. In the context of this study, PPV is the proportion of those patients identified as AD patients by the classifier, who actually have AD.

*NPV*: Not used as commonly, NPV is the probability that the patient does not have the disease, given that the test result is negative. It is calculated as the proportion of the sample population that is correctly identified by the test to belong to control group, among all those identified as target, correctly or otherwise. In the context of this study, NPV is the proportion of those patients identified as normal by the classifier, who are in fact cognitively normal.

In Table 1, *A* is the number of patients classified as AD, who are in fact diagnosed as probable AD, by the clinical evaluation, *B* is the number of patients, also classified as AD (albeit incorrectly), who are in fact cognitively normal, *C* is the number of patients classified as normal (again, incorrectly), who were originally diagnosed as probable AD, and *D* is the number of patients who are (correctly) classified as cognitively normal, and are in fact clinically determined to be cognitively normal.  $A + B + C + D$  is the total number of patients. Then,

$$\text{Overall performance} = \frac{A + D}{A + B + C + D}, \quad (18)$$

$$\text{Sensitivity} = \frac{A}{A + C}, \quad (19)$$

$$\text{Specificity} = \frac{D}{B + D}, \quad (20)$$

$$\text{PPV} = \frac{A}{A + B}, \quad (21)$$

$$\text{NPV} = \frac{D}{C + D}. \quad (22)$$

#### 4. Results

As mentioned above, six sets of ERPs were obtained from each patient (three electrodes, two types of stimulus), analyzed at five levels of frequency bands (0–1, 1–2, 2–4, 4–8 and 8–16 Hz, constituting individual feature sets) for each set of ERPs, using each of four types of wavelets. Diagnostic classification performances, along with sensitivity, specificity, PPVs and NPVs were obtained for each of the above-mentioned combinations. Furthermore, considering that ERPs are known to occupy primarily the 0–4 Hz range, we have also included this frequency range as the sixth feature set, obtained by concatenating the first three sets of coefficients.

Presenting the results for every combination would be impractical, and unnecessarily lengthen this paper. Summary results are therefore provided here. Specifically, the results corresponding to Daubechies four wavelets are provided in most detail, including the performance of each frequency band for each of the six sets of ERPs. Db4 was chosen due to its common use in broad range applications, including analysis of biological signals. Then, the performance figures for each set of ERP using each of the four wavelets are provided, but only for the highest performing frequency band. For all cases, we provide the overall classification performance obtained by a single classifier, as well as an ensemble of five classifiers. Both single and ensemble performances are averages of three independent 52-fold leave-one-out trials, whereas the *best ensemble* is the best performing leave-one-out trial out of the three. As we discuss below, the ensembles performed, on average, better than individual classifiers. The sensitivity, specificity and PPVs are therefore provided for ensemble performances.

Tables 2–4 summarize the performance figures obtained when responses to target tones were processed with Db4 wavelet, for each of the three electrodes. The best performing spectral band for all electrodes was the 2–4 Hz range (corresponding to level  $D_6$  in Figs. 4 and 5), indicated in boldface in Tables 2–4. Of the three electrodes, the best performance across all categories was obtained with the  $C_z$  electrode with an average ensemble performance of 72.4%, best ensemble performance of 75%, with sensitivity, specificity, PPV and NPV values of 68.6%, 69.2%, 72.5% and 65.3%, respectively.

Tables 5–7 summarize the classification performances obtained when responses to novel tones were processed with Db4 wavelet, for each of the three electrodes. The performance obtained with novel tones were significantly better, particularly for the  $P_z$  electrode than the performances obtained with the target tones. This is perhaps one of the most surprising outcomes of this study, as novel tones were not originally intended to be used for AD versus normal discrimination, but rather to

Table 2  
Spectral-specific performances obtained from  $C_Z$  electrode—target response (Db4)

Target $C_Z$	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
0–1 Hz	55.7	57.7	61.5	48.6	63.3	61.6	51.2
1–2 Hz	50.0	48.7	51.9	47.9	47.5	51.5	43.8
2–4 Hz	<b>62.8</b>	<b>72.4</b>	<b>75.0</b>	<b>68.6</b>	<b>69.2</b>	<b>72.5</b>	<b>65.3</b>
4–8 Hz	56.4	54.5	57.7	51.4	47.5	53.9	45.0
8–16 Hz	53.2	55.1	61.5	57.1	49.2	56.7	49.7
0–4 Hz	55.8	57.1	57.7	51.4	58.3	59.5	50.5

Table 3  
Spectral-specific performances obtained from  $P_Z$  electrode—target response (Db4)

