12-31-2006

Multiresolution wavelet analysis of event-related EEG potentials using ensemble of classifier data fusion techniques for early diagnosis of Alzheimer's disease

Apostolos Topalis
Rowan University

Multiresolution wavelet analysis of event-related EEG potentials using ensemble of classifier data fusion techniques for early diagnosis of Alzheimer's disease

by

Apostolos Topalis

A Thesis Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Department: Electrical and Computer Engineering
Major: Engineering (Electrical Engineering)

Approved

Members of the Committee

In Charge of Major Work

For the Major Department

For the College

Rowan University
Glassboro, New Jersey
2006

ABSTRACT

Apostolos Topalis
MULTIRESOLUTION WAVELET ANAYLSIS OF EVENT-RELATED EEG POTENTIALS USING ENSEMBLE OF CLASSIFIER DATA FUSION TECHNIQUES FOR THE EARLY DETECTION OF ALZHEIMER'S DISEASE
2005/06
Dr. Robi Polikar
Master of Science in Electrical Engineering

The recent advances and knowledge in medicine and nutrition have greatly improved our average life expectancy. An unfortunate consequence of this longer life span, however, is a dramatic increase in the number of individuals suffering from dementia, and more specifically, from Alzheimer's disease (AD). Furthermore, AD remains under-diagnosed and under-treated until its more severe stages due to lack of standard diagnostic tools available to community clinics. A search for biomarkers that will allow early diagnosis of the disease is therefore necessary to develop effective medical treatments. Such a biomarker should be non-invasive, simple to obtain, safe, inexpensive, accurate, and most importantly, must be made available to local health clinics for maximum effectiveness. Event related potentials (ERPs) of the electroencephalogram have the potential to become such a diagnostic biomarker for AD.

This work investigates the use of ERP signals for the early detection of AD. The analysis of the ERP signals is accomplished through multiresolution wavelet decomposition, producing time-frequency features in successive spectral bands. In previous studies, these feature sets were concatenated and used as inputs to a neural

network classifier. This contribution investigates training an ensemble of classifiers on each feature set separately, and combining the ensemble decisions in a data fusion setting. Comparisons of intra-signal and inter-signal ensemble combinations are presented in along with the benefits of using an ensemble of classifiers in data fusion.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

The advancement of medicine and nutrition has greatly improved the human lifespan. Over the last century, there has been a ten-fold increase in the number of Americans over the age of 65 because of an increasing population and the elderly living longer. In 1900 there were about 3.1 million people over the age of 65 living in the U.S. which has grown to more than 35 million today. By the year 2030, the number of over 65 year olds is expected to double to 70 million [1]. Unfortunately, the longer life expectancy brings an increased probability of individuals suffering from dementia.

Dementia includes a group of disorders causing a progressive decline in a person's mental functioning. Symptoms include loss of memory, intellect, rationality, social skills and normal emotional reactions. Some of the most common causes of dementia include Alzheimer's disease (AD), vascular dementia (VaD), and Parkinson's disease. Alzheimer's disease is the most common cause of dementia accounting for 50%-70% of all cases [2]. Its prevalence increases with age where an estimated 5% of 65 year olds are affected and over 30% of 85 year olds suffer from Alzheimer's. In the U.S. alone, over 4.5 million suffer from Alzheimer's disease, and this number is expected to reach 12 – 16 million by 2050. Furthermore, it is estimated that there are currently 18 million people affected by AD worldwide, and this too is projected to double in the next 20 years. Alzheimer's disease not only incapacitates its victim, but it causes grief for the victim's caregiver and a devastating financial toll on the society with an annual cost of over $100 billion [3, 4].

The symptoms of AD and several different types of dementia can be confused with symptoms resulting from various medical conditions, psychological problems and

drug reactions. It is very important to recognize signs of dementia and differentiate between conditions that imitate this disorder. Conditions that mimic the symptoms could be reversible if treated early. Additionally, progressive dementias like AD respond better to treatment if caught in their early stages [5].

Alzheimer's disease disrupts the brain's complex signaling system. The brain is made up of billions of nerve cells, called neurons, connected in an immeasurable network. This network of interconnected neurons distributes chemical messages as electrical impulses by means of neurotransmitters. A neuron relays a message only when the timed combination of incoming impulses has met a certain threshold. Multiply this by millions of neurons firing simultaneously and the result is the control network which allows us to speak, move, see, think, and remember.

Alzheimer's disease disturbs the intricate messaging network by means of two abnormal structures in the brain: amyloid plaques and neurofibrillary tangles [6]. Plaques are made of beta-amyloid, a toxic molecule, which originates from a normal protein called amyloid precursor protein (APP). Enzymes breakdown APP and cause the beta-amyloid fragments to cluster together into damaging plaques. The beta-amyloid plaques block neurons from sending messages compromising the brain's communication network [4].

Neurofibrillary tangles form due to the collapsing of protein tau. Normally, protein tau stabilizes the internal support structure of neurons. In Alzheimer's disease, changes cause threads of the protein tau to become entangled resulting in the decomposition of the neuron [4, 7]. The relationship between amyloid plaques and neurofibrillary tangles is illustrated in Figure 1.1.

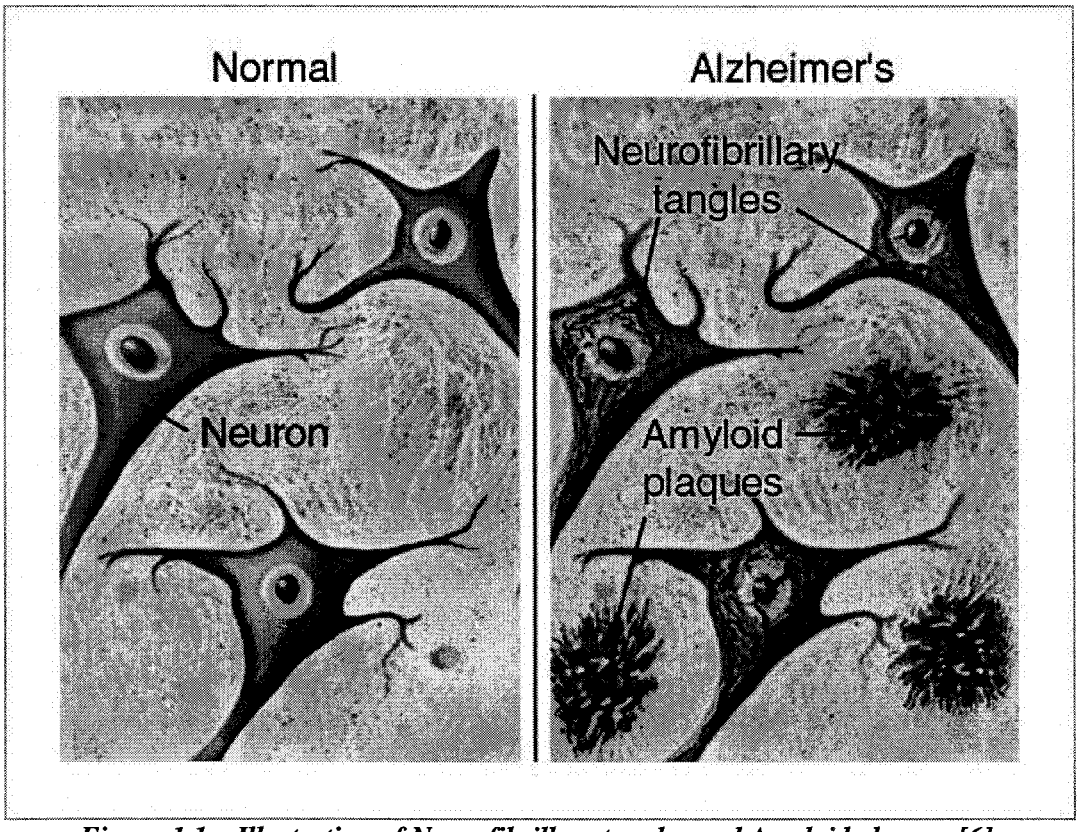

Figure 1.1 – Illustration of Neurofibrillary tangles and Amyloid plaques [6].

These plaques and tangles can only be identified by examining the brain tissue under a microscope, leaving autopsy as the only method for positive diagnosis. The progression of AD is typically divided into three stages: mild, moderate and severe; or early-stage, mid-stage and late-stage Alzheimer's. Janssen characterizes the different phases of the disease, where the mild stage typically results in impairments of mental abilities such as memory loss as well as mood swings. In the moderate stage, behavioral disturbances usually further develop, whereas physical problems are dominant in the advanced stage, and the patients cannot care for themselves. However, the individual course of the disease is very variable [8].

The cause of Alzheimer's disease remains a mystery and there is no single factor that leads to the diagnosis of AD. With the combination of multiple tests and measures,

professionals believe they can diagnose AD with a high accuracy [5]. These tests include a physical exam, in-depth interviewing of the patient and family, along with various medical and psychological tests. There is currently no test that can diagnose AD, therefore examinations are intended to rule out other causes of dementia, such as nutritional deficiencies, infection, metabolic disorders, and drug effects. Neuropsychological tests enable a clinician to analyze a patient's cognitive status. Specifically, the mini mental state exam (MMSE) uses a scoring system that is helpful in following mental status over time. Even though other causes of the symptoms can be ruled out, the diagnosis is "probable" or "possible" Alzheimer's because the only definitive test is postmortem examination of the brain [9].

1.1 MOTIVATION

Diagnostic markers are potential evidence in physiological systems that a person may have AD or is likely to get the disease. The incubation period of the disease is unknown, although it is clear that AD begins many years before characteristic symptoms appear. Studies show that non-demented individuals had sufficient plaques for postmortem diagnosis of AD, although there were no signs of cognitive impairment [10, 11]. In other words, the harmful plaques begin to form decades before the symptoms of the disease prevail in its victims. Since it is easier to prolong or even stop the progression of its effects than it is to reverse the development, it is essential to identify biomarkers that will detect the disease at an early stage. The discovery of markers with a correlation to AD will not only be a useful tool in diagnosis, but could also identify early stage AD patients for the development of new medical treatments [12].

The search of these types of biomarkers results in many challenges due to the inherent difficulty of the problem. Clinical diagnosis can be accurate but is only possible for symptomatic patients. Experimental findings of potential biomarkers can only be validated after lengthy longitudinal studies that follow a patient's condition from a pre-symptomatic stage to autopsy verification of AD. Ideally, the biomarker or group of biomarkers for AD should detect the illness at an early stage. The biomarker's diagnostic test must also result in a high sensitivity for determining AD patients and a high specificity for distinguishing between other dementias. The physiological measure of this marker should be non-invasive, simple to perform, safe and acceptable to both patients and physicians, and inexpensive to allow for its availability at local health clinics.

1.2 SCOPE OF THE THESIS

The objectives of this work include the analysis of event related potentials (ERPs), obtained from the electroencephalogram (EEG), for the early diagnosis of Alzheimer's disease. AD is the most common dementia in older patients (over 60), and remains under-diagnosed and under-treated since there is no current test or procedure that is diagnostic. Various experiments were run to find the best combinations of ERP biomarkers for the early diagnosis of the disease. The ERP signals were invoked using an oddball paradigm and recorded at three electrodes: Cz, Fz, and Pz. These signals were generated on responses to two tones (target and novel). The analysis of the ERP signals is accomplished through a wavelet decomposition producing coefficient sets in successive spectral bands. Coefficients of each spectral band were used to train and test a neural network ensemble of classifiers. The ensemble decisions were then combined in

various experiments to increase the overall generalization performance of the system. The novelty of this work comes in the use of individual frequency bands separately rather than the concatenation of all coefficients. Furthermore, the ensemble decisions based on individual spectral bands are combined. This training and combination method reduces a multi-dimensional problem into multiple lesser-dimensional problems by means of data fusion.

1.3 ORGANIZATION OF THE THESIS

Chapter 2 provides a detailed literature review on a background to various methods for early diagnosis of AD. Specific methods used in this study are provided in detail in Chapter 3. These methods include the acquisition of ERP signals through an oddball paradigm, the wavelet transform for feature extraction, the Learn++ algorithm for data fusion, and several performance measures. Chapter 4 provides results on a variety of experiments based on the methods described in Chapter 3. Furthermore, data fusion tests and their results are described in detail in Chapter 4 including intra-signal and inter-signal combinations. Chapter 5 consists of a discussion of the results along with conclusions about the study and suggestions for future work.

CHAPTER 2 – BACKGROUND

Several studies have been conducted to identify distinguishing biomarkers for the diagnosis of Alzheimer's disease [12]. These studies can be grouped into two categories: those investigating chemical biomarkers and those investigating signals obtained from the brain.

Studies on spinal fluids have lead to some progress in determining biochemical markers for Alzheimer's disease. The beta-amyloid molecule, a 42-amino acid peptide, can self-assemble into small clumps known as beta-amyloid-derived diffusible ligands (ADDLs), as well as the plaques that were described in the introduction. Experiments show that an increased level of ADDL has been directly related to neurological dysfunctions such as memory loss. Researchers have also found above normal ADDL levels from samples of people with AD. Furthermore, there is evidence of a correlation between the levels of ADDL in the spinal fluid of AD patients and the concentration of ADDL in the subject's brain. Thus, an ADDL spinal fluid measure could become a potential diagnostic biomarker [13]. The cerebrospinal fluid tau protein (CSF-tau) has also been linked to AD as an increase in the protein is found in most victims' brain. These increased levels of CSF-tau, on the other hand, can also be found in other dementias, more specifically, vascular dementia [14].

Cerebrospinal fluid samples for both of these potential markers are collected using a lumbar puncture, also known as the spinal tap procedure, which can be very painful, is available only in hospitals and some clinics, and is also very expensive. There is an easier obtained chemical biomarker for AD called F2-isoprostane. F2-isoprostane is measured through urine analysis, but has not been as dependable as CSF analysis or

ultimately postmortem tissue analysis for accurate diagnosis. Several studies show a considerable increase in F2-isoprostane in patients with AD and even mild cognitively impaired (MCI) subjects, compared to controls. However, additional independent studies found no correlation between AD, MCI and control subjects [15-17].

Since the disease affects the communication network of the brain, measures of electrical and functional brain activity should result in some correlation with the disease. Structural brain imaging and positron emission tomography (PET) scans have been used primarily for symptomatic AD, but evidence shows that imaging can help predict the progression of the disease through its pre-symptomatic stages [12, 18, 19]. Data from magnetic resonance imaging (MRI) studies show a relationship between a decline in brain volume and a regression of cognitive performance. More specifically, the hippocampal volume decreases in both AD and healthy individuals with age, but the rate of change is more than 2.5 times faster in AD patients [20, 21]. Babiloni used low resolution brain electromagnetic tomography (LORETA) to compare mild Alzheimer's disease with vascular dementia and cognitively normal subjects. Furthermore, the study investigated the sensitivity to different stages of severity. Even though the results show promising correlations, evaluation of the clinical usefulness of the technique has not been explored [22].

The electroencephalogram (EEG) is a medical device used to measure the electrical activity of the brain through sensitive electrodes applied to the scalp. More specifically, the EEG is the sum of the extra-cellular current flows of a large group of neurons. The EEG was originally developed as a method for investigating mental processes but later became a tool in medical diagnosis. Several studies focus on the

analysis of discriminate features between the EEG signals of patients with mild AD and control subjects. EEG spectral analysis has also been popular in not only distinguishing between Alzheimer's disease and controls [23, 24], but also between Alzheimer's and other degenerative dementias [25, 26]. Variations of EEG analysis have also been explored such as subjects with eyes open compared to eyes closed [27], or comparisons of task difficulty [28]. Hogan investigated the use of a memory test while recording EEG signals. Although, the performance on the memory test did not differ between patients with AD and cognitively normal subjects , the EEG measurements showed some distinguishing factors with AD patients possibly resulting from brain atrophy [29].

Since EEG signals are inherently non-stationary, Petrosian suggested that a combined wavelet and recurrent neural network (RNN) approach may be useful in a time-frequency analysis of long-term continuous EEGs for early recognition of AD. However, the study was administered using a small cohort and should be extended on larger patient populations before its clinical diagnostic value can be established. Further lines of investigation might also require that EEGs be recorded from patients engaged in certain mental (cognitive) activities [30].

Event related potentials (ERPs) are triggered based on the subject's reactions to different stimuli, mainly auditory, visual or somatosensory. Auditory stimuli could consist of single tones of pre-determined frequency, visual stimuli are produced by the presence of a light or a change in a pattern like a checkerboard and somatosensory stimuli are results of electrical stimulation of peripheral nerves. Additional measures involve the patient's reaction to these stimuli by means of simple tasks. A procedure can produce no-task evoked potentials, where the subjects are relaxed and instructed to perceive the

stimuli without performing any task. Alternatively, the oddball paradigm consists of two different stimuli: a non-target (standard stimulus) and a less frequent target stimulus. The subjects are instructed to react to the target stimulus by means of a simple task, such as pressing a button.

ERPs have the potential to become a diagnostic biomarker for Alzheimer's disease. Acquiring the signals simply involves an EEG recording along with an oddball paradigm stimulus procedure. Recording ERP signals is a well-established, reliable, non-invasive practice and can be made available at local community clinics. EEG signals alone are said to be affected by coexisting medical illnesses, varying levels of anxiety and drowsiness during recordings, amplifying the difficulty in identifying Alzheimer's patients. On the other hand, components of ERP signals, which are based on cognitive reactions, are believed to be insensitive to these concerns [31- 35].

ERP signals contain positive and negative peaks corresponding to the oddball stimulus reactions. Neurological disorders affecting the temporal-parietal regions of the brain, such as AD, have been known to deplete a specific positive peak, with an approximate latency of 300ms, called the P300 [36]. Polich et al. have shown that increased latency and decreased amplitude of P300, is associated with AD [31, 37]. Several other efforts, such as [36, 34, 38-42] have later confirmed the strong link between Alzheimer's disease and the P300.

Further studies include the use of unexpected novel sounds in the oddball paradigm. Findings of these experiments show promise in distinguishing AD from various forms of dementia and also cognitively normal subjects [36]. However, despite the statistical differences, the analysis of the P300, as a result of target or novel tones, in

amplitude and latency does not characterize individual patients as having AD. ERPs from cognitively normal subjects may have a delayed or reduced P300, where those with Alzheimer's disease, could produce a prominent P300, as shown in Figure 2.1. The inability of classical statistical approaches in individually identifying specific cases demands more sophisticated approaches for such individual identification.

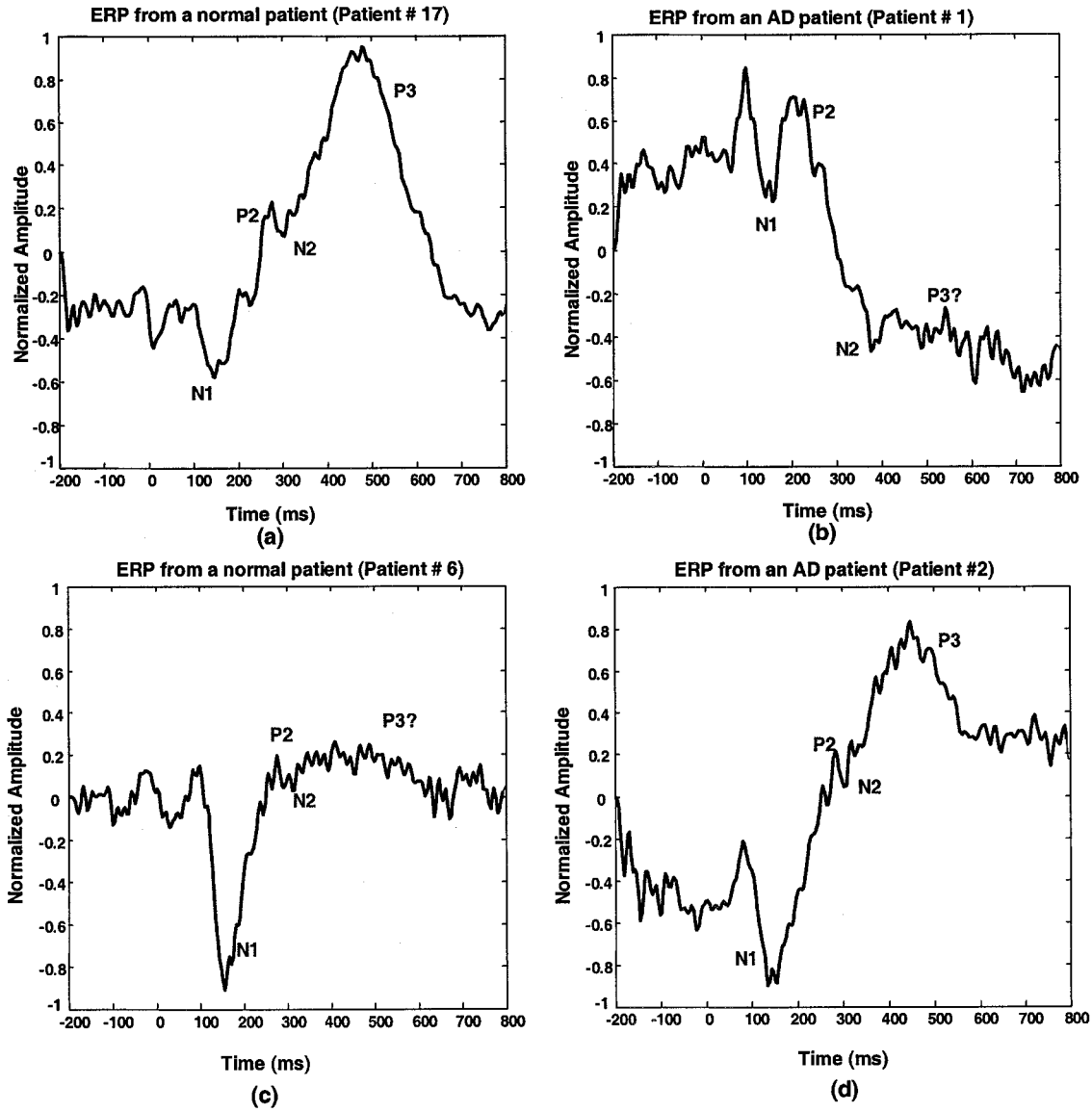

Figure 2.1 – (a) evident P300 in normal patient (b) missing P300 in AD patient, (c) & (d) not all individual cases follow this expected behavior.

