12-31-2004

Daubechies and quadratic b-spline wavelets for automated early diagnosis of Alzheimer's disease

Genevieve Jacques
Rowan University

Daubechies and Quadratic B-spline Wavelets for automated early diagnosis of Alzheimer's disease

by

Genevieve Jacques

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)

Members of the Committee:

_____ _____

In Charge of Major Work

_____ _____

For the Major Department

_____

For the College

Rowan University
Glassboro, NJ
2004

ABSTRACT

Genevieve Jacques
DAUBECHIES AND QUADRATIC B-SPLINE WAVELETS FOR AUTOMATED EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE
2004/05
Dr. Robi Polikar
Master of Science in Electrical Engineering

Alzheimer's disease is a neurological disorder characterized by nerve degeneration and neuronal death. The diagnosis of Alzheimer's disease at an early stage is a major concern due to the lack of a standard and effective diagnosis procedure available to community healthcare providers, as well as the increasing numbers of the elderly population affected. Clinical evaluation, the standard AD diagnostic procedure conducted at major university hospitals and research clinics, achieves, on average, a positive predictive value of 93%, with a sensitivity of 83%, specificity of 55% and an overall accuracy of 75%. An effective and objective tool for early diagnosis of the disease is important to have a meaningful impact on healthcare. Such a procedure must be inexpensive, non-invasive and available to community physicians, who often provide the first line of intervention, particularly at the early stages of the disease.

Studies performed using wavelets or other signal processing methods to analyze EEG signals in an attempt to find a non-invasive biomarker for Alzheimer's disease have had varying degrees of success. Two types of wavelets have commonly been used for analyzing the event-related potentials of EEG signals: Daubechies 4 wavelets and Quadratic B-spline

wavelets. Analysis was performed using these two types of wavelets and the coefficients obtained from this analysis were then used with several algorithms for automated classification. Cross validation was performed with a multilayer perceptron (MLP) neural network and an ensemble of MLP networks for classification of the coefficients. The classification algorithms are compared as well as the two types of wavelets. Noting that different information may be available from two different types of wavelets, a data fusion method with an ensemble of classifiers was implemented to combine relevant information and possibly boost performance of the algorithm.

The use of an automated algorithm is a feasible approach for diagnosing AD in a community health clinic. The accuracy of each method used in this study is similar to that of a clinical evaluation and as an automated algorithm provides a diagnostic tool for the detection of AD that can be made available to community physicians.

The results for data fusion indicate that the wavelets are not likely extracting complementary information from the signals since the combination of these two wavelet feature sets does not consistently produce a more informed decision than either set individually. The use of data fusion for combining features is however, a feasible approach to this classification problem if the chosen feature sets provide complementary information.

ACKNOWLEDGEMENTS

First of all, I would like to thank my family and friends for their support and understanding throughout the course of my graduate studies. I am very grateful for everyone who has been there for me through this challenging time in my life.

I would like to thank Dr. Polikar, my advisor, for his guidance throughout this project. Thank you very much for providing me with this opportunity. This project was part of a collaborative effort between the University of Pennsylvania, Drexel University and Rowan University. I would like to thank Dr. Clark from the University of Pennsylvania for his work with recruiting the patients for this project and Dr. Kounios and his associates from Drexel University for providing the EEG data. Without their efforts this project would not have been possible.

Special thanks to my committee members, Dr. Mandayam, Dr. Farrell and Dr. Clark. I would also like to thank Apostolos Topalis and Devi Parikh for their help in different areas of the research. Finally, I would like to thank the National Institute on Aging of the National Institutes of Health (under grant number P30 AG10124 - R01 AG022272) for supporting my graduate studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Early diagnosis of Alzheimer's disease (AD) is a major public health issue since it affects large portions of the elderly population. Clinical evaluation, the standard procedure for diagnosis of AD, has an overall accuracy of 75%, with a sensitivity of 83%, specificity of 55% and positive predictive value of 93% (for definitions see appendix A) [1]. Unfortunately, this approach requires the expertise of neuropsychologists and is available only at major research hospitals. There is currently no standard or effective diagnosis procedure available to community healthcare providers who essentially serve as the first line of intervention for the disease. To have a meaningful impact on healthcare, an effective and objective procedure for the diagnosis of AD must be inexpensive, non-invasive and available to community physicians. The goal of this study has been to create an automated algorithm for the early diagnosis of AD that can be made available to community health clinics and is as accurate as a clinical evaluation.

Magnetic resonance imaging (MRI), urinalysis, spinal fluids and analysis of electroencephalographic (EEG) signals are among different methods that have been considered diagnostic tools, however the results remain inconclusive. More recently, several studies have been performed using wavelets and other signal processing methods to analyze EEG signals in an attempt to establish a biomarker for AD. These attempts have shown varying degrees of success. However, the nonstationary nature of such signals and the ability of wavelets to localize time-frequency information make wavelet analysis a potentially promising tool. Therefore the diagnostic value of wavelet analysis of EEG

signals is explored in this study. Two types of wavelets have been used: Daubechies 4 wavelets and Quadratic B-spline wavelets. The coefficients obtained from this analysis were then used to train an automated classifier. Cross validation was performed with a multilayer perceptron (MLP) neural network and an ensemble of MLP networks for classification of the coefficients. The classification algorithms are compared as well as the two types of wavelets. Noting that different information may be available from two different types of wavelets and that an ensemble method may perform better than a single classifier, a data fusion method with an ensemble of classifiers was also implemented to combine relevant information and boost performance of the algorithm.

1.1 ALZHEIMER'S DISEASE

Alzheimer's disease (AD) is the most common form of dementia, which is associated with nerve degeneration and neuron death. The disease was first described in 1906 by German physician Dr. Alois Alzheimer. Although the disease was considered rare at one time, research has shown that it is the leading cause of dementia, accounting for more than half of all dementia cases [2, 3]. The early stages of the disease show a person's recognition of their lost abilities and a demonstration of concern and anxiety due to these losses. As the disease progresses, problems are developed in attention and concentration to the point that simple tasks become huge obstacles [4].

Increasing age is the greatest risk factor for the development of AD. Approximately one in ten individuals over the age of 65 and nearly half of those over 85 are affected [2]. Approximately 4.5 million people in the United States have AD according to data based on the number of cases detected in an ethnically diverse population sample in

the 2000 U.S. census. By 2050, this number could range from 11.3 to 16 million cases in the U.S. as more people live to reach their 80's and 90's [2].

1.1.1 BIOLOGICAL ASPECTS OF ALZHEIMER'S DISEASE

As mentioned above, AD is associated with nerve degeneration and neuron death. The damage of the nerve cells typically begins with cells involved in learning and memory and then eventually spreads to cells that control every aspect of thinking, judgment, and behavior. As the disease progresses, the damage reaches the cells that control and coordinate movement.

The specific cause of AD is not clear, however the disease is characterized by the formation of amyloid plaques and neurofibrillary tangles in the brain. Amyloid plaques are clumps of protein fragments that accumulate outside of cells. Beta amyloid, a normally harmless protein, is believed to cause these deposits of plaque that form between neurons early in the disease process, before neurons begin to die and symptoms develop [5]. The role of the amyloid deposits as part of the pathology of the disease is uncertain due to a variant form of the disease which has no amyloid deposits present [3].

Neurofibrillary tangles, on the other hand, are clumps of altered proteins inside cells, namely the tau protein [2]. In AD, threads of tau protein undergo alterations that cause them to become twisted forming these tangles. Some researchers believe this may cause serious damage to neurons, causing them to die [5]. However, the role of tau in AD has also been questioned since mutations in the tau gene have been linked to a variety of neurodegenerative diseases other than AD [3].

Other research in effort to identify the cause of the disease has indicated that about 10% AD cases may be genetic, causing several cases of early-onset forms of Alzheimer's disease. Three proteins genetically predisposed to mutation, amyloid precursor (APP), presenilin 1 (PS1), and presenilin 2 (PS2) can result in the production of amyloid plaques [3, 5]. The cause of the disease is not necessarily as important as the fact that the course of the disease remains the same regardless of its cause [6].

1.1.2 DIAGNOSIS

Unfortunately, the only definitive means of diagnosing the disease is via an autopsy, so there has been substantial effort in finding a method for diagnosing this illness before the most debilitating stages of the disease set in. The current method for diagnosis is a series of clinical interviews with the subject and their caregiver, performed by a neuropsychologist. This method is used to determine if there are changes in cognitive status and basically assesses if the person probably has AD. It is difficult to determine if a person is suffering from AD, other forms of dementia (i.e. vascular dementia) or if the symptoms in question are those just associated with normal aging [4]. The positive predictive value of clinical diagnosis remains around 93%, while there is a sensitivity of 83%, specificity of 55% and overall accuracy of 75% [7] (See Appendix A for calculation of sensitivity, specificity and positive predictive value).

There are other methods being considered for diagnosing the illness such as a spinal tap to extract cerebrospinal fluid tau, a known biomarker of AD, however this is a fairly invasive procedure. Other methods such as magnetic resonance imaging (MRI) scans which can capture images of the lesions in the brain have also been used but are

very expensive procedures and not always available. The abnormalities in the brain tend to disrupt the brain's electrical signals and can be detected through electroencephalographic (EEG) signals which are obtained in a noninvasive manner and their acquisition is fairly inexpensive. EEG signals can provide information about the overall integrity of the cerebral cortex as well as giving evidence of specific disorders [4]. EEG signals are also more readily available for use in diagnosis and with the use of different processing techniques for the signals, may prove to provide more accurate results. EEG analysis has not been used for AD diagnosis, however, this is mainly due to difficulties in distinguishing between EEG changes that could be attributed to AD from those due to normal aging, other medical illnesses, and other factors associated with physiology [8].

A reliable method for diagnosing this illness early is needed so that medicines may be administered in a timely fashion, decreasing the chances of the illness progressing to the latter stages. Although there is currently no cure for Alzheimer's disease, researchers have made progress in developing treatments that may help improve the quality of life for people with AD, which makes early diagnosis all the more important [5].

1.2 ELECTROENCEPHALOGRAPHY (EEG)

One of the chief methods for determining the specific functions of particular elements of the nervous system is the analysis of electroencephalographic or EEG signals. EEG signals represent the electrical activity of the brain as voltage over time and are recorded using a series of electrodes connected to the scalp. These signals have been used in some cases to determine illness associated with brain activity, such as epilepsy. Changes in the brain's electrical activity can reflect changes in cognitive status and therefore one goal of this study is to determine whether changes in EEG signals reflect changes due to AD [5].

The specific alterations in the EEG signals of AD patients observed in different experiments will be outlined in section 1.2.4.

Richard Caton (1875) was the first to observe the EEG phenomenon with his work involving rabbits and monkeys, however Hans Berger, a neuropsychiatrist, is the noted discoverer of the human EEG. Berger started studying human EEGs in 1924 using various galvanometers and during the course of the years that followed he showed different features within the brain signals such as sleep spindles, fluctuations of consciousness, first evidence of the alpha rhythms, as well as evidence of a variety of disorders [9].

The EEG was developed as a method for investigating mental processes however clinical applications quickly became evident, particularly in epilepsy. EEG became more popular with the introduction of event-related potentials (ERPs) which are components of the EEG resulting from specific sensory and cognitive processes (further explained in section 1.2.4).

Many advances in EEG studies led to breakthroughs in neurophysiology and the idea that different neurological disorders could be explored further through the use of EEGs caused research to shift in that direction [9, 10].

1.2.1 EEG RECORDINGS

EEG recordings are acquired using electrodes placed in different locations on the scalp with electric potentials recorded between pairs of active electrodes (bipolar recordings) or with respect to a passive electrode or reference (monopolar recordings). These measures are primarily performed on the surface of the scalp (scalp EEG) but special electrodes can also be placed on the surface of the brain during a surgical operation (intracranial EEG). Better resolution can be achieved with intracranial implanted electrodes, however a sur-

gical procedure is required for placement of the electrodes, which makes it impractical for most human studies [10].

For scalp recordings, the standard system in use for electrode placement is the International 10-20 system, developed to keep a consistent placement for comparison of studies. The system involves a number of electrodes connected to key scalp locations generally referenced to 2 electrodes in the earlobes to obtain signals from particular regions in the brain.

Figure 1. 1 shows 20 electrodes and their positions on the scalp, however many more electrodes may be placed in between the labeled electrodes for a more thorough analysis. The notation F for frontal, C for central (cortex), P for parietal, T for temporal, O for occipital (and A for auditory reference), indicate the regions of the brain where the electrodes are placed. Convention calls for odd numbers on left and even numbers on right [9].

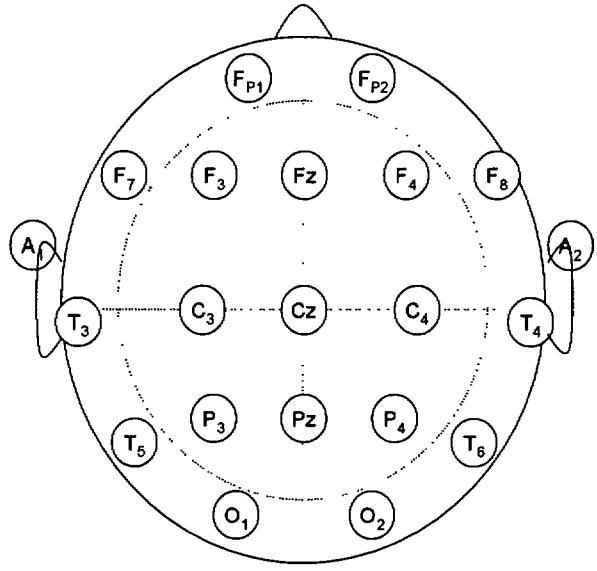


Figure 1. 1: International 10-20 system for placement of electrodes on the scalp [8, 9].

One problem with scalp electrodes is the artifacts that alter the EEG signal due to head movements, eye blinking, muscle activity, etc. Since the EEG signals have low amplitude, artifacts contaminate the recordings, often making analysis difficult. These artifacts are usually removed as part of preprocessing procedure. Synchronized and repeated signals are then averaged to make the components of interest within the signal more pronounced [10].

1.2.2 SPECTRAL CONTENT OF THE EEG

Researchers have been attuned to the study of different types of brain oscillations and their relationship with different pathologies since the beginning of electroencephalography. It has been declared that the oscillatory activity of the brain signals is a valid measure of cognitive processing [11]. An EEG signal can be broken down into different frequency bands that exist within the signal and each band has been shown to be associated with different brain functions. The five main frequency bands are as follows:

- The delta band (0.5 – 3.5 Hz) is characteristic of deep sleep stages and has been correlated with different pathologies [10]. The amplitude of the delta response is also increased during experiments using the oddball paradigm (the data collection protocol used in this study--see Section 1.2.3) hinting that the response may be linked to signal detection and the corresponding decision [12, 13].
- The theta band (3.5-7.5 Hz) has been correlated with higher cognitive and associative brain processes [13, 14]. The event-related potential (ERP) components (see Section 1.2.4) in the theta band are prolonged after presentation of the target stimulus in ex-

periments using the oddball paradigm indicating a relationship with selective attention [12, 13].

- The alpha band (7.5-12.5 Hz) is sometimes divided into subsets as alpha1 (7.5-10 Hz) and alpha2 (10-12.5 Hz). The oddball paradigm has been shown to influence the alpha responses in the P300 (Section 1.2.4) [12]. Some results indicate that the alpha oscillations are associated with working memory [12, 15]. Schurmann and Basar, 2001, extend that the alpha response is made up of parallel processes in relation to memory, movement and sensation [16].
- The beta band (12.5 – 30 Hz) is also divided into subsets of beta1 (12.5-20 Hz) and beta2 (20-30 Hz). The beta rhythms are stronger in central and frontal electrodes and are found to be enhanced during states of expectancy or tension [10].
- The gamma band (30 – 60 Hz) became more popular after experiments at the cellular level showed a relationship with the linking of stimulus features into perceived information (binding theory), but have not been of much interest otherwise [10]. Basar-Eroglu *et al.*, 1996 suggested that gamma-band activity is part of the common language elements of the brain, which may also be associated with mutual information transfer between subcomponents of the brain as it is with other oscillations such as theta, alpha or beta [17]. The gamma response is, however, likely to be linked with sensory perceptual phenomenon as indicated in a study by Karakas and Basar, 1998 [18]. Slobounov *et al.*, 2000, performed a study akin to the visual paradigm (see Section 1.2.3) which involved the recognition of non-stable postures of computer-animated human body models. Results indicated that the subject's specific judgment or awareness postural instability influenced gamma-band activity [19].

1.2.3 DATA COLLECTION PROTOCOLS

In certain protocols for EEG acquisition, the patient is usually exposed to a sensory stimulus in order to elicit a particular response, known as the event-related potential (see Section 1.2.4). There are three modalities typically used as stimuli: auditory, visual and somatosensory. For the auditory modality, the stimuli are single tones of a preset frequency, or clicks with a broadband frequency distribution. For the visual modality, stimuli are produced by a single light or sometimes by the reversal of a pattern such as that of a checkerboard. For the somatosensory, peripheral nerves are stimulated using electrical stimuli. Bimodal stimulation, used in different experiments, is the combination of the visual and auditory modalities [10].

Sequences of stimuli are arranged in paradigms in order to study the responses to tasks that can test such factors as memory, reaction time, awareness, etc. The tasks involved could be anything from a simple task such a pressing a button to memorization of extensive lists. Some of the most common paradigms are explained below:

1. No-task evoked potentials paradigm involves the subject in a relaxed state, instructed to perceive the stimuli without performing any task.