Target $P_Z$	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
0–1 Hz	54.5	53.8	55.8	46.4	57.5	55.9	48.1
1–2 Hz	55.7	62.2	65.4	59.3	60.8	63.7	56.4
2–4 Hz	<b>60.9</b>	<b>66.0</b>	<b>67.3</b>	<b>62.1</b>	<b>65.0</b>	<b>67.6</b>	<b>59.5</b>
4–8 Hz	49.4	53.9	55.7	51.4	50.8	55.2	47.0
8–16 Hz	53.2	58.3	59.6	53.6	60.8	61.5	52.9
0–4 Hz	58.3	60.9	65.4	56.4	62.5	63.7	55.2

Table 4  
Spectral-specific performances obtained from  $F_Z$  electrode—target response (Db4)

Target $F_Z$	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
0–1 Hz	58.3	59.0	63.5	49.3	64.2	61.7	52.0
1–2 Hz	59.0	55.8	61.5	50.0	51.7	54.3	47.4
2–4 Hz	<b>59.0</b>	<b>64.7</b>	<b>65.4</b>	<b>62.9</b>	<b>61.7</b>	<b>66.0</b>	<b>58.7</b>
4–8 Hz	60.3	62.8	67.3	54.3	63.3	63.6	54.2
8–16 Hz	57.7	58.3	59.6	50.7	61.7	60.7	51.8
0–4 Hz	55.1	54.5	57.7	52.1	48.3	53.7	46.8

Table 5  
Spectral-specific performances obtained from  $C_Z$  electrode—novel response (Db4)

Novel $C_Z$	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
0–1 Hz	53.8	54.5	55.8	49.3	56.7	67.0	49.0
1–2 Hz	51.9	51.3	57.7	49.3	48.3	53.0	44.4
2–4 Hz	56.4	54.5	55.8	45.7	62.5	58.7	49.7
4–8 Hz	<b>62.8</b>	<b>64.1</b>	<b>65.4</b>	<b>57.9</b>	<b>68.3</b>	<b>68.4</b>	<b>58.1</b>
8–16 Hz	55.8	57.7	57.7	50.0	62.5	61.5	51.6
0–4 Hz	60.9	57.7	59.6	48.6	64.2	61.3	51.7

Table 6  
Spectral-specific performances obtained from  $P_Z$  electrode—novel response (Db4)

Novel $P_Z$	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
0–1 Hz	61.5	66.0	67.3	57.8	73.3	71.7	59.9
1–2 Hz	<b>75.0</b>	<b>78.2</b>	<b>80.8</b>	<b>67.1</b>	<b>78.3</b>	<b>78.8</b>	<b>67.0</b>
2–4 Hz	63.5	66.0	69.2	54.3	72.5	69.9	57.7
4–8 Hz	62.8	65.4	69.2	60.0	61.7	65.2	56.6
8–16 Hz	63.5	67.3	71.2	52.1	78.3	73.9	58.4
0–4 Hz	70.5	73.7	75.0	65.0	80.8	79.9	66.5

Table 7  
Spectral-specific performances obtained from  $F_Z$  electrode—novel response (Db4)

Novel $F_Z$	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
0–1 Hz	51.9	50.6	51.9	42.8	55.0	52.3	45.4
1–2 Hz	52.6	51.9	55.8	47.8	50.8	53.2	45.5
2–4 Hz	<b>54.5</b>	<b>58.3</b>	<b>59.6</b>	<b>47.1</b>	<b>66.7</b>	<b>52.4</b>	<b>51.9</b>
4–8 Hz	50.6	56.4	57.7	47.1	57.5	56.4	48.3
8–16 Hz	53.8	57.7	59.6	53.6	59.2	60.4	52.5
0–4 Hz	51.9	49.4	50.0	37.1	54.2	48.1	42.6

Table 8  
Best spectral performances obtained using Db4 wavelet

Db4	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
TC 2–4 Hz	62.8	72.4	75.0	68.6	69.2	72.5	65.3
TP 2–4 Hz	60.9	66.0	67.3	62.1	65.0	67.6	59.5
TF 2–4 Hz	59.0	64.7	65.4	62.9	61.7	66.0	58.7
NC 4–8 Hz	62.8	64.1	65.4	57.9	68.3	68.4	58.1
<b>NP 1–2 Hz</b>	<b>75.0</b>	<b>78.2</b>	<b>80.8</b>	<b>67.1</b>	<b>78.3</b>	<b>78.8</b>	<b>67.0</b>
NF 2–4 Hz	54.5	58.3	59.6	47.1	66.7	52.4	51.9