Automated classification algorithms, such as neural networks, have been implemented for individual patient diagnosis. The accuracy of such classification algorithms, as in any other application of neural networks, depends on the quantity and the quality of the training data. The available data must be sufficiently large to represent the unknown, true distribution of specific classes. The features extracted from the data must also carry discriminatory information between different classes, such as AD and normal; or AD and VaD. Published methods obtain the features of the ERPs in the time domain by means of amplitude and latency of signal peaks such as the P300 [36, 40-47]. Other authors explore the frequency domain, more specifically, the power of different spectral bands of the ERP [29, 48-55].

Since the ERP is a time and frequency varying (non-stationary) signal, a time-frequency based analysis may be more suitable in identifying relevant features with discriminating properties. The wavelet transform satisfies these requirements resulting in various studies that apply such techniques to EEG analysis. Previous studies focus more on non-AD analysis specifically for structural analysis of the P300 [56-68]. Other studies investigated a wavelet analysis of EEG signals, along with a neural network classification approach, but ERPs were not examined [51, 69] or the studies did not specifically target the diagnosis of AD. For individual AD-specific diagnosis, there have been very few studies that use an appropriate time-frequency analysis, such as discrete wavelet transform, followed by neural network classification. The results of similar studies, such as [30, 70-72], differ from this contribution due to several reasons: minimal sized cohorts were used, sometimes between 10 and 30 subjects, making it difficult to resemble the existing Alzheimer's disease class distribution. The clinically diagnosed AD patients

were not targeted at their earliest stages, and in some cases, raw EEG signals were used rather than ERP signals.

While most of the methods have proven to be inconclusive, the ones showing promise remain primarily university and research hospital based tools. While clinical and neuropsychological evaluations achieve an average positive predictive value of 93%, this level of expertise is typically available only at university or research hospitals, can be very expensive, and remains beyond reach of most patients. Most patients unable to use university or research facilities are evaluated at local community healthcare centers, where the expertise and accuracy of AD specific diagnosis remains uncertain. A recent study reported that despite the advantage of longitudinal follow up, a group of health maintenance organization-based physicians had 83% sensitivity, and 53% specificity with an overall accuracy of 75% for the clinical diagnosis of AD [73].

CHAPTER 3 – METHODS

3.1 EXPERIMENTAL SETUP

The current basis for AD diagnosis is a clinical evaluation, using a series of neuropsychological tests, including interviews with the patient and their caregivers. Ninety patients have been recruited by the Memory Disorders Clinic and Alzheimer's Disease Research Center of University of Pennsylvania, according to the following inclusion and exclusion criteria. The two test cohorts include probably AD patients and cognitively normal subjects.

Inclusion criteria for cognitively normal cohort

- Age must be greater than 60
- Clinical Dementia Rating score = 0 (explained below)
- Mini Mental State Exam Score > 26 (explained below)
- No evidence of cognitive decline during the two years prior to enrollment based on an interview with the subject's informant

Exclusion criteria for cognitively normal cohort

- Evidence of any central nervous system neurological disease (e.g. stroke, multiple sclerosis, Parkinson's disease, or other form dementia) evaluated by history or exam
- The use of sedative, anxiolytic or anti-depressant medications within 48 hours of ERP acquisition

Inclusion criteria for AD cohort:

- Age must be greater than 60

- Clinical Dementia Rating score >0.50;
- Mini Mental State Exam Score ≤ 26;
- Presence of functional and cognitive decline over the previous 12 months based on detailed interview with an informant
- Satisfaction of National Institute of Neurological and Communicative Disorders and Stroke - Alzheimer's Disease and Related Disorders Association NINCDS-ADRDA criteria for probable AD [74].

Exclusion criteria for AD cohort:

- Same as for the cognitively normal controls.

All subjects received a through medical history and neurological exam. Key demographic and medical information, including current medications (prescription, over-the-counter, and alternative medications) were noted. The evaluation included standardized assessments for overall impairment, functional impairment, extrapyramidal signs, behavioral changes and depression.

The inclusion criteria for AD cohort were designed to ensure that subjects were at the earliest stages of the disease. One measure is the Clinical Dementia Rating (CDR) scale, used to gauge AD progression. The test rating is based on six cognitive-functional categories: memory, orientation, judgment, community affairs, home & hobbies, and personal care. A five point scale is used to rate the function in each category: normal (0), questionable impairment (0.5), mild impairment (1), moderate impairment (2), and severe impairment (3). An interview with the subject and informant is used to rate each category and special scoring rules are used to determine the overall CDR score.

Another metric is the Mini Mental State Exam (MMSE), a standardized test widely used to assess cognitive mental status. The test assesses orientation, attention, immediate and short-term recall, language, and the ability to follow simple verbal and written commands. In addition, MMSE provides a total score on a scale of cognitive function. Cognitive performance as measured by the MMSE shows an inverse relationship between MMSE scores and age. The MMSE medians range from 29 for those 18 to 24 years of age, to 25 for individuals 80 years of age and older. Education corresponds directly to the MMSE scores, where the median score is 29 for individuals with at least 9 years of schooling, 26 for those with 5 to 8 years of schooling, and 22 for those with 0 to 4 years of schooling [75, 76]. A score of less than 19 is considered to indicate 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 NINCDS-ADRDA criteria for probable AD.

3.2 ODDBALL PARADIGM

An auditory oddball paradigm was used to obtain the ERP signals. The protocol described by Yamaguchi et al, with slight modifications, was used [36]. Binaural audiometric thresholds were first determined for each subject using a 1000 Hz tone. The evoked response stimulus was presented to both ears at 60 dB above each subject's auditory threshold. This procedure removes any bias to variations in hearing between each patient. The stimulus consists of tone bursts 100 ms in duration, including 5 ms onset and offset envelopes. Tones at a frequency of 1000 Hz were presented 65% of the trials, and tones of 2000 Hz were presented 20% of the trials in a random sequence. The remaining 15% of the trials consisted of unexpected novel sounds that were presented

randomly. All trials totaled 1000 stimuli, and were delivered to each subject with an inter-stimulus interval of 1.0 to 1.3 seconds. The subjects were instructed to press a button each time they heard the 2000 Hz oddball (target) tone. Data collection typically took less than 30 minutes, which included frequent breaks, approximately three minutes of rest every five minutes. The experimental session was preceded by a 1-minute practice session where the subjects sampled the difference between target and standard tones.

ERPs were recorded from 19 electrodes embedded in an elastic cap. The electrode impedances were kept below 20 kΩ. Recordings which contained obvious artifacts were identified and rejected by the EEG technician. Each ERP potential was then amplified, sampled at 256 Hz for each channel, and low-pass filtered. The ERPs for each channel were then averaged, notched filtered at 59-61Hz and base-lined with respect to the pre-stimulus interval. The final signals consisted of 257 samples with a one-second duration. The signals, for each tone, of multiple trials, were synchronized and averaged to remove variations in cortical activity. Furthermore, all responses with artifacts, responses to missed targets, etc. were removed from the average, by the EEG technician, to obtain consistent ERP responses.

In P300 studies, the ERPs were typically obtained from the P_Z, C_Z or F_Z electrodes of the 10-20 EEG electrode placement system, as seen in Figure 3.1. The Pz electrode was analyzed since ERPs are known to be most prominent in the central parietal regions of the cortex [64]. The Pz, Cz, and Fz electrodes measure the cognitive activity in response to the tones rather than temporal lobe electrodes that would measure activity in hearing the tones. Furthermore, the P300 can only be observed in response to target

tones, and usually only responses to this tone are typically analyzed. The justification for analyzing the remaining two electrodes is the relative and symmetric proximity of Cz and Fz electrodes to the Pz electrode, see Figure 3.1. Additionally, the reason for analyzing the responses to the novel tones is that there may be relevant information present in the other components of the ERP, such as the P3a. Furthermore, these components may be more prominent in responses to the novel tones.

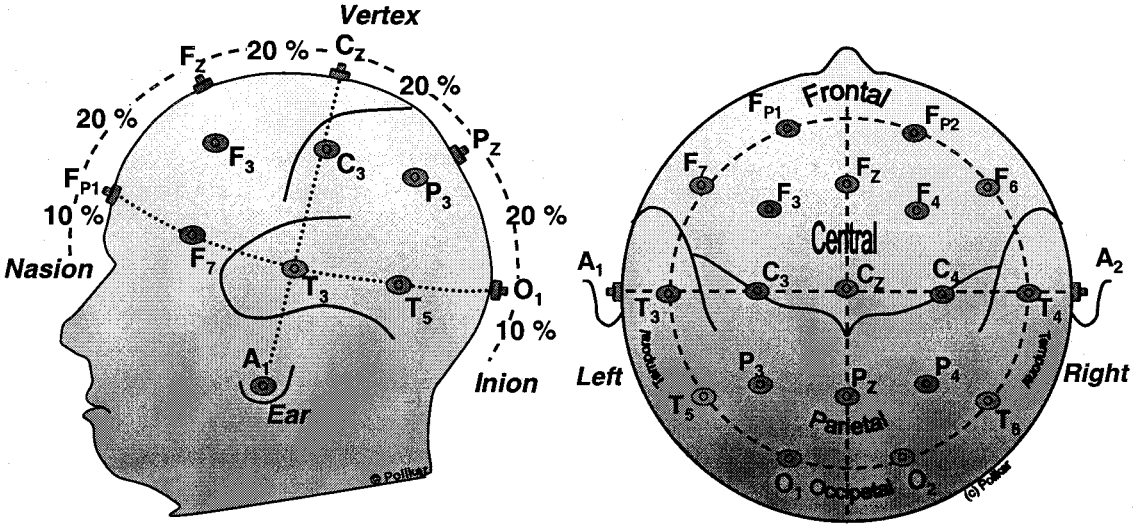

Figure 3.1 – The 10-20 International EEG electrode placement system [94].

The ERPs have been separated into two groups, those coming from control patients, and those coming from possible AD patients. The following figures show the grand averages for each type of tone and each electrode. Standard Cz, Fz, and Pz are shown in Figure 3.2 (a), (b) and (c) respectively. Target responses from the three electrodes are shown in Figure 3.3, and novel responses are shown in Figure 3.4, in the same format. The recording process begins 200ms before the stimulus is presented to the subject. The vertical line in each of the figures represents the beginning of the stimulus.

Also, all ERP plots are presented with an inverted y-axis to follow standard procedure in viewing these types of signals.

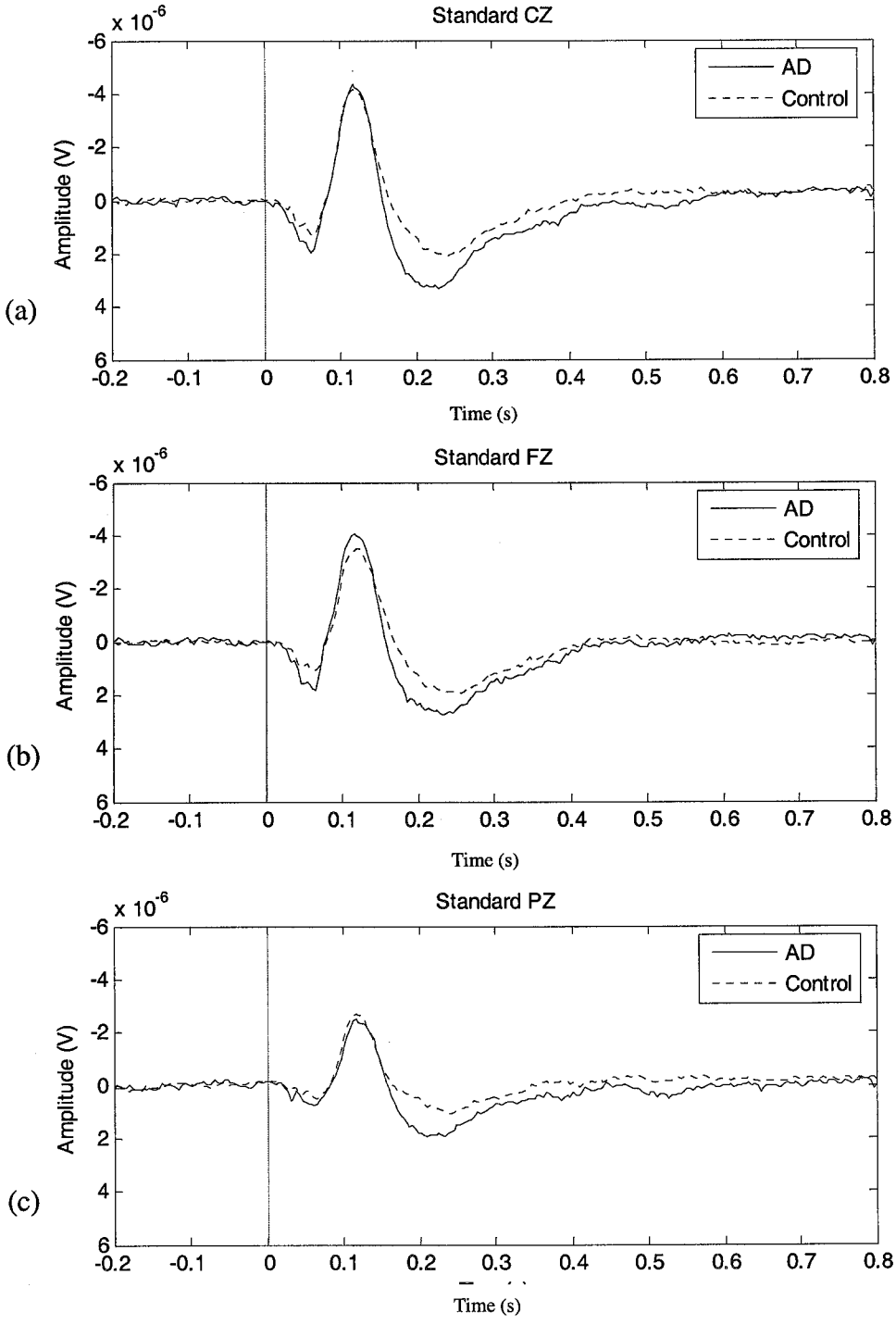

Figure 3.2 – Standard ERP signals for AD and Control patients from (a) Cz, (b) Fz and (c) Pz electrodes.

On average, the ERP responses to standard tones show little or no difference between AD patients and the control patients.

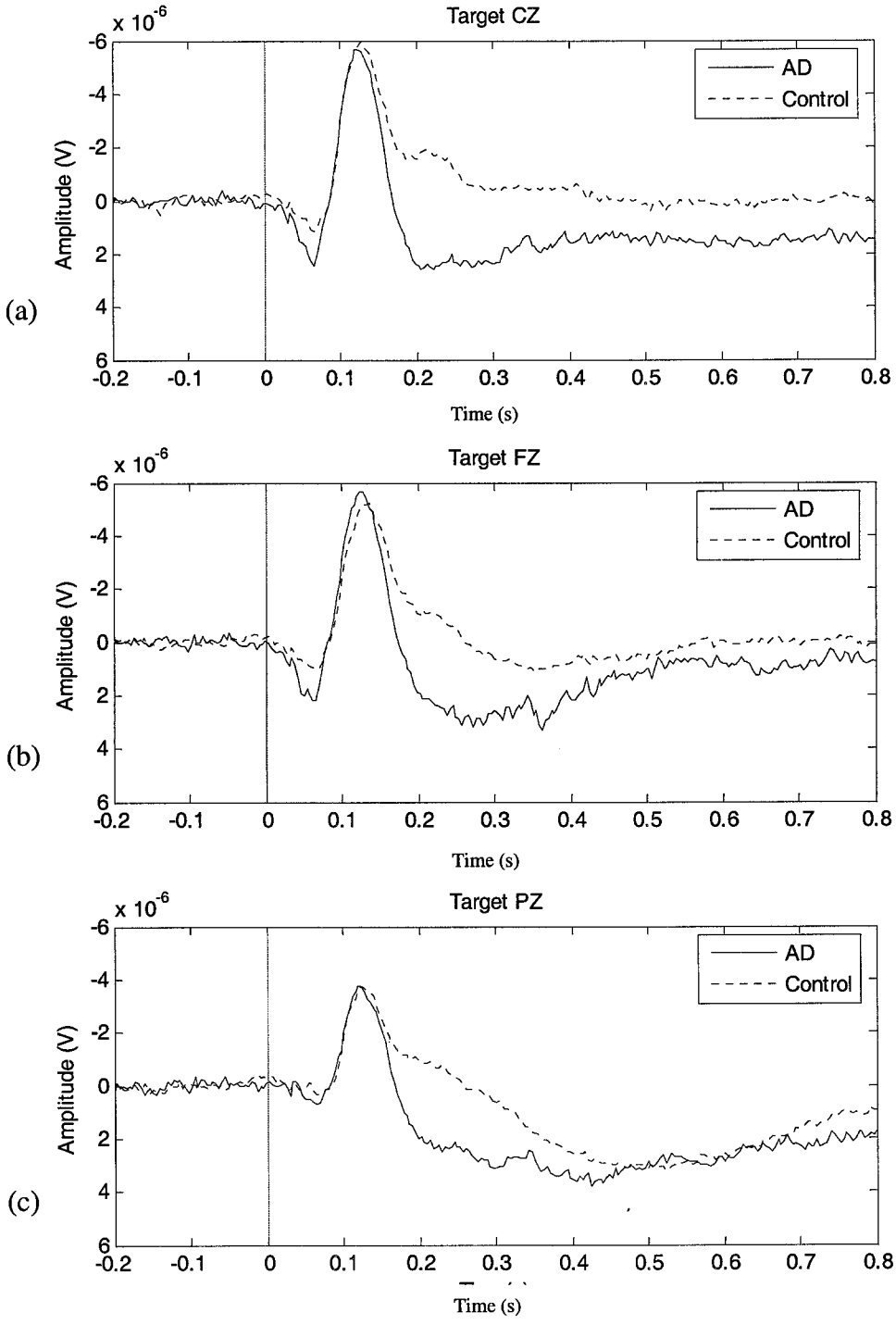

Figure 3.3 – Target ERP signals for AD and Control patients from (a) Cz, (b) Fz and (c) Pz electrodes.

The grand averages of ERP signals based on target tones produced a difference between AD patients and cognitively normal subjects in the 0-600ms range.

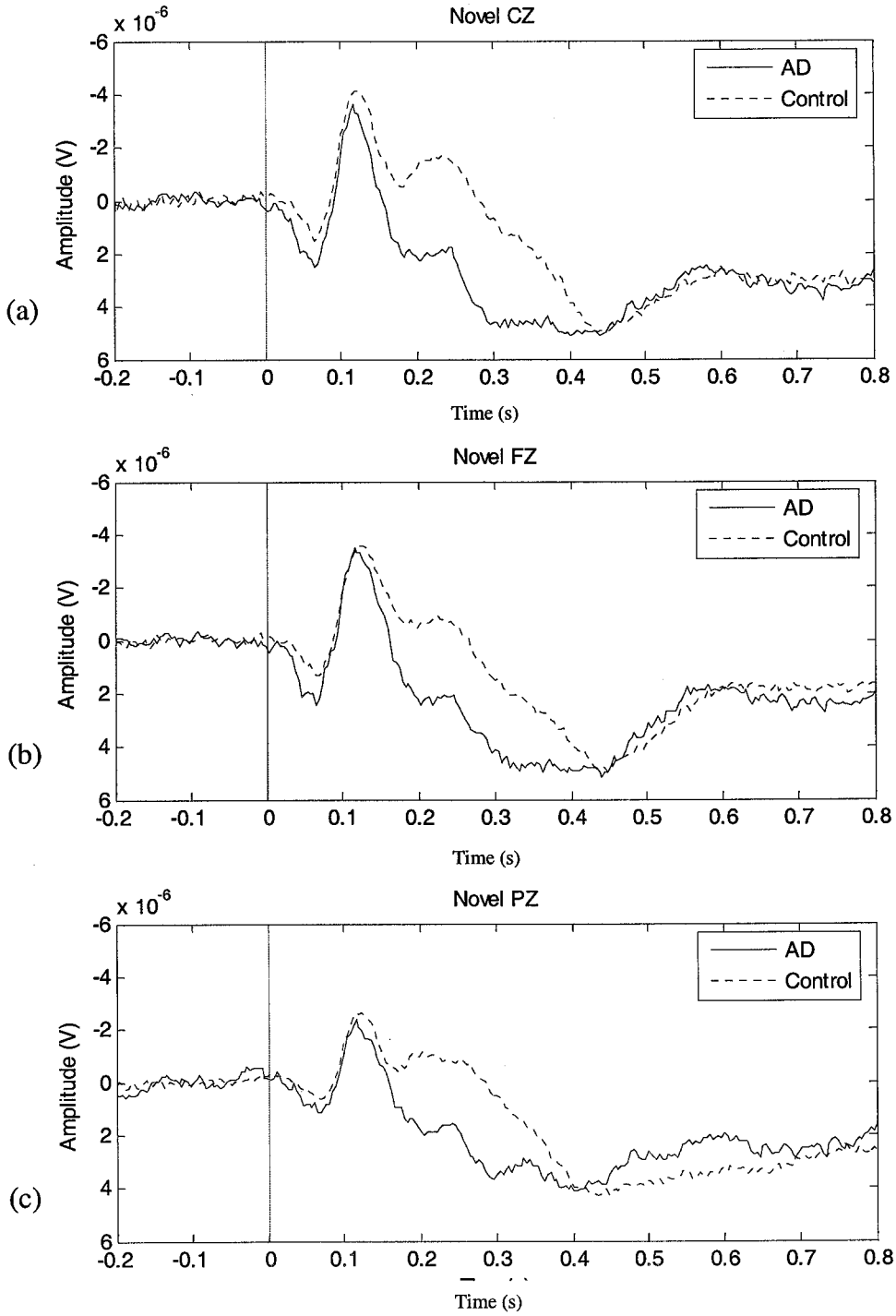

Figure 3.4 – Novel ERP signals for AD and Control patients from (a) Cz, (b) Fz and (c) Pz electrodes.