2. The **oddball paradigm** involves two different stimuli presented in a pseudorandom order. The oddball tone (or target tone) is presented randomly in a series of frequently occurring (standard) tones. The standard tone is presented in 75-80% of the trials, and the oddball stimuli in the remaining 20-25% of the trials. The oddball stimulus is usually a different frequency from the standard, set far enough apart to be distinguishable from the frequent stimulus [8, 10]. For example, for the visual modality, the oddball could be a different color flash of light. The oddball tones are generally fairly different from the

regular tones, however for some studies, the tones maybe close in frequency making the test 'harder' than usual [18]. The test subjects are instructed to do a simple task after hearing each oddball tone such as pressing a button, keeping a mental count of the number of oddball tones, etc [8, 10, 20].

3. Mismatch negativity paradigm involves a regular stimulus and a deviant stimulus like the oddball paradigm, but the subjects are asked to perform an irrelevant task not acknowledging the oddball tones, with attention devoted elsewhere [18].

4. The "three-stimulus" paradigm involves a typical oddball paradigm with novel or "distractor" stimuli randomly added. These novel tones consist of a disruptive sound such as dog barking, color forms, etc. that disrupt the regular oddball paradigm routine [21].

Yamaguchi *et al.*, 2000, developed a variation of this paradigm with the use of novel tones consisting of 60 unique environmental sounds, recorded from Disney movies and edited to be 200ms in duration. Frequent stimuli occur 65% of the time, oddball stimulus 20% and the novel tones 15%. Again, the subjects are asked to respond only to the oddball stimulus by performing a simple task defined at the onset of the experiment. This type of experiment is performed in efforts to differentiate between different types of dementia [22, 23].

1.2.4 EVENT RELATED POTENTIALS (ERPs)

The change in the ongoing EEG due to stimulation is called an event related potential (ERP) or sometimes an evoked potential (EP) [10]. ERPs are a series of positive and negative peaks that occur in response to a specific event to which the subject is usually

asked to respond (detecting the target tone). Each element of the ERP has a name that denotes its sign and its latency after the stimulus is perceived by the subject.

P50 is a positive peak occurring around 50ms following the stimulus. N1 is a negative peak occurring at approximately 100ms and according to a study by Golob and Starr, 2000, the amplitude and latency changes of this peak may be observed during memorization tasks [24]. Polich *et al.*, 1997, speculate that differences in N1 amplitude/latency between target and standard stimulus responses in an oddball paradigm may originate from attentional processing [23].

P2 is a positive peak at approximately 200ms after the stimulus. P2 response is stronger due to the standard stimulus in an oddball paradigm as opposed to target stimulus, implying that it contains a component due to sensitivity of the sensory processes other than cognitive processes [23].

N2 is a negative peak at approximately 200ms. This response is found to be stronger in response to target tones in the oddball paradigm, however, given its close proximity to the P3 component, it is hypothesized that the amplitude and latency of the N2 may be affected by the P3 generation [23]. P3 or P300 is a positive peak occurring around 300ms. The P300 has been shown to occur only in response to oddball tones and has also been associated with mental activity.

The P3 is measured by quantifying its amplitude and latency, where amplitude is defined as the voltage difference between a prestimulus baseline and the largest peak with latency between 250-400ms. The latency is the time measured from stimulus onset to the point of maximum positive amplitude within the particular latency window [23]. The latency and the amplitude of the P300 component have been shown to be related to

age and the mental ability of the individual [9, 23]. Figure 1. 2 shows a typical ERP in response to a target tone (b) and in response to a non-target tone (a). Note that the P300 appears in response to the target tone [20].

The P300 has a dominant delta response oscillation [11], which appears independent of the modality of the stimulus [25]. The P300 can be attributed to a manifestation of central nervous system activity involved with the processing of new information when attention is engaged in updating memory. The latency of the P300 in the discrimination task provides an indication of individual variability in mental processing capability and speed [8].

In an experiment by Katayama and Polich in 1999, 12 cognitively normal young adults were subject to EEG recording using a three stimulus oddball paradigm in both the visual and auditory modalities. The P300 component was largest over the parietal and midline electrodes and occurred in response to the target and non-target stimuli for both modalities [26].

In a study by Polich and Herbst, 2000, the P300 was found to be comparable to other clinical test procedures considering its underlying measurement and variability. The only catch is that the test procedures to obtain the measurement must be standardized, such that the methods and conditions under which the data are collected are identical. This is something that has yet to be achieved [27].

For experiments involving the novel tones, a second P300 is created. The target P3 or P3b is the traditional P300 with the strongest area of detection being the parietal electrode. The novelty P3 or P3a is in response to an alarming or novel stimulus and originates in the frontal region. The use of the novel tones is said to increase the P3b and,

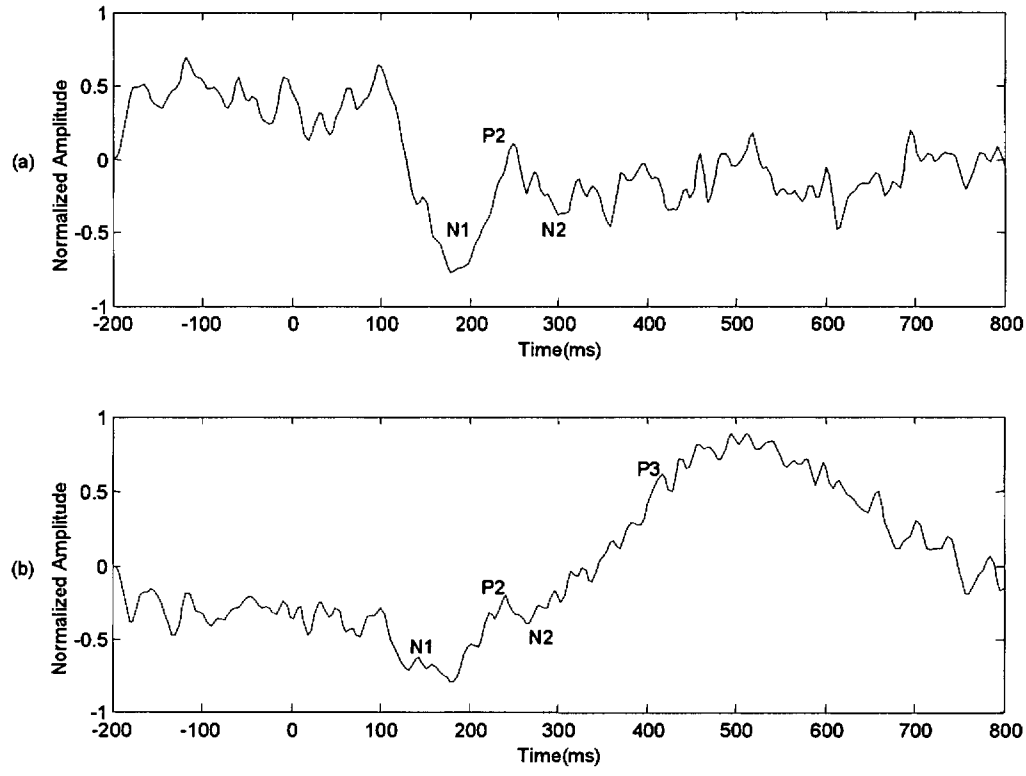


Figure 1. 2: ERPs, (a): Nontarget (standard) Response, (b): Target response.

of course, elicit a P3a. The P3a, however, is only readily observed in about 20% of normal subjects which, although this peak may be the most sensitive to changes in cognitive function, tends to limit its use [8, 21, 22].

Due to the low amplitude of the ERPs compared with the ongoing EEG, averaging several of the responses is a common practice in effort to visualize the ERP. The ERPs have a similar pattern of response which is more or less predictable under similar conditions [10]. Sufficient numbers of artifact-free trials have been shown to stabilize ERP measures (amplitude and latency) [21].

1.3 EEG FOR THE DIAGNOSIS OF ALZHEIMER'S DISEASE

Different studies have shown abnormalities in the EEG of AD patients. Mildly affected AD patients have increased theta activity compared to control groups and more severely affected patients show decreased alpha activity and increased delta activity [28]. AD patients have been found to have increased delta and theta power but decreased alpha and beta mean power. The severity of dementia has also been indicated in EEG rhythms where the delta and/or theta rhythms can be increased even in the earlier stages of AD and may predict the development of the disease [29, 30].

In an EEG study by Babaloni *et al.* a decline in the alpha1 activity in central, parietal, temporal and limbic region electrodes is apparent when comparing mild AD patients with vascular dementia patients and elderly control subjects. The authors conclude that this decline may be specifically linked with mild AD. Data also suggested that increased delta activity is not associated specifically with the slowing of cortical function which may be attributed to the impairment of cortical connectivity in AD [29]. A linear trend between theta activity and hippocampal atrophy in a study by Grunwald *et al.*, 2001, suggests a connection between the two [31].

Claus *et al.* performed a study solely on a group of AD patients with EEG and Computed Tomography (CT) scans taken initially and at a follow-up of 4 years indicated that decreased beta and decreased alpha activity are predictors of survival in patients with early AD [32]. Studies by Buchwald *et al.*, 1989 and Green *et al.*, 1992, have shown the P50 ERP to be diminished or nonexistent for AD patients [33, 34], however, Fien, Biggins and Van Dyke, 1994, contrast both previous studies by finding no changes in this peak due to AD [35].

The P300 as mentioned in the previous sections has been related to cognitive processes that require attentional allocation and immediate memory processes. It has been said that the P300 latency is prolonged and that the amplitude is decreased in AD patients and sometimes this can occur so that the peak is not at all obvious as shown in Figure 1. 3 [21, 36]. The P300 directly reflects currents triggered by cortical postsynaptic potentials and seems to be primarily generated in the temporo-parietal cortex which makes sense because this area shows pronounced synaptic loss in AD [37].

The P300, being derived from neural activity, is necessarily affected by the physical state of its underlying physiology. With this in mind, the P300 has been found to be affected by dementia but there are many factors that affect this particular peak. Figure 1. 4 shows a nearly nonexistent P300 for a normal subject and a prominent component for an AD subject.

With most studies, the factors known to affect the P300 are kept to a minimum, as much as can be controlled, such as no food intake prior to EEG, no medication of certain types can be taken within 48 hours of EEG, time of day/time of year consistent through group, etc. Some guidelines are put in place to keep variations due to factors outside the scope of the study to a minimum [8, 38].

Criteria to control the factors known to affect the P300 response in this study were added to limit the effects of factors other than AD on the P300. These factors are explained in Section 4.1 in Table 4.1. Even with such parameters in place, the P300, as shown in Figure 1. 4 is most likely affected by other factors beyond those controlled in this study. The P300 component is not analyzed specifically, however the frequency range in which the P300 and other ERP components occur are explored.

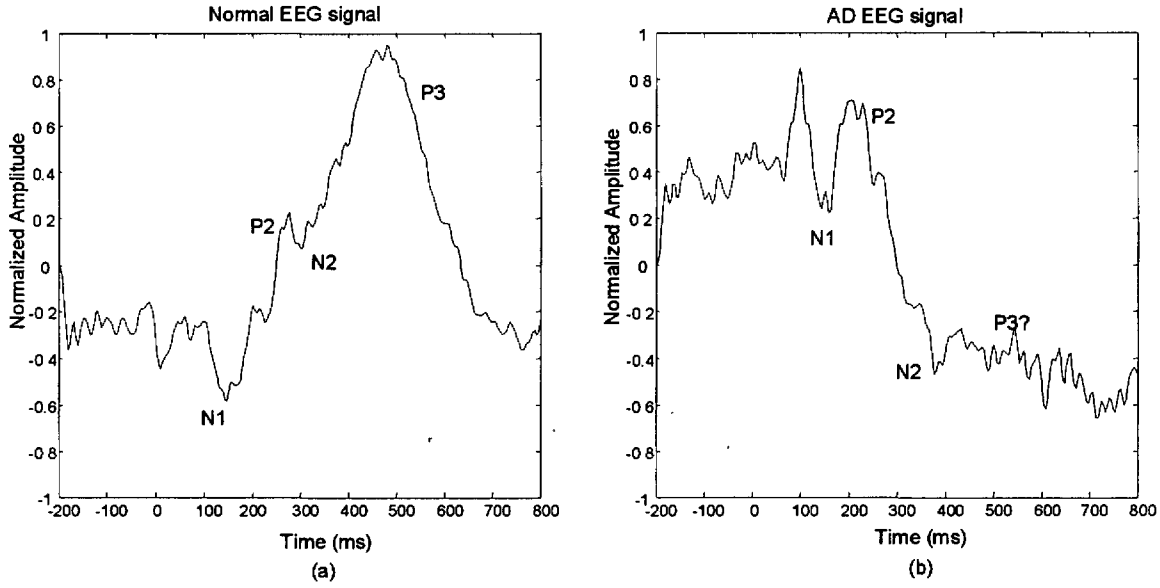


Figure 1. 3: (a) Normal subject EEG with obvious P3, (b) AD subject EEG with missing P3 component.

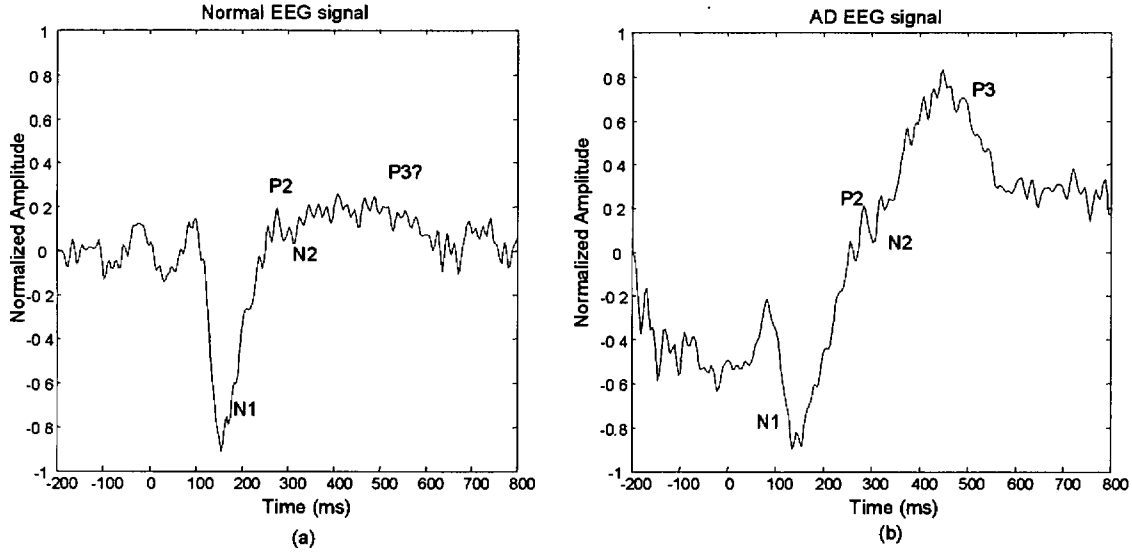


Figure 1. 4: (a) Normal subject EEG with P3 not obvious, (b) AD subject EEG with prominent P3 component.

1.3.1 SCOPE AND ORGANIZATION OF THESIS

1.3.1a SPECIFIC OBJECTIVES

1. Determine if multiresolution wavelet analysis is a feasible approach for the early diagnosis of Alzheimer's disease through the use two different types of wavelets where classification performance is determined with different pattern recognition techniques.
2. Compare the performance of the two different types of wavelets used in the analysis as well as the performance of the different pattern recognition algorithms.
3. Combine the two types of wavelets using a data fusion method in effort to provide a more informed decision for classification.
4. To create an automated algorithm for the early diagnosis of Alzheimer's disease through the analysis of EEG signals that is comparable in accuracy to a clinical diagnosis.

Multiresolution wavelet analysis, as will be shown in Section 2.3, is a suitable method for the analysis of nonstationary signals such as EEG signals. This indicates that wavelets are likely a feasible approach for the early diagnosis of Alzheimer's disease. Two different types of wavelets are used to analyze the signals due to their use in previous studies. It is initially unknown if one wavelet will perform better than the other.

The different pattern recognition techniques were also compared. The Learn++ algorithm had been shown in previous studies to perform about the same, if not better in some cases than a single MLP neural network. The results of these methods are compared and are described further in Chapter 5 and Chapter 6. The data fusion of the two types of wavelets was performed based on results from each of the types of wavelets individually. The two types of wavelets appeared to be extracting different information from the signal implying that the combination of the two wavelets may provide a more

informed classification decision. It was expected that the performance of the data fusion would be greater than the performance values from either wavelet individually.

1.3.1b ORGANIZATION

The organization of this thesis is as follows. The background detailing the reasons for using EEG signals for the diagnosis of AD has been described in this chapter. Chapter 2 describes previous use of several different techniques for the analysis of EEG signals to provide a background into the basis of this project. Several methods in analyzing EEG signals described here have been further explored to advance research in this area. Methods particularly in the time-frequency domain are utilized along with different automated classification techniques. The theory involved in these particular techniques is described in detail in Chapter 3.

The overall expectation of this research is to explore the feasibility of using wavelets with several pattern recognition techniques for the early diagnosis of AD. First of all, two methods for analyzing EEG signals are explored in detail in Chapter 3 (power spectral density and wavelet analysis with focus on two types of wavelets). Results for these methods using a single MLP neural network and using the Learn++ algorithm with an ensemble of MLP networks are shown in Chapter 5. After results are obtained for the different feature sets with the MLP and Learn++, the features are then combined using data fusion in order to determine whether a better classification performance than any of the previous methods on their own can be achieved. Figure 1. 5 shows an overview of the analysis while the details are discussed in Chapter 4

All simulations of the results indicate that wavelets, used with either Learn++ or a single MLP, provide a diagnostic performance comparable to that of a clinical diagnosis. A comparison of results for the two methods-single MLP and Learn++ is given in Chapter 5 along with the specific results. Data fusion of several features using an ensemble approach was expected to be able to give a more informed classification decision than a single feature set with a single classifier or ensemble of classifiers. Results are shown for the combination of the two wavelet feature sets with a data fusion method using an ensemble of classifiers approach. Overall, the results obtained in this study indicate that wavelet analysis is a feasible tool for the early detection of AD but also that the data fusion of several features (different wavelets or other feature) could be used to allow for a more informed decision for classification.