Table 9  
Best spectral performances obtained using Db8 wavelet

Db8	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
TC 2–4 Hz	50.6	53.8	59.6	48.6	51.7	54.1	46.2
TP 2–4 Hz	52.6	55.8	57.7	51.4	56.7	58.1	50.0
TF 2–4 Hz	46.8	53.8	53.8	48.6	55.8	56.3	48.2
NC 4–8 Hz	59.0	64.1	67.3	57.1	68.3	57.9	57.7
<b>NP 1–2 Hz</b>	<b>69.2</b>	<b>73.7</b>	<b>75.0</b>	<b>66.4</b>	<b>77.5</b>	<b>77.5</b>	<b>66.5</b>
NF 2–4 Hz	53.2	50.6	51.9	42.9	53.3	51.7	44.4

Table 10  
Best spectral performances obtained using Sym5 wavelet

Sym5	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
TC 2–4 Hz	53.8	55.1	57.7	51.4	55.8	57.8	49.5
TP 2–4 Hz	51.3	46.8	50.0	47.1	41.7	48.6	40.1
TF 2–4 Hz	55.8	62.2	67.3	60.7	57.5	62.6	55.6
NC 4–8 Hz	64.1	62.8	69.2	58.6	60.0	63.6	55.0
<b>NP 1–2 Hz</b>	<b>67.9</b>	<b>79.5</b>	<b>84.6</b>	<b>73.5</b>	<b>79.2</b>	<b>80.4</b>	<b>72.1</b>
NF 2–4 Hz	54.5	51.9	53.8	47.9	52.5	53.9	46.4

improve the robustness of the P300 component generated in response to the target tones. Furthermore, the spectral band that provides the best performance is 1–2 Hz, as opposed to 2–4 Hz that performed well with target tones. With the  $P_Z$  electrode, ERPs in response to novel tones decomposed with Db4 wavelet obtained overall generalization performances of 75%, 78.2% and 80.8% using averaged single classifier, average ensemble classifier and best ensemble classifier, respectively. The sensitivity was 67.1%, specificity 78.3% and PPV was 78.8%.

Also interesting to note that 0–4 Hz that includes both the 1–2 and 2–4 Hz coefficients did not perform as well, confirming that the information provided by 0–1 Hz coefficients is not only non-discriminatory on their own right, but their existence has a deteriorating effect.

The same set of experiments was repeated with three additional wavelets, Db8, Sym5 and quadratic B-splines. The results for all four wavelets are summarized in Tables 8–11, where performances for only the best performing spectral bands are

Table 11  
Best spectral performances obtained using Qbs wavelet

Qbs	Single classifier OGP	Ensemble OGP	Best ensemble	Ensemble sensitivity	Ensemble specificity	Ensemble PPV	Ensemble NPV
TC 2–4 Hz	50.0	52.6	53.8	47.9	50.8	52.8	45.7
TP 2–4 Hz	52.6	57.1	61.5	57.8	49.2	56.8	50.5
TF 2–4 Hz	56.4	55.8	55.8	55.7	49.2	56.0	49.0
NC 4–8 Hz	59.6	57.7	63.5	52.1	57.5	59.0	50.6
<b>NP 1–2 Hz</b>	<b>67.3</b>	<b>65.4</b>	<b>65.4</b>	<b>57.1</b>	<b>71.7</b>	<b>70.4</b>	<b>59.1</b>
NF 2–4 Hz	47.4	50.0	50.0	43.6	54.2	52.7	45.0

included. Each row in Table 8 is therefore the best row from the above six tables, and included here for completeness. Several interesting observations can be made from these results: first, the best performing frequency band was the same regardless of the wavelet used: 2–4 Hz for responses to target tones from all three electrodes, and for novel tones from  $F_Z$  electrode; 1–2 Hz for responses to novel tones from  $P_Z$  electrode and 4–8 Hz for responses to novel tones from  $C_Z$  electrode. This confirms the reasonable proposition that the choice of wavelet does not change the amount of information provided by each frequency band. Second, while the best performing frequency bands do not change, the actual performance figures do vary depending on the wavelet chosen. Tables 8–11 indicate that symlet wavelets perform better than all other wavelets, whereas quadratic B-splines provide the lowest performance. Furthermore, whereas the average single classifier performance at 67.9% is lower with symlets than that obtained with Db4 (75%), the average ensemble performance at 79.5% and best ensemble performance at 84.6% outperform the Db4 wavelet. Finally, the medical diagnostic figures are also higher with the symlets, providing 73.5% sensitivity, 79.2% specificity and 80.4% PPV.

## 5. Conclusions and discussions

The application presented in this work is concerned with automated early diagnosis of AD, using a non-invasive and cost-effective biomarker that can be measured in a community healthcare clinic setting. This problem is widely recognized as a particularly difficult one, not only for machine learning algorithms, but even for most neurophysiologists and neuropsychologists. The difficulty of the problem is exacerbated with our requirement to diagnose the disease at its earliest possible stage, during which the symptoms are often not much different than those that are associated with normal aging.