The averages of the ERP signals due to unexpected, novel sounds seem to carry distinguishing information between AD and control patients.

3.3 TIME-FREQUENCY ANALYSIS

A standard procedure in signal processing is the frequency domain analysis obtained through the Fourier transform (FT). The FT basically calculates the correlation between the signal to be analyzed and complex sinusoids of different frequencies and provides specific information on how much of each frequency exists in the signal. The continuous FT can be derived by the function:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt \tag{3.1}$$

where $X(f)$ represents $x(t)$ in the frequency domain. The function $x(t)$ has limitations such as being continuous and sufficiently smooth. The inverse transform can be calculated by:

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{i2\pi ft} dt \tag{3.2}$$

for every real number t. The inverse FT is the recombination of all frequency components of $x(t)$. Limitations on the FT are that it gives no information about time and requires signals to be stationary. A stationary signal is a signal that repeats within the sampled region.

The short-time Fourier transform (STFT), also known as the Gabor transform, was developed to provide a time-frequency representation of the signal. This transform uses the FT to determine frequency and phase content of localized sections of the signal. The signal to be transformed is multiplied by a window function, which is usually one near the origin and decays to zero at the edges. This type of window is used to suppress boundary issues inherent in rectangular windows. The STFT applies the windowing

function to the complex sinusoids of the Fourier transform and slides the window throughout the signal. The STFT can be calculated by the equation:

$$\bar{X}(\tau, f) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-i2\pi ft} dt \tag{3.3}$$

where $x(t)$ is the signal to be analyzed, $w(t)$ is the windowing function, and τ is the translation of the window throughout the signal. An apparent difficulty is choosing the width of the analysis window. If the window is too narrow, poor frequency resolution is obtained, and if the window is too wide, the time localization becomes imprecise.

Grossmann and Morlet [77] introduced the wavelet transform in order to overcome this problem. The main advantage of the technique is that the window size varies. A wide window size captures the low frequencies while a narrow window size captures the high frequencies. The varying window size allows the wavelet transform to provide desirable time-frequency resolution in all frequency ranges. Furthermore, it does not require the original signal to be stationary. Wavelet transforms can be classified into two types: continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The difference between the two types is the CWT operates over every possible scale and translation, while the DWT uses a specific set of scale and translation values. The CWT is written as:

$$\gamma(s, \tau) = \int_{-\infty}^{\infty} x(t) \psi_{s,\tau}^{*} dt \tag{3.4}$$

where * denotes the complex conjugation, and $x(t)$ is the function being decomposed into a set of basis functions. The function, $\psi_{s,\tau}(t)$, is generated from the scale (s) (inverse of frequency) and translation (τ) of the *mother wavelet* $\psi(t)$ by Equation 3.5.

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \tag{3.5}$$

The factor $1/\sqrt{s}$ is for energy normalization across different scales. The wavelet transform only deals with general properties of its calculation. Unlike the FT, the wavelet basis functions are not specified, moreover the analysis can be done using custom wavelets specific to particular applications. In choosing a particular wavelet, two main conditions must be satisfied, which give wavelets their name. First, the average value of the wavelet in the time domain must be zero, more formally written as:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{3.6}$$

In other words, $\psi(t)$ must be a wave. Second, the wavelet must have finite energy over infinity and thus satisfy Equation 3.7.

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt < \infty \tag{3.7}$$

The CWT is computed by continuously shifting a continuously scalable function over an entire signal, while calculating the correlation between the signal and the wavelet function. The CWT is therefore extremely redundant and unpractical for many applications. The DWT, on the other hand, overcomes this problem by scaling and translating in discrete steps. In other words, the time-scale space is sampled at discrete intervals.

The DWT utilizes two sets of functions, a scaling function, $\phi(t)$, and a wavelet function, $\psi(t)$, each associated with low-pass and high-pass 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:

$$\phi(t) = \sum_n h[n]\phi(2t-n) \tag{3.8}$$

$$\psi(t) = \sum_n g[n]\phi(2t - n) \tag{3.9}$$

The scaling function $\phi_{j,k}(t)$ and 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:

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k) \tag{3.10}$$

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right) \tag{3.11}$$

Different scale and translations of these functions allow differentiation of frequency and time localizations of the signal. The coefficients (weights), $h[n]$ and $g[n]$, that satisfy Equations (3.8) and (3.9) constitute the impulse responses of the low-pass and high-pass 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 high-pass and low-pass filtering of the time domain signal.

The continuous ERP signals are considered to be the original time domain signal $\boldsymbol{x(t)}$, where $\boldsymbol{x(t)}$ was sampled at 256 samples/second to form the discrete time signal $\boldsymbol{x[n]}$. The discrete time signal $\boldsymbol{x[n]}$ is first passed through a half-band high-pass filter, $g[n]$, and a low-pass filter, $h[n]$. The original signal was sampled at 256 samples/second, and therefore could only contain frequencies between 0-128 Hz (0-π radians), according to the Nyquist-Shannon sampling theorem. After half-band filtering, the spectral bands of the discrete signals are divided in half ($0 - \pi/2$) and ($\pi/2 - \pi$), thus allowing half the samples to be removed. This is accomplished by down sampling with a factor of 2. Filtering followed by down sampling or subsampling constitutes one level of decomposition, and it can be expressed as follows:

$$d_1[k] = y_{high}[k] = \sum_n x[n] \cdot g[2k-n] \tag{3.12}$$

$$a_1[k] = y_{low}[k] = \sum_n x[n] \cdot h[2k-n] \tag{3.13}$$

where $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the high-pass and low-pass filters, respectively, after the down sampling. The output of the high-pass filter, $y_{high}[k]$, represents *Level 1 DWT coefficients* (d_1: *level 1 detail coefficients*). The output of the low-pass filter, a_1: *the level 1 approximation coefficients*, is further decomposed by passing $y_{low}[k]$ through another set of high-pass and low-pass filters, to obtain level 2 detail coefficients d_2 and level 2 approximation coefficients, a_2, respectively.

At each level, the procedure results in half the time resolution and double the frequency resolution, allowing the signal to be analyzed at different frequency ranges with different resolutions. Even though the time resolution is down sampled by a factor of two for each level of decomposition, higher time resolution is unnecessary when analyzing low spectral bands. Figure 3.5 displays the spectral bands for each level of the wavelet transform, and the trend for further decomposition.

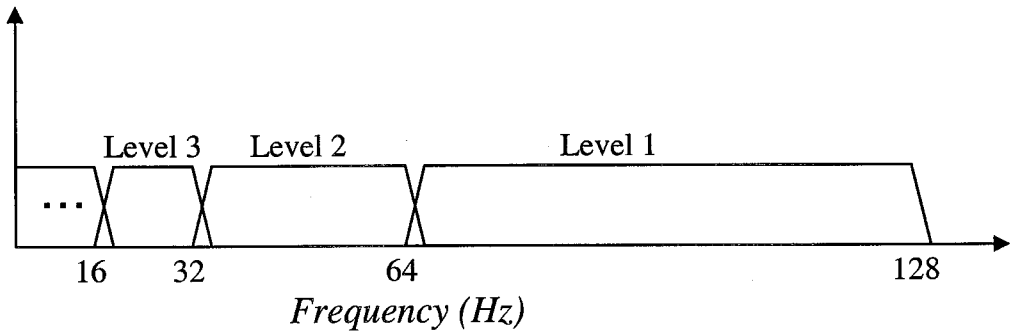

Figure 3.5 – Frequency domain representation of DWT decomposition.

The overall wavelet decomposition algorithm is illustrated in Figure 3.6, 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 decomposition in Figure 3.6 is obtained using the Daubechies wavelets with 4 vanishing moments, whose corresponding filters $h[n]$ and $g[n]$ are of length 8. The number of coefficients in each decomposition level is determined by the filter convolved with the original signal, then down sampled. In the first level of decomposition, the signal length is 257, and the filter length is 8, resulting in a filtered signal with 264 points. After down sampling, it is reduced to 132 points and the process repeats. An approximation signal $A_j(t)$ and a detail signal $D_j(t)$ can be reconstructed from level j approximation and detail coefficients:

$$A_j(t) = \sum_k a_j[k] \cdot \phi_{j,k}(t) \tag{3.14}$$

$$D_j(t) = \sum_k d_j[k] \cdot \psi_{j,k}(t) \tag{3.15}$$

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:

$$\begin{aligned} x(t) &= A_j(t) + \sum_{j=-\infty}^{j} D_j(t) \\ &= \sum_k a_j[k] \cdot \phi_{j,k}(t) + \sum_{j=-\infty}^{j} \sum_k d_j[k] \cdot \psi_{j,k}(t) \end{aligned} \tag{3.16}$$

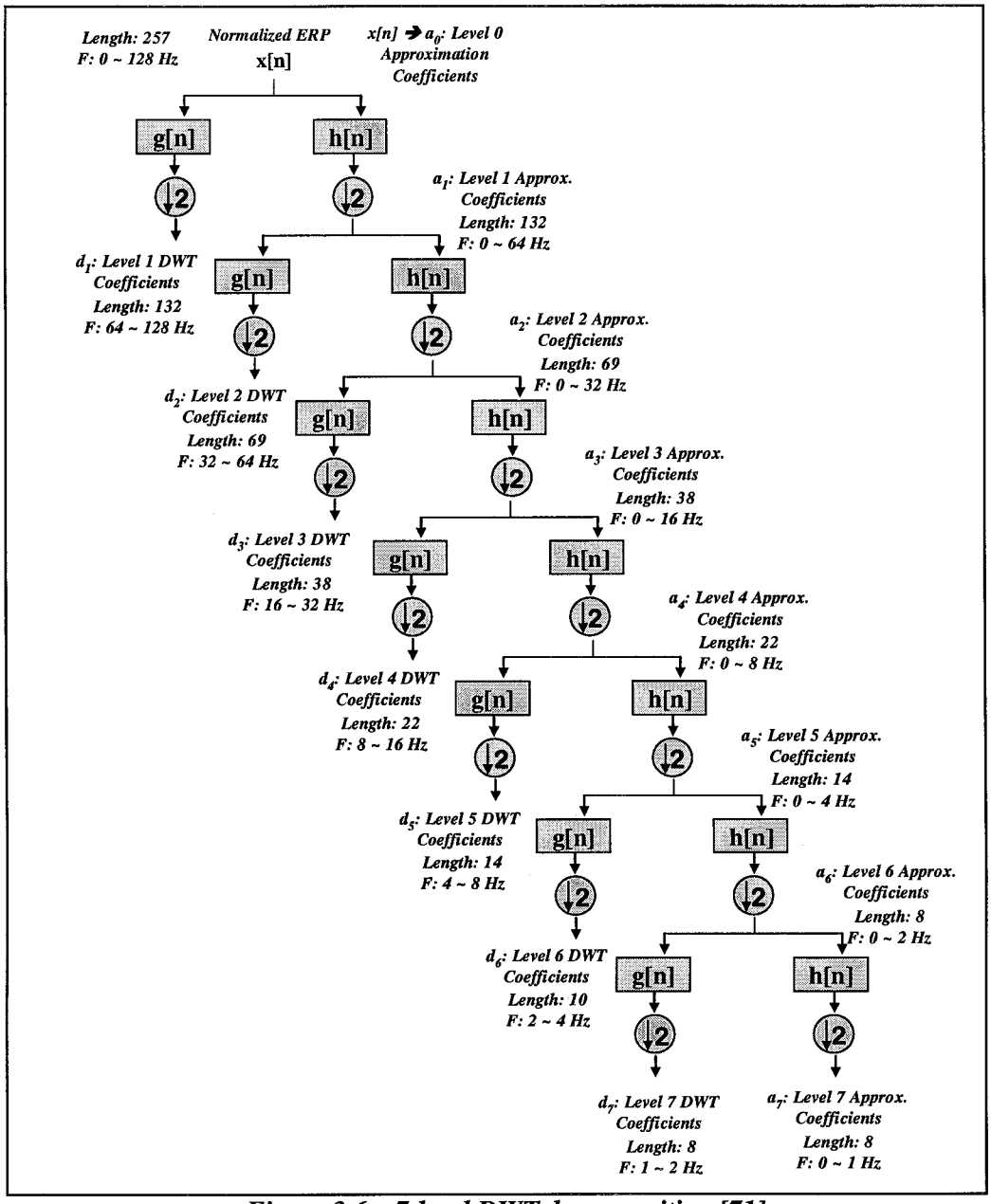

Figure 3.6 – 7-level DWT decomposition [71].

In this study, the Daubechies wavelet with 4-vanishing moments (Db4) was used based on previous studies of EEG analysis [71, 78]. Figure 3.7 shows the scaling function $\varphi(t)$ as well as the wavelet function $\psi(t)$ for the Db4 wavelet. The decomposition low and high-pass filters as well as the reconstruction low and high-pass filters are also illustrated

in Figure 3.7. Figure 3.8 illustrates the reconstructed signals obtained at each level, of a 7-level decomposition, for the analysis of ERP signals.

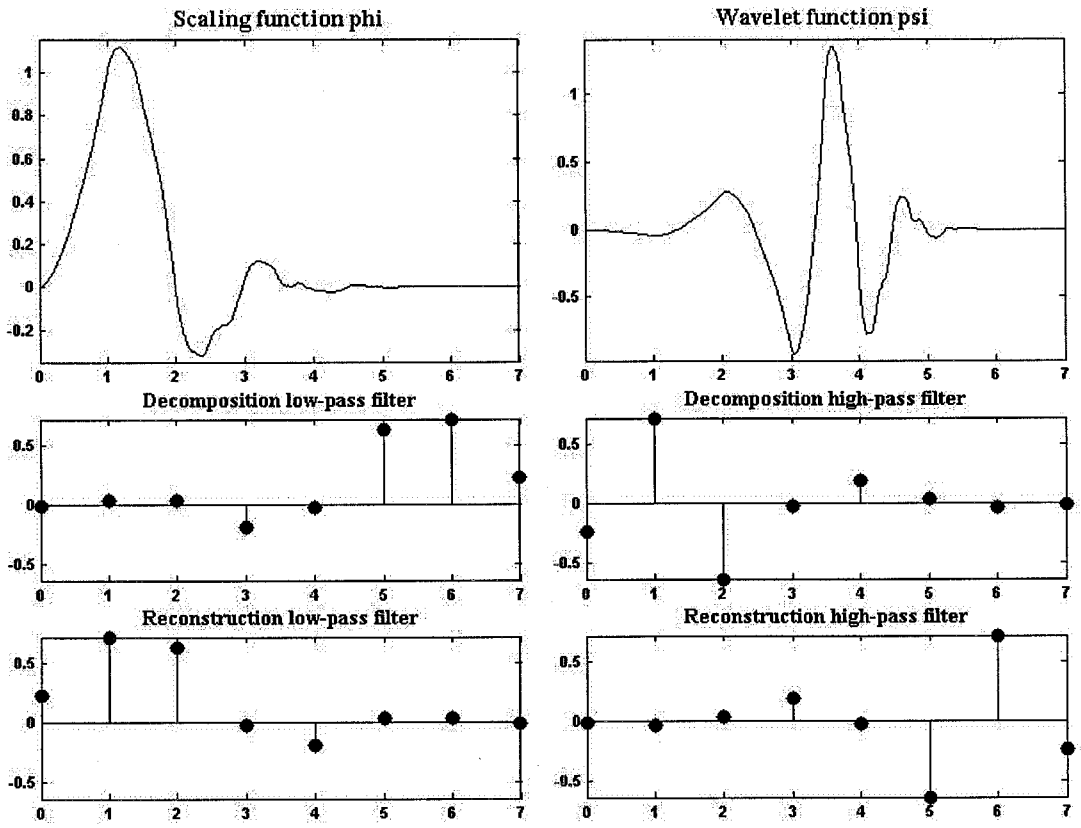

Figure 3.7 – Daubechies-4 wavelet scaling function $\varphi(t)$ and wavelet function $\psi(t)$.

Whereas signal reconstruction is not required as part of this work, the reconstructed signals in Figure 3.8 for both probable AD and cognitively normal patients indicate that majority of the signals' energy lie in $D_6 \sim D_7$ and A_7. Comparing the time domain signals of both patient types in Figure 3.8, it is difficult to distinguish between the AD and normal patient: both show peaks at 300ms mimicking the P300. However, spectral bands of the wavelet decomposition show differences that could contribute to the correct classification of each signal. The results presented in Chapter 4 will later confirm this observation.

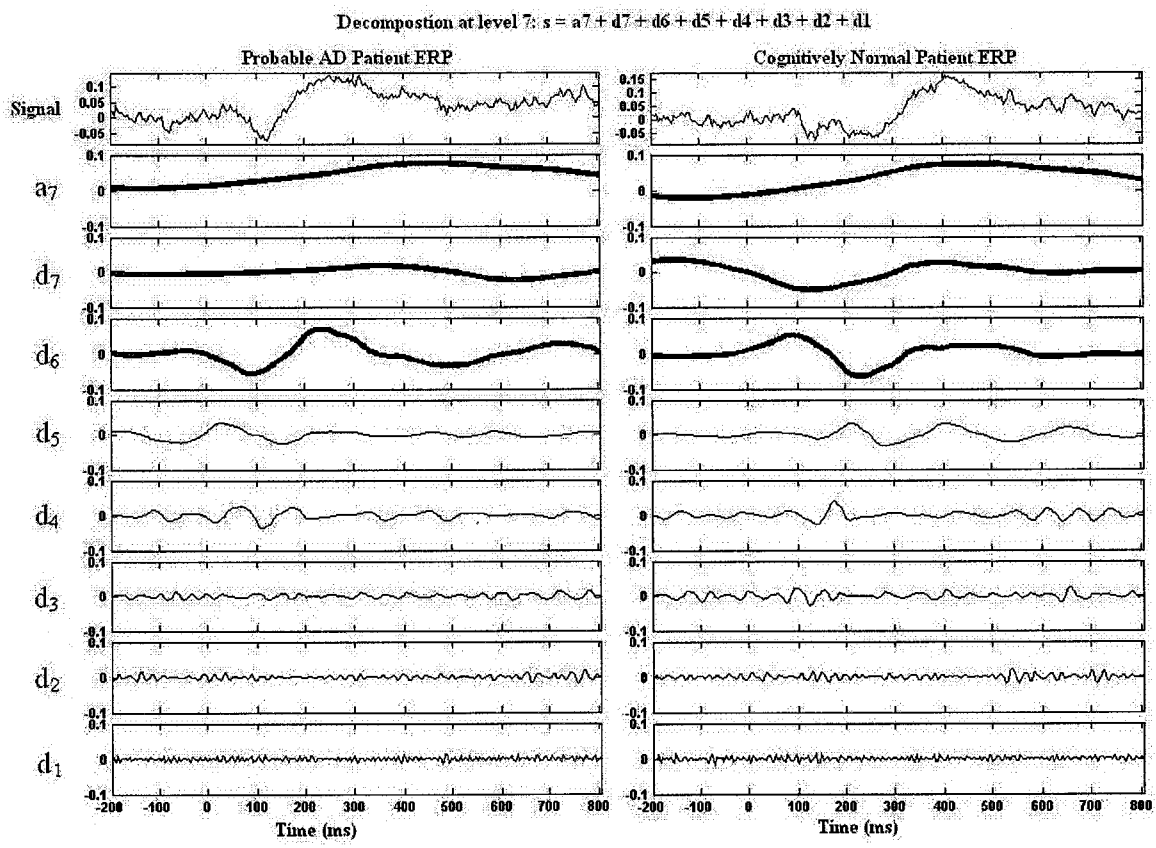

Figure 3.8 – Reconstructed signals of a 7-level wavelet decomposition for a probable AD patient and a cognitively normal patient.

The EEG recording was initiated 200ms before the stimulus (to provide a baseline signal) and continued for 800ms after the stimulus. Figures 3.3 and 3.4 show most of the distinguishing information lays between 0 and 600ms of the stimulus, therefore the coefficients of the wavelet decomposition falling within this interval were used in the analysis.

3.4 CLASSIFICATION

The classification process will be explained in detail in this section. The coefficients resulting from the wavelet decomposition are used to train an artificial neural network (ANN). An artificial neural network is a neural network model in artificial intelligence. ANNs are mathematical models for defining a function $f(\boldsymbol{x}) = y$. The description *network* is used since the structure of an ANN resembles simple interconnected elements. In this application, the inputs $\boldsymbol{x}$ are the wavelet coefficients, and the output y is the classification of the patient. In training an ANN, the presented features are mapped to the correct outputs by minimizing a cost function. The cost function is simply the misclassification of any instance (patient). Testing the 'trained' network involves presenting an instance to the ANN, that has not been used in training, and comparing the output of the ANN with the known classification. Each trained ANN is referred to as a classifier.