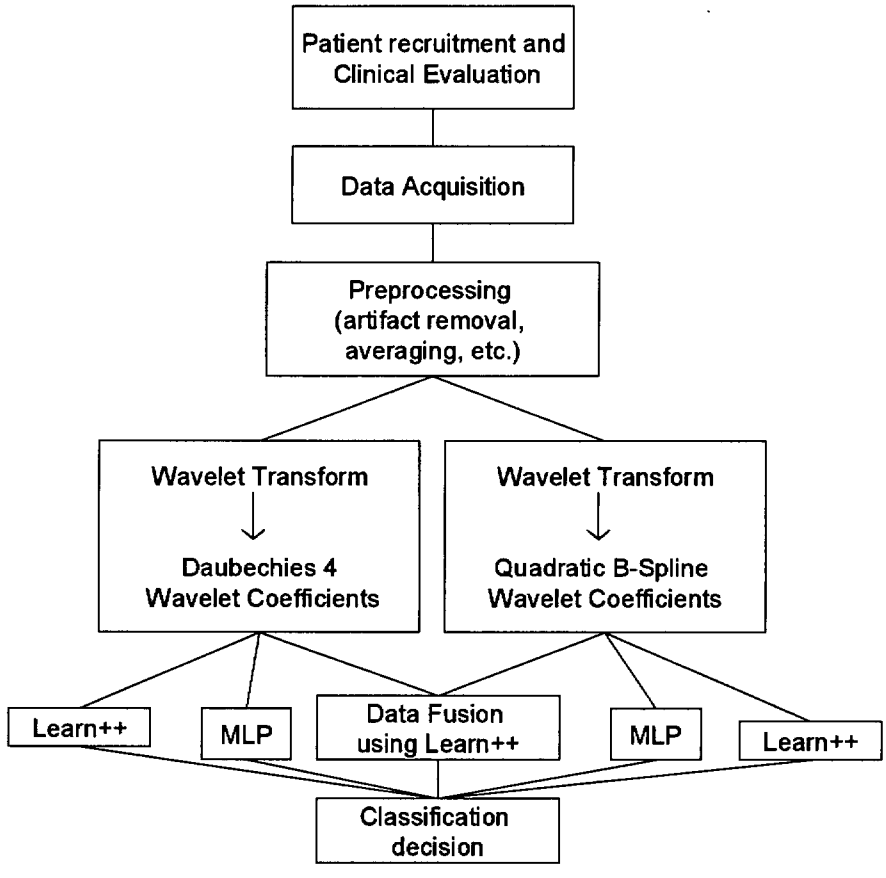


Figure 1. 5: Overview of project.

CHAPTER 2

BACKGROUND

The techniques described in this chapter have all been used in the analysis EEG signals. Some studies focus specifically on the analysis of ERPs, while others explore the frequency bands without focusing on the ERPs. Some of these techniques have been found to give promising results in EEG analysis for the detection of AD, however, many others have not been applied specifically for the detection of AD, such as the quadratic B-Spline wavelets. Several of these techniques were chosen for further exploration based on different factors, such as feasibility, dataset characteristics, etc. The power spectral density and two wavelets- Daubechies4 and Quadratic B-spline are further described in Chapter 3.

2.1 FREQUENCY DOMAIN TECHNIQUES

Spectral analysis based on the Fourier Transform has been the most widely used quantitative method for the analysis of EEG signals. It allows the separation and study of different EEG rhythms, which is difficult to perform visually since several rhythms occur simultaneously. Power spectral density and coherence are further discussed in Chapter 3, while dipole source analysis is briefly described in Appendix D.

2.1.1 POWER SPECTRAL DENSITY

This method involves the computation of the power spectral density for the individual frequency bands of interest in the signal. A formal definition of power spectral density

can be found in Section 3.1.2. Several studies have been performed using this method for detection of AD.

Pucci *et al.*, 1998, performed a study with 25 AD patients, 25 non-AD demented patients and 50 normal control subjects. EEG data were collected in resting eyes closed/eyes opened conditions and the data were subject to spectral analysis where the power was computed in five 5.5 Hz bands in the range of 1-28.5 Hz. Five unconventional frequency bands were explored: 1-6.5 Hz (B1), 6.5-12 Hz (B2), 12-17.5 Hz (B3), 17.5-23 Hz (B4), 23-28.5 Hz (B5). The absolute and relative powers were calculated for all frequency bands and derivations and then a logarithmic transform were applied. Three different power spectrum profiles were observed: (i) type A spectra, characterized by single or multiple peaks in the 6.5-12 Hz range, (ii) Type B spectra, characterized by increased power below 6.5 Hz and no peaks in the 6.5-12 Hz range, and (iii) type C spectra, having a low overall power and no dominant peaks in the 6.5-12 Hz range or elsewhere. They found a lack of dominant activity in the 6.5-12 Hz band in a significant number of the AD patients. In conclusion, the study provides further evidence that EEG spectral analysis allows for the identification of a particular group of AD patients from other AD patients or other types of dementia [39].

Pucci *et al.*, 1999, evaluated the relationship between various EEG spectral parameters and the age of onset of AD. The study involved EEG recordings during wakefulness with eyes opened and again with eyes closed. The test involved 150 AD subjects and 52 normal controls. The AD patient group was subdivided into three groups according to age of onset (OA): 1) $OA \leq 60$, 2) $61 < OA < 69$, 3) $OA \geq 70$. The frequencies explored consisted of five 5.5 Hz bands in the range of 1-28.5 Hz as in the previous

study. EEG spectral profiles were obtained from plots of the power spectral density values and were considered substitutes for the absolute powers. The three different power spectrum profiles were observed: (i) type A spectra, (ii) Type B spectra, and (iii) type C spectra, as described in the previous study by Pucci *et al.* The results confirm that this method can discriminate between AD patients and normal controls although the intention was to compare between AD groups for this study. In early AD, the EEG spectrum is characterized by lack of a dominant peak in the 6.5-12 Hz frequency range. The age of onset was found to correlate inversely with the 1-6.5 Hz relative powers and positively with the 6.5-12 Hz relative power. Evidence was shown for EEG changes due to early onset of AD when all three subgroups were compared [40].

Rodriguez *et al.*, 1999, performed a study to determine whether a parameter such as the power spectral profile may be useful in the detection of AD. Forty-eight patients with probable AD were recruited and divided into four groups based on global deterioration scores (GDR) which is a scale for assessment of dementia. 18 healthy control subjects were also recruited for the study. The EEG was recorded with eyes closed, in a resting state. The spectral profile for each patient was expressed by the relative power of seven different frequency bands (2-3.5 Hz, 4-5.5 Hz, 6-7.5 Hz, 8-9.5 Hz, 10-11.5 Hz, 12-13.5 Hz, and 14-22.5 Hz). Overall, the findings showed that 4-5.5 Hz and 10-11.5 Hz band powers displayed the highest significance statistically in differentiating between the groups. They conclude that spectral morphology significantly changes with the progression of AD and since it is widely available may be useful in a clinical setting for staging AD [41].

Polich, 1999, outlined several EEG/ERP studies using similar recording and analysis methods. EEG was recorded for resting eyes open and eyes closed conditions and spectral analysis was used to extract spectral power and mean frequency for six bands: delta (0.25-4 Hz), theta (4-8 Hz), alpha-1 (7.5-9.5 Hz), alpha-2 (9.5-12.5 Hz), beta-1 (12-20 Hz), and beta-2 (20-70 Hz). The overall or global (0.25-70 Hz) EEG power and mean frequency were also computed. An oddball paradigm was performed in different modalities to elicit ERPs for the different experiments. The spectral power data were subject to logarithmic transformations and then analyzed statistically while the mean frequency data were assessed directly. The P300 amplitude and latency were also analyzed for some of the studies [38]. These studies, 1-4, are described below in the order they were discussed in Polich, 1999.

Study 1, on EEG/ERP by Intriligator and Polich, 1995, explored individual variability in EEG and its relationship to the P300 variation in a group of normal young adult subjects. The background EEG was quantified and correlated with the P300 ERP measures to assess the effects of alpha changes on both the background EEG and P300. This was done primarily because a relationship between the EEG and P300 would help to account and control ERP variation between individuals and improve the utility of the P300 as a measure of cognition. The findings showed that EEG spectral power and P300 amplitude are positively correlated, where the strongest relationships were observed for the delta, theta and alpha-1 bands. The frequency of the EEG is associated with P300 latency in a variable manner across the different frequency bands and recording sites. Finally, the relationship between the P300 and EEG was strongest with the eyes open condition [38].

Study 2, by Spencer and Polich, 1998, performed on a group of young adults, analyzed the P300 while manipulating variables known to affect its measures. Spectral analysis was used to examine the individual EEG trials that were later averaged to produce the ERP. The amplitude/latency as well as spectral power/mean frequency measures were obtained. In order to manipulate attention-requiring task demands, the oddball paradigm was performed several times with varying tasks: 1) ignore the target, 2) count the target and 3) count the target and standard tones. The alpha and delta bands demonstrated greater spectral power for the count target and count both conditions. The findings show that ERP task demands differentially affect the EEG power spectrum as a function of the attention requirements of the task. In summary of these studies, task demands affect the alpha band activity while task-induced changes that affected the P300 also produced changes in the delta and theta bands [38].

Study 3, Polich, 1997, conducted a comprehensive study on 120 subjects ages 20-70+ years, with 10 males/10 females per each age decade. EEG spectral and ERP P300 measures were used to assess changes between the age groups by correlating the EEG power from each band with the P300 latency. Age produced systematic decreases in power and some increase in frequency for the specific frequency bands. As age increases, delta, theta and alpha-2 power decrease with an increase noted for the beta-2 power. The decrease in EEG power for the delta, theta and alpha bands was correlated with the decline in P300 amplitude as age increases. These effects were strongest for the eyes open condition with auditory stimulus. No reliable associations were found for P300 latency and the EEG mean frequency.

Study 4, Polich, 1996, reviewed measures between the P300 and age by comparing several studies. The correlations between the P300 and age vary across the studies because of differences in sample characteristics and task parameters. The relationship between the P300 and EEG may be mediated by event-related desynchronization (ERD) when the task information is processed. ERDs are believed to originate from stimulus and task related shifts in the alpha band power and have been linked to attentional processes [38].

Hibino *et al.*, 2001, performed analysis on EEG data collected (resting eyes closed condition) from 26 AD patients and 17 normal control subjects but with multiple EEG recordings for some of the AD patients, giving a total of 35 recordings for AD patients and the 17 from the control group. Power spectra was calculated using the fast Fourier transform and separated into 9 frequency bands: delta (3-4 Hz), theta 1 (4-6 Hz), theta 2 (6-8 Hz), alpha 1 (8-9 Hz), alpha 2 (9-11 Hz), alpha 3 (11-13 Hz), beta 1 (13-15 Hz), beta 2 (15-20 Hz), and beta 3 (20-25 Hz). The relative power of each frequency band was calculated and the results for the alpha waves were averaged. The delta and beta waves were not used for the analysis because influences due to eye movement in the delta waves were not negligible and no effects on the beta waves due to cognitive deterioration have been reported. The theta 1, theta 2 and averages of the alpha waves were used as input into a fuzzy neural network (FNN), which uses fuzzy inference to make a decision. Parameters were defined for two models, one for discrimination between AD and control subjects and the other to estimate MMS (Mini-mental state, see Appendix C) scores into two groups, above 24 and less than or equal to 24. Three controls were misclassified as having AD in the first model and for the second model, two of the patients

MMS scores were incorrectly estimated. The authors conclude that this method provides promising results for staging the level of AD as well as diagnosis [42].

A study by Babaloni *et al.*, 2004 was conducted for the purpose of defining EEG sources specific for mild AD compared to vascular dementia (VaD) and to normal elderly subjects. The study extended previous research by providing topographical localization of EEG sources at specific frequency bands to characterize the mild AD group. EEG was recorded with 19 electrodes in resting eyes closed condition. FFT-based power spectrum analysis was performed to compute the power density of the different frequency bands: delta (2-4 Hz), theta (4-8 Hz), alpha 1 (8-10.5 Hz), alpha 2 (10.5-13 Hz) beta 1 (13-20 Hz) and beta 2 (20-30 Hz). The low resolution brain electromagnetic tomography (LORETA-KEY) is a functional imaging technique that gives a 3-D distribution of electric neuronal activity, modeling the cortex as a collection of volume elements called voxels. This study used LORETA to compute the power spectral profiles of cortical EEG sources. Mild AD showed occipital, temporal and limbic delta sources have a greater magnitude then those of normal aging. However the magnitude of the delta sources in VaD was greater than mild AD. Occipital alpha 1 sources showed a decline in mild AD compared to VaD. They conclude that the LORETA approach is a valid potential method for illustrating the power spectrum profiles as the level of cortical EEG sources and further studies should be performed with neural networks or discriminant analysis [29].

2.1 .2 OTHER METHODS

COHERENCE

Besthorn *et al.*, 1994 recorded EEG signals from 50 AD patients and 42 normal controls with resting, eyes closed condition from 17 electrodes. Coherence is the normalized crosspower spectrum per frequency of two signals that are recorded simultaneously from different electrodes on the scalp. Coherence is discussed in more detail in Section 3.1.3. Coherence was computed for 6 frequency bands: delta (1.5-3.5 Hz), theta (3.5-7.5 Hz), alpha1 (7.5-9.5 Hz), alpha2 (9.5-12.5 Hz), beta1 (12.5-17.5 Hz), and beta2 (17.5-25.0 Hz) and the spatially averaged coherence was calculated. The results show differences between the AD patients and controls in all frequencies except for the delta band. The effect was most pronounced in the frontal and central derivations of the theta, alpha, and beta frequency bands. The authors indicate that the results show the effects of neuronal loss and neocortical disconnection [43].

Locatelli *et al.*, 1998 analyzed EEG signals from 10 probable AD patients and 10 normal controls. EEG was recorded in resting, eyes closed condition from 16 electrodes with ears linked as reference. 50 epochs with a duration of 1 second were used for analysis. Coherences for the delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz) and beta (12-30 Hz) bands were calculated as the mean coherence values of the 50 epochs selected. Both 'local' and 'far' coherences were calculated to determine whether the changes due to AD were associated with the impairment of the short or long axonal fibers.

Their findings are in accordance with previous results showing a decreased coherence for EEG high frequencies in AD. For the regional coherence, alterations were more pronounced over the left tempero-parietal-occipital areas. For the long distance coher-

ence, a decrease between pairs of 'far' electrodes confirms that long axons are more compromised than short ones in AD. Results also show that for the AD group there is a significant decrease of alpha interhemispheric coherence, with these results more pronounced for patients with more severe cognitive impairment. The coherence decrease of alpha and beta bands begins in the earliest stages of the disease, while the increase of slow band coherence is evident in the later stages [44].

Jelic *et al.*, 2000, examined the EEG signals from 27 subjects with mild cognitive impairment (MCI). The purpose of this study was to evaluate the clinical course of patients with MCI and the incidence of dementia within the group as well as the EEG changes following cognitive deterioration and the potential of neurophysiological measures for predicting dementia.

EEG was recorded in resting awake condition with eyes closed from 20 electrodes. Relative power was calculated for four conventional frequency bands: delta (2-4 Hz), theta (4-8 Hz), alpha (8-13 Hz, and beta (13-20 Hz). Coherence was analyzed through statistical analysis. 14 patients were diagnosed with AD, while the other 13 remained clinically stable in the follow up analysis after an average of 21 months. The patients who progressed to AD had a significantly higher theta relative power and lower mean frequency than the rest of the group. The temporal and temporo-occipital derivations were shown to be the most sensitive to these effects. The results support the idea that a subgroup among MCI patients will develop AD [45].

Hogan *et al.*, 2003, examined memory-related EEG power and coherence over temporal and central recording sites in 10 mild AD patients and 10 normal controls. A memory test was performed during EEG data collection with 5 electrodes (Fz, Cz, Pz, T3

and T4, with auditory reference). The following spectrum bands were defined for further analysis: delta (0.5-3 Hz), theta (3-5 Hz), lower alpha 1 (5-7 Hz), lower alpha 2 (7-9 Hz), upper alpha (9-11 Hz), beta 1(15-20 Hz), and beta 2 (20-25 Hz). The frequency band 12-14 Hz was not analyzed. Spectral power and coherence for the named frequency bands were analyzed.

For both AD patients and controls, the power in the delta to upper alpha range was greater over the three central sites as opposed to the two temporal sites. For the beta frequency range, the opposite occurred, where the power was greater over the two temporal sites. This could be due to the reference electrodes detecting activity from the underlying source, thus inflating the coherence.

AD patients were found to have more power in lower alpha 1 than controls. This difference occurred primarily in central as opposed to temporal sites. For coherence, the upper alpha band produced significant group differences. The overall results are consistent with previous research where it is suggested that upper alpha has a role in memory processes and beta has a role in task demand. The authors suggest that spectral power and coherence measures may offer insight into underlying cortical disruption [46].