The proposed approach seeks a synergistic combination of some well-established techniques, such as ERP analysis using multiresolution wavelets, with more recent developments in machine learning, such as ensemble systems. Specifically, we analyze the discriminatory ability of ERPs obtained in response to novel tones, as well as commonly used target tones, acquired from three different electrodes. One of the most surprising outcomes of this study is the ability of novel tones to discriminate the ERPs of cognitively normal people from those with earliest form of AD. It was found that novel tones at 1–2 Hz, acquired from the  $P_Z$  electrode, provide the best performance, regard-

less of the wavelet used, though the best performances were obtained when these signals were decomposed and analyzed using the symlet wavelets with five vanishing moments. Daubechies wavelets with four vanishing moments closely followed Sym5. It should be noted that symlets have similar properties to that of Daubechies wavelets, and in fact they look similar; however, symlets are near symmetric, whereas Daubechies wavelets are not.

Ensemble performances were in general higher than single classifier performances, and sometimes by wide margins, demonstrating the usefulness of the ensemble approach. However, not all ensemble performances were better than those of individual classifiers, indicating that the ensemble classifiers must be constructed with care: if all classifiers that constitute the ensemble provide similar information, then there is nothing to be gained from using an ensemble approach. However, if the classifiers are negatively correlated, that is, they make errors on different instances, their combination can provide a performance boost.

Recall that a recent study estimates the community clinic-based physicians' diagnostic performances with 83% sensitivity, 53% specificity and 75% overall classification performance. 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. Furthermore, with sensitivity, specificity and PPVs also reaching 80% ranges, these results are particularly promising, and provide clinically useful outcomes.

A particularly interesting observation can also be made with sensitivity and specificity figures: at 73%, the sensitivity of the novel  $P_Z$  at 1–2 Hz is the only metric that is lower than that obtained by the community physicians (83%); however, the specificity of the approach at 79.2% is significantly better than the 55% obtained by the physicians. Therefore, while the approach is promising, and outperforms the community physicians in general diagnostic performance, it is particularly useful in discriminating cognitively normal individuals from early AD patients, as measured by specificity.

Overall, the results presented above are significant because an EEG-based automated classification system is non-invasive, objective, substantially more cost-effective than clinical

evaluations, and can be easily implemented at community clinics, where most patients get their first intervention.

Our current and future work include analyzing additional wavelets, as well as combining the discriminatory information provided by individual frequency bands in a data fusion setting using the ensemble of classifiers approach. We note that simply concatenating the coefficients from different frequency bands do not necessarily provide better performance, as demonstrated by the poor performance of the 0–4 Hz coefficients. However, combining individual ensemble of classifiers, each trained with signals at a particular frequency range, may prove to be more effective.

## 6. Summary

The rapidly growing proportion of elderly population, combined with lack of standard and effective diagnostic procedures that are available to community healthcare providers, makes early diagnosis of Alzheimer's disease (AD) a major public healthcare concern. Several signal processing-based approaches—some combined with automated classifiers—have been proposed for the analysis of EEG signals, which have resulted in varying degrees of success. To date, the final outcomes of these studies remain largely inconclusive primarily due to lack of adequate study cohort, as well as the inherent difficulty of the problem. This paper describes a new effort using multiresolution wavelet analysis on event-related potentials (ERPs) of the EEG to investigate whether EEG can be a reliable biomarker for AD.

Several factors set this study apart from similar prior efforts: (i) a larger cohort recruited through a strict inclusion/exclusion protocol, and diagnosed through a thorough and rigorous clinical evaluation process; (ii) data from three different electrodes and two different stimulus tones (target and novel) are analyzed; (iii) different mother wavelets have been employed in analysis of the signals; (iv) performances of six frequency bands (0–1, 1–2, 2–4, 4–8, 8–16 and 0–4 Hz) have been individually analyzed; (v) an ensemble of classifier-based decision is implemented and compared to a single classifier-based decision and, most importantly; (vi) the earliest possible diagnosis of the disease is targeted. Some expected and some interesting outcomes were observed, with respect to each parameter analyzed.

*Individual frequency bands:* In general, 1–2 and 2–4 Hz bands provide the most discriminatory information. This was expected as the spectral content of the primary ERP component of interest, the P300, is known to reside in these intervals.

*The choice of wavelet:* The best performing bands remained the same when the data was analyzed using different wavelets. However, the specific performances of frequency bands varied with the choice of the wavelet, Sym5 providing the best results.

*Choice of classifier:* In general, ensemble-based systems outperformed single classifier-based systems, sometimes with wide margins, indicating that such systems should be examined more carefully.