A novelty of this work is the investigation of an ensemble of classifiers for the classification of individual ERP signals. An ensemble based system, combines the hypotheses of several diverse classifiers, rather than a single classifier decision. The diversity in the classifiers is contributed to the use of different subsets of the training data. This difference in training, results in varying decision boundaries which are eventually combined to create a general ensemble decision. Numerous studies have shown that such an approach can often outperform a single classifier system, is usually resistant to over-fitting problems, and can often provide more stable results. Additionally, an ensemble of classifiers can be used for data fusion. A single classifier can be trained on data produced by a single source, where additional classifiers can be trained on added sources of data. In data fusion, an ensemble of classifiers could be

generated based on data from a single source. The fusion would involve the combination of multiple ensembles creating a meta-ensemble of classifiers. For further understanding of multiple classifier systems, refer to [79-90].

A single classifier system, using the multilayer perceptron (MLP), was used as an initial classification method. Multiple experiments with several architectural parameters, such as the number of hidden layer nodes ranging from 5 to 50, and error goals in the 0.005 to 0.1 range were used to find the optimal neural network architecture. After several independent trials, 10 hidden layer nodes and an error goal of 0.01 were chosen for all experiments. The Learn++ algorithm was used to generate ensemble systems, consisting of MLP base classifiers, with a varying number of classifiers. In general, a 10-classifier ensemble provided good results.

3.4.1 LEARN++ FOR DATA FUSION

Learn++, originally inspired in part by AdaBoost, [86] generates an ensemble of classifiers, where each classifier is trained on a strategically updated distribution of the training data that focuses on unseen and misclassified instances. Unlike AdaBoost, which is designed to choose instances that have not been seen or learned by only the previous classifier, Learn++ specifically targets these instances based on the ensemble's decision. Furthermore, Learn++ has the ability to generate an ensemble of classifiers for each data set coming from different sources, and appropriately combine the classifier outputs to take advantage of the additional information in subsequent data sources. The Learn++ algorithm for data fusion is discussed in detail below.

In data fusion, each new source of features is treated as a new data set. In other words, an ensemble of classifiers is trained on each feature set separately. The data sets are denoted as $\boldsymbol{DS}_k$, k =1, 2, …, K, where K is the total number of feature sources. For each dataset, $\mathbf{DS}_k$, the algorithm inputs are: the training data $\boldsymbol{S}_k$ comprised of m_k instances $\boldsymbol{x}_i$ along with their correct labels $y_i \in \Omega = \{\omega_1, \cdots \omega_C\}$, i = 1, 2, …, m_k, for C number of classes, a supervised classification algorithm BaseClassifier, generating individual classifiers (a multi-layer perceptron in this case), and an integer T_k, denoting the number of classifiers to be generated for the k^{th} dataset. The pseudocode of the algorithm and its block diagram, for training with data from the k^{th} data source, are provided in Figures 3.9 and 3.10, respectively. The BaseClassifier can be any supervised learning algorithm, where the parameters can be adjusted to ensure adequate diversity. Specifically, different decision boundaries are generated with the use of subsets of the training data. Even though each classifier is trained based on a subset of the training data, the generated classifier must result in a minimum performance of ½ on the entire training data set. If this condition is not met, the newly generated classifier is discarded and retrained. If classifier outputs are class-conditionally independent, then the overall error monotonically decreases as new classifiers are added. Originally known as the Condorcet Jury Theorem (1786) [91-93], this condition is necessary and sufficient for a two-class problem (C = 2), and it is sufficient, but not necessary, for C > 2.

At each iteration t, the Learn++ algorithm generates a new classifier. The algorithm trains the BaseClassifier on a strategically selected subset $\boldsymbol{TR}_t$ of the current training data to generate hypothesis h_t. The training subset $\boldsymbol{TR}_t$ is chosen from the training data according to a distribution $\boldsymbol{D}_t$. The distribution $\boldsymbol{D}_t$ is updated based on the

misclassified instances and normalized over the entire training data $\boldsymbol{S}_k$. The distribution $\boldsymbol{D}_t$ determines which instances of the training data are more likely to be selected into the training subset $\boldsymbol{TR}_t$. This distribution is initially set to be uniform, by initializing $w_1(i) = 1/m_k, \forall i,\ i = 1,2,\cdots,m_k$ giving equal probability for each instance to be selected into $\boldsymbol{TR}_1$. At each iteration t, the weights previously adjusted at iteration t-1 are normalized in step 1. All references to a *step* in the algorithm refer to the pseudo-code in Figure 3.9.

$$\boldsymbol{D}_t = \boldsymbol{w}_t \Big/ \sum_{i=1}^{m} w_t(i) \tag{3.17}$$

to ensure a proper distribution. Training subset $\boldsymbol{TR}_t$ is drawn according to $\boldsymbol{D}_t$ (step 2), and the BaseClassifier is trained on $\boldsymbol{TR}_t$ (step 3). A hypothesis h_t is generated by the t^{th} classifier. This classifier produces an error ε_t, which is computed as the sum of the distribution weights of the misclassified instances on the current dataset $\boldsymbol{S}_k$ (step 4)

$$\varepsilon_t = \sum_{i:h_t(\mathbf{x}_i)\neq y_i} \boldsymbol{D}_t(i) = \sum_{i=1}^{m} \boldsymbol{D}_t(i)\left[\!\left[h_t(\mathbf{x}_i) \neq y_i \right]\!\right] \tag{3.18}$$

where, $[\![\cdot]\!]$ evaluates to 1, if the predicate holds true, and 0 otherwise. If the condition $\varepsilon_t < ½$ is met, the hypothesis h_t is accepted and its error is normalized (step 5) to obtain

$$\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}, \quad 0 < \beta_t < 1 \tag{3.19}$$

If $\varepsilon_t > ½$, the current hypothesis is discarded, and a new training subset is selected by returning to step 2. For the addition of each classifier all hypotheses generated are combined using weighted majority voting, to obtain the composite hypothesis H_t (step 6). The contribution of each classifier to the composite hypothesis is based on a weight inversely proportional to its normalized error. The hypotheses with smaller training

errors are awarded a higher voting weight in the decision of H_t. The composite hypothesis H_t is formally written as:

$$H_t = \arg\max_{y\in\Omega} \sum_{t:h_t(\mathbf{x})=y} \log(1/\beta_t) \tag{3.20}$$

It can be shown that the weight selection of $\log(1/\beta_t)$ is optimum for weighted majority voting [88]. Similar to the individual classifier, the error of the composite hypothesis H_t is computed as the sum of the distribution weights of the instances that were misclassified by the ensemble decision H_t (step 6)

$$E_t = \sum_{i:H_t(\mathbf{x}_i)\neq y_i} D_t(i) = \sum_{i=1}^{m} D_t(i)\left[\left| H_t(\mathbf{x}_i) \neq y_i \right|\right] \tag{3.21}$$

Since individual hypotheses that make up the composite hypothesis all have individual errors less than ½, the composite error satisfies the expression $0 \leq E_t < ½$. The normalized composite error B_t is obtained as

$$B_t = \frac{E_t}{1-E_t}, \quad 0 < B_t < 1 \tag{3.22}$$

and used for updating the distribution weights assigned to individual instances

$$w_{t+1}(i) = w_t(i) \times B_t^{1-\left[\left|H_t(\mathbf{x}_i)\neq y_i\right|\right]} = w_t(i) \times \begin{cases} B_t, & \text{if } H_t(\mathbf{x}_i) = y_i \\ 1, & \text{otherwise} \end{cases} \tag{3.23}$$

Equation (3.23) 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. The final hypothesis H_{final} is obtained by combining all hypotheses that have been generated thus far from all K data sources.

Readers familiar with the AdaBoost algorithm have noticed the overall similarities, but also the key difference between the two algorithms. The weight update rule of Learn++ targets instances that have not yet been correctly classified by the

ensemble, whereas AdaBoost focuses on instances that have been misclassified by the previous classifier. Specifically, in AdaBoost, the weight distribution is updated based on the decision of a single previously generated hypothesis h_t [86], whereas Learn++ updates its distribution based on the decision of the current ensemble through the use of the composite hypothesis H_t.

Figure 3.10 conceptually represents the system level organization of the overall algorithm for generating ensembles. Figure 3.11 illustrates the combination of ensembles for data fusion applications where each feature set comes from different sources. An ensemble of classifiers is generated for each of the feature sets, which are then combined through weighted majority voting. For data fusion applications, however, the contribution of each ensemble is adjusted based on the observed training performance on each data source. This adjusted weight is then used to obtain the final hypothesis H_{final}.

$$H_{final}(\mathbf{x}) = \arg\max_{y\in\Omega} \sum_{k=1}^{K} \sum_{t:h_t(\mathbf{x})=y} \log\left(\frac{1}{\beta_t \alpha_k}\right) \tag{3.24}$$

where, α_k was chosen as the error, calculated as the misclassification ratio of the final composite hypothesis on S_k:

$$\alpha_k = \left(\sum_{i=1}^{m_k} \left[\left| H_{T_k}(\mathbf{x}_i) \neq y_i \right|\right]\right) \Big/ m_k \tag{3.25}$$

where H_{Tk} indicates the final composite hypothesis generated from the k^{th} training data S_k of feature set FS_k.

During training, particular parameters were monitored to avoid classifiers with infinite voting-weights. When a classifier perfectly learns the entire training data, it results in $\varepsilon_t = 0$, which causes $\beta_t = 0$, and hence the voting weight of h_t to be infinite. A

classifier with infinite voting-weight outvotes all other classifiers and creates a decision h_t based on a single classifier. This situation can be avoided either by making classifiers weaker so that the training error exceeds zero, or by adjusting ε_t to be greater than zero. Additionally, the number of classifiers generated for each dataset should be the same. If one data set contains additional classifiers, the ensemble trained on this data set has a greater contribution to the final hypothesis. The number of classifiers of a particular data set should only differ when prior knowledge about each data set is known. Furthermore, the addition of classifiers to each ensemble might not result in a considerable performance increase, but when the final hypothesis is calculated, their addition in data fusion could improve the performance significantly.

Input: For each feature set DS_k, $k = 1, 2, \ldots, K$

- Training data $S_k = [(\mathbf{x_i}, y_i)]$, i=1, …, m_k with labels $y_i \in \Omega = \{\omega_1, \cdots \omega_C\}$.
- Supervised algorithm BaseClassifier.
- Integer T_k specifying the number of classifiers.

Do for each k = 1, 2, …, K

Initialize $w_1(i) = 1/m_k, \forall i,\ i = 1,2,\cdots,m_k$

Do for $t = 1, 2, \ldots, T_k$:

1. Set $\boldsymbol{D_t} = w_t \Big/ \sum_{i=1}^{m_k} w_t(i)$ so that $\boldsymbol{D_t}$ is a distribution.
2. Draw training $\boldsymbol{TR_t}$ subset from $\boldsymbol{D_t}$.
3. Obtain h_t by training with data $\boldsymbol{TR_t}$
4. Calculate the error of h_t on S_k, $\varepsilon_t = \sum_{i:h_t(\mathbf{x}_i) \neq y_i} \boldsymbol{D_t}(i)$

   If $\varepsilon_t > ½$, discard h_t and ***Go to*** Step 2.
5. Set $\beta_t = \varepsilon_t / (1-\varepsilon_t)$. Obtain the composite hypothesis through weighted majority voting

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

6. Compute error of H_t: $E_t = \sum_{i:H_t(\mathbf{x}_i) \neq y_i} \boldsymbol{D_t}(i)$
7. Set $B_t = E_t/(1-E_t)$, and update the weights:

$$w_{t+1}(i) = w_t(i) \times \begin{cases} B_t, & \text{if } H_t(\mathbf{x}_i) = y_i \\ 1, & \text{otherwise} \end{cases}$$

Compute voting weights adjustment factor:

$$\alpha_k = \left(\sum_{i=1}^{m_k} \left[\left| H_{T_k}(\mathbf{x}_i) \neq y_i \right| \right] \right) \Big/ m_k$$

Output the final hypothesis:

$$H_{final}(\mathbf{x}) = \arg\max_{y \in \Omega} \sum_{k=1}^{K} \sum_{t:h_t(\mathbf{x})=y} \log\left(\frac{1}{\beta_t \alpha_k} \right)$$

Figure 3.9 – Learn++ pseudocode for data fusion [94].

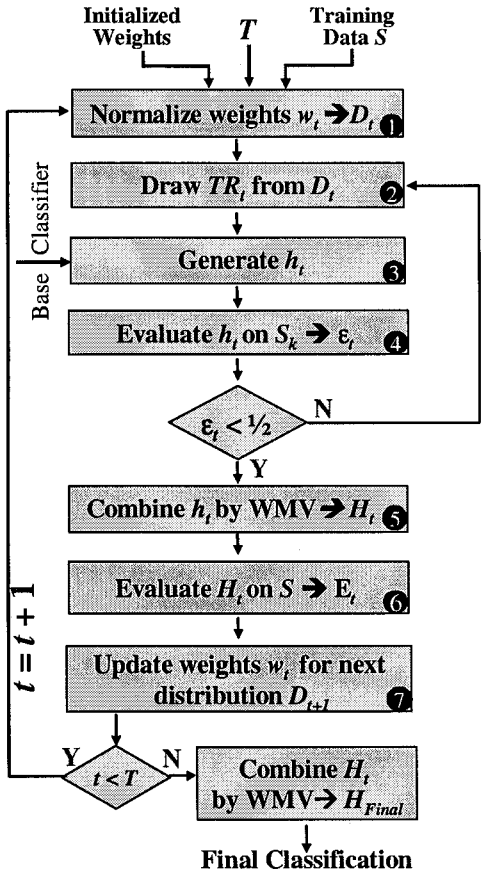

Figure 3.10 – Block diagram of Learn++ algorithm [94].

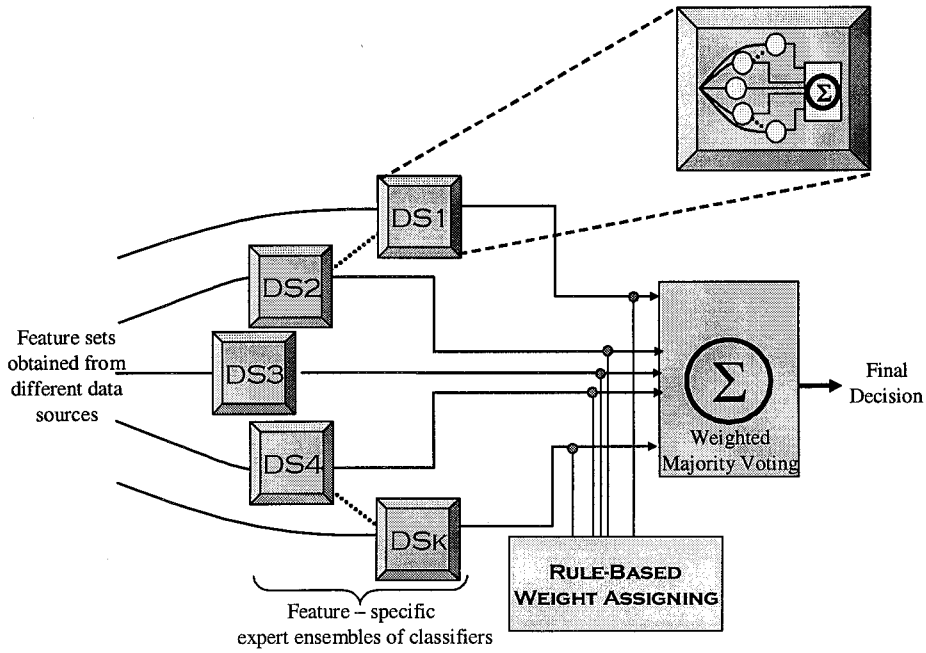

Figure 3.11 – Data fusion illustration [94].

3.4.2 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. The classification 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. Although computationally costly, the leave-one-out process is considered to be the most rigorous, reliable, conservative and unbiased estimate of the true performance of the system, as it removes the bias of choosing particularly easy or difficult instances into training or test datasets. 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 datasets, 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 five times. Since, the system that results in the single best leave-one-out performance is implemented, the single best leave-one-out results along with the average of five repetitions are provided.

3.4.3 DIAGNOSTIC PERFORMANCE FIGURE

The overall performance measure used in gauging typical learning algorithms lacks the class specific performances. For example, if the test data included 90 instances from class *A* and 10 instances from class *B*, and the algorithm correctly classified all instances from class *A*, but only half the instances from class *B*, the overall performance would be 95%. In medical tests and procedures, class descriptive measures are necessary for patient evaluation. Sensitivity, specificity, positive and negative predictive values are four commonly used quantities in medical diagnostics. Table 3.1 summarizes the concepts described below.

Overall Generalization Performance (OGP): For learning algorithms, the overall generalization performance is the average leave-one-out validation performance of the classifiers on test data. Specifically, the OGP represents the average probability of a correct decision. With medical tests, OGP refers to the accuracy of the test. OGP describes the ratio of patients the classification system has correctly identified.

Sensitivity: Can be described as the probability of a positive diagnosis when the condition is present in the patient. In other words, sensitivity is the ability of a medical test to correctly identify the target group. In this study, sensitivity is the ratio of the number of AD patients correctly identified by the classification system.

Specificity: Can be described as the probability of a negative diagnosis when the condition is not present in the patient. In other words, specificity is the ability of a test to correctly identify the control group. In this study, specificity is the ratio of cognitively normal patients correctly identified by the classification system.

Positive Predictive Value (PPV): PPV is the probability that the patient has the disease of those patients who test positive. It is calculated as the proportion of the sample population that is identified by the test as the target group, who in fact belong to the target group. In this study, PPV is the proportion of patients identified as AD patients by the classification system, who actually have AD.

Negative Predictive Value (NPV): NPV is the probability that the patient does not have the disease of those patients whose test results are negative. It is calculated as the proportion of the sample population that is identified by the test as the control group, who in fact belong to control group. In this study, NPV is the proportion of patients identified as cognitively normal by the classification system, who are actually cognitively normal.

Table 3.1 – Category labels for defining diagnostic quantities.

	Number of Patients ↘	True Condition	
		Probable AD	Cognitively Normal
Classification Decision	Probable AD	*A*	*B*
	Cognitively Normal	*C*	*D*

In Table 3.1, *A* is the number of patients classified as AD, and diagnosed as probable AD by the clinical evaluation. On the other hand, *B* is the number of patients who are also classified as AD, but clinically diagnosed as cognitively normal. Furthermore, *C* is the number of patients who are classified as normal, but clinically diagnosed as probable AD. Finally, *D* is the number of patients who are classified as cognitively normal, and

clinically diagnoses as cognitively normal. $A+B+C+D$ is the total number of patients. Then,

$$Overall\ Performance = \frac{A+D}{A+B+C+D} \tag{3.26}$$

$$Sensitivity = \frac{A}{A+C} \tag{3.27}$$

$$Specificity = \frac{D}{B+D} \tag{3.28}$$

$$PPV = \frac{A}{A+B} \tag{3.29}$$

$$NPV = \frac{D}{C+D} \tag{3.30}$$

3.5 METHOD OVERVIEW

The experimental setup included criteria to recruit patients at the earliest stages of Alzheimer's disease along with age-matched cognitively normal subjects. The cohort was subject to the oddball paradigm procedure, which invoked event-related potentials. The ERP signals were then decomposed, using the wavelet transform, into successive spectral bands. The coefficients from each spectral band were used as inputs into the Learn++ algorithm for data fusion. A leave-one-out validation process and various diagnostic measures were used to determine the feasibility of the approach.

CHAPTER 4 – RESULTS

This chapter describes the results of various experiments included in this work. Section 4.1 provides details about the patient cohort and the extracted features. It also presents results based on single ensemble classifiers. Section 4.2 focuses on intra-signal data fusion combinations, while Section 4.3 presents inter-signal fusion. The best two-ensemble systems are then combined in Section 4.4 followed by further analysis of data fusion in Section 4.5.

4.1 – ENSEMBLE RESULTS

Ninety patients that have been recruited and 71were used in this analysis. Reasons for excluding 19 of the total patients vary from the patient not being able to complete the test to equipment failure. The final cohort consists of 34 Alzheimer's patients (μ_{AGE}=75, μ_{MMSE}=24.7), and 37 cognitively normal (μ_{AGE}=76, μ_{MMSE}=29.2). The oddball paradigm provided the ERP signals from two types of stimuli tones (target and novel), and our focus is on three specific EEG electrodes: Cz, Fz, and Pz.

Inherently, the DWT decomposition process results in coefficients from several successive frequency bands within the 0-128 Hz frequency range. The relevant information is known to reside between 0-16 Hz. Therefore, the entire coefficient feature set of the DWT contains unnecessary components and only d_4-d_7 and a_7 coefficients are analyzed. Figure 4.1 illustrates the coefficients resulting from a 7-level decomposition, moreover the number of coefficients of each level along with the spectral band. In the 0-16 Hz frequency range, the wavelet decomposition produces five feature sets divided into a_7:(0-1 Hz), d_7:(1-2 Hz), d_6:(2-4 Hz), d_5:(4-8 Hz), and d_4:(8-16 Hz). In previous efforts

these features were concatenated, forming a 62-dimensional feature set, and used for training of a classifier.