Brunovsky *et al.*, 2003, examined EEGs from 38 AD patients with varying degrees of dementia (mild, moderate, severe) using EEG spectra and coherences. Spectrum analysis was performed for 6 frequency bands: delta (0.5-3.5 Hz), theta (3.5-7.5 Hz), alpha 1 (7.5-9.5 Hz), alpha 2 (9.5-12.5 Hz), beta 1(12.5-17.5 Hz), and beta 2 (15.5-25 Hz). The mean amplitudes (square roots of the power values) were calculated for each band. For each of the 6 frequency bands, intrahemispheric ('local') coherences were calculated for the anterior and posterior brain regions and the interhemispheric ('far') coherences

consisted of pairs of distant electrodes on either side of the head. The results of statistical evaluation indicated decreased 'far' coherences in the alpha frequency bands were highly correlated with the degree of dementia. This indicates that the lower the coherence in this range, the higher the degree of dementia for the patient. The authors conclude that a combination of EEG spectral values and coherence may be used to supplement clinical evaluations [47].

DIPOLE SOURCE ANALYSIS

In a study by Musha *et al.*, 2002, 25 patients with mild AD, 33 patients with moderately severe AD and 56 control subjects underwent EEG recording in resting eyes closed conditions with 21 electrodes. Uniform scalp potential distributions are approximated well with single equivalent current dipoles within the brain. The degree of uniformity is measured by a quantity known as dipolarity. Cortical lesions result in randomly oriented dipoles which lower the dipolarity of the scalp potential distribution. A dipole is pair of equal and opposite electric charges of opposite sign separated by a small distance. In this study they examine the mean alpha dipolarity to measure the degree of uniformity of the alpha EEG component in mild and moderately severe AD patients with a control group of normal aging subjects. The alpha frequencies were obtained using the FFT and a 3 Hz-wide, digital bandpass filter. The results show a progressive decline in the mean alpha dipolarity and an increase in its variability as AD becomes more severe. The authors hypothesize that the degree of non-uniformity of the electrical activity of the brain shows the degree of neuronal death in the brain. Attempts were made to model the neurophysiological state of each individual's brain. The method used in this study was named Diagnosis Method of Neuronal Dysfunction or DIMENSION. The authors suggest that fur-

ther research on this method would be necessary, also mentioning that an account of medications taken by patients should be considered [48].

2.2 TIME DOMAIN TECHNIQUES

In some cases analysis is performed without using any transformations and calculation are performed in the time domain. Some of the studies discussed previously in ERP section, conducted to determine correlations between the P300 amplitude and latency across different groups used time domain analysis. Mutual information analysis is discussed in this section and is further described in Appendix E.

CROSS MUTUAL INFORMATION

Jeong, Gore and Peterson, 2001, performed a study using baseline EEGs for 15 AD patients and 15 healthy controls using 16 channels. The purpose of the study was to determine the information transmission between cortical areas of the brain of an AD patient by estimating the average Cross-Mutual Information (CMI) between EEG electrodes. CMI is used to quantify information passed from one time series to another. Auto mutual information (AMI) estimates how much on average, the value of the time series can be predicted from the values of the time series at preceding points. Mutual information gives a measure of the linear and nonlinear dependencies between two time series. Once the CMI was calculated and the AMI was estimated, statistical analysis was performed to compare the two subject groups. The results showed that the CMIs as a function of distance between electrodes dropped off more rapidly for AD patients than for the controls. This was more obvious for increased distance between electrodes. It is speculated that reduced CMIs in AD patients could be due to dissociation between cortical and subcorti-

cal structures in the brain or by degeneration of subcortical structures. In the AD patients, the AMIs decreased much slower with time delay than they did for the controls. The rate of decline of the AMI is positively correlated with entropy, so the slower decline of the AMIs of the AD patients implies that their EEG activity is less complex than the activity of the controls. The authors suggest for further research to be geared towards examining the associations of mutual information measures with cognitive variables and comparing the transmission of information in AD patients with other types of dementia possibly to develop a means for early diagnosis [49].

2.3 TIME-FREQUENCY DOMAIN TECHNIQUES

The use of wavelets in the analysis of EEG signals has become more popular over the years. The time-frequency localization is one of its many valued properties, which are discussed in more detail in Section 3.2. Most of the studies below use wavelets to determine if they provide adequate information about ERPs that would make them beneficial in further research.

2.3.1 DAUBECHIES 4 WAVELET

Polikar *et al.*, 1997 applied the Daubechies 4 wavelet to EEG data collected from 14 normal subjects and 14 subjects diagnosed with probable AD. The ERP response in the oddball paradigm was analyzed to determine if the use of the wavelet transform was feasible for the detection of AD with a multilayer perceptron (MLP) neural network. Half of the signals, 7 AD and 7 normal, were used for training while the rest were used for testing the network. The generalization performance of the network was 93%. The results

confirm that the approach is feasible for classifying ERPs but a more diverse database with a larger variety of signals would allow for statistically valid generalizations [20].

Petrosian, *et al.*, 2001, explored wavelet transform by using specifically designed and trained recurrent neural networks (RNNs) to discriminate between EEGs of ten mild AD patients and ten age-matched control subjects. The EEG recordings were taken during resting state without the use of a paradigm. The Daubechies 4 wavelet was chosen due to its good localizing properties in the time and frequency domain. The RNNs used in the study belong to a type of discrete-time recurrent MLPs. This type of network has better temporal capabilities than that of a regular feedforward MLP and is capable of representing and encoding strongly hidden states. Training on three AD subjects and three controls and testing on the remaining controls yielded performance that was better than chance with 80% sensitivity and 100% specificity. Five out of seven of the AD subjects were correctly classified. The authors suggest that their approach may be extended to include more classes such as other types of dementia [50].

2.3.2 QUADRATIC B-SPLINE WAVELET

The quadratic B-spline wavelets have been widely used in the analysis of ERPs, but have not in previous research been applied to ERPs with the purpose of differentiating between AD and normal cohorts. In a study by Ademoglu *et al.*,1997, quadratic B-spline wavelets were used in the analysis of pattern-reversal visual evoked potentials of normal and demented patients. This study was primarily performed to further explore the components of the ERPs, not to distinguish between classes. The wavelet was shown to capture particular events occurring within the ERP and was found suitable for further applications as

a discrimination technique for cases of clinical problems associated with peak latency [57].

Demiralp *et al.*, 1999a, performed analysis of EEG signals from 10 healthy volunteers, aged 18-25 years, obtained through the use of the oddball paradigm. The purpose was to investigate the effectiveness of the wavelet transform in the analysis of ERPs. The transform is applied to both averaged and single-sweep responses. The results show that the delta, theta, and alpha processing levels all have some involvement in time during the oddball task. Analysis involving the topography pattern of the electrodes indicates that the wavelet coefficients reflect functional components at specific regions of the brain. The authors conclude that the wavelet transform is a powerful method for decomposing ERPs on the basis of frequency, temporal position and scalp topography. They suggest that signals of the same frequency ranges but with specific temporal position and/or scalp topography can be distinguished. It is also possible that processes overlapping in time can be separated by frequency and/or scalp topography. The wavelet transform is useful for its capability of separating various simultaneously occurring frequencies that could not be detected in the time domain signals [51].

Demiralp *et al.*, 1999b, also performed an ERP study on 10 healthy volunteers between 18 and 55 years of age using an auditory oddball paradigm. The goal of this study was to classify single ERPs according to their characteristic response properties. Quadratic b-spline wavelets were used for their near optimal time-frequency localization properties. Wavelet decomposition yields six sets of coefficients, each belonging to different frequency bands, 62-125 Hz (high gamma), 31-62 Hz (gamma), 16-31 Hz (beta), 8-16 Hz (alpha), 4-8 Hz (theta), and 0.5-4 Hz (delta). The grand average of the target and

non-target response were calculated and compared. The delta response amplitude was found to be increased for the target response. Also for the theta and alpha frequency ranges showed increased amplitude and prolonged target responses. The P300 response was detected in single trials using the wavelet coefficients for the delta frequency range. Instead of just using single trials, the averaging technique is still the most reliable technique for ERP estimation although it is inefficient. Another approach would be to classify the patterns in single trials building more homogeneous subgroups of sweeps and then average those in order to obtain ERPs with lower variability[52].

Quiroga *et al.* in 2001 used quadratic B-spline wavelets and Fourier transform to analyze evoked potentials and compared results of the analysis. The evoked potentials were elicited through the use of two techniques: pattern visual (to review alpha band) and bimodal (to review gamma band). The results for analysis of the alpha band indicate that they are best localized in the occipital region. The gamma amplitude was the largest in response to bimodal stimulation as compared to auditory or visual stimulation alone. The responses to bimodal stimulation indicate that the response may be something other than the response to the two forms of stimulation combined but a new response. For the comparison of the wavelet transform to the Fourier transform, the researchers found the wavelets to have two advantages. The first is that the wavelet transform is capable of analyzing non-stationary signals which is critical for avoiding spurious results of the Fourier transform considering that brain signals are highly non-stationary. The second is that better time-frequency resolution can be achieved so the ERP responses can be analyzed closely [53].

CHAPTER 3

APPROACH

The theory involved in this study is discussed in this section. The Fourier transform and coherence analysis were not explicitly used in this study but are included here for completeness. Power spectral density and wavelet analysis are discussed in detail in this chapter to provide a thorough background into this study.

3.1 FREQUENCY DOMAIN TECHNIQUES

Frequency analysis is an alternate and informative technique for describing time domain signals. An advantage of the frequency domain representation over the time representation is its ability to give a clear visualization of the specific frequencies and their contributions within the signal. The Fourier Transform (FT) is computationally attractive since it can be calculated using the Fast Fourier Transform, an extremely efficient algorithm. The FT is often used for the analysis of frequency bands and the analysis of evoked responses in EEG signals [10, 54].

3.1.1 THE FOURIER TRANSFORM

The Fourier Transform describes a signal *x(t)* as a linear superposition of complex exponentials characterized by their frequency *f*.

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

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

Equation (3.1) is the continuous Fourier Transform of the signal *x(t)* and Equation (3.2) is the inverse Fourier transform of *X(f)*. *X(f)* can be seen as an inner product of the signal *x(t)* with complex exponential $e^{-2\pi ft}$, also known as the kernel function.

$$X(f) = < x(t), e^{-i2\pi ft} > \tag{3.3}$$

Since the kernel functions, $e^{-i2\pi ft}$ are orthogonal, the FT is not redundant. Consider the following signal, *x[n]* consists of N discrete values, sampled at Δ seconds apart

$$x[n] = \{x_0, x_1, \ldots, x_{N-1}\} = \{x_j\} \tag{3.4}$$

where x_j is the measurement taken at a time $t_j = t_0 + j\Delta$. The discrete Fourier Transform (DFT) of this signal is defined as:

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-i2\pi kn/N} \qquad k = 0, \ldots, N-1 \tag{3.5}$$

The signal can be reconstructed from its DFT coefficients:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i2\pi kn/N} \tag{3.6}$$

where $X(k) = X^*(N-k)$ holds for real signals and the discrete frequencies defined as:

$$f_k = \frac{k}{N\Delta} \tag{3.7}$$

Note that the discrete Fourier Transform gives *N/2* independent complex coefficients, thus giving a total of *N* values as in the original signal and therefore is not redundant. The frequency resolution is then:

$$\Delta f = \frac{1}{N\Delta} \tag{3.8}$$

The frequency $f_N = \frac{1}{2\Delta}$, corresponding to $k = \frac{N}{2}$ in Equation (3.7) is called the Nyquist frequency and it is the highest frequency that can be detected with a sampling period Δ. If the signal *x(n)* has frequencies above the Nyquist frequency, the frequencies are processed as though they are in the range $f_k < f_N$, thereby producing aliasing effects in the signal. The greater the spurious effect, the greater the errors and the less adequate the sampling rate Δ for representing the signal [10, 54].

3.1.2 POWER SPECTRAL DENSITY

The successive discrete values in a signal such as the EEG have a certain degree of interdependence. The joint probability distribution of the signal can be calculated to describe this interdependence. When applied to a pair of values at two discrete moments, $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$, the definition of the joint probability can be considered for N realizations of the signal and the number of times that at t_1 a value v and at t_2 a value u are encountered is considered equal to n_{12}. So the joint probability may be defined as:

$$p(\mathbf{x}(t_1) = v, \mathbf{x}(t_2) = u) = \lim_{N \to \infty} \frac{n_{12}}{N} \tag{3.9}$$

The properties of a signal generated by a random process can be described by specifying the joint probability density function

$$\rho\big((\mathbf{x}(t_1), \mathbf{x}(t_2) \ldots \mathbf{x}(t_n)\big) \tag{3.10}$$

for all discrete time samples $t_1, t_2, \ldots, t_n$ and for all finite values of n. This computation can be rather complex so an alternative approach is to compute a number of averages characteristic of the signal such as the covariance, correlations and spectra which are useful for general description of the signal.

The covariance at two time samples $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ between two random variables is given by

$$E[(\mathbf{x}(t_1)-E(\mathbf{x}(t_1)))(\mathbf{x}(t_2)-E(\mathbf{x}(t_2)))] \tag{3. 11}$$

where E represents the expected value of the variable. To estimate the covariance between any two variables $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$, it is necessary to take averages over a number of realizations of an ensemble. The covariance for a stationary ergodic signal (see Appendix C) can also be estimated using the time average as shown:

$$\Phi_{xx}(\tau_k) = \langle \mathbf{x}(t)\mathbf{x}(t+\tau_k) \rangle = \frac{1}{T}\sum_{i=1}^{N} x(t_i)x(t_i+\tau_k) \tag{3. 12}$$

where T is the time for one realization of the signal and $\tau_k = k \cdot \Delta t$. If every sample function is representative of the whole, the time average $\Phi_{xx}(\tau)$ for one realization $\underline{x}(t)$ is an estimate of the ensemble average $R_{xx}(\tau)$:

$$R_{xx}(\tau) = E[\mathbf{x}(t)\mathbf{x}(t+\tau)] \tag{3. 13}$$

which is also known as the autocorrelation function. With this equation, assuming that the signal $\mathbf{x}(t)$ has zero mean, then for the value $\tau=0$,

$$R_{xx}(0) = E[x^2(t)] = \lim_{t\to\infty} \frac{1}{2T}\int_{-T}^{T} x^2(t)dt \tag{3. 14}$$

gives the average power or variance, σ, of the signal. The Fourier transform of the autocorrelation function is the power density spectrum or power spectrum, S_{xx}, a commonly used parameter for EEG quantification.

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j2\pi f\tau}d\tau = FT[R_{xx}(\tau)] \tag{3. 15}$$

The power spectrum gives the distribution of the squared amplitude of different frequency components within the signal.

For calculation of the power spectrum of a discrete EEG signal, the epoch $[\mathbf{x}(t_1)]$ is considered as a signal sampled at the intervals Δt, $x(n\Delta t)$ with a total of N samples $(n=1...N)$. By using the discrete Fourier transform, the periodogram, $F_{xx}(f_i)$ can be calculated as:

$$F_{xx}(f_i) = \frac{\Delta t}{N}\left|\sum_{n=1}^{N} x(t_n)e^{-j2\pi i\cdot\Delta f\cdot n\Delta t}\right|^2 \qquad (3.16)$$

where $f_i = i\cdot\Delta f$ with $i=0,1,2...,N$.

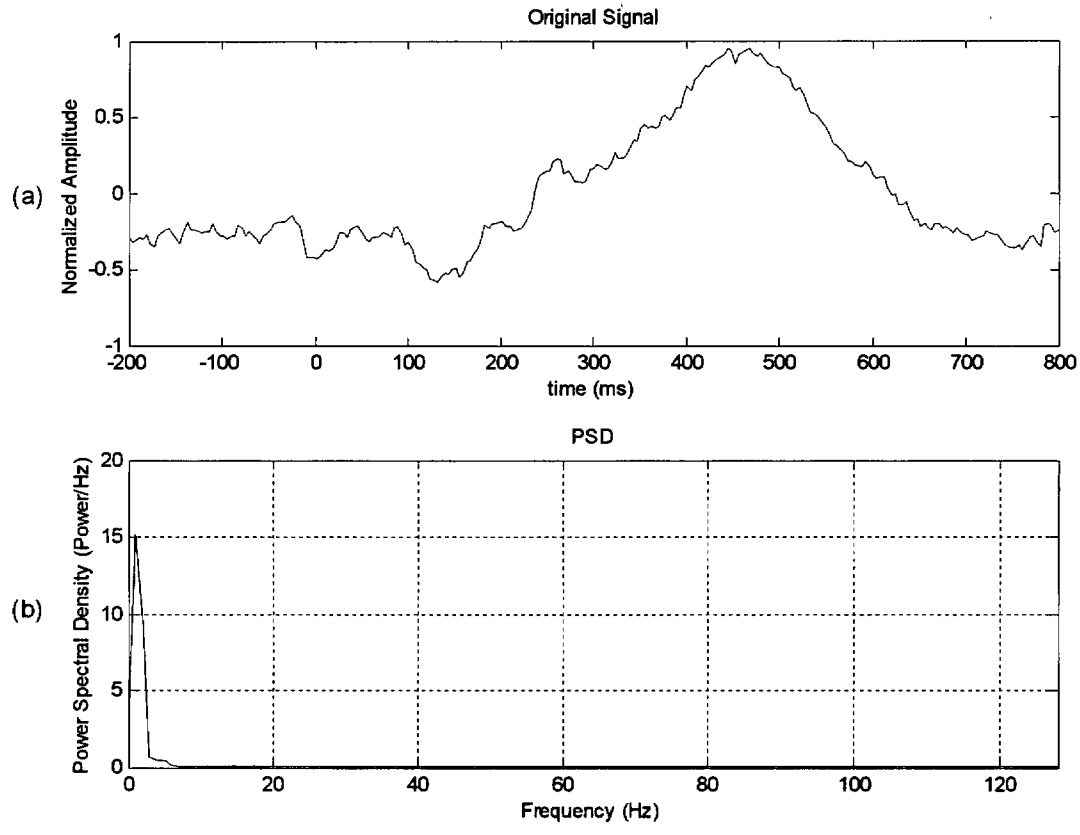


Figure 3. 1: Power spectral density of an EEG signal (a) original signal, (b) PSD of signal.