*Choice of electrode:* As expected, the  $P_z$  electrode provided the best performance, confirming results of previous efforts.

*Choice of stimulus:* Most surprisingly, the ERPs obtained in response to novel tones provided a better diagnostic performance than the traditionally used responses to target tones.

*Diagnostic performance:* Perhaps the most significant outcome of this study are the promising results obtained through the proposed approach. At around 80%, the overall performance of the proposed approach exceeded that of the trained community clinic physicians, and closely approached the gold standard performance of the university hospital-based clinical evaluation. While the algorithm did well on all diagnostic performance figures, such as sensitivity, specificity and PPV, its performance on specificity was particularly promising.

Considering the most challenging nature of diagnosing AD at its earliest stages, the results of this study justify the feasibility of this technique as a low-cost, objective, non-invasive approach that can be easily made available to the community clinics.

## Acknowledgments

This work is supported by National Institute on Aging of the National Institutes of Health under Grant number P30 AG10124-R01 AG022272, and by National Science Foundation under Grant number ECS-0239090.

## References

- [1] Alzheimer's Disease International, About Alzheimer's disease. Available at: (<http://www.alz.co.uk/alzheimers/faq.html#howmany>), (last accessed: 12-14-2005).
- [2] Alzheimer's Association, Basic facts and statistics. Available at: (<http://www.alz.org/Resources/TopicIndex/BasicFacts.asp>) (last accessed: 12-14-2005).
- [3] T. Sunderland, R.E. Gur, S.E. Arnold, The use of biomarkers in the elderly: current and future challenges, *Biol. Psychiatry* 58 (2005) 272–276.
- [4] M.J. de Leon, W. Klunk, Biomarkers for the early diagnosis of Alzheimer's disease, *Lancet Neurol.* 5 (2006) 198–199.
- [5] A. Lim, W. Kukull, D. Nochlin, J. Leverenz, W. McCormick, J. Bowen, L. Teri, J. Thompson, E. Peskind, M. Raskind, E. Larson, Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series, *J. Am. Geriatr. Soc.* 47 (1999) 564–569.
- [6] J. Polich, P300 and Alzheimer's disease, *Biomed. Pharmacother.* 43 (1999) 493–499.
- [7] D.S. Goodin, Clinical utility of long latency cognitive event-related potentials (P3): the pros, *Electroencephalogr. Clin. Neurophysiol.* 76 (1990) 2–5.
- [8] J.M. Ford, N. Askari, J.D.E. Gabrieli, D.H. Mathalon, J.R. Tinklenberg, V. Menon, J. Yesavage, Event-related brain potential evidence of spared knowledge in Alzheimer's disease, *Psychol. Aging* 16 (2001) 161–176.
- [9] J.M. Olichney, D.G. Hillert, Clinical applications of cognitive event-related potentials in Alzheimer's disease, *Phys. Med. Rehabil. Clin. North Am.* 15 (2004) 205–233.
- [10] D.J. Linden, The p300: where in the brain is it produced and what does it tell us?, *Neuroscientist* 11 (2005) 563–576.
- [11] S. Yamaguchi, H. Tsuchiya, S. Yamagata, G. Toyoda, S. Kobayashi, Event-related brain potentials in response to novel sounds in dementia, *Clin. Neurophysiol.* 111 (2000) 195–203.
- [12] J. Polich, C. Ladish, F.E. Bloom, P300 assessment of early Alzheimer's disease, *Electroencephalogr. Clin. Neurophysiol.* 77 (1990) 179–189.
- [13] Y. Mochizuki, M. Oishi, T. Takasu, Correlations between P300 components and regional cerebral blood flows, *J. Clin. Neurosci.* 8 (2001) 407–410.
- [14] B.A. Ardekani, S.J. Choi, G.A. Hossein-Zadeh, B. Porjesz, J.L. Tanabe, K.O. Lim, R. Bilder, J.A. Helpem, H. Begleiter, Functional magnetic