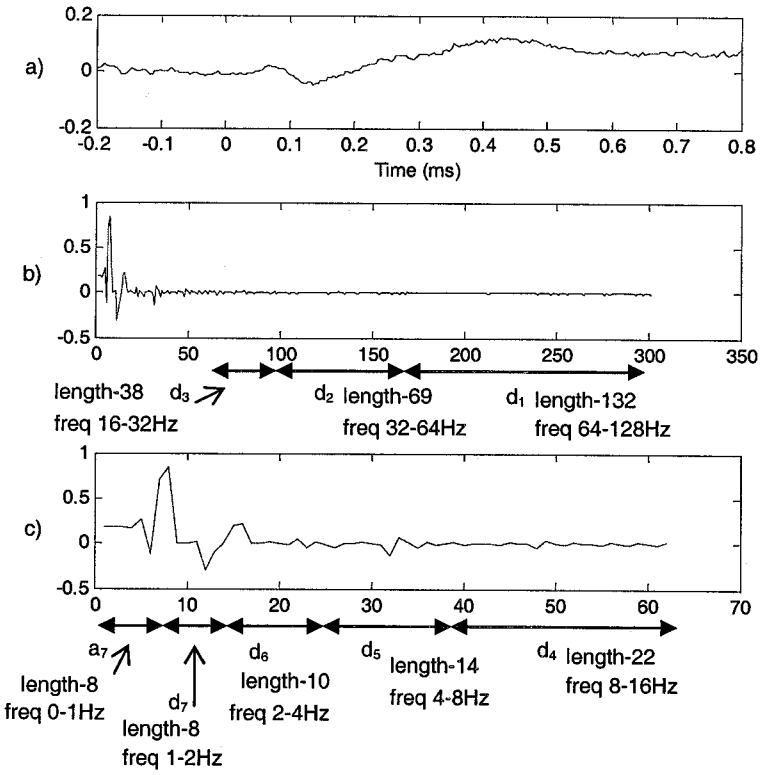

Figure 4.1 – (a) Time domain ERP signal, (b) coefficients of a 7-level wavelet decomposition, (c) wavelet coefficients representing 0-16 Hz.

The Learn++ algorithm, described in Chapter 3, has been used to generate the ensemble of classifiers. Results using a 10-classifier ensemble, on the above mentioned frequency bands, for six sets of ERP signals obtained from each patient (3 electrodes, 2 types of stimulus) are shown below. H_{10} describes the average performance of five leave-one-out trials using a 10-classifier ensemble. Since the implemented system will be based on a single leave-one-out trial, the column labeled, B-L1O, denotes the best single leave-one-out performance. CI is the confidence interval on five independent leave-one-out trials, and Sn, Sp, PPV, NPV describe the above mentioned sensitivity,

specificity, positive predicted value, and negative predicted value respectively. The coefficients obtained from a DWT decomposition of an ERP signal, measured at a specific electrode, in response to a particular tone will be referred to as [tone-electrode].

Table 4.1 – Performance values on ensemble systems trained on target-Cz coefficients for five frequency bands.

Target CZ	B-L1O	H_{10}	±CI	Sn	Sp	PPV	NPV
(0-1 Hz)	63.38%	**62.54%**	1.56%	64.71%	60.54%	60.13%	65.18%
(1-2 Hz)	54.93%	**52.11%**	3.50%	52.94%	51.35%	50.00%	54.30%
(2-4 Hz)	63.38%	**59.72%**	3.18%	60.59%	58.92%	57.64%	62.07%
(4-8 Hz)	59.15%	**53.52%**	5.10%	54.71%	52.43%	51.56%	55.59%
(8-16 Hz)	46.48%	**42.25%**	7.10%	42.35%	42.16%	39.96%	44.50%

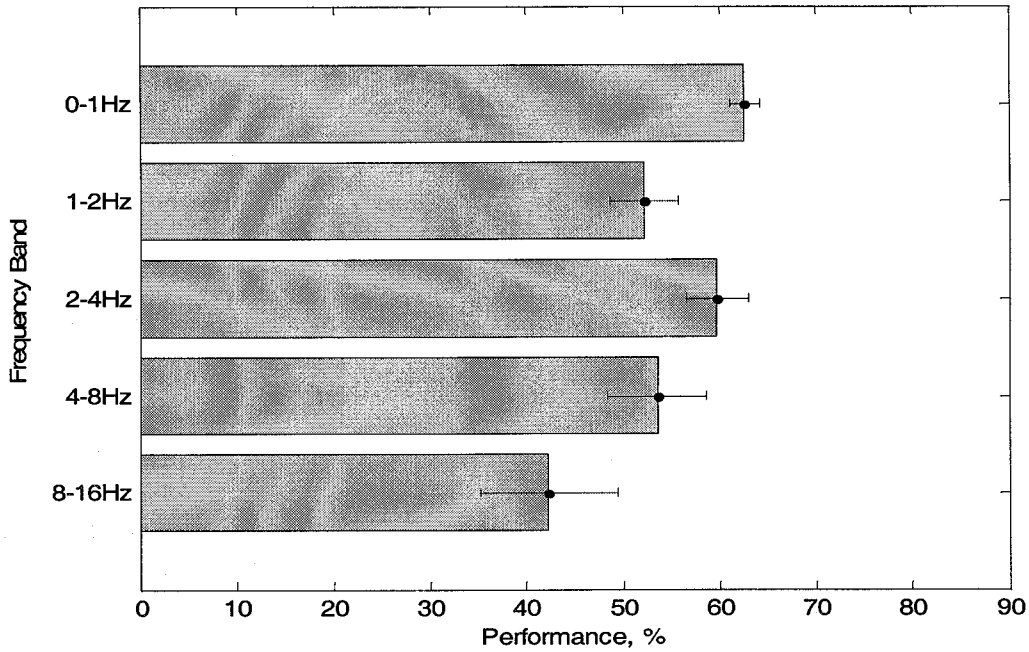

Figure 4.2 – Ensemble performance results trained on target-Cz coefficients.

Table 4.1 shows the generalization performances of the 10-classifier ensembles trained on target-Cz coefficients. Figure 4.2 illustrates the performance values along with corresponding CI for each spectral band. The table indicates that the highlighted frequency band (0-1 Hz) is the most accurate. In addition, this band produced the

smallest confidence interval, and highest Sn, Sp, PPV, and NPV results compared to other levels of decomposition.

Table 4.2 – Performance values on ensemble systems trained on target-Fz coefficients for five frequency bands.

Target FZ	B-L1O	H_{10}	±CI	Sn	Sp	PPV	NPV
(0-1 Hz)	59.15%	**53.52%**	5.93%	49.41%	57.30%	51.42%	55.29%
(1-2 Hz)	66.20%	**62.54%**	3.18%	62.35%	62.70%	60.61%	64.45%
(2-4 Hz)	63.38%	**58.03%**	5.85%	60.59%	55.68%	55.58%	60.72%
(4-8 Hz)	59.15%	**54.08%**	7.68%	52.35%	55.68%	52.08%	55.98%
(8-16 Hz)	56.34%	**53.52%**	2.47%	51.18%	55.68%	51.48%	55.39%

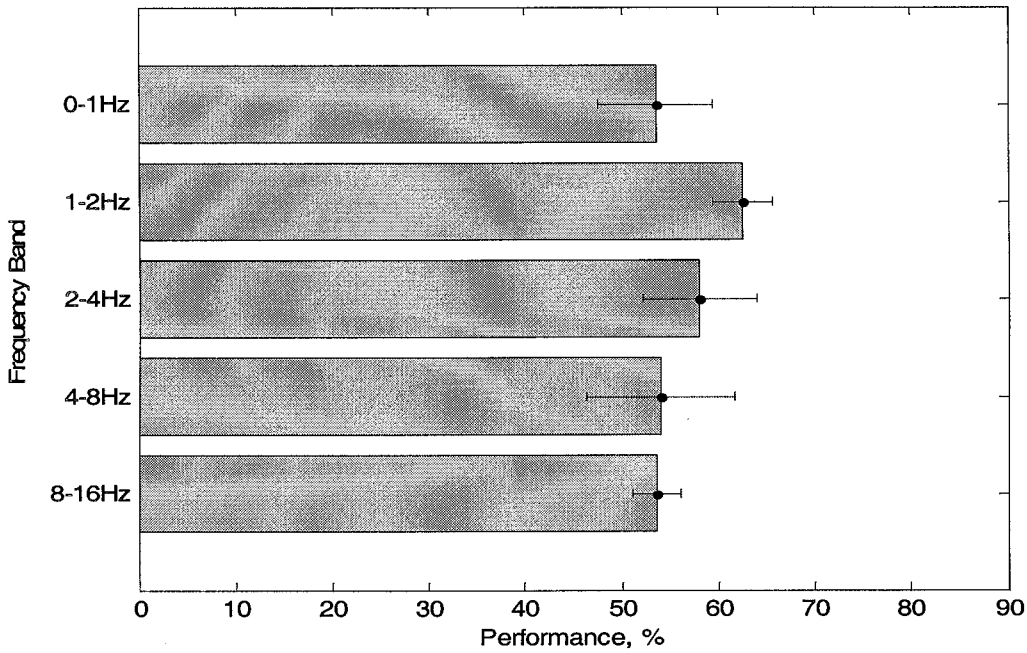

Figure 4.3 – Ensemble performance results trained on target-Fz coefficients.

Table 4.2 lists the generalization leave-one-out performances of the ensembles trained on target-Fz features; here the best performing frequency band is (1-2 Hz). Figure 4.3 illustrates the performances listed in Table 4.2.

Table 4.3 – Performance values on ensemble systems trained on target-Pz coefficients for five frequency bands.

Target PZ	B-L1O	H_{10}	±CI	Sn	Sp	PPV	NPV
(0-1 Hz)	63.38%	**61.69%**	1.92%	52.94%	69.73%	61.63%	61.73%
(1-2 Hz)	61.97%	**56.62%**	5.30%	54.12%	58.92%	54.78%	58.32%
(2-4 Hz)	64.79%	**63.66%**	0.78%	64.71%	62.70%	61.46%	65.98%
(4-8 Hz)	52.11%	**49.58%**	2.28%	48.24%	50.81%	47.37%	51.67%
(8-16 Hz)	50.70%	**45.92%**	6.01%	48.82%	43.24%	44.38%	47.52%

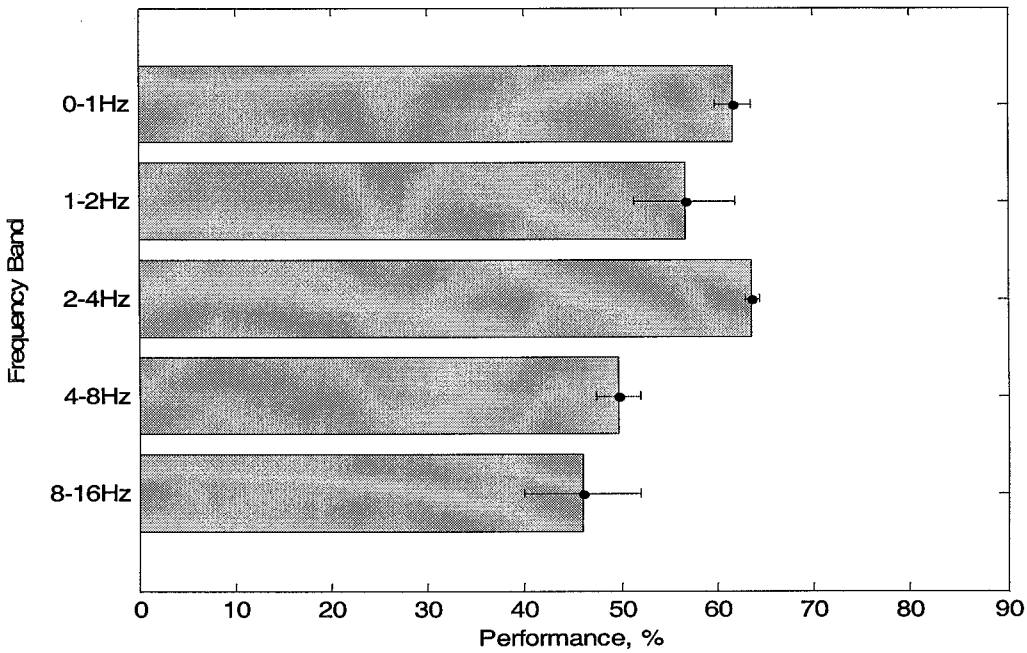

Figure 4.4 – Ensemble performance results trained on target-Pz coefficients.

Generalization performances obtained by using target-Pz coefficients are shown in Table 4.3. Average results with corresponding CI are illustrated in Figure 4.4. In this case, the frequency band (2-4 Hz) is most accurate. It is important to note for target signals, each electrode contributes its best distinguishing features from different frequency bands. Furthermore, even though the best performing electrode for target signals has been the Pz electrode, the differences are not statistically significant.

Tables 4.4-4.6 summarize the generalization performance for novel-Cz, novel-Fz, and novel-Pz feature sets respectively. The results are in the same format as Tables 4.1-4.3, where the best performing frequency band is highlighted. The performance results for novel tones are also illustrated in Figures 4.5-4.7.

Table 4.4 – Performance values on ensemble systems trained on novel-Cz coefficients for five frequency band.

Novel CZ	B-L1O	H_{10}	±CI	Sn	Sp	PPV	NPV
(0-1 Hz)	57.75%	**53.80%**	3.36%	54.12%	53.51%	51.72%	55.92%
(1-2 Hz)	60.56%	**58.31%**	3.18%	57.06%	59.46%	56.46%	60.10%
(2-4 Hz)	63.38%	**59.15%**	3.71%	52.94%	64.86%	58.13%	59.96%
(4-8 Hz)	54.93%	**50.42%**	6.23%	48.82%	51.89%	48.16%	52.54%
(8-16 Hz)	49.30%	**45.92%**	2.65%	46.47%	45.41%	43.75%	48.08%

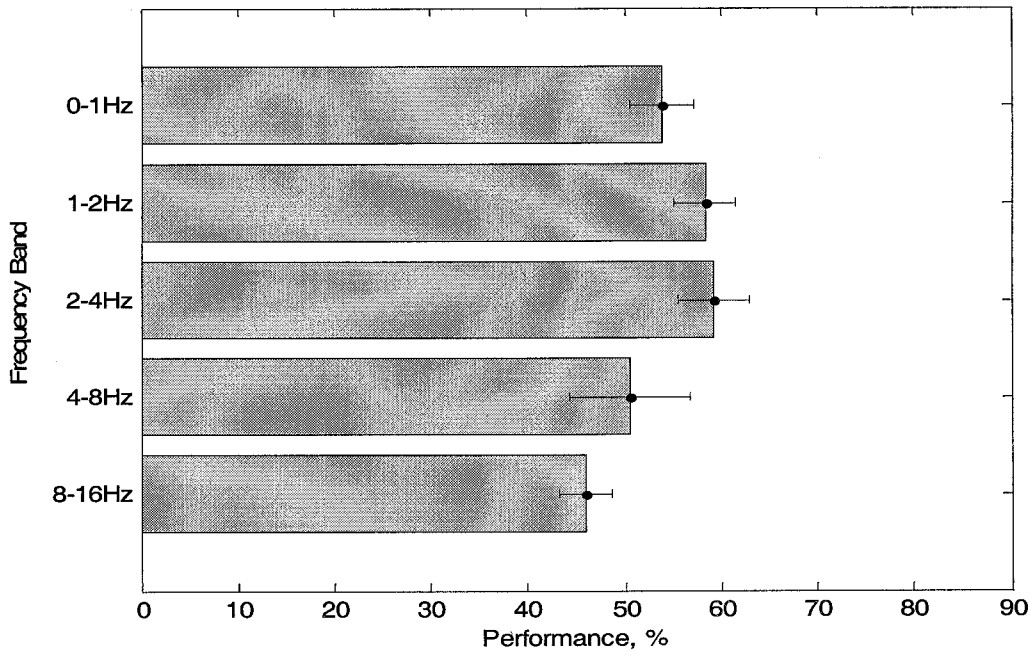

Figure 4.5 – Ensemble performance results trained on novel-Cz coefficients.

Table 4.5 – Performance values on ensemble systems trained on novel-Fz coefficients for five frequency bands.

Novel FZ	B-L1O	H_{10}	±CI	Sn	Sp	PPV	NPV
(0-1 Hz)	64.79%	**58.31%**	4.88%	55.88%	60.54%	56.47%	59.99%
(1-2 Hz)	60.56%	**55.77%**	4.03%	55.88%	55.68%	53.69%	57.88%
(2-4 Hz)	61.97%	**58.59%**	3.83%	55.88%	61.08%	56.90%	60.21%
(4-8 Hz)	52.11%	**49.01%**	4.69%	52.35%	45.95%	47.18%	51.08%
(8-16 Hz)	53.52%	**52.68%**	1.56%	54.71%	50.81%	50.57%	54.96%

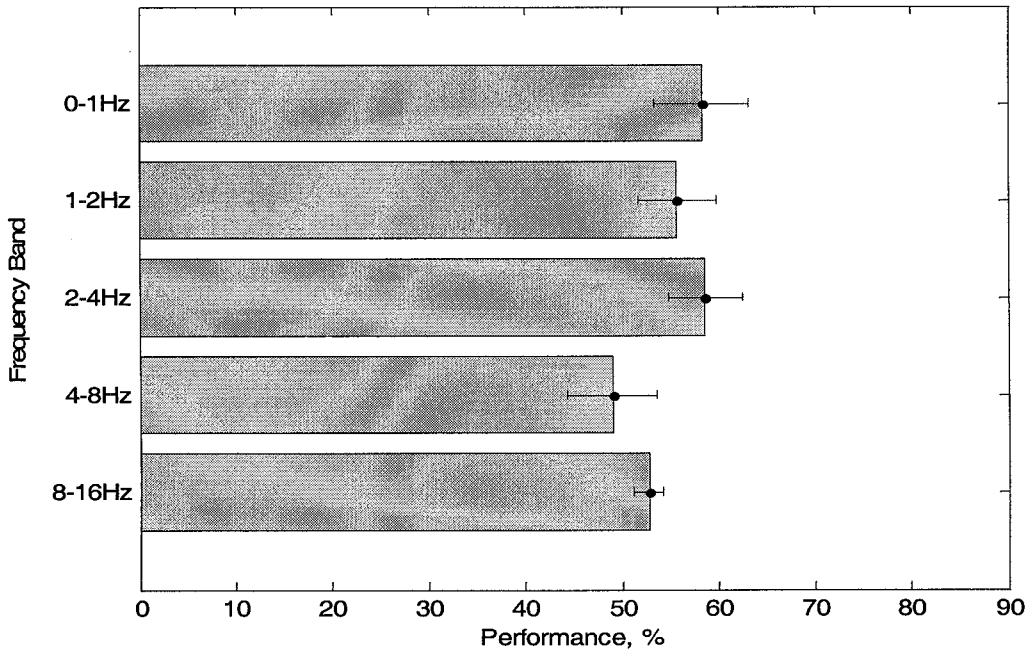

Figure 4.6 – Ensemble performance results trained on novel-Fz coefficients.

Table 4.6 – Performance values on ensemble systems trained on novel-Pz coefficients for five frequency bands.

Novel PZ	B-L1O	H_{10}	±CI	Sn	Sp	PPV	NPV
(0-1 Hz)	74.65%	**71.83%**	2.14%	65.88%	77.30%	72.72%	71.19%
(1-2 Hz)	74.65%	**72.68%**	2.35%	71.76%	73.51%	71.57%	73.96%
(2-4 Hz)	71.83%	**68.17%**	3.18%	64.12%	71.89%	67.65%	68.74%
(4-8 Hz)	70.42%	**66.48%**	5.16%	65.29%	67.57%	64.90%	68.06%
(8-16 Hz)	57.75%	**53.52%**	4.95%	51.76%	55.14%	51.39%	55.54%

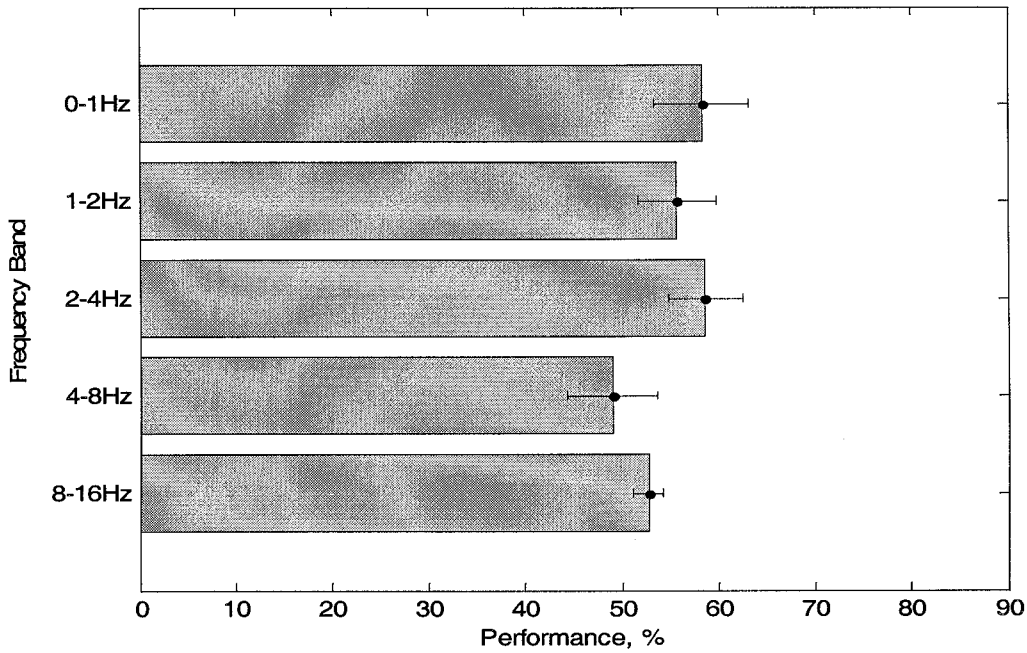

Figure 4.7 – Ensemble performance results trained on novel-Pz coefficients.

On average, the best performing novel-Cz and novel-Fz frequency bands do not out-perform the any of the target signal accuracies. On the other hand, performances on novel-Pz features have significantly achieved better results than each of the other ensembles trained on coefficients in response to target tones.

Even though the highlighted frequency band performs best in each table, other frequency bands contain relevant features for this problem. This can be seen in Table 4.6

and Figure 4.7, where the spectral band (0-1 Hz) does not show statistical difference in its performance, when compared to the (1-2 Hz) frequency range. From the above results, it is also evident that, target signals have a higher sensitivity, while novel signals have higher specificity accuracy. In other words, target signals perform better in correctly classifying AD patients as having the disease, while novel signals perform better in correctly identifying control patients as not having the disease. From these observations, we can justify the inference that a combination of the relevant features may result in a better overall performance.