With the use of a window, $W(f_k)$, the periodogram may be smoothed in order to obtain $P_{xx}(f_k)$ which is a better estimate of the real power spectrum $S_{xx}(f)$:

$$P_{xx}(f_i) = \sum W(f_k)F_{xx}(f_{i+k}) \tag{3. 17}$$

where $W(f_k)$ is the smoothing window with a duration of $(2p+1)$ samples [9]. The periodogram is actually the discrete-time Fourier transform (DTFT) of the estimate of the correlation function [55]. Figure 3. 1 shows the power spectral density of an EEG signal using the periodogram method.

3.1 .3 COHERENCE ANALYSIS

Coherence is a quantitative measure of the consistency of phase between two signals. Coherencies measured between different distinct scalp locations provide measures of mutual influences, however the magnitudes of these influences in complex, non-linear dynamic systems can be significantly different at different frequencies or scales. Data suggests that in the neocortex, global dynamics are displayed at certain frequencies (usually near 10 Hz) while local behavior is evident at other frequency ranges (often near 40 Hz). Amplitude and coherency are generally independent measures of waveforms. Increased EEG scalp amplitude is normally associated with increased local 'synchrony' (increased coherence of local current sources, typically within a few centimeters), however this can occur independently of longer range coherence changes which may have a small effect on amplitude.

Each epoch, n, of channel x yields the complex Fourier transform $X_{xn}(f)$. The single epoch power spectrum or auto spectral density function is:

$$S_{xn}(f) = X_{xn}(f)X^*_{xn}(f) \tag{3. 18}$$

where * indicates the complex conjugate. It follows that the cross spectral density function for channels x and y is:

$$S_{xyn}(f) = X_{xyn}(f)X^{*}_{xyn}(f) \tag{3. 19}$$

Accurate estimates are provided by these equations if many epochs *(N)* are averaged where the cross spectral density function estimate is:

$$\hat{S}_{xy}(f) = \frac{1}{N}\sum_{n=1}^{N} S_{xyn}(f) \tag{3. 20}$$

with a similar expression for the power spectrum $S_x(f)$ for each channel x. The coherency estimate based on N epochs is given as:

$$\hat{\gamma}^2_{xy}(f) = \frac{\left|\hat{S}_{xy}(f)\right|^2}{\hat{S}_x(f)\hat{S}_y(f)} \tag{3. 21}$$

[46, 56]. For a 95% confidence interval, the power or coherence estimates can be expressed as:

$$\frac{\hat{S}(f)}{1+2e} \le S(f) \le \frac{\hat{S}(f)}{1-2e} \tag{3. 22}$$

(this applies to auto or cross spectral density and coherency). The confidence interval depends on the normalized root mean square error, e. For spectral density estimates based on statistically independent epochs, the error is estimated by:

$$e_S = \frac{1}{\sqrt{N}} \tag{3. 23}$$

if the epoch is greater than 25. The error term for the coherence estimates is given as:

$$e_{\gamma^2} \cong \sqrt{2/N}\,\frac{1-\hat{\gamma}^2}{|\hat{\gamma}|} \tag{3. 24}$$

This depends on the coherence itself and may be estimated using the estimated coherence given that the error is small (less than 0.2) [56].

3.2 TIME-FREQUENCY DOMAIN TECHNIQUES

Fourier Transform involves a correlation between the signal to be analyzed and complex exponentials of different frequencies. The appropriate interpretation of the FT requires the signal to be stationary as it results in a loss of time information. 'Windowing' the complex exponential kernel functions of the FT gives a time evolution of the frequencies by sliding the windows throughout the signal.

$$STFT_X^{(\omega)}(t,f) = \int [x(\tau)\cdot \omega^*(\tau - t)]\cdot e^{-i2\pi f\tau} d\tau \qquad (3.25)$$

This procedure, the Short-time Fourier Transform (STFT) or Gabor Transform, consists of correlating the original signal with modulated complex exponentials. *x(t)* represents the original signal while *ω(t)* is the windowing function and Equation (3.25) shows that the STFT is simply the multiplication of the FT of the signal by a windowing function. This procedure gives an optimal time-frequency representation, but is limited by the Heisenberg Uncertainty Principle (see Appendix B) which gives a lower bound of $\Delta t \Delta f \geq \frac{1}{4\pi}$, where Δt is the time resolution and Δf is the frequency resolution. Δt indicates how well two spikes in time can be separated from one another in the transform domain and Δf indicates how well two spectral components can be separated from each other in the transform domain [57, 60]. If the window is too narrow, there will be poor frequency resolution; and if the window is too wide, there will be poor time localization. For the best resolution, low frequencies require a wide window and high frequencies re-

quire a narrow window. The STFT is not suitable for analyzing signals involving different range of frequencies because of its fixed window properties. For example a wide window would provide good localization for the low frequencies within the signal but poor localization of any high frequency components [59].

3.2.1 THE WAVELET TRANSFORM

The Wavelet Transform was developed to overcome the fixed resolution problems of the STFT. The main advantage of wavelets is their varying window size, wide for low frequencies and narrow for the high frequencies, leading to an optimal time-frequency resolution in all frequency ranges [10, 59]. The continuous wavelet transform is discussed first leading into the discrete wavelet series and transform. Multiresolution analysis and the subband coding algorithm are then explained before the types of wavelets chosen for this analysis are presented.

3.2.2 CONTINUOUS WAVELET TRANSFORM

A wavelet family $\psi_{b;a}$ is a set of functions created by dilations and translations of a unique mother wavelet $\psi(t)$:

$$\psi_{b,a}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \tag{3.26}$$

where $b, a \in \Re$, $a \neq 0$ are the translation and scale parameters, respectively. Translation is a time shift and scale is essentially the inverse of the frequency where large scale gives global behavior and small scale gives local behavior (similar to a map scale). As a increases the wavelet becomes narrower and by varying b, the mother wavelet is displaced in time. The wavelet functions must satisfy two conditions:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{3. 27}$$

and

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \tag{3. 28}$$

The continuous wavelet transform of a signal $x(t) \in L^2(\Re)$ is defined as the correlation between the signal and the wavelet functions $\psi_{b,a}$:

$$CWT_x^{\psi}(b,a) = \Psi_x^{\psi}(b,a) = |a|^{-1/2} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt = < x(t), \psi_{b,a} > \tag{3. 29}$$

where * denotes complex conjugation. Then, different correlations of $< x(t), \psi_{b;a} >$ indicate how closely the wavelet function correlates with the signal at each scale a. If the signal contains a component of the frequency at the particular scale, then the wavelet basis function at that scale will be similar to the signal at the location where that frequency occurs. Since the correlation is made with different scales of a single function, the wavelet transform is said to give a time-scale representation [10, 58, 59].

3.2.3 DISCRETE WAVELET SERIES

It is more practical to define the Wavelet Transform at discrete scales, a, and discrete times, b since the CWT can not be electronically computed. The conventional scheme for discretizing the time-scale parameters is called dyadic grid sampling. For this process, time remains continuous but the time scale parameters are sampled by choosing $a_j = 2^j$, $b_{j,k} = k2^j$, with $j,k \in Z$ [10, 57].

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \qquad j,k \in Z \tag{3.30}$$

For wavelet series, $\psi_{j,k}(t)$ are required to be orthonormal, biorthogonal or frame (see Appendix C), however, for the orthonormal case, shown below, the analysis and synthesis wavelets are the same.

$$\Psi_x^{\psi_{j,k}} = \int x(t)\psi_{j,k}^*(t)dt \tag{3.31}$$

or

$$x(t) = c_\psi \sum_j \sum_k \Psi_x^{\psi_{j,k}} \psi_{j,k}(t) \tag{3.32}$$

where c_ψ is a constant that depends on the wavelet used [57, 59].

3.2.4 DISCRETE WAVELET TRANSFORM

The discrete wavelet series is in essence just a sampled version of the CWT which provides redundant information during reconstruction of the signal. The Discrete Wavelet Transform (DWT) is similar to the series but applies to discrete-time signals, *x[n]*. The DWT is typically implemented using a method called Multiresolution Analysis.

3.2.5 MULTIRESOLUTION ANALYSIS

Contracted versions of the wavelet function match high frequency components of the original signal and dilated versions match low frequency components. Correlating the wavelet functions of different scales with the original signal gives the "detail components" of the signal at different scale levels. Multiresolution analysis is a hierarchical scheme that uses the DWT to analyze the signal at different frequency bands with different resolutions through the decomposition of the signal.

If the subspaces of $\mathcal{L}^2$ generated by the wavelets $\psi_{j,k}$ are denoted as W_j for each level j, the space $\mathcal{L}^2$ can be decomposed as a direct sum of the subspaces W_j,

$$\mathcal{L}^2 = \sum_{j \in Z} W_j \tag{3.33}$$

where $Z = \{...,-1,0,1,...\}$. The closed subspaces are defined as:

$$V_j = W_{j+1} \oplus W_{j+2} \oplus \ldots \qquad j \in Z \tag{3.34}$$

The subspaces V_j are a multiresolution approximation of $\mathcal{L}^2$ and are generated by scales and translations of the scaling (basis) function $\phi_{j,k}$. For the subspaces V_j, the orthogonal complementary subspaces W_j are given as:

$$V_{j-1} = V_j \oplus W_j \qquad j \in Z \tag{3.35}$$

A discrete signal $x[n]$, denoted as x_0, given a sampling rate of $\Delta t = 1$, can be successively decomposed with the following recursive scheme

$$x_{j-1}(n) = x_j(n) \oplus r_j(n) \tag{3.36}$$

where the terms $x_j(n) \in V_j$ give a coarse representation of the signal and $r_j(n) \in W_j$ give the details for each scale $j = 0, 1, ..., N$. For any resolution level $N > 0$, the decomposition of the signal, $X(n)$ can be written as:

$$X(n) = \sum_k x_N(k)\phi(2^{-N}n - k) + \sum_{j=1}^{N}\sum_k C_j(k)\psi_{j,k}(n) \tag{3.37}$$

where $\phi(\cdot)$ is the scaling function, $C_j(k)$ are the wavelet coefficients, and the sequence $\{x_N(k)\}$ represents the approximation of the signal at resolution level N. The second term is the wavelet expansion. The wavelet coefficients $C_j(k)$ can be interpreted as the residual errors between successive approximations of the signal at scales $j-1$ and j, where

$$r_j(n) = \sum_k C_j(k)\psi_{j,k}(n) \tag{3.38}$$

is the detail signal at scale j.

Multiresolution decomposition starts by correlating the signal with shifted versions of a contracted wavelet function where the coefficients obtained give the detail of the high frequency components. The remaining part will be an approximation of the original signal that can be obtained by correlating the signal with the scaling function, which is orthogonal to the wavelet function. The wavelet function is then dilated and from the coarser signal the procedure is repeated, thus giving a decomposition of the signal in different scale levels. This method gives a decomposition of the signal that can be implemented with very efficient algorithms since it can easily be computed through a series of simple filters, as shown in the following section for the subband coding algorithm [10, 59].

3.2.6 SUBBAND CODING ALGORITHM

The subband coding algorithm is a filter bank implementation of the discrete wavelet transform (DWT). The DWT analyzes the signal at different frequency bands with different resolutions through the decomposition of the signal (multiresolution analysis). The DWT utilizes two sets of functions: scaling functions and wavelet functions, each associated with lowpass and highpass filters, respectively. Decomposition is performed by passing the original signal, *x[n]* through a half band high-pass filter, *g[n],* and a half band low-pass filter, *h[n]*, then sub sampling each half by two by discarding every other sample. Half of the samples can then be eliminated according to the Nyquist's rule, since the highest frequency in the signal is now decreased by half.

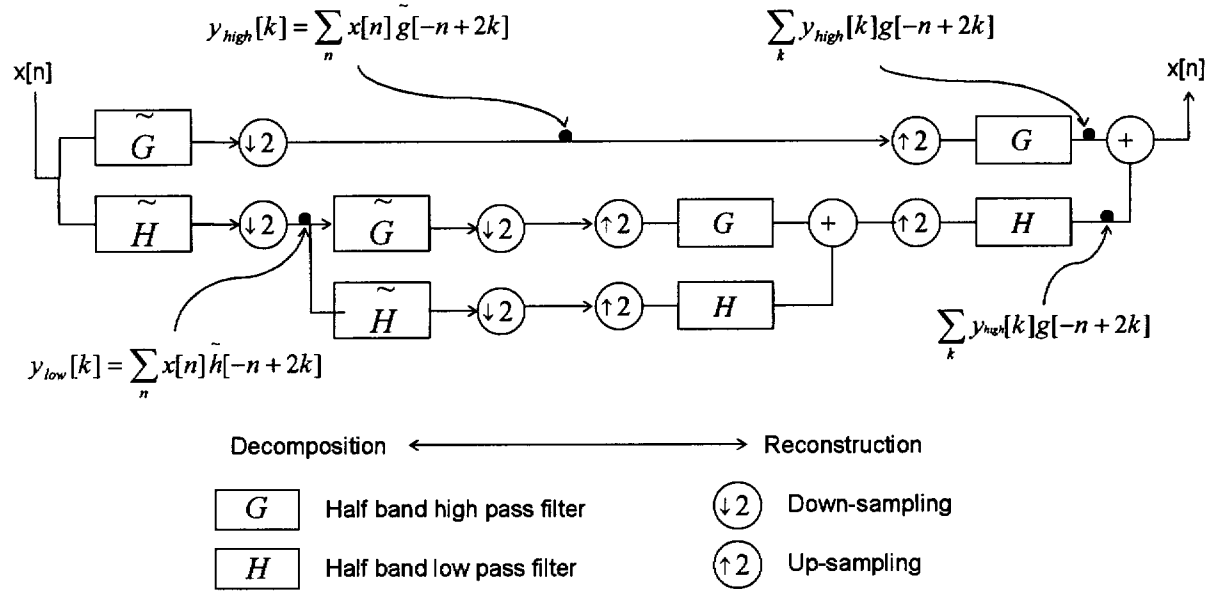


Figure 3. 2: 2-level DWT decomposition [52].

One level of decomposition is given as:

$$y_{high}[k] = \sum_n x[n] \cdot g[2k - n] \tag{3. 39}$$

and

$$y_{low}[k] = \sum_n x[n] \cdot h[2k - n] \tag{3. 40}$$

where $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the filters after sub sampling by two [20, 60]. The outputs of the high-pass filters are referred to as detail coefficients and are denoted as d_i, where $i = 1, 2, \ldots, log_2 N$, where N is the total number of samples in the signal. The outputs of low-pass filters are referred to as approximation coefficients, a_i, and they represent the approximation of the original signal at the current resolution level. The original signal can be reconstructed through the summation of the coefficients at all levels. At each level, the filtering procedure results in half the time resolution and double the frequency resolution. The block diagram in Figure 3. 2 illustrates the algorithm [20].

3.2.7 WAVELET CHOICES

The particular type of wavelet used for any application is normally chosen according to the similarity of the wavelet to the signal to be analyzed. This is done to minimize noise in the reconstruction of the signal as well as to better localize the structures of interest within the signal. The two types of wavelets chosen for this study have been used in different studies for analyzing ERPs—the Daubechies 4 wavelet and the Quadratic B-Spline wavelet.

3.2.7a DAUBECHIES 4 WAVELETS

Ingrid Daubechies invented the so-called compactly supported orthonormal wavelets, which made discrete wavelet analysis practical. Compact support is given by the size of the window varying throughout the signal so that for high frequency the window is smaller and for low frequency the window is larger. This gives good time resolution at high frequencies and good frequency resolution at lower frequencies [59, 61].
The Daubechies family wavelets are denoted as dbN, where N is the order, and db the "surname" of the wavelet. The first wavelet, db1, is the Haar wavelet. Figure 3. 3 shows the wavelet functions ψ of the next nine members of the family.

The Daubechies 4 wavelet has been used in several studies for analyzing ERPs in general, as well as for the detection of AD [20, 50]. The Daubechies mother wavelet

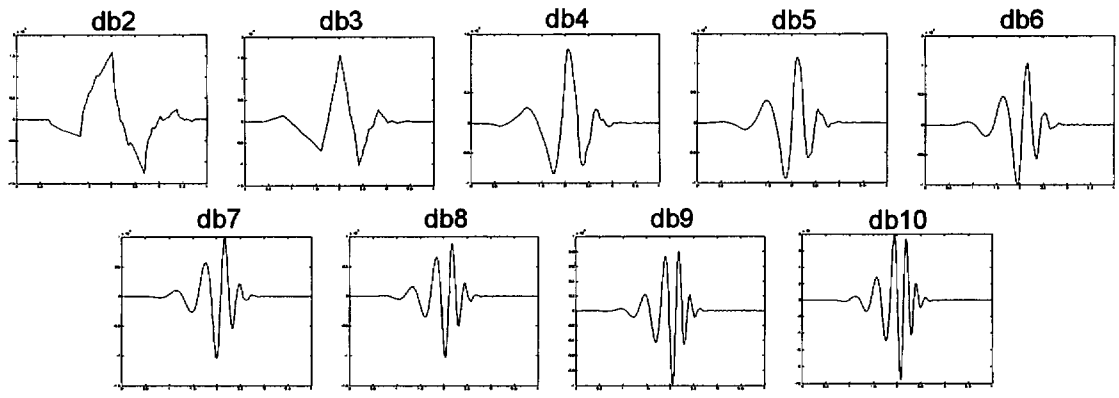


Figure 3. 3: Daubechies family functions [60].

has a 'fractal structure' and has good localizing properties in both time and frequency domains [50]. The scaling function for the db4 wavelet is given by

$$H(z)=\frac{1}{8}\left(\left(1-\sqrt{3}\right)+\left(3-\sqrt{3}\right)z^{-1}+\left(3+\sqrt{3}\right)z^{-2}+\left(1+\sqrt{3}\right)z^{-3}\right) \tag{3. 41}$$

and the scaling function coefficients are $\begin{Bmatrix} 0.48296 \\ 0.83652 \\ 0.22414 \\ -0.12941 \end{Bmatrix}$ [62].