- resonance imaging of brain activity in the visual oddball task, *Cognit. Brain Res.* 14 (2002) 347–356.
- [15] E.J. Golob, J.K. Johnson, A. Starr, Auditory event-related potentials during target detection are abnormal in mild cognitive impairment, *Clin. Neurophysiol.* 113 (2002) 151–161.
- [16] N.A. Phillips, H. Chertkow, M.M. Leblanc, H. Pim, S. Murtha, Functional and anatomical memory indices in patients with or at risk for Alzheimer's disease, *J. Int. Neuropsychol. Soc.* 10 (2004) 200–210.
- [17] B.A. Ally, G.E. Jones, J.A. Cole, A.E. Budson, The P300 component in patients with Alzheimer's disease and their biological children, *Biol. Psychol.* 72 (2006) 180–187.
- [18] E.J. Golob, A. Starr, Effects of stimulus sequence on event-related potentials and reaction time during target detection in Alzheimer's disease, *Clin. Neurophysiol.* 111 (2000) 1438–1449.
- [19] C. Mulert, L. Jager, O. Pogarell, P. Bussfeld, R. Schmitt, G. Juckel, U. Hegerl, Simultaneous ERP and event-related fMRI: focus on the time course of brain activity in target detection, *Methods Find. Exp. Clin. Pharmacol.* 24 (Suppl. D) (2002) 17–20.
- [20] J. Benvenuto, Y. Jin, M. Casale, G. Lynch, R. Granger, Identification of diagnostic evoked response potential segments in Alzheimer's disease, *Exp. Neurol.* 176 (2002) 269–276.
- [21] C.D. Morgan, C. Murphy, Olfactory event-related potentials in Alzheimer's disease, *J. Int. Neuropsychol. Soc.* 8 (2002) 753–763.
- [22] R.M. Chapman, G.H. Nowlis, J.W. McCrary, J.A. Chapman, T.C. Sandoval, M.D. Guillily, M.N. Gardner, L.A. Reilly, Brain event-related potentials: diagnosing early-stage Alzheimer's disease, *Neurobiol. Aging*, 2006, in press, doi: 10.1016/j.neurobiolaging.2005.12.008.
- [23] M.J. Hogan, G.R. Swanwick, J. Kaiser, M. Rowan, B. Lawlor, Memory-related EEG power and coherence reductions in mild Alzheimer's disease, *Int. J. Psychophysiol.* 49 (2003) 147–163.
- [24] C. Huang, L. Wahlund, T. Dierks, P. Julin, B. Winblad, V. Jelic, Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study, *Clin. Neurophysiol.* 111 (2000) 1961–1967.
- [25] V. Knott, E. Mohr, C. Mahoney, V. Ilivitsky, Quantitative electroencephalography in Alzheimer's disease: comparison with a control group, population norms and mental status, *J. Psychiatry Neurosci.* 26 (2001) 106–116.
- [26] C. Besthorn, H. Sattel, F. Hentschel, S. Daniel, R. Zerfass, H. Forstl, Quantitative EEG in frontal lobe dementia, *J. Neural Transm. Suppl.* 47 (1996) 169–181.
- [27] S.Y. Cho, B.Y. Kim, E.H. Park, J.W. Kim, W.W. Whang, S.K. Han, H.Y. Kim, Automatic recognition of Alzheimer's disease with single channel EEG recording, a new beginning for human health, *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, September 17–21, 2003, pp. 2655–2658.
- [28] M. Lindau, V. Jelic, S.E. Johansson, C. Andersen, L.O. Wahlund, O. Almkvist, Quantitative EEG abnormalities and cognitive dysfunctions in frontotemporal dementia and Alzheimer's disease, *Dement. Geriatr. Cognit. Disord.* 15 (2003) 106–114.
- [29] T. Kai, Y. Asai, K. Sakuma, T. Koeda, K. Nakashima, Quantitative electroencephalogram analysis in dementia with Lewy bodies and Alzheimer's disease, *J. Neurol. Sci.* 237 (2005) 89–95.
- [30] D. Solo, R. Hornero, P. Espino, J. Poza, C.I. Nchez, R. de la Rosa, Analysis of regularity in the EEG background activity of Alzheimer's disease patients with approximate entropy, *Clin. Neurophysiol.* 116 (2005) 1826–1834.
- [31] K. Bennys, G. Rondouin, C. Vergnes, J. Touchon, Diagnostic value of quantitative EEG in Alzheimer's disease, *Neurophysiol. Clin./Clin. Neurophysiol.* 31 (2001) 153–160.
- [32] V. Kolev, T. Demiralp, J. Yordanova, A. Ademoglu, U. Alkac, Time–frequency analysis reveals multiple functional components during oddball P300, *Neuroreport* 8 (1997) 2061–2065.
- [33] A. Ademoglu, T. Demiralp, J. Yordanova, V. Kolev, M. Devrim, Decomposition of event-related brain potentials into multicomponents using wavelet transform, *Appl. Signal Process.* 5 (1998) 142–151.
- [34] E. Basar, M. Schurmann, T. Demiralp, C. Basar-Eroglu, A. Ademoglu, Event-related oscillations are 'real brain responses'—wavelet analysis and new strategies, *Int. J. Psychophysiol.* 39 (2001) 91–127.
- [35] T. Demiralp, A. Ademoglu, Y. Stefanopoulos, C. Basar-Eroglu, E. Basar, Wavelet analysis of oddball P300, *Int. J. Psychophysiol.* 39 (2001) 221–227.
- [36] T. Demiralp, A. Ademoglu, Decomposition of event-related brain potentials into multiple functional components using wavelet transform, *Clin. Electroencephalogr.* 32 (2001) 122–138.
- [37] R.Q. Quiroga, O.W. Sakowitz, E. Basar, M. Schurmann, Wavelet transform in the analysis of the frequency composition of evoked potentials, *Brain Res. Protocols* 8 (2001) 16–24.
- [38] R.Q. Quiroga, H. Garcia, Single-trial event-related potentials with wavelet denoising, *Clin. Neurophysiol.* 114 (2003) 376–390.
- [39] S. Aviyente, L.A.W. Brakel, R.K. Kushwaha, M. Snodgrass, H. Shevrin, W.J. Williams, Characterization of event related potentials using information theoretic distance measures, *IEEE Trans. Biomed. Eng.* 51 (2004) 737–743.
- [40] B.H. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho, K. Sundaresan, An exploratory study of factors affecting single trial P300 detection, *IEEE Trans. Biomed. Eng.* 51 (2004) 975–978.
- [41] M. Fatourechi, S.G. Mason, G.E. Birch, R.K. Ward, A wavelet-based approach for the extraction of event related potentials from EEG, in: *Proceedings—IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2004, pp. 737–740.
- [42] A.K. Ozdemir, S. Karakas, E.D. Cakmak, D.I. Tufekci, O. Arikan, Time–frequency component analyser and its application to brain oscillatory activity, *J. Neurosci. Methods* 145 (2005) 107–125.
- [43] E.M. Bernat, W.J. Williams, W.J. Gehring, Decomposing ERP time–frequency energy using PCA, *Clin. Neurophysiol.* 116 (2005) 1314–1334.
- [44] M. Morup, L.K. Hansen, C.S. Herrmann, J. Parnas, S.M. Arnfred, Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG, *NeuroImage* 29 (2006) 938–947.
- [45] M. Karrasch, M. Laine, J.O. Rinne, P. Rapinaja, E. Sinerva, C.M. Krause, Brain oscillatory responses to an auditory–verbal working memory task in mild cognitive impairment and Alzheimer's disease, *Int. J. Psychophysiol.* 59 (2006) 168–178.
- [46] A.A. Petrosian, D.V. Prokhorov, W. Lajara-Nanson, R.B. Schiffer, Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG, *Clin. Neurophysiol.* 112 (2001) 1378–1387.
- [47] S. Yagneswaran, M. Baker, A. Petrosian, Power frequency and wavelet characteristics in differentiating between normal and Alzheimer EEG, 2002 IEEE Engineering in Medicine and Biology 24th Annual Conference and the 2002 Fall Meeting of the Biomedical Engineering Society (BMES/EMBS), vol. 1, October 23–26, 2002, pp. 46–47.
- [48] R. Polikar, F. Keinert, M.H. Greer, Wavelet analysis of event related potentials for early diagnosis of Alzheimer's disease, in: A. Petrosian, F.G. Meyer (Eds.), *Wavelets in Signal and Image Analysis, From Theory to Practice*, Kluwer Academic Publishers, Boston, 2001.
- [49] G. Jacques, J. Frymiare, C. Kounios, C. Clark, R. Polikar, Multiresolution analysis for early diagnosis of Alzheimer's disease, 26th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS2004), 2004, pp. 251–254.
- [50] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, 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* 34 (1984) 939–944.
- [51] J.R. Cockrell, M. Folstein, Mini Mental State Examination (MMSE), *Psychopharmacology* 24 (1988) 689–692.
- [52] R.M. Crum, J.C. Anthony, S.S. Bassett, M. Folstein, Population-based norms for the Mini-Mental State Examination by age and educational level, *J. Am. Med. Assoc.* 269 (1993) 2386–2391.
- [53] M. Unser, The gallery at wavelet.org. Available at: (<http://www.wavelet.org/phpBB2/gallery.php>) (last accessed: 03-03-2006).
- [54] T. Demiralp, J. Yordanova, V. Kolev, A. Ademoglu, M. Devrim, V.J. Samar, Time–frequency analysis of single-sweep event-related potentials by means of fast wavelet transform, *Brain and Lang.* 66 (1999) 129–145.