4.2 – INTRA-SIGNAL DATA FUSION

While Learn++ was originally developed as an incremental learning algorithm, its ensemble structure allows it to be used in data fusion applications as well. This is because the algorithm can accept a new dataset even if it contains completely different features as compared to the data the algorithm had previously seen. When used in data fusion mode, Learn++ seeks to incrementally learn novel information content from datasets that come from the same application but are composed of different features. A comparison of this approach and the concatenation of features sets can be seen in the Appendix. The following results are based on the best performing combination of two frequency bands for each ERP signal.

Table 4.7 – Ensemble combination of best performing target-Cz coefficients.

Target CZ	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
(0-1 Hz)	63.38%	**62.54%**	1.56%	64.71%	60.54%	60.13%	65.18%
(2-4 Hz)	63.38%	**59.72%**	3.18%	60.59%	58.92%	57.64%	62.07%
DF (0-1,2-4Hz)	69.01%	**65.07%**	3.36%	65.88%	64.32%	62.93%	67.25%

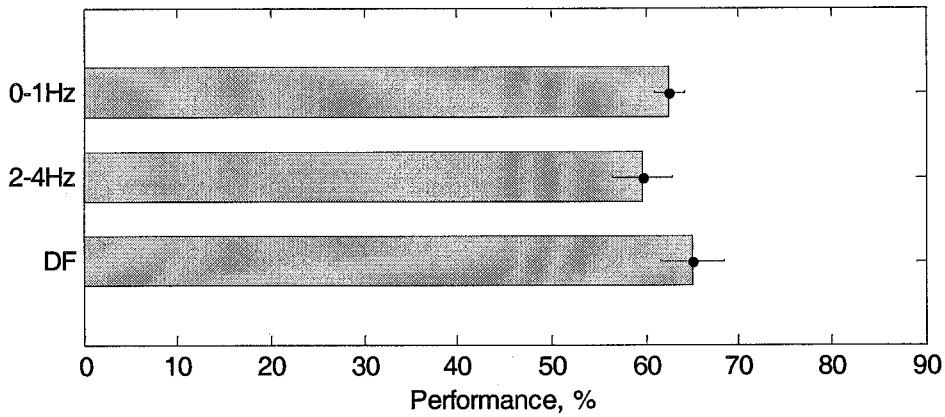

Figure 4.8 – Individual ensemble performance on target-Cz coefficient and Data Fusion combination.

The ensemble combination of target-Cz coefficients of frequencies (0-1 Hz) and (2-4 Hz) resulted in the best increase of accuracy, although the difference in performances is not statistically significantly. Table 4.7 and Figure 4.8 show the individual performance values with corresponding CI along with the data fusion results.

Table 4.8 – Ensemble combination of best performing target-Fz coefficients.

Target FZ	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
(1-2 Hz)	66.20%	**62.54%**	3.18%	62.35%	62.70%	60.61%	64.45%
(2-4 Hz)	63.38%	**58.03%**	5.85%	60.59%	55.68%	55.58%	60.72%
DF (1-4Hz)	67.61%	**63.66%**	5.16%	60.00%	67.03%	62.63%	64.63%

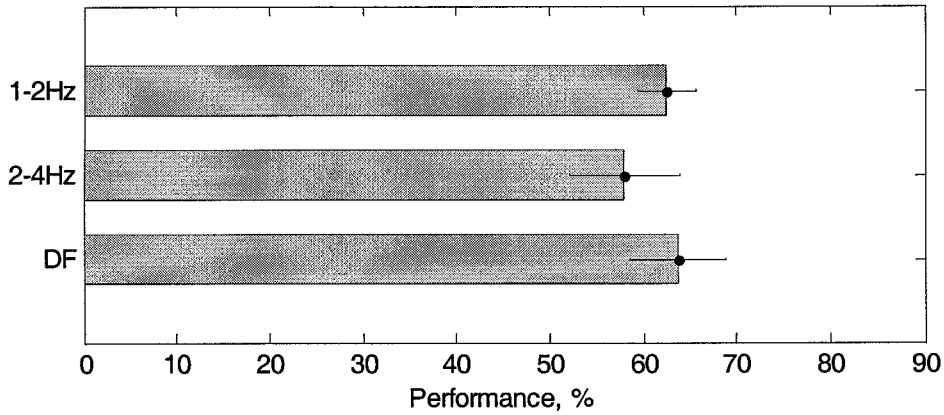

Figure 4.9 – Individual ensemble performance on target-Fz coefficients and Data Fusion combination.

Table 4.8 shows the results of ensemble data fusion between target-Fz coefficients at frequency bands of (1-2 Hz) and (2-4 Hz). This combination does not result in a statistically significant performance difference, as seen in Figure 4.9. Additionally, the fusion of these two feature sets sacrifices the sensitivity performance (correctly classified Alzheimer patients) for the increase of specificity performance (correctly classified control subjects).

Table 4.9 – Ensemble combination of best performing target-Pz coefficients.

Target PZ	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
(0-1 Hz)	63.38%	**61.69%**	1.92%	52.94%	69.73%	61.63%	61.73%
(1-2 Hz)	61.97%	**56.62%**	5.30%	54.12%	58.92%	54.78%	58.32%
(2-4 Hz)	64.79%	**63.66%**	0.78%	64.71%	62.70%	61.46%	65.98%
DF (1-4Hz)	69.01%	**67.89%**	1.46%	67.65%	68.11%	66.19%	69.64%
DF (0-1,2-4Hz)	69.01%	**67.32%**	1.92%	68.82%	65.95%	65.02%	69.74%

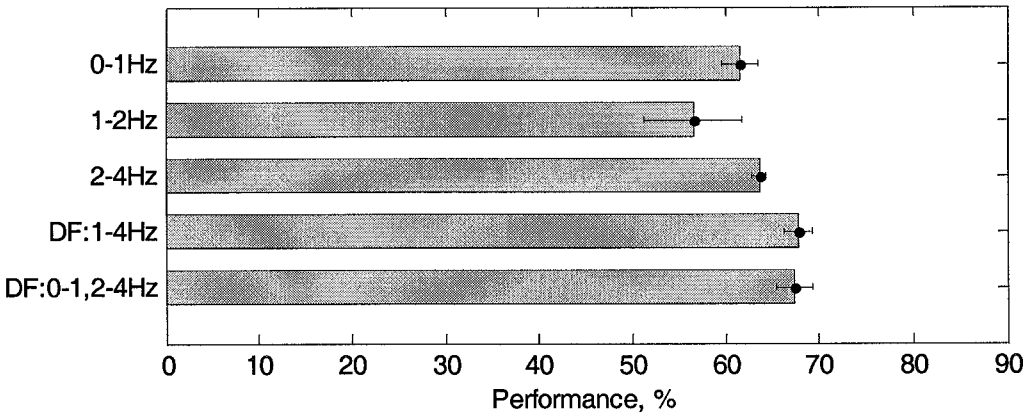

Figure 4.10 – Individual ensemble performance on target-Pz coefficients and Data Fusion combination.

The data fusion of ensembles trained on target-Pz coefficients, led to two well performing combinations, seen in Table 4.10. Many observations can be made by this experiment. First, the fusion between frequency bands (1-2 Hz) and (2-4 Hz) resulted in an increase in both sensitivity and specificity. In this case, the ensemble fusion of coefficients between (1-4 Hz) produced a statistically significant increase compared to the individual ensembles, as seen in Figure 4.10. On the other hand, the fusion between frequency bands (0-1 Hz) and (2-4 Hz), although resulting in a comparable overall accuracy, sacrificed specificity accuracy to increase its sensitivity. More specifically, even though there were more correctly classified AD patients, the number of correctly classified control patients decreased.

Table 4.10 – Ensemble combination of best performing novel-Cz coefficients.

Novel CZ	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
(0-1 Hz)	57.75%	**53.80%**	3.36%	54.12%	53.51%	51.72%	55.92%
(1-2 Hz)	60.56%	**58.31%**	3.18%	57.06%	59.46%	56.46%	60.10%
DF (0-2Hz)	63.38%	**56.34%**	5.39%	55.29%	57.30%	54.09%	58.60%

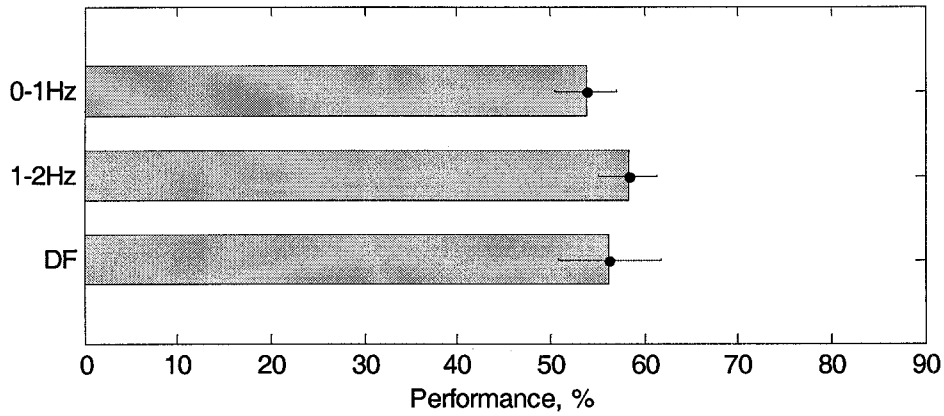

Figure 4.11 – Individual ensemble performance on novel-Cz coefficients and Data Fusion combination.

Results from the ensemble combination of the best performing novel-Cz coefficients, in Table 4.10, show that data fusion does not always increase performance. In this case, a single feature set containing frequencies (0-1 Hz) and (1-2 Hz) performs better than its combination with (0-2 Hz) coefficients, see Figure 4.11. This experiment shows that if there is no relevant information in each element of data fusion, the combination does not produce improved results.

Table 4.11 – Ensemble combination of best performing novel-Fz coefficients.

Novel FZ	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
(1-2 Hz)	60.56%	**55.77%**	4.03%	55.88%	55.68%	53.69%	57.88%
(2-4 Hz)	61.97%	**58.59%**	3.83%	55.88%	61.08%	56.90%	60.21%
DF (1-4Hz)	61.97%	**59.72%**	2.65%	55.29%	63.78%	58.48%	60.81%

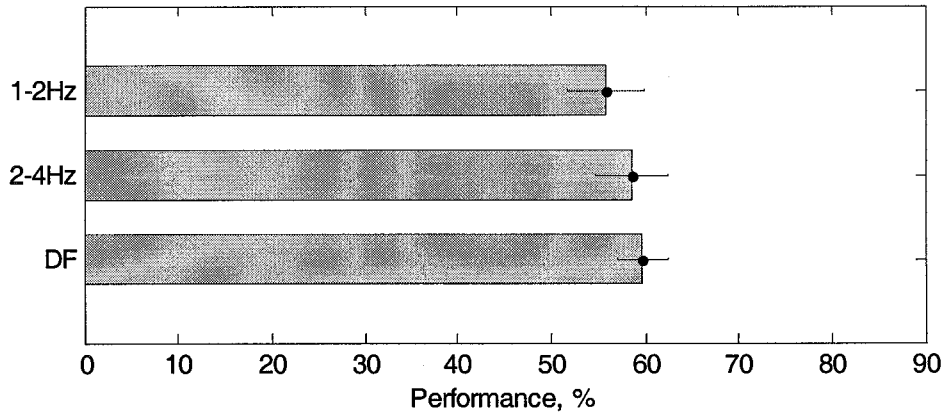

Figure 4.12 – Individual ensemble performance on novel-Fz coefficients and Data Fusion combination.

Table 4.11 reflects the ensemble combination performance of novel-Fz coefficients. There is only a slight difference, which can be seen in Figure 4.12, in the fusion of coefficients from frequency bands (1-2 Hz) and (2-4 Hz). In this case, there is an increase in specificity at the cost of sensitivity. In other words, the fusion correctly classified additional cognitively normal patients, but misclassified AD patients that were previously correctly classified.

Table 4.12 – Ensemble combination of best performing novel-Pz coefficients.

Novel PZ	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
(0-1 Hz)	74.65%	**71.83%**	2.14%	65.88%	77.30%	72.72%	71.19%
(2-4 Hz)	71.83%	**68.17%**	3.18%	64.12%	71.89%	67.65%	68.74%
DF (0-1,2-4Hz)	78.87%	**77.75%**	1.46%	75.29%	80.00%	77.66%	77.94%

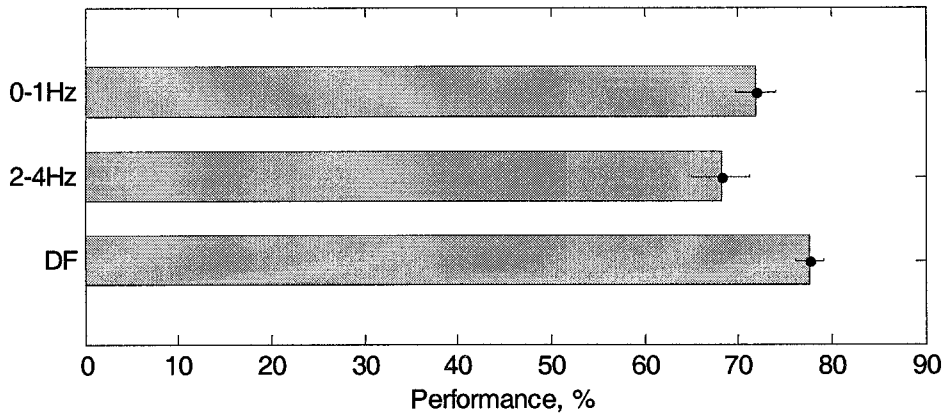

Figure 4.13 – Individual ensemble performance on novel-Pz coefficients and Data Fusion combination.

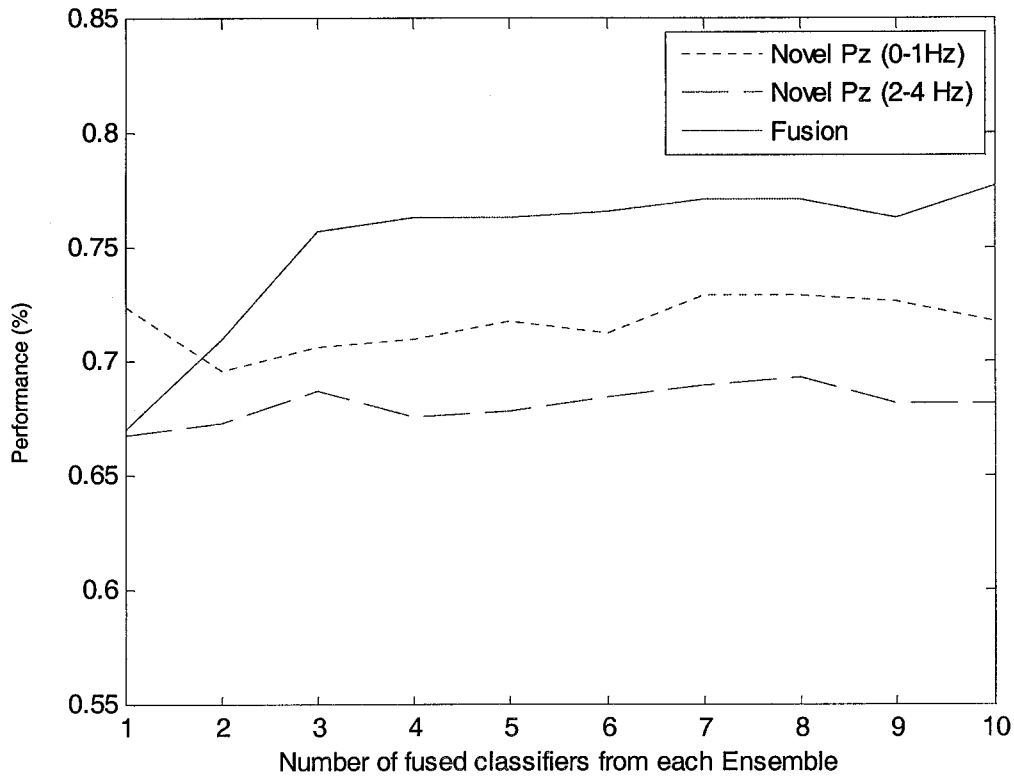

Figure 4.14 –Data fusion performance from novel-Pz coefficients at (0-1 Hz) and novel-Pz coefficients at (2-4 Hz).

The data fusion between novel-Pz coefficients of frequency bands (0-1 Hz) and (2-4 Hz), resulted in a statistically significant increase, compared to both individual ensembles. This difference can be seen in Figure 4.13. Figure 4.14 illustrates the performance of two individual ensembles as well as their combination through data fusion. As previously noted, the final performance of 77.8% has been achieved using both 10-classifier ensembles, resulting in a twenty classifier system. Therefore, each number on the x-axis represents the number of classifiers each ensemble contributes to the fusion. The graph shows the improvement resulting from the addition of each classifier.

The advantages of data fusion can be seen when there is an increase in both sensitivity and specificity resulting in an overall increase in performance. Specifically, if there is an increase in both sensitivity and specificity, patients that were misclassified by individual ensembles are correctly classified after the combination of these ensembles. On the other hand, if the sensitivity accuracy increases while specificity decreases the fused ensemble system correctly classifies previously misclassified AD patients, but also misclassifies previously correctly classified control subjects. Experiments in Section 4.2 provide examples where the combination of the best performing ensembles does not result in the best fused system. For example, in Table 4.12, the best performing fused ensemble system was trained on feature sets containing frequencies (0-1 Hz) and (2-4 Hz), but the best performing single-ensemble contains frequencies (1-2 Hz), shown in Table 4.6.

4.3 – INTER-SIGNAL DATA FUSION

Unlike intra-signal fusion, where information from different spectral bands of the same signal are fused, in inter-signal data fusion, information is combined from different spectral bands of different signal sources. Section 4.2 focused on the fusion of ensembles trained from the same signal but with different frequency bands. Section 4.3 will focus on combining ensembles based on signals generated from novel and target toness. Section 4.2 experiments resulted in a best two-ensemble fused overall performance of 77.8% (average of 5 leave-one-out trials), with a single best leave-one-out performance of 78.9%. The combined ensemble performance of coefficients from novel tone responses between electrodes did not show an improvement when compared to the individual ensembles. The best performing single-ensemble classifiers were generated using novel-Pz coefficients, thus various combinations of these coefficients fused with target single-ensemble classifiers are compared.

Table 4.13 – Data Fusion performance for the ensemble combination of novel-Pz coefficients between (0-1 Hz) and target-Pz coefficients between (2-4 Hz).

	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
Novel PZ (0-1 Hz)	74.65%	**71.83%**	2.14%	65.88%	77.30%	72.72%	71.19%
Target PZ (2-4 Hz)	64.79%	**63.66%**	0.78%	64.71%	62.70%	61.46%	65.98%
Data Fusion	78.87%	**73.24%**	4.28%	70.59%	75.68%	72.78%	73.79%

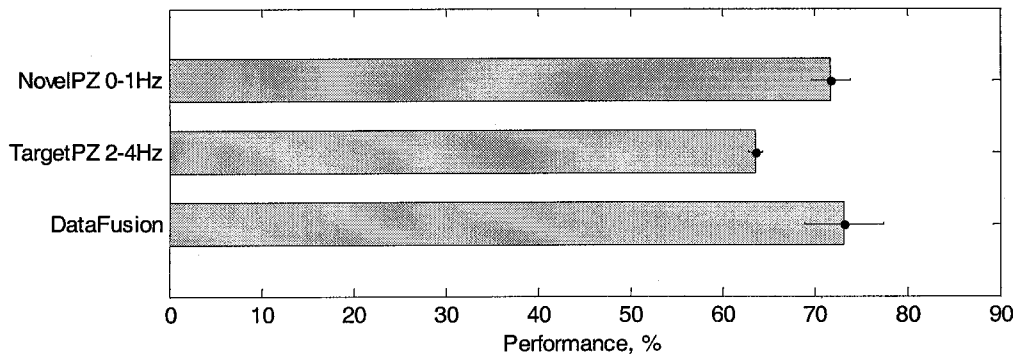

Figure 4.15 – Data Fusion performance of novel-Pz coefficients at (0-1 Hz) and target-Pz coefficients at (2-4 Hz).

Table 4.13 and Figure 4.15 summarize the fused generalization performance of novel-Pz and target-Pz coefficients at (0-1 Hz) and (2-4 Hz) respectively. There is a slight increase in the performance due to an increased sensitivity measure, but also a decrease in the specificity compared to the individual ensembles.

Table 4.14 – Data Fusion performance for the ensemble combination of novel-Pz coefficients between (1-2 Hz) and target-Pz coefficients between (2-4 Hz).

	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
Novel PZ (1-2 Hz)	74.65%	**72.68%**	2.35%	71.76%	73.51%	71.57%	73.96%
Target PZ (2-4 Hz)	64.79%	**63.66%**	0.78%	64.71%	62.70%	61.46%	65.98%
Data Fusion	78.87%	**76.34%**	1.92%	78.82%	74.05%	73.69%	79.19%

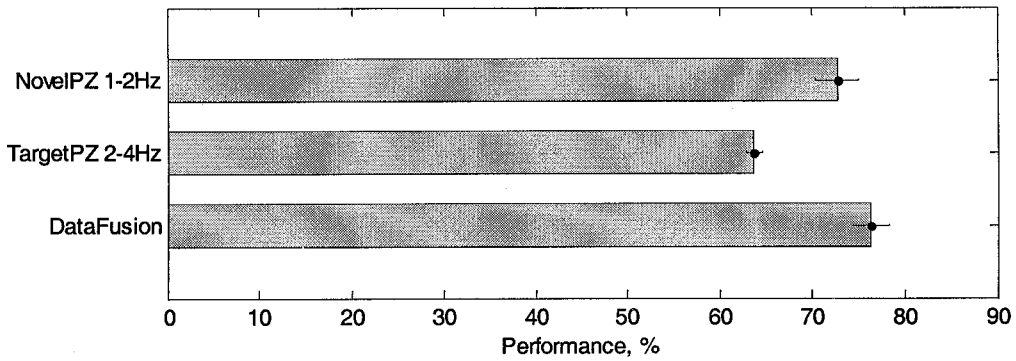

Figure 4.16 – Data Fusion performance of novel-Pz coefficients at (1-2 Hz) and target-Pz coefficients at (2-4 Hz).