The highpass and lowpass filter coefficients for the Daubechies 4 wavelet are given in Table 3. 1.

Table 3. 1: Daubechies 4 filter coefficients.

h	g
0.2304	-0.0106
0.7148	-0.0329
0.6309	0.0308
-0.0280	0.1870
-0.1870	-0.0280
0.0308	-0.6309
0.0329	0.7148
-0.0106	-0.2304

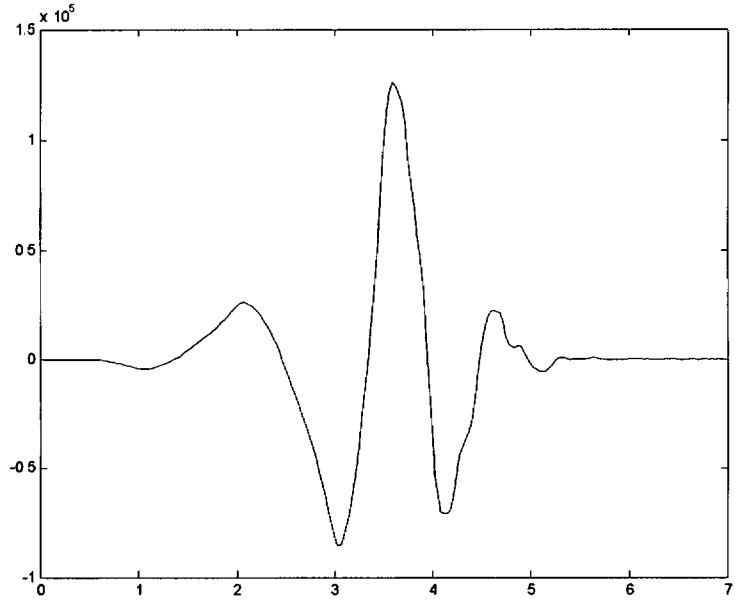


Figure 3. 4: Daubechies 4 wavelet function [61].

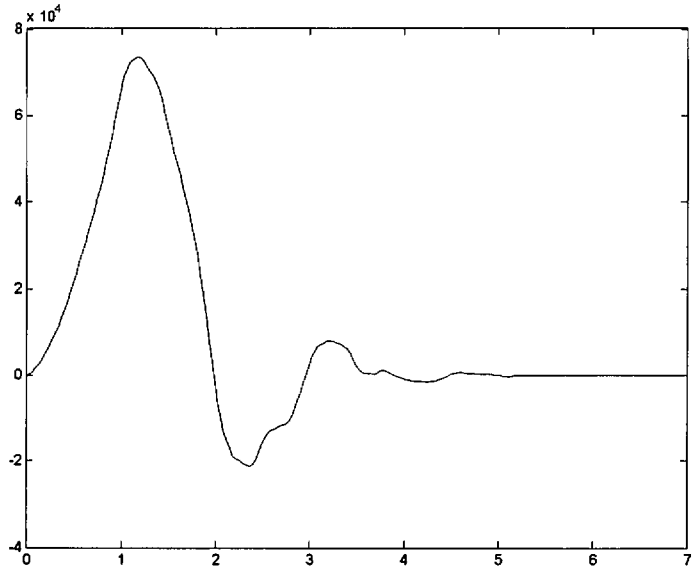


Figure 3. 5: Daubechies 4 scaling function [61].

Plots of the scaling function and wavelet function for the Daubechies 4 wavelet are shown in Figure 3. 4 and Figure 3. 5, respectively.

3.2 .7b QUADRATIC B-SPLINES WAVELETS

Quadratic B-spline wavelets have been used for analyzing ERPs in general but have not been used in particular for the detection of AD. Spline wavelets have been used for their near optimal time-frequency localization properties [51, 52, 63]. These wavelets are semi-orthogonal (See Appendix C) and have compact support in time [57]. The scaling function and wavelet for the quadratic B-splines are shown in Figure 3. 6 and Figure 3. 7. The filter coefficients for the quadratic B-spline wavelets are given in Table 3. 2.

Table 3. 2: Filter coefficients from [47, 49] for quadratic B-spline.

Coefficients of the Truncated Decomposition Filters h, g (IIR) and Reconstruction Filters p^2, q^2 (FIR) for Quadratic Spline Filters				
k	$h(k)$	$g(k)$	$p^2(k)$	$q^4(k)$
-10	+0.00157	-0.00388		
-9	+0.01909	-0.03416		
-8	-0.00503	+0.00901		
-7	-0.04440	+0.07933		
-6	+0.01165	-0.02096		
-5	+0.10328	-0.18408		
-4	-0.02593	+0.04977		+1/480
-3	-0.24373	+0.42390		-29/480
-2	+0.03398	-0.14034	+1/4	+147/380
-1	+0.65523	-0.90044	+3/4	-303/480
0	+0.65523	+0.90044	+3/4	+303/480
1	+0.03398	+0.14034	+1/4	-147/380
2	-0.24373	-0.42390		+29/480
3	-0.02593	-0.04977		-1/480
4	+0.10328	+0.18408		
5	+0.01165	+0.02096		
6	-0.04440	-0.07933		
7	-0.00503	-0.00901		
8	+0.01909	+0.03416		
9	+0.00157	+0.00388		

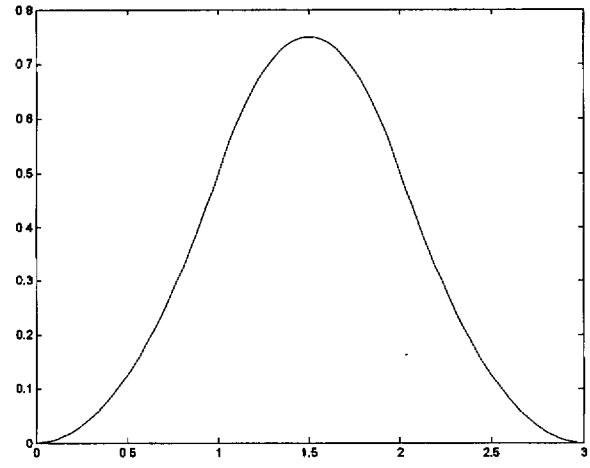


Figure 3. 6: Quadratic B-spline scaling function [61].

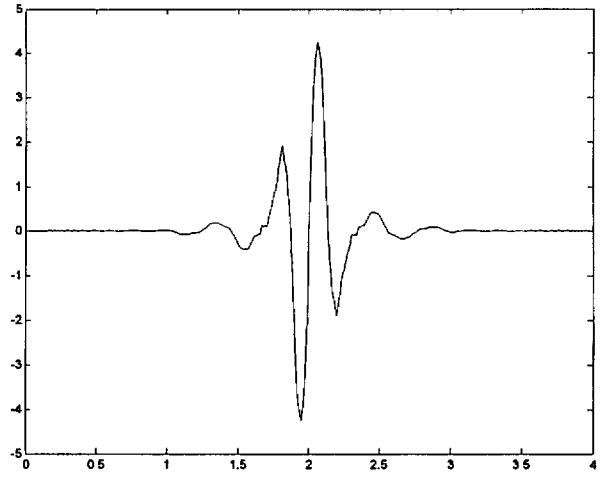


Figure 3. 7: Quadratic B-spline wavelet.

Quadratic B-spline wavelets are discussed in more detail in Appendix G.

3.3 PATTERN RECOGNITION TECHNIQUES

Many studies use statistical methods in the analysis of EEG signals for classification purposes, however, automated classification algorithms have become more widely used. Some of these automated methods are discussed in this section. In automated classifica-

tion applications, it is common to obtain the distinctive features (e.g. wavelet coefficients) of the signals to be identified then present the classification algorithm with training data consisting of those features. The trained system is then evaluated on data not seen during the training [64].

3.3.1 MLP NEURAL NETWORK

The Multilayer Perceptron (MLP) neural network is a popular and powerful classifier used in classification of signals whose classes are separated by a complex decision boundary. Neural networks are typically trained with a subset of the data with the class information available and then tested with another subset of data where the classes are then determined by the network. The training allows for the network to identify key features in the signals that are characteristic of the particular classes. Since it is possible that some unknown information or characteristics may exist within the signal, a neural network may be able to detect these characteristics. An MLP is a feed-forward network that has one or more hidden layers between the input and output layers. The MLP is trained using the supervised backpropagation learning rule (see section 3.4.1a) to distinguish between the different classes (two in this study, AD and normal).

The dimensionality of the signals to be classified determines the number of input nodes (wavelet coefficients in this study). The number of output nodes is equal to the number of classes. The input, I_i, $i = 1, 2, \ldots,$ *(62 or 121 for this study)* depends on the wavelet type or other feature. The number of hidden layers and the number of nodes, H_j, $j = 1, 2, \ldots J$, at each layer depend on the application. O_k, $k = 1, 2$ are the values of the output nodes.

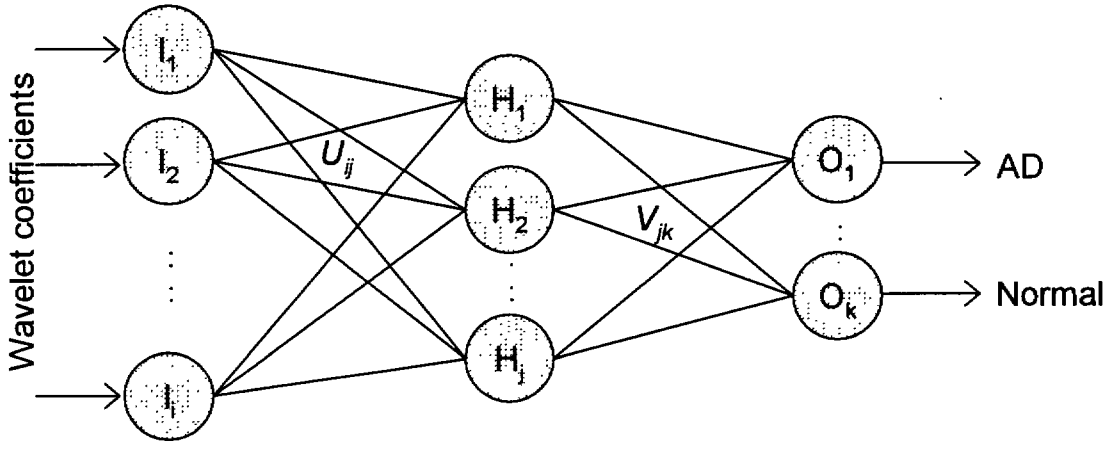


Figure 3. 8: MLP Neural Network architecture.

The nodes are all fully interconnected to the nodes of the adjacent layers by a set of weights. The weights connecting the input nodes to the hidden layer nodes are denoted by U_{ij}, and the weights connecting the hidden layer to output nodes are denoted by V_{jk}. The weights are determined through the backpropagation training algorithm described in the next section [20, 65, 66].

3.3.1a BACKPROPAGATION LEARNING RULE

Backpropagation is one of the most general and simple methods for supervised training of multilayer networks. During training, the desired outputs for the given input patterns are known and therefore the input to hidden weights can be adjusted to approximate the outputs. The credit assignment problem exists since there is no set weight for the hidden unit. Backpropagation allows for the calculation of an effective error for each hidden unit enabling the derivation of a learning rule for the weights between the input and hidden layers.

During the learning phase, the untrained network is presented with patterns and the outputs of the network are compared with the known class information for those patterns. The difference between the outputs and the target values is the error or criterion function which is a scalar function of the weights. This function is minimized when the output of the network equals the desired output and thus the weights are adjusted to reduce the error. The training error is the sum over the output units of the squared difference between the desired output, t_k and the actual output, z_k and is given as:

$$J(\mathbf{W}) \equiv \frac{1}{2}\sum_{k=1}^{c}(t_k - z_k)^2 = \frac{1}{2}\|\mathbf{t} - \mathbf{z}\|^2 \tag{3. 42}$$

where c is the length of the output vectors and $\mathbf{W}$ is the weight matrix. The backpropagation learning rule is based on the gradient decent algorithm where the weights are initialized randomly and then changed in the negative direction of the gradient in order to reduce the error:

$$\Delta\mathbf{W} = -\eta\frac{\partial J}{\partial \mathbf{W}} \tag{3. 43}$$

where η is the learning rate, indicating the relative step size of the change in weights. The algorithm demands taking a step in weight space that lowers the corresponding value of the criterion function. The algorithm takes a weight vector at iteration m and updates it according to $\mathbf{W}(m+1) = \mathbf{W}(m) + \Delta\mathbf{W}(m)$ where m is an index of the pattern presentation.

For a three layer network, considering the hidden to output weights, w_{jk},

$$\frac{\partial J}{\partial w_{kj}} = -\delta_k \frac{\partial net_k}{\partial w_{kj}} \tag{3. 44}$$

where δ_k is defined as $\delta_k = -\partial J / \partial net_k$. Assuming that the activation function (a nonlinear function emitted by the hidden unit), $f(\cdot)$ is differentiable, Equation (3.42) can be differentiated so that δ_k becomes $\delta_k = (t_k - z_k) f' net_k$. Together these results give the weight update rule for the hidden to output weights:

$$\Delta w_{kj} = \eta \delta_k y_j = \eta (t_k - z_k) f'(net_k) y_j \tag{3.45}$$

The sensitivity for a hidden unit is defined as

$$\delta_j \equiv f'(net_j) \sum_{k=1}^{c} w_{kj} \delta_k \tag{3.46}$$

and the learning rule for the input to hidden layer is given as

$$\Delta w_{ji} = \eta x_i \delta_j = \eta \left[\sum_{k=1}^{c} w_{kj} \delta_k \right] f'(net_j) x_i \tag{3.47}$$

Equations (3.45) and (3.47) along with the training protocols give the backpropagation algorithm which may also be referred to as the "backpropagation of errors" algorithm because the error propagates back from the output layer to the hidden layer in order to update the input to hidden weights [65].

3.3.2 LEARN++

Learn++ is an incremental learning algorithm, inspired in part by AdaBoost, which can learn additional information from new data such even when such data introduce new classes. For this particular study, there are only two classes, AD and Normal control subjects, however, new classes representing the stages of AD could be introduced for further analysis making this algorithm suitable for more advanced study of this particular

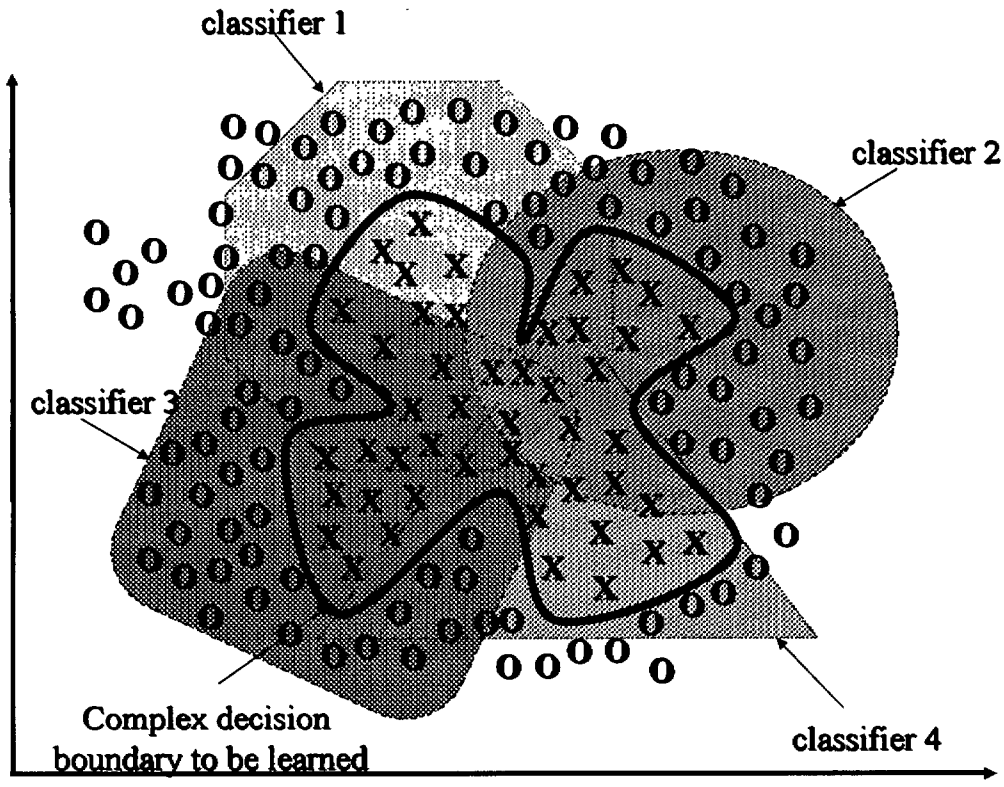


Figure 3. 9: Weak Classifiers.

problem. For each database that becomes available, Learn++ generates a number of relatively weak classifiers, whose outputs are combined through a weighted majority voting scheme to obtain the final classification. The weak classifiers (or weak hypotheses) are created using strategically chosen subsets of the available database and are shown in Figure 3. 9 to learn only the region of the subset for which they are trained. A dynamically updated distribution over the training data instances is calculated in such a way that the distribution is biased towards instances that have not been adequately learned by the previous hypotheses.