- [55] A.P. Bradley, W.J. Wilson, On wavelet analysis of auditory evoked potentials, *Clin. Neurophysiol.* 115 (2004) 1114–1128.
- [56] A. Subasi, E. Ercelebi, Classification of EEG signals using neural network and logistic regression, *Comput. Methods Programs Biomed.* 78 (2005) 87–99.
- [57] B.V. Dasarathy, B.V. Sheela, Composite classifier system design: concepts and methodology, *Proc. IEEE* 67 (1979) 708–713.
- [58] L.K. Hansen, P. Salamon, Neural network ensembles, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (1990) 993–1001.
- [59] R.E. Schapire, The strength of weak learnability, *Mach. Learn.* 5 (1990) 197–227.
- [60] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, G.E. Hinton, Adaptive mixtures of local experts, *Neural Comput.* 3 (1991) 79–87.
- [61] J.A. Benediktsson, P.H. Swain, Consensus theoretic classification methods, *IEEE Trans. Syst. Man Cybern.* 22 (1992) 688–704.
- [62] D.H. Wolpert, Stacked generalization, *Neural Networks* 5 (1992) 241–259.
- [63] T.K. Ho, J.J. Hull, S.N. Srihari, Decision combination in multiple classifier systems, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (1994) 66–75.
- [64] Y. Freund, R.E. Schapire, Decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55 (1997) 119–139.
- [65] J. Kittler, M. Hatef, R.P.W. Duin, J. Mates, On combining classifiers, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (1998) 226–239.
- [66] L.I. Kuncheva, *Combining Pattern Classifiers, Methods and Algorithms*, Wiley, New York, 2005.
- [67] J. Ghosh, Multiclassifier systems: back to the future, in: F. Roli, J. Kittler (Eds.), *Third International Workshop on Multiple Classifier Systems*, Lecture Notes in Computer Science, vol. 2364, 2002, pp. 1–15.
- [68] T.K. Ho, Multiple classifier combination: lessons and the next steps, in: A. Kandel, H. Bunke (Eds.), *Hybrid Methods in Pattern Recognition*, World Scientific Publishing, 2002.
- [69] Various Authors, *Proceedings of International Workshop on Multiple Classifier Systems, 2000–2005*.
- [70] R. Polikar, L. Udpa, S.S. Udpa, V. Honavar, Learn<sup>++</sup>: an incremental learning algorithm for supervised neural networks, *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 31 (2001) 497–508.
- [71] M. Muhlbaier, A. Topalis, R. Polikar, Learn<sup>++</sup>.MT: a new approach to incremental learning, in: F. Roli, J. Kittler, T. Windeatt (Eds.), *Fifth International Workshop on Multiple Classifiers Systems*, Lecture Notes in Computer Science, vol. 3077, 2004, pp. 52–61.
- [72] L. Shapley, B. Grofman, Optimizing group judgmental accuracy in the presence of interdependencies, *Public Choice (Hist. Arch.)* 43 (1984) 329–343.
- [73] P.J. Boland, Majority system and the Condorcet jury theorem, *Statistician* 38 (1989) 181–189.
- [74] D. Berend, J. Paroush, When is Condorcet's jury theorem valid?, *Soc. Choice Welfare* 15 (1998) 481–488.