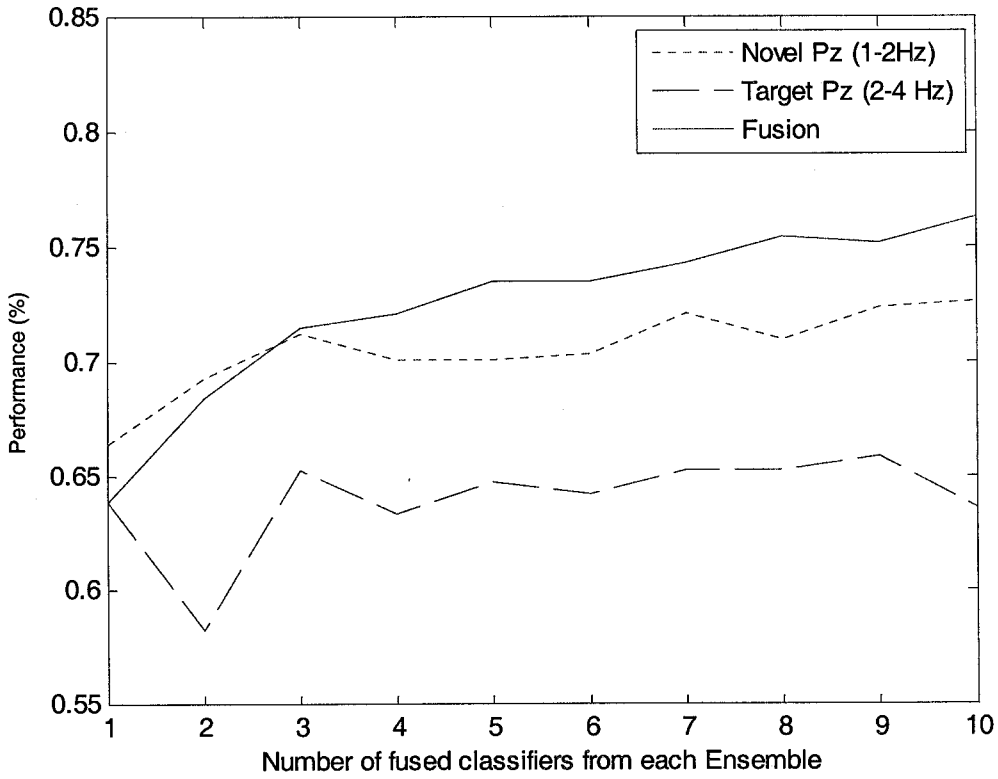

Figure 4.17 – Data fusion performance from novel-Pz coefficients at (1-2 Hz) and target-Pz coefficients at (2-4 Hz).

The combination of novel-Pz coefficients at (1-2 Hz) with target-Pz coefficients at (2-4 Hz), resulted in a noteworthy increase of overall performance. Furthermore, the fused performance is comparable to the best intra-signal accuracy of 77.8%. Figure 4.16 illustrates the performance results in Table 4.14 including respective confidence intervals. Similar to Figure 4.14, Figure 4.17 also demonstrates the point that each classifier added to the ensemble might not make a difference within its own ensemble, but contributes to the overall fusion. In this case, the first classifier, trained on target-Pz coefficients at (2-4 Hz) performs about the same as the entire ensemble at 64%. In contrast, its decision-making significantly increases the overall accuracy of the fused system.

Table 4.15 – Data Fusion performance for the ensemble combination of novel-Pz coefficients between (2-4 Hz) and target-Pz coefficients between (2-4 Hz).

	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
Novel PZ (2-4 Hz)	71.83%	**68.17%**	3.18%	64.12%	71.89%	67.65%	68.74%
Target PZ (2-4 Hz)	64.79%	**63.66%**	0.78%	64.71%	62.70%	61.46%	65.98%
Data Fusion	74.65%	**69.01%**	4.46%	69.41%	68.65%	67.10%	70.98%

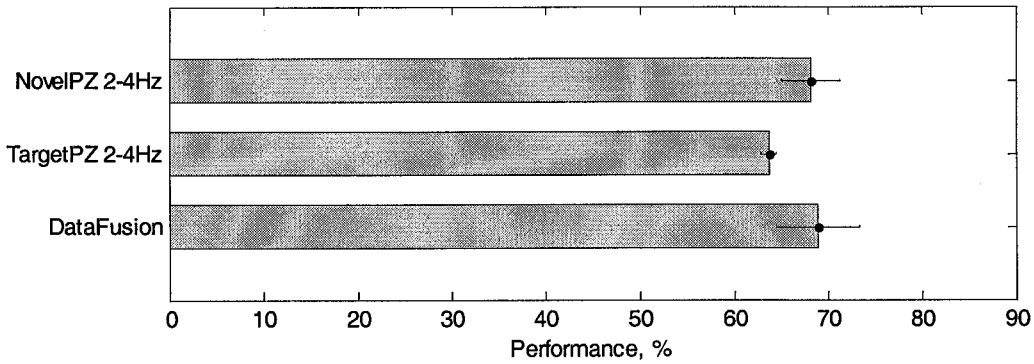

Figure 4.18 – Data Fusion performance of novel-Pz coefficients at (2-4 Hz) and target-Pz coefficients at (2-4 Hz).

Table 4.15 and Figure 4.16 present a slight performance increase, after the fusion of ensembles generated with novel-Pz coefficients and target-Pz coefficients of the same frequency range (2-4 Hz). Additional various combinations did not show improvement of the data fusion systems compared to the individual ensembles.

4.4 – FOUR-ENSEMBLE SYSEMS

Experiments from both Sections 4.2 and 4.3 produced similar performance values. Simulations in Section 4.2, comparing intra-signal combinations, resulted in a best data fusion performance of 77.8%, when combining novel-Pz coefficients at (0-1 Hz) and novel-Pz coefficients at (2-4 Hz). Tests in Section 4.3, comparing inter-signal

combinations, resulted in a best data fusion performance of 76.3 %, when combining novel-Pz coefficients at (1-2 Hz) and target-Pz coefficients at (2-4 Hz). The sensitivity values are 75.3% and 78.8% from the best experiments in Sections 4.2 and 4.3, respectively. The specificity measures are 80% and 74% respectively. The next intuitive experiment would suggest combining all four ensembles to increase the generalization performance.

Table 4.16 – Four-ensemble data fusion.

	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
NP(0-1 Hz, 2-4 Hz)	78.87%	**77.75%**	1.46%	75.29%	80.00%	77.66%	77.94%
NP(1-2Hz),TP(2-4Hz)	78.87%	**76.34%**	1.92%	78.82%	74.05%	73.69%	79.19%
DF- 4 ensemble	78.87%	**77.18%**	1.46%	72.94%	81.08%	77.98%	76.56%

Table 4.16 includes the results of data fusion with four ensembles. The four ensembles used were the best performing combinations from sections 4.2 and 4.3. In this case, the combination did not result in an increase of generalization performance. A slight increase of specificity resulted in the combined system at the cost of sensitivity performance.

4.5 – ENSEMBLE DATAFUSION

Results from Section 4.1 are based on 10-classifier systems. In most cases, the performance of the entire system is not significantly better than a single classifier. In other words, the first classifier of the system produces comparable results to the 10-classifier system. Although additional classifiers, in an ensemble, do not provide improvement with this application, the number of classifiers in a data fusion setting does

affect the accuracy of the system. Graphical representations of the best results from Sections 4.2 and 4.3 are further analyzed below.

In both experiments, illustrated in Figures 4.14 and 4.17, the addition of the 10th classifier increases the overall data fusion performance. The upward slope of the data fusion performance suggests that additional classifiers could continue to enhance the performance.

Table 4.17 – Data Fusion performance of 20-classifer ensembles trained on novel-Pz coefficients at (0-1 Hz) and target-Pz coefficients at (2-4 Hz).

	B-L1O	H_{10}	CI	Sn	Sp	PPV	NPV
Novel PZ (0-1 Hz)	74.65%	**72.96%**	1.92%	67.06%	80.54%	76.01%	72.77%
Novel PZ (2-4Hz)	71.83%	**68.17%**	3.18%	64.12%	71.89%	67.65%	68.74%
20-classifier-DF	81.69%	**79.15%**	1.92%	75.29%	82.70%	80.08%	78.47%

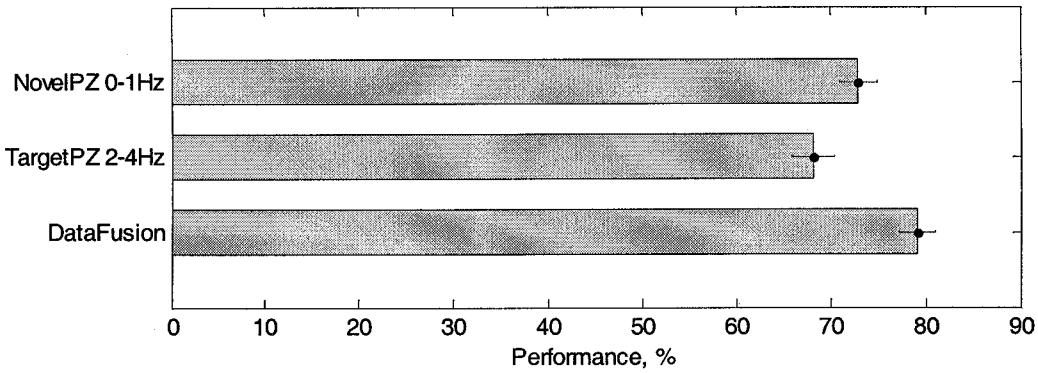

Figure 4.19 – Data Fusion performance of 20-classifer ensembles trained on novel-Pz coefficients at (0-1 Hz) and target-Pz coefficients at (2-4 Hz).

Table 4.17 tabulates the results of two 20-classifier ensembles, trained on novel-Pz coefficients at (0-1 Hz) and target-Pz coefficients at (2-4 Hz), and their decision fused combination. Figure 4.19 illustrates the performance of each ensemble, the combined system along with confidence intervals for each.

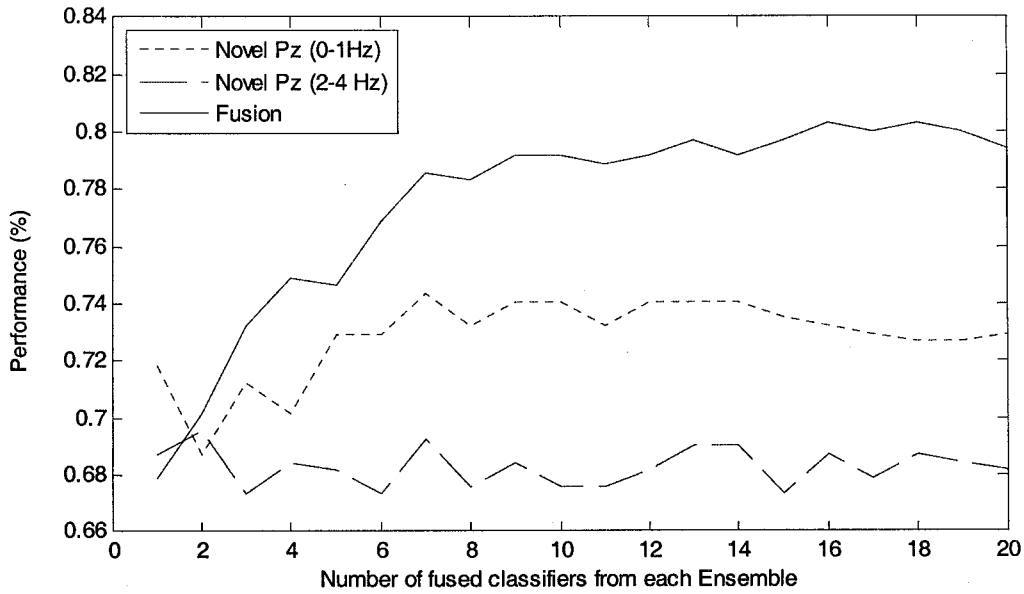

Figure 4.20 – Data fusion performance from novel-Pz coefficients at (0-1 Hz) and target-Pz coefficients at (2-4 Hz) using 20-classifer systems.

This experiment resulted in an overall increase of generalization performance (five leave-one-out trials) of 79.2%, as well as the best single leave-one-out trial at 81.7%. The system resulted in medical diagnostic measures of 75.3% for sensitivity, 82.7% for specificity and 80.1% for PPV. The increasing trend in Figure 4.14 is extended in Figure 4.20, with the combination of two 20-classifier ensembles. The fusion performance in Figure 4.20 begins to stabilize after the addition of about 16 classifiers. A slight decrease is noticed with the addition of the 20^{th} classifier, which is believed to be a result of the low number of trials. Furthermore, the peak average performance exceeds 80% with data fusion using 18-classifier ensembles. This system also produced the maximum single leave-one-out performance of 83.1%. The overall performance exceeds the 75% diagnostic performance of trained physicians at community based healthcare clinics. This comparison suggests that the procedures and approaches described in this work show promise in the development of a tool for the early detection of AD.

CHAPTER 5 – CONCLUSIONS

The application of this work consisted of an automated approach for the early diagnosis of Alzheimer's disease. The approach has used a non-invasive and cost-effective biomarker whose measure can be made available in a community healthcare clinic setting. The difficulty of this problem lies in the early diagnosis of the disease, during which obvious symptoms are not evident in the patients. Strict inclusion and exclusion criteria were followed in the subject recruitment to remove any unwanted bias during the signal acquisition. The procedure includes an oddball paradigm, used to invoke event related potentials in the EEG signals. ERP signals from the task related target tones and unexpected novel tones are recorded then analyzed using multi-resolution wavelet decomposition. The decomposition results in a set of coefficients for successive frequency bands. Feature level fusion and decision level fusion were compared in combining information from the multiple feature sets. On average, decision level fusion out-performed the individual ensembles and feature level fusion. Therefore, each set of coefficients were used to train an ensemble of classifiers. Classification was then achieved by combing multiple ensembles of classifiers trained on different feature sets. This technique, as presented, determines the discriminatory information in the ERP signals and uses a decision fusion to increase generalization performance. It was found that novel tones at 1-2 Hz, acquired from the Pz electrode, provide the best single-ensemble performance. Furthermore, the best intra-signal fusion performance again came from the novel-Pz coefficients, combining frequency bands of (0-1 Hz) and (2-4 Hz). Similar performance values resulted in the inter-signal combination of the target-Pz coefficients at (2-4 Hz) and the novel-Pz coefficients in the (1-2 Hz) frequency range. In

many cases the addition of classifiers in a single ensemble did not increase the generalization performance as expected. On the other hand, when the two ensembles were combined with data fusion, the additional number of classifiers in each ensemble did contribute to the increased performance of the entire system. The upward trend of data fusion performance led to the addition of more classifiers ultimately increasing the overall performance.

6.1 ACCOMPLISHMENTS

The objectives of this work included the analysis of ERP signals for the early diagnosis of Alzheimer's disease. The approach demonstrated the feasibility of a non-invasive, simple to perform and inexpensive procedure of an EEG recording as a diagnostic tool for AD. The contributions that were presented in this work are outlined below:

1) A multi-resolution wavelet analysis on ERP signals in response to target and novel tones from three EEG electrodes: Cz, Fz, Pz. Each combination of tone and electrode was analyzed separately at five spectral bands to find which individual feature set contained the most relevant information.
2) A decision level data fusion of ensembles was investigated with intra-signal combinations to include novel information from multiple spectral bands. Similarly, inter-signal combinations were considered to include additional relevance from multiple signals.
3) Further analysis on the fused system resulted in an automated classification algorithm capable of outperforming the overall accuracy of community healthcare clinics in diagnosing AD.

6.2 FUTURE WORK

The approach described in this work involves various combinations of ERP signals measured at different electrode locations. The final algorithm resulted in a diagnostic tool with a high specificity in distinguishing cognitively normal subjects from AD patients. Future work should assess the performance of the algorithm in distinguishing between AD and different types of non-AD patients, specifically patients with other forms of dementia such as VaD or Parkinson's disease. Furthermore, a detailed analysis of various ERP signals could result in specific components that provide relevant information for distinguishing between other forms of dementia. Since the algorithm had a high specificity of 82%, it could also aide the community healthcare clinician in their diagnosis where the specificity is 53%.

The combination method described in this work, weighted each ensemble based on its performance on the training data. Future work should include a weight adjustment on each ensemble based on the diagnostic performances and the ensemble's decision. In other words, if an ensemble, trained on a particular feature set, is known to perform well in correctly classifying AD patients, its weight should be adjusted accordingly when the ensemble's decision is AD. An analysis should also determine which patients are misclassified by each feature set. This information would aide in the combination by including feature sets that correctly classify instances that were originally misclassified. Finally, a strategic contribution measure of all relevant components would provide the optimal combination rule for practical use of this diagnostic tool.

REFERENCES

[1] J. Murphy, "The nation's aging population deserves better medical care" [online], June 1998, available from World Wide Web: <http://www.brown.edu/Administration/News_Bureau/Op-Eds/Murphy.html>, Last accessed: 3-14-2006.

[2] Alzheimer's Australia, "Dementia – different types," [online], Feb 2006, available from World Wide Web: <http://www.betterhealth.vic.gov.au/bhcv2/bhcarticles.nsf/pages/Dementia_different_types?OpenDocument>, Last accessed: 3-14-2006.

[3] Alzheimer's Disease International, "About Alzheimer's Disease," [online], Jan 2006, available from World Wide Web: <http://www.alz.co.uk/alzheimers/faq.html#howmany> Last accessed: 3-14-2006.

[4] Alzheimer's Association, "Basic Facts and Statistics," [online], 2006 available from World Wide Web: <http://www.alz.org/Resources/TopicIndex/BasicFacts.asp>, Last accessed: 3-14-2006.

[5] Help-guide, "Aging Issues- Expert, Non-Commercial Information on Mental Health & Lifelong Wellness," [online], 2006, available from World Wide Web: <http://www.helpguide.org/elder/alzheimers_dementias.htm>, Last accessed: 3-14-2006.

[6] American Health Assistance Foundation, "Alzheimer's Disease Research," [online], 2006, available from World Wide Web: <http://www.ahaf.org/alzdis/about/AmyloidPlaques.htm>, Last accessed: 3-14-2006.

[7] Lieberman, A. "Alzheimer's Disease," [online], 2005, available from World Wide Web: <http://www.liebermanparkinsonclinic.com/content/view/415/1/> Last accessed: 3-14-2006.

[8] Janssen, Cilag, "Round-the-Clock Resources for Dementia," [online], March 2006, available from World Wide Web: <http://www.dementia.com/bgdisplay.jhtml?itemname=dementia_about>, Last accessed: 3-14-2006.

[9] Help-guide, "Aging Issues- Diagnosis and Risk," [online], 2006, available from World Wide Web: <http://www.helpguide.org/elder/alzheimers_diagnosis_risk_factors.htm> Last accessed: 3-14-2006.

[10] Morris JC, Price AL. "Pathologic correlates of nondemented aging, mild cognitive impairment, and early-stage Alzheimer's disease," J Mol Neurosci. 2001, Oct;17(2):101-18.

[11] Goldman WP, Price JL, Storandt M, et al. "Absence of cognitive impairment or decline in preclinical Alzheimer's disease," *Neurology*. 2001 Feb 13;56(3):361-7.

[12] Reckess, G.Z., "Antecedent Biomarkers in Alzheimer's Disease," [online], Nov 2003, available from World Wide Web: <http://www.alzforum.org/res/enab/workshops/biomarkers.asp> Last accessed: 3-14-2006.

[13] Klein, B., "Nanosensor could help Alzheimer's diagnosis," [online], May 2004, available from World Wide Web: <http://nanotechweb.org/articles/news/3/5/6/1> 11 May 2004, Last accessed: 3-14-2006.

[14] Andreasen, N., Vanmechelen, E., Van de Voorde, A., Davidsson, P., Hesse, C., Tarvonen, S., Raiha, I., Sourander, L., Winblad, B., Blennow, K., "Cerebrospinal fluid tau protein as a biochemical marker for Alzheimer's disease: a community based follow up study," J Neurol Neurosurg Psychiatry. 1998 Mar;64(3):288. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9527138&dopt=Abstract

[15] Montine TJ, Beal MF, Cudkowicz ME, et al. "Increased CSF F2-isoprostane concentration in probable AD," *Neurology*. 1999 Feb;52(3):562-5.

[16] Yao Y, Zhukareva V, Sung S, Clark CM, Rokach J, Lee VM, Trojanowski JQ, Pratico D. "Enhanced brain levels of 8,12-iso-iPF(2α)-VI differentiate AD from frontotemporal dementia," *Neurology*. 2003 Aug 26;61(4):475-478.

[17] Praticò D, Lee VMY, Trojanowski JQ, et al. "Increased F2-isoprostanes in Alzheimer's disease: evidence for enhanced lipid peroxidation in-vivo," FASEB J. 1998 Dec;12(15):1777-83.

[18] T. Sunderland, R.E. Gur, and S.E. Arnold, "The Use of Biomarkers in the Elderly: Current and Future Challenges, Biological Psychiatry," 58, 272-276 (8-15-2005).