The Learn++ algorithm is outlined in Figure 3. 10. For each database D_k, $k=1,...,K$ made available to Learn++, the inputs to the algorithm are:

(i) a sequence of m training data instances x_i with their correct labels y_i,

(ii) a weak classification algorithm WeakLearn, for use as a base classifier, and (iii) an integer T_k that specifies the number of classifiers (hypotheses) to be generated. The WeakLearn can be any supervised classifier that can achieve at least 50% classification performance on the instances of its own training dataset. A neural network with appropriate parameters can serve as a weak classifier (weak Multilayer Perceptron, MLP, used in this case).

Learn++ starts by initializing a weight distribution, according to which a training subset TR_t and test subset TE_t are drawn at the t^{th} iteration of the algorithm. This distribution is initially uniform, giving equal probability to each instance to be selected into the first training subset. If an additional database is introduced, the weights are then adjusted by the Init_dist subroutine (described later). At each iteration t, the weights w_t from previous iteration are normalized to give a legitimate distribution D_t (step 1). Training and test subsets are then selected according to D_t (step 2), and the weak classifier is then trained with the training subset (step 3). A hypothesis h_t is obtained as the t^{th} classifier, whose error is then computed on $TR_t + TE_t$ (entire database D_k) by adding the distribution weights of the misclassified instances,

$$\varepsilon_t = \sum_{i:h_t(x_i)\neq y_i} D_t(i) \tag{3.48}$$

The WeakLearn requirement for classifying at least half of the training data correctly is enforced by requiring that the error be less than ½. If this is the case, the error is normalized to obtain the normalized error

$$\beta_t = \varepsilon_t / (1-\varepsilon_t), \quad 0 < \beta_t < 1 \tag{3.49}$$

Algorithm Learn++

Input: For each dataset drawn from $\mathcal{D}_k$ $k=1,2,...,K$

- Sequence of m examples $S=[(x_1,y_1),(x_2,y_2),...,(x_m,y_m)]$.
- Weak learning algorithm **WeakLearn**.
- Integer T_k, specifying the number of iterations.

Do for each $k=1,2,...,K$:

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

Call Init_dist if k>1 for distribution initialization when a new database becomes available

Do for $t = 1,2,...,T_k$:

1. Set $D_t = w_t \Big/ \sum_{i=1}^{m} w_t(i)$ so that D_t is a distribution.

2. Draw training TR_t and testing TE_t subsets from D_t

3. Call **WeakLearn** to be trained with TR_t.

4. Obtain a hypothesis h_t : X -> Y, and calculate the error of h_t: $\varepsilon_t = \sum_{i:h_t(x_i)\neq y_i} D_t(i)$ on $TR_t + TE_t$.

If $\varepsilon_t > 1/2$, discard h_t and go to step 2. Otherwise, compute normalized error as $\beta_t = \varepsilon_t / (1-\varepsilon_t)$

5. Call weighted majority voting and obtain the composite hypothesis, $H_t = \arg\max_{y\in Y} \sum_{t:h_t(x)=y} \log(1/\beta_t)$

6. Compute the error of the composite hypothesis

$$E_t = \sum_{i:H_t(x_i)\neq y_i} D_t(i) = \sum_{i=1}^{m} D_t(i)\left[\left[H_t(x_i) \neq y_i\right]\right]$$

7. Set $B_t = E_t / (1-E_t)$ and update the weights:

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

$$= w_t(i) \times B_t^{1-\left[\left[H_t(x_i)\neq y_i\right]\right]}$$

Call Weighted majority voting and **Output** the final hypothesis:

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

Figure 3. 10: Algorithm Learn++.

If the error is more than ½, the current hypothesis is discarded, and a new training subset is selected (step 4). All hypotheses generated thus far are then combined using the weighted majority voting to obtain the composite hypothesis H_t (step 5).

$$H_t = \arg\max_{y\in Y} \sum_{t:h_t(x)=y} \log(1/\beta_t) \qquad (3.\ 50)$$

The weighted majority voting chooses the class receiving the highest vote from all hypotheses. The hypotheses with good performance records are awarded higher voting weights. The error of the composite hypothesis is then computed as the sum of distribution weights of the instances misclassified by H_t (step 6)

$$E_t = \sum_{i H_t(x_i) \neq y_i} D_t(i) = \sum_{i=1}^{m} D_t(i) \left[\left| H_t(x_i) \neq y_i \right|\right] \tag{3.51}$$

where $\left[\left|\bullet\right|\right]$ evaluates to 1, if the predicate holds true. The normalized composite error B_t is then computed as

$$B_t = E_t / (1 - E_t), \quad 0 < B_t < 1 \tag{3.52}$$

which is then used in updating instance weights (step 7):

$$w_{t+1}(i) = w_t(i) \times \begin{cases} B_t, & \text{if } H_t(x_i) = y_i \\ 1, & \text{otherwise} \end{cases} \tag{3.53}$$

$$= w_t(i) \times B_t^{I - \left[\left| H_t(x_i) \neq y_i \right|\right]}$$

The weight update rule in Equation (3.53) reduces the weights of the instances that are correctly classified, making them less likely to be selected into the next training subset. The probability of misclassified instances being selected into the next training subset is then increased. The algorithm is essentially forced to focus on instances that are difficult to classify. In an incremental learning setting the newly introduced instances quickly gain the focus of the algorithm. When a new database is introduced, the Init_dist routine initializes the weight distribution and the routine jumps to step 5 and calls the weighted majority with the new database, determines which instances are misclassified by the current composite hypothesis, and initializes the weight distribution accordingly. This works effectively, particularly when instances from a previously unknown class are

introduced. At any point, a final hypothesis H_{final} can be obtained by combining all hypotheses generated thus far using the weighted majority voting

$$H_{final}(x) = \arg\max_{y \in Y} \sum_{k=1}^{K} \sum_{t:h_t(x)=y} \log \frac{1}{\beta_t} \tag{3. 54}$$

The weight update rule allows effective acquisition of new information when new databases are introduced [64, 67].

3.3.3 BACKGROUND FOR DATA FUSION

The results from the different feature sets with the MLP and also with the ensemble of MLP networks have repeatedly been similar to those of a clinical evaluation.
The results for each feature with an MLP neural network and the Learn++ algorithm are shown in Chapter 5. However, since Learn++ is an incremental learning algorithm, it also has the capability of learning additional information provided by additional features, hence data fusion.

Data fusion performed using the Learn++ algorithm on other datasets has been shown to improve the performance compared to that of a single feature set. The use of this method for combining the features yields results based on more information than those achieved with a single set of features, therefore providing a more informed classification decision [68]. Intuitively, we expect the performance of the data fusion algorithm to be at least as accurate as the best feature set performance.

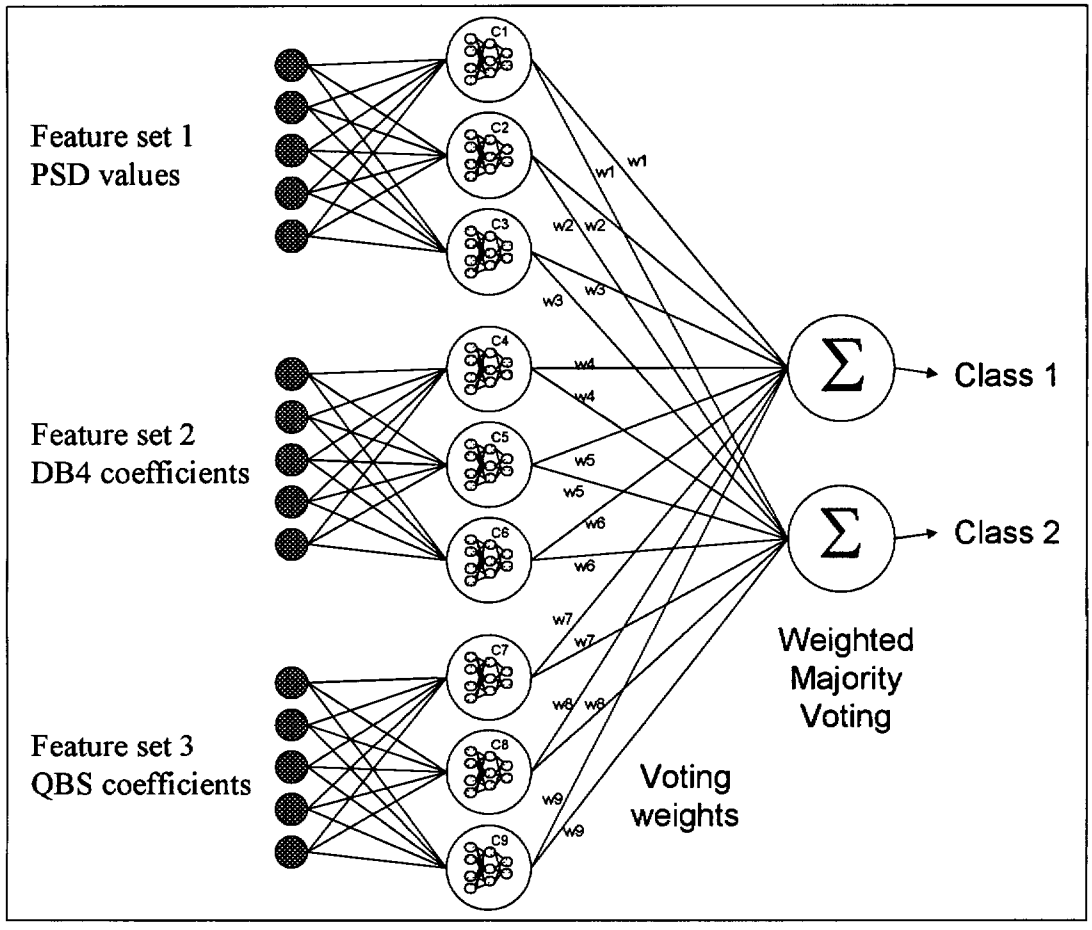


Figure 3. 11: Data fusion using an ensemble of classifiers.

CHAPTER 4

IMPLEMENTATION

4.1 EEG DATA

The data set used for this project consisted of EEG acquired at Drexel University from the patients recruited by the University of Pennsylvania for the development of an algorithm for the early detection of Alzheimer's disease. Table 4.1 and Table 4.2 show the inclusion and exclusion criteria, respectively, for the study.

Thirty-two elderly test subjects, 14 diagnosed with probable AD and 18 cognitively normal controls, have so far been recruited for the study. ERP recordings were obtained from each subject using the oddball paradigm while they were comfortably seated facing a computer screen in a specially designated room. The novel paradigm was used to take advantage of the paradigm's ability to elicit responses in the time-domain signal that have been indicative of mental status and the differentiation between forms of dementia.

Table 4. 1: Inclusion Criteria for the study (for information on any test criteria Appendix F includes a list of resources for further information.

Inclusion Criteria for Research Subjects		
	Normal subjects	AD subjects
Age	60 years or older	60 years or older
Clinical Dementia Rating (CDR) score	0	Greater than or equal to 0.5
Mini-mental Status (MMS) exam	Greater than or equal to 24	Less than 24
Clinical assessment results for AD	Negative	Positive
NINCDS-ADRDA criteria	does not satisfy	satisfies

Table 4. 2: Exclusion Criteria for study.

Exclusion Criteria for both groups
(i) evidence of any central nervous system neurological disease
(ii) use of sedatives or antidepressant medications within 48 hours of ERP acquisition

The protocol originally described by [22], with slight modifications was used. Binaural audiometric thresholds were determined for each subject using a 1000 Hz tone. The evoked response stimulus was presented to both ears using stereo speakers at an amplitude determined to be comfortable for their particular hearing level. The stimulus consisted of tone bursts 100 ms in duration, including 5 ms inset and offset envelopes. Tones of 1000 and 2000 Hz were presented in a random sequence with the tones occurring in 65% and 20% of the trials respectively. The remaining 15% of the trials consisted of novel sounds presented randomly. These included 60 unique environmental sounds that were recorded digitally and edited to 200 ms duration.

A total of 1000 stimuli, including frequent 1000 Hz (n=650), infrequent 2000 Hz tones (n=200) and novel sounds (n=150) were delivered to each subject with an inter-stimulus interval of 1.0-1.3 seconds. The subjects were instructed to click a mouse each time they heard the 2000 Hz tone. With frequent breaks (e.g. approximately three minutes of rest every five minutes), the data collection process lasted about 30 minutes per subject with each session preceded by a 1 minute practice session without the novel sounds. Each recording is 1 second in duration with a 200 ms prestimulus interval.

The ERPs were recorded from 19 tin electrodes embedded in an elastic cap, using linked mastoids as reference. The electrode impedances were kept below 20 kΩ to yield a good signal with the high-impedance MANSCAN amplifier system used in the study.

Artifacts were identified and rejected by the EEG technician. The remaining scalp potentials were amplified, digitized at 256 Hz/channel (19 channels) and stored.
The averaging protocol involved averaging 30~85 recordings per patient yielding 1~3 recordings per patient. The variability inherent in ERPs ensures that no two recordings are ever identical and therefore separate averages were all included in the analysis to provide additional data. All averages have been notched filtered at 59-61 Hz and baselined with the prestimulus interval.

Four of the original 32 subjects were unable to complete the study due to various reasons, and hence their recordings were excluded from further analysis. All multiple averages from the 28 remaining subjects, a total of 75 recordings, 26 from AD and 49 from normal controls, were therefore analyzed. In the analysis of these signals, "multiple averages" refers to those 1~3 averages (30~40 recordings per average) per patient while "overall average" refers to the average of all signals (80~90 recordings per average). These recordings have been utilized in different training/testing schemes in order to gain an understanding of the dynamics of the data. The training/testing sets are defined in different ways according to the averages used and level of cross validation. This is explained further in Section 5.4.

4.2 APPROACH

Data from the Pz electrode in response to the target tones were used since previous studies have indicated this particular electrode to have the strongest response to the target tones in the oddball paradigm. The data were analyzed using several methods which have been described in detail in the previous sections. Prior to analysis the data were

normalized by subtracting the mean and dividing by the maximum. The wavelet transform and PSD have been implemented to extract features from the signal.

The preprocessed signals were then used to train an MLP neural network to obtain initial classification performance. The Learn++ algorithm is then implemented on only the wavelet data because the performance for the PSD values with the MLP were poor. Learn++ has the ability to learn new data without forgetting previously acquired knowledge and without requiring access to any of the previously seen data, even when the new data introduce new classes. The classification performance of Learn++ is similar to that of a single MLP, however, Learn++ is equipped with the ability to learn incrementally which for this particular application may prove useful in the future if more classes are added. With this in mind, after the initial analysis with an MLP and also with the Learn++ algorithm, data fusion was performed with the two sets of wavelet data using a data fusion method with Learn++.

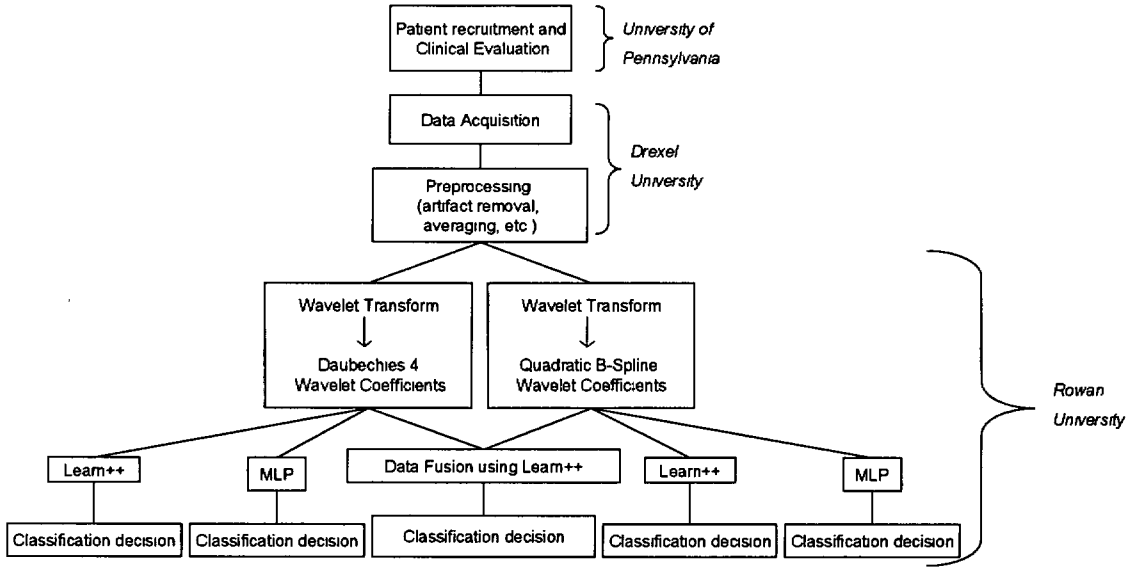
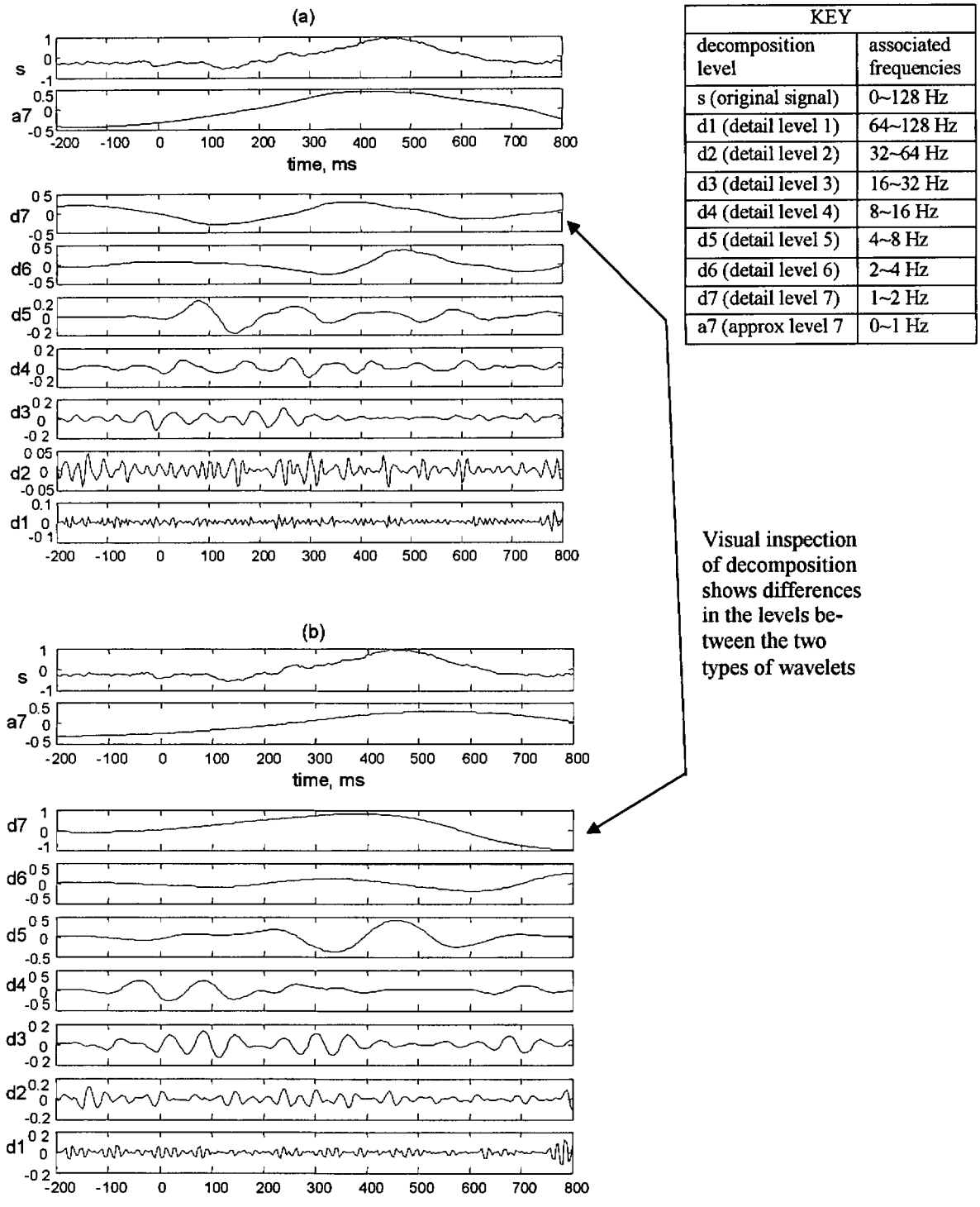


Figure 4. 1: Overview of analysis.

4.3 OUTLINE OF METHOD AND REASONING

Analysis of the wavelet decomposition shows that complimentary information may be extracted for each different choice of mother wavelet. Figure 2 (a) and (b) show a 7-level decomposition of the same signal, decomposed using the Daubechies 4 and Quadratic B-Spline wavelets. The differences between the decomposition of the same signal using the two different types of wavelets imply that they may be extracting different information, albeit very minor, from the same signal. The analysis with the MLP and Learn++ show that even though different information may be extracted from the signals, the two wavelets give similar performance. This may be the case if each wavelet was extracting different, relevant and complementary information from the signal. If this is indeed the case, then combining the feature sets could yield a better classification performance and therefore could increase the predictive value of a classifier. Figure 3.11 shows a diagram involving an ensemble of classifiers that are each trained with a set of features and combined through a weighted majority voting scheme to obtain classification.



KEY	
decomposition level	associated frequencies
s (original signal)	0~128 Hz
d1 (detail level 1)	64~128 Hz
d2 (detail level 2)	32~64 Hz
d3 (detail level 3)	16~32 Hz
d4 (detail level 4)	8~16 Hz
d5 (detail level 5)	4~8 Hz
d6 (detail level 6)	2~4 Hz
d7 (detail level 7)	1~2 Hz
a7 (approx level 7	0~1 Hz

Figure 4. 2: Wavelet decomposition comparison (normal subject)—(a) Daubechies 4 (DB4), (b) Quadratic B-Spline (QBS).

CHAPTER 5

RESULTS

5.1 K-FOLD CROSS VALIDATION

A cross-validation procedure is implemented to estimate the true generalization performance of each of the classifiers or ensemble of classifiers. This method can be used to estimate any statistical parameter with a measure of certainty in the original estimate [65, 66]. The dataset is divided into k-blocks as shown in Figure 5. 1. The network is trained with all *k-1* blocks and the k^{th} block is used as the testset. This procedure is repeated k times so that each block is used as a testset one time. The average of the performance values on each of the testsets is the k-fold cross-validation performance on the dataset.

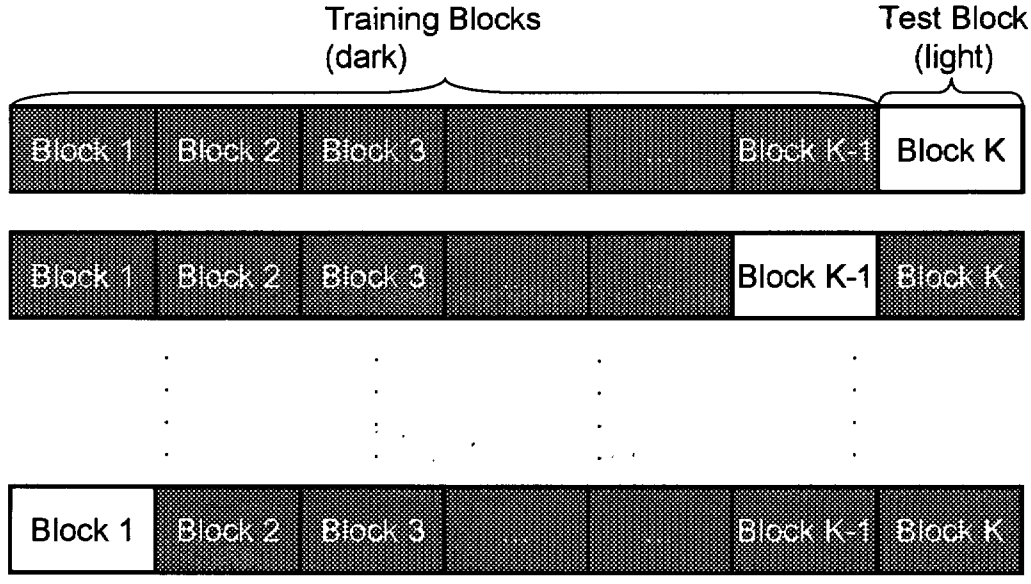


Figure 5. 1: K-fold cross validation [58].

5.2 OVERVIEW OF RESULTS

Wavelet analysis was performed on the dataset using the Daubechies 4 and quadratic B-spline wavelets. These different feature sets were used to train and test an MLP neural network and Learn++. The power spectral analysis was also performed on the signal however classification was attained solely with an MLP.

Performance figures are provided in this chapter. The performance of both types of wavelets was compared using the Learn++ algorithm and an MLP. Overall results for Learn ++ and MLP were generally similar. It was expected that through the use of Learn++ with these features and further, through the use of data fusion with Learn++, the performance could increase, therefore the results for data fusion of the two types of wavelets are shown.

5.3 RESULTS USING PSD

Five-fold cross validation was performed on power spectral density values for the following breakdown of the frequencies:

[1] (0.25 - 1.50 Hz) delta1
[2] (1.25 - 2.50 Hz) delta2
[3] (2.25 - 3.50 Hz) delta3
[4] (3.25 - 4.75 Hz) theta1
[5] (4.50 - 6.00 Hz) theta2
[6] (5.75 - 7.50 Hz) theta3
[7] (7.25 - 9.00 Hz) alpha1
[8] (8.75 - 10.50 Hz) alpha2
[9] (10.25 - 12.00 Hz) alpha3
[10] (11.75 - 13.50 Hz) alpha/beta
[11] (13.25 - 16.00 Hz) beta1
[12] (15.75 - 18.50 Hz) beta2
[13] (18.25 - 21.00 Hz) beta3

Table 5. 1: MLP performance from 5-fold crass validation with PSD values from frequencies 18.5 Hz and below.

HIDDEN LAYER NODES ↓	ERROR GOAL →	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
20		68.0	68.0	60.0	53.3	63.3	67.3	56.0	63.3	**75.3**	56.7
25		64.0	70.7	61.3	60.7	56.0	60.7	60.7	60.7	64.0	70.7
30		67.3	71.3	64.0	63.3	60.0	63.3	64.0	60.0	74.7	68.0
35		64.0	67.3	68.0	57.3	64.7	68.0	60.0	70.7	67.3	56.0

Only the overall averages were used for this analysis. The frequency bands were left slightly overlapping so that no information was lost from the signal. The conventional EEG bands were broken down to explore the frequencies thoroughly. The power spectral density was estimated using a modified periodogram method. The frequency bands were explored systematically with the best results achieved using the first 12 frequency bands together (shown in Table 5. 1).

An MLP was used for classification with different numbers of hidden layer nodes set to 20, 25, 30, 35 and error goal 0.01 to 0.1 in increments of 0.01. The best results obtained from this method was 75.3% using the first 12 frequency bands and with the network having 20 hidden layer nodes and an error goal of 0.09.

The results for the power spectral density range from 53% to 75% depending on the network parameters. These results indicate that the power spectral density values may not be adequate for this classification task.

5.4 RESULTS USING WAVELET ANALYSIS

For the signals used in the analysis, the following diagram summarizes the details of the decomposition using the DWT. The original signals were 257 samples long and the high-

highest frequency in the signals was 128 Hz. The breakdown of the signal at each level of the DWT can be seen in the representation in Figure 5. 2.

Extra coefficients are added on to the wavelet coefficients during the filtering implementation as a results of the convolution operation. The number of coefficients for each detail level and the highest level approximation is stored in a length vector, L as: L = [L(a7) L(d7) L(d6)...L(d1) L(s)] (where L(s) is the length of the entire signal). The actual number of coefficients at each level varies with each type of wavelet and these values are given below in the section for each wavelet.

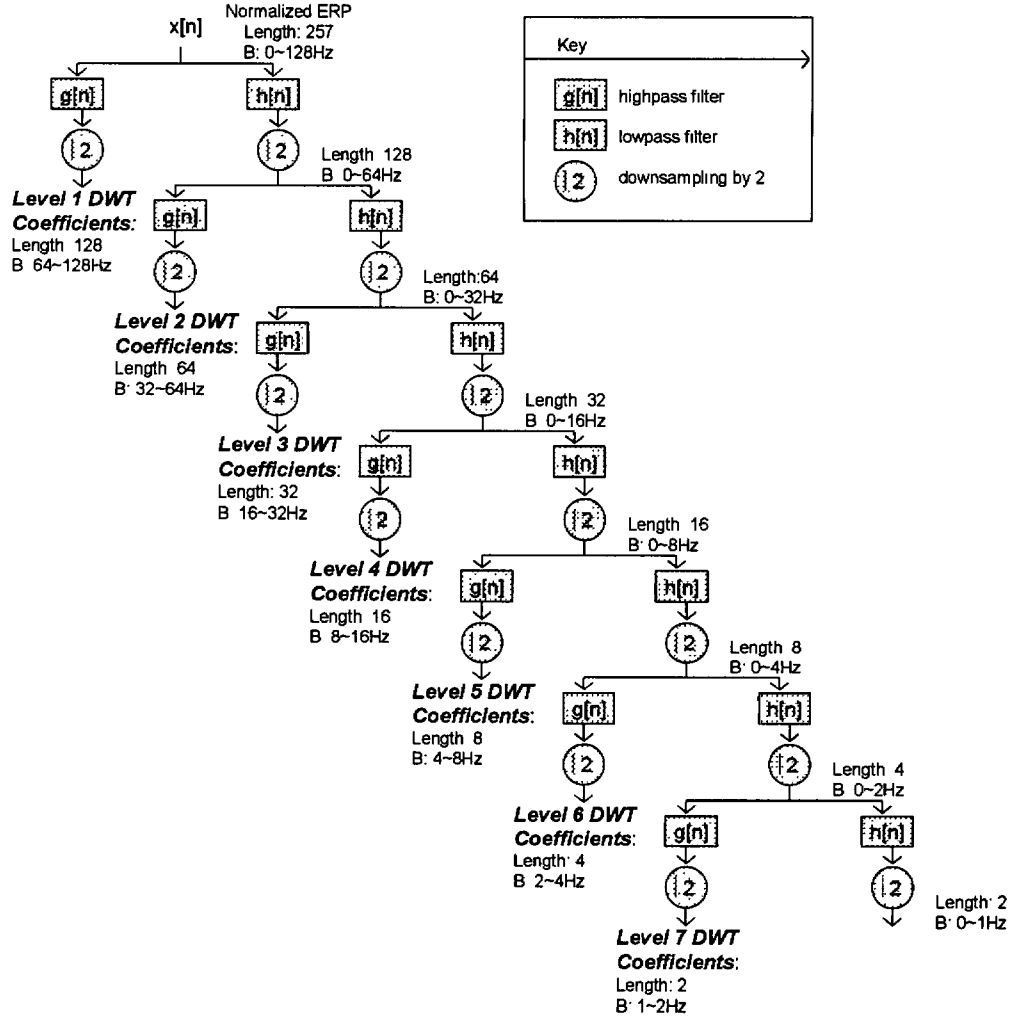


Figure 5. 2: 7-level wavelet decomposition.

5.4.1 DAUBECHIES 4

The number of coefficients per decomposition level was determined and is outlined in Table 5. 2. The coefficients for the levels a7, d7, d6, d5, d4 were used which correspond to the frequencies below 16 Hz. This means that only the first 62 coefficients were used for the analysis. The breakdown of the frequency bands is according to the values calculated in the block diagram shown previously.

Omitting the highest frequencies of the signal results in a good amount of data reduction, but how much is too much? The approximation of the signal with the exclusion of the superfluous coefficients can be seen by plotting the reconstruction of the signal. Figure 5. 3 shows the original signal and its reconstruction with only the specified number of coefficients as a comparison. The approximation using 62 coefficients closely follows the original signal indicating that there is no significant loss of information.

Figure 5. 4 shows the decomposition of a signal from (a) a normal control and (b) AD patient. Variations in the signals when compared to each other are obvious, however this is not always the case when just using visual inspection methods. The decomposition of the signal makes the components occurring at different frequencies more obvious when plotted in this manner.

Table 5. 2: Number of coefficients in each decomposition level for Daubechies 4.

approximation level 7	detail level 7	detail level 6	detail level 5	detail level 4	detail level 3	detail level 2	detail level 1
8	8	10	14	22	38	69	132

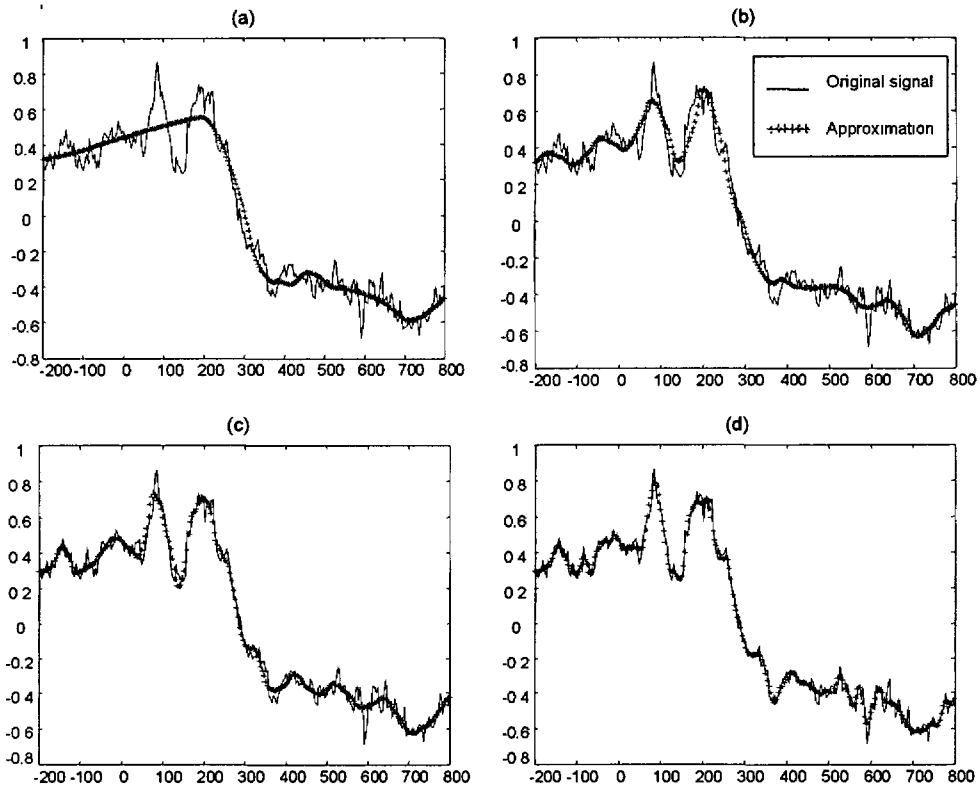