**Robi Polikar** is an assistant professor of Electrical and Computer Engineering at Rowan University, in Glassboro, NJ. He has received his B.Sc. degree in electronics and communications engineering from Istanbul Technical University, Istanbul, Turkey in 1993, and his M.Sc. and Ph.D. degrees, both comajors in electrical engineering and biomedical engineering, from Iowa State University, Ames, IA in 1995 and 2000, respectively. His current research interests are in pattern recognition, ensemble systems, computational models of learning, incremental learning and their applications in biomedical engineering.

**Apostolos Topalis** is a graduate student with the Electrical and Computer Engineering Department at Rowan University. His areas of interest include signal processing, pattern recognition and microprocessor system design.

**Deborah Green** is the lab manager for the EEG Lab at Drexel University, Philadelphia, PA. She has a B.A. degree in psychology from University of Delaware in 2003. Her research interests are in the neuropsychology of healthy aging and dementia, specifically in early diagnostic tools and possible intervention strategies, and the affects of mood and overall health on cognition. She will soon start her Ph.D. studies at Drexel.

**John Kounios** is a professor of Psychology at Drexel University, and the director of the EEG Lab. He received his dual major B.A. degree in psychology and music theory/composition from Haverford College in 1978, and his Ph.D. in experimental psychology from University of Michigan in 1985. His research focuses on the neural and cognitive bases of semantic information processing, problem solving and creativity.

**Christopher M. Clark**, a board certified neurologist, is an associate professor of neurology, associate director of the Alzheimer's Disease Center, director of the Memory Disorders Clinic (MDC) and director of the recently initiated Center of Excellence for Research on Neurodegenerative Diseases. He is a fellow of the University of Pennsylvania's Institute on Aging. His research interests focus on AD and the development of diagnostically specific markers, the identification and evaluation of new treatments, the development of new instruments to measure rates of change and studies of the relationship between Parkinson's disease and AD.