[19] M.J. de Leon and W. Klunk, "Biomarkers for the early diagnosis of Alzheimer's disease," *The Lancet Neurology*, 5, 198-199 (2006).

[20] Jack CR Jr, Dickson DW, Parisi JE, et al. "Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia," *Neurology*. 2002 Mar 12;58(5):750-7

[21] Bobinski M, de Leon MJ, Wegiel J, et al. "The histological validation of postmortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease," *Neuroscience*. 2000;95(3):721-5.

[22] C. Babiloni, G. Binetti, E. Cassetta, D. Cerboneschi, "Mapping distriputed sources of cotical rhythms in mild Alzheimer's disease. A multicentric EEG study," *NeuroImage*, vol. 22, pp. 57-67, 2004.

[23] G. Rodriguez, F. Copello, P. Vitali, G. Perego, F. Nobili, "EEG spectral profile to stage Alzheimer's disease," *Clinical Neurophysiology*, vol. 110, pp. 1831-1837, 1999.

[24] E. Pucci, N. Belardinelli, G. Cacchio, M. Signorino, F. Angeleri, "EEG power spetrum differences in early and late onset forms of Alzheimer's disease," *Clinical Neurophysiology*, vol. 110, pp. 621-631, 1999.

[25] E. Pucci, G. Cacchio, R. Angeloni, N. Berlardinelli, G. Nolfe, M. Signorino, F. Angeleri, "EEG spectral analysis in Alzheimer's disease and different degenerative dementias," *Archives of Gerontology and Geriatrics*, vol. 26, pp 283-291, 1998.

[26] M. Signorino, E. Pucci, E. Brizioli, G. Cacchio, G. Nolfe, F. Angeleri, "EEG power spectrum typical of vascular dementia in subgroup of Alzheimer patients," *Archives of Gerontology and Geriatrics*, vol. 23, pp. 139-151, 1996.

[27] M. Signorino, E. Brizioli, L. Amadio, N. Belardinelli, E. Pucci, F. Angeleri, "A EEG power index (eyes open vs. eyes closed) to differentiate Alzheimer's from vascular dementia and healthy ageing," *Archives of Gerontology and Geriatrics*, vol. 22, pp 245-260, 1996.

[28] J. Polich, L. Hoffman, "Alzheimer's disease and P300: evaluation of modality and task difficulty," *Brain Topography Today*, pp. 527-536, 1998.

[29] M.J. Hogan, G. Swanwick, J. Kaiser, M. Rowan, B. Lawlor, "Memory-related EEG power and coherence reductions in mild Alzheimer's disease" *International Journal of Psychophysiology*, vol. 49, pp. 147-163, 2003.

[30] 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" *Clinical Neurophysiology*, vol. 112, pp. 1378-1387, 2001.

[31] J. Polich, P300 and Alzheimer's disease, Biomed.Pharmacother., 43, 493-499 (1989).

[32] D.S. Goodin, "Clinical utility of long latency; event-related potentials (P3): the pros, Electroencephalography and Clinical Neurophysiology," 76, 2-5 (1990).

[33] J.M. Ford, N. Askari, J.D.E. Gabrieli, D.H. Mathalon, J.R. Tinklenberg, V. Menon, and J. Yesavage, "Event-Related Brain Potential Evidence of Spared Knowledge in Alzheimer's Disease," *Psychology and Aging*, 16, 161-176 (2001).

[34] J.M. Olichney and D.G. Hillert, "Clinical applications of cognitive event-related potentials in Alzheimer's disease," *Phys.Med.Rehabil.Clin.N.Am.*, 15, 205-233 (2004).

[35] D.J. Linden, The p300: where in the brain is it produced and what does it tell us?, *Neuroscientist*, 11, 563-576 (2005).

[36] S. Yamaguchi, H. Tsuchiya, S. Yamagata, G. Toyoda, and S. Kobayashi, "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, 111, 195-203 (2-1-2000).

[37] J. Polich, C. Ladish, and F.E. Bloom, "P300 assessment of early Alzheimer's disease," *Electroencephalogr.Clin.Neurophysiol.*, 77, 179-189 (1990).

[38] Y. Mochizuki, M. Oishi, and T. Takasu, "Correlations between P300 components and regional cerebral blood flows," *Journal of Clinical Neuroscience*, 8, 407-410 (2001).

[39] B.A. Ardekani, S.J. Choi, G.A. Hossein-Zadeh, B. Porjesz, J.L. Tanabe, K.O. Lim, R. Bilder, J.A. Helpern, and H. Begleiter, "Functional magnetic resonance imaging of brain activity in the visual oddball task," *Cognitive Brain Research*, 14, 347-356 (2002).

[40] E.J. Golob, J.K. Johnson, and A. Starr, "Auditory event-related potentials during target detection are abnormal in mild cognitive impairment," *Clinical Neurophysiology*, 113, 151-161 (2002).

[41] N.A. Phillips, H. Chertkow, M.M. Leblanc, H. Pim, and S. Murtha, "Functional and anatomical memory indices in patients with or at risk for Alzheimer's disease," *J.Int.Neuropsychol.Soc.*, 10, 200-210 (2004).

[42] B.A. Ally, G.E. Jones, J.A. Cole, and A.E. Budson, "The P300 component in patients with Alzheimer's disease and their biological children," *Biological Psychology*, In Press, (2006)

[43] E.J. Golob and A. Starr, "Effects of stimulus sequence on event-related potentials and reaction time during target detection in Alzheimer's disease," *Clinical Neurophysiology*, 111, 1438-1449 (8-1-2000).

[44] C. Mulert, L. Jager, O. Pogarell, P. Bussfeld, R. Schmitt, G. Juckel, and 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 Supplement D, 17-20 (2002).

[45] J. Benvenuto, Y. Jin, M. Casale, G. Lynch, and R. Granger, "Identification of diagnostic evoked response potential segments in Alzheimer's disease," *Exp.Neurol*, 176, 269-276 (2002).

[46] C.D. Morgan and C. Murphy, "Olfactory event-related potentials in Alzheimer's disease," *J.Int.Neuropsychol.Soc*, 8, 753-763 (2002).

[47] R.M. Chapman, G.H. Nowlis, J.W. McCrary, J.A. Chapman, T.C. Sandoval, M.D. Guillily, M.N. Gardner, and L.A. Reilly, "Brain event-related potentials: Diagnosing early-stage Alzheimer's disease," *Neurobiology of Aging*, In Press, Corrected Proof, (2006).

[48] C. Huang, L. Wahlund, T. Dierks, P. Julin, B. Winblad, and V. Jelic, "Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study," *Clinical Neurophysiology.*, 111, 1961-1967 (2000).

[49] V. Knott, E. Mohr, C. Mahoney, and V. Ilivitsky, "Quantitative electroencephalography in Alzheimer's disease: comparison with a control group, population norms and mental status," *Psychiatry Neuroscience*, 26, 106-116 (2001).

[50] C. Besthorn, H. Sattel, F. Hentschel, S. Daniel, R. Zerfass, and H. Forstl, "Quantitative EEG in frontal lobe dementia," *J.Neural Transm.Suppl*, 47, 169-181 (1996).

[51] S.Y. Cho, B.Y. Kim, E.H. Park, J.W. Kim, W.W. Whang, S.K. Han, and 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, Sep 17-21 2003, Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, vol. 3, pp. 2655-2658

[52] M. Lindau, V. Jelic, S.E. Johansson, C. Andersen, L.O. Wahlund, and O. Almkvist, "Quantitative EEG abnormalities and cognitive dysfunctions in frontotemporal dementia and Alzheimer's disease," *Dementia Geriatric Cognitive Disorder*, 15, 106-114 (2003).

[53] T. Kai, Y. Asai, K. Sakuma, T. Koeda, and K. Nakashima, "Quantitative electroencephalogram analysis in dementia with Lewy bodies and Alzheimer's disease," *Journal of the Neurological Sciences*, 237, 89-95 (10-15-2005).

[54] D. Solo, R. Hornero, P. Espino, J. Poza, C.I. nchez, and R. de la Rosa, "Analysis of regularity in the EEG background activity of Alzheimer's disease patients with Approximate Entropy," *Clinical Neurophysiology*, 116, 1826-1834 (2005).

[55] K. Bennys, G. Rondouin, C. Vergnes, and J. Touchon, "Diagnostic value of quantitative EEG in Alzheimer's disease," *Neurophysiologie Clinique/Clinical Neurophysiology*, 31, 153-160 (2001).

[56] V. Kolev, T. Demiralp, J. Yordanova, A. Ademoglu, and U. Alkac, "Time-frequency analysis reveals multiple functional components during oddball P300," *Neuroreport*, 8, 2061-2065 (5-27-1997).

[57] A. Ademoglu, T. Demiralp, J. Yordanova, V. Kolev, and M. Devrim, "Decomposition of event-related brain potentials into multicomponents using wavelet transform," *Applied Signal Processing*, 5, 142-151 (1998).

[58] E. Basar, M. Schurmann, T. Demiralp, C. Basar-Eroglu, and A. Ademoglu, "Event-related oscillations are 'real brain responses' - wavelet analysis and new strategies," *International Journal of Psychophysiology*, 39, 91-127 (2001).

[59] T. Demiralp, A. Ademoglu, Y. Istefanopulos, C. Basar-Eroglu, and E. Basar, "Wavelet analysis of oddball P300," *International Journal of Psychophysiology*, 39, 221-227 (2001).

[60] T. Demiralp and A. Ademoglu, "Decomposition of event-related brain potentials into multiple functional components using wavelet transform," *Clinical Electroencephalography*, 32, 122-138 (2001).

[61] R.Q. Quiroga, O.W. Sakowitz, E. Basar, and M. Schurmann, "Wavelet Transform in the analysis of the frequency composition of evoked potentials," *Brain Research Protocols*, 8, 16-24 (2001).

[62] R.Q. Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clinical Neurophysiology*, 114, 376-390 (2003).

[63] S. Aviyente, L.A.W. Brakel, R.K. Kushwaha, M. Snodgrass, H. Shevrin, and W.J. Williams, "Characterization of Event Related Potentials Using Information Theoretic Distance Measures," *IEEE Transactions on Biomedical Engineering*, 51, 737-743 (2004).

[64] B.H. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho, and K. Sundaresan, "An exploratory study of factors affecting single trial P300 detection," *IEEE Trans. on Biomedical Eng*, 51, 975-978 (2004).

[65] M. Fatourechi, S.G. Mason, G.E. Birch, and R.K. Ward, "A wavelet-based approach for the extraction of event related potentials from EEG, Proceedings - IEEE International Conference on Acoustics," *Speech, and Signal Processing*, vol. 2, pp. 737-740 (2004).

[66] A.K. Ozdemir, S. Karakas, E.D. Cakmak, D.I. Tufekci, and O. Arikan, "Time-frequency component analyser and its application to brain oscillatory activity," *Neuroscience Methods.*, 145, 107-125 (2005).

[67] E.M. Bernat, W.J. Williams, and W.J. Gehring, "Decomposing ERP time-frequency energy using PCA," *Clinical Neurophysiology*, 116, 1314-1334 (2005).

[68] M. Morup, L.K. Hansen, C.S. Herrmann, J. Parnas, and S.M. Arnfred, "Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, 29, 938-947 (2-1-2006).

[69] M. Karrasch, M. Laine, J.O. Rinne, P. Rapinoja, E. Sinerva, and C.M. Krause, "Brain oscillatory responses to an auditory-verbal working memory task in mild cognitive impairment and Alzheimer's disease," *International Journal of Psychophysiology*, 59, 168-178 (2006).

[70] S. Yagneswaran, M. Baker, and 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), Oct 23-26 2002, vol. 1, pp. 46-47

[71] R. Polikar, F. Keinert, and M.H. Greer, "Wavelet analysis of event related potentials for early diagnosis of Alzheimer's disease," Petrosian, A. and Meyer F.G. (Ed.) Wavelets in Signal and Image Analysis, From Theory to Practice, Kluwer Academic Publishers, Boston, 2001.

[72] Jacques G, Frymiare J, C. Kounios, C. Clark, and R. Polikar, "Multiresolution analysis for early diagnosis of Alzheimer's disease," 26th Annual Int. Conf. of IEEE Engineering in Medicine and Biology Soc.(EMBS2004), pp. 251-254

[73] A. Lim, W. Kukull, D. Nochlin, J. Leverenz, W. McCormick, J. Bowen, L. Teri, J. Thompson, E. Peskind, M. Raskind, and E. Larson, "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series," Journal of the American Geriatrics Society, 47, (1999), 564-569.

[74] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E.M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, 34, 939-944 (1984).

[75] J.R. Cockrell and M. Folstein, Mini Mental State Examination (MMSE), Psychopharmacology, 24, 689-692 (1988).

[76] R.M. Crum, Anthony J.C., Bassett S.S., and M. Folstein, "Population-based norms for the Mini-Mental State Examination by age and educational level," *Journal of American Medical Association*, 269, 2386-2391 (1993).

[77] Grossmann and Morlet. Decomposition of Hardy Functions into square integrable wavelets of constant shape. SIAM J. Math. Anal., 1984; 15: 723-736.

[78] C. Yamaguchi, "FFT and Wavelet Analyses of EEGs Measured During Simple Arithmetic Calculations and E-mail keying on Cell-phone Keypad," *International Journal of Sciences*, vol 15 (2005), pp. X-Y.

[79] B.V. Dasarathy and B.V. Sheela, "Composite classifier system design: concepts and methodology," *Proceedings of the IEEE*, 67, 708-713 (1979).

[80] L.K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 993-1001 (1990).

[81] R.E. Schapire, "The Strength of Weak Learnability," *Machine Learning*, 5, 197-227 (1990).

[82] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, 3, 79-87 (1991).

[83] J.A. Benediktsson and P.H. Swain, "Consensus theoretic classification methods," IEEE Transactions on Systems, *Man and Cybernetics*, 22, 688-704 (1992).

[84] D.H. Wolpert, Stacked generalization, Neural Networks, 5, 241-259 (1992).

[85] T.K. Ho, J.J. Hull, and S.N. Srihari, "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 66-75 (1994).

[86] Y. Freund and R.E. Schapire, "Decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, 55, 119-139 (1997).

[87] J. Kittler, M. Hatef, R.P.W. Duin, and J. Mates, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 226-239 (1998).

[88] L.I. Kuncheva, "Combining Pattern Classifiers, Methods and Algorithms", (2005).

[89] J. Ghosh, "Multiclassifier Systems: Back to the Future, 3rd Int. Workshop on Multiple Classifier Systems," in Roli, F. and Kittler, J. (Ed.), *Lecture Notes in Computer Science*, vol. 2364, pp. 1-15

[90] T.K. Ho, "Multiple classifier combination: Lessons and the next steps," in Kandel, A. and Bunke, H. (Ed.) *Hybrid Methods in Pattern Recognition*, World Scientific Publishing, 2002.

[91] L. Shapley and B. Grofman, "Optimizing group judgmental accuracy in the presence of interdependencies," *Public Choice (Historical Archive)*, 43, (1984), 329-343.

[92] P.J. Boland, "Majority system and the Condorcet jury theorem," *Statistician*, 38, (1989), 181-189.

[93] D. Berend and J. Paroush, "When is Condorcet's Jury Theorem valid?," Social Choice and Welfare, 15, (1998), 481-488.

[94] Parikh D., Stepenosky N., Topalis A., Green D., Kounios J. , Clark C. and Polikar R. , "Ensemble based data fusion for early diagnosis of Alzheimer's disease," Proc. of 27th Int. Conf. of IEEE Engineering in *Medicine and Biology* Soc., Shanghai, China, September 2005.

APPENDIX

There are two methods of data fusion for this application. The first method, referred to as feature level fusion, involves the concatenation of features sets resulting from the wavelet decomposition. This approach combines the coefficients of multiple frequency bands and uses them as inputs into the neural network. The second method, called decision level fusion, generates an ensemble of classifiers on each of the feature sets and combines the decisions of the ensembles for the final classification. Figure A.1 and A.2 illustrate the results of two experiments comparing the two methods of data fusion.

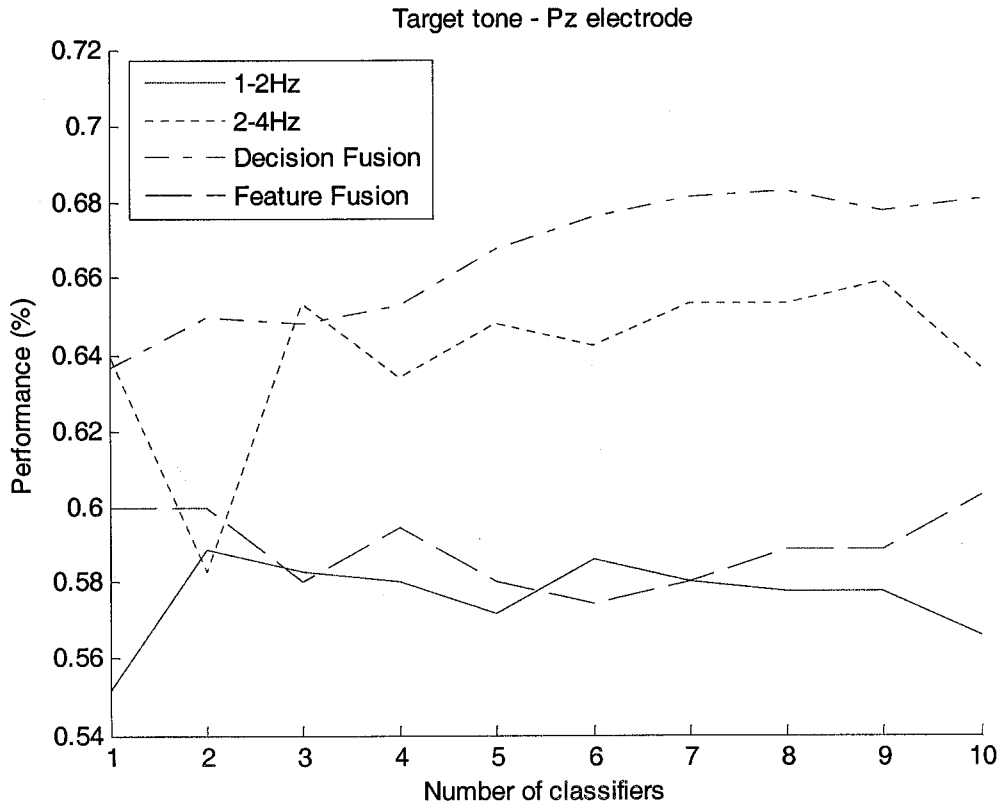

Figure A.1 – Performance comparison of decision and feature level fusion with target-Pz coefficients

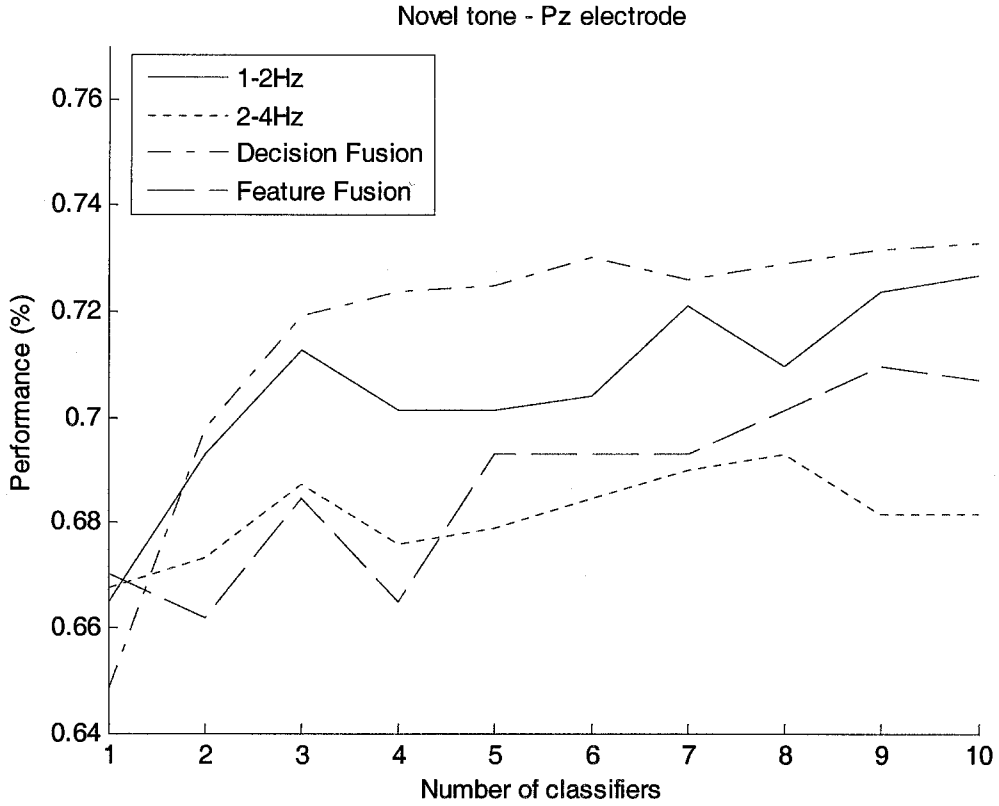

Figure A.2 – Performance comparison of decision and feature level fusion with novel-Pz coefficients

In Figure A.1, the performance of the target-Pz coefficients at (1-2 Hz) combined with (2-4 Hz) using feature level fusion does not out-perform the individual ensembles. On the other hand, the decision level fusion performs better than each of the ensembles. The performance results are similar in Figure A.2, where decision level fusion out-performs each of the ensembles and also feature level fusion. In feature level fusion, the features are concatenated creating a high dimensional problem. Decision level fusion, divides the dimensionality into smaller classification problems, and combines the decisions of the classifiers increasing the performance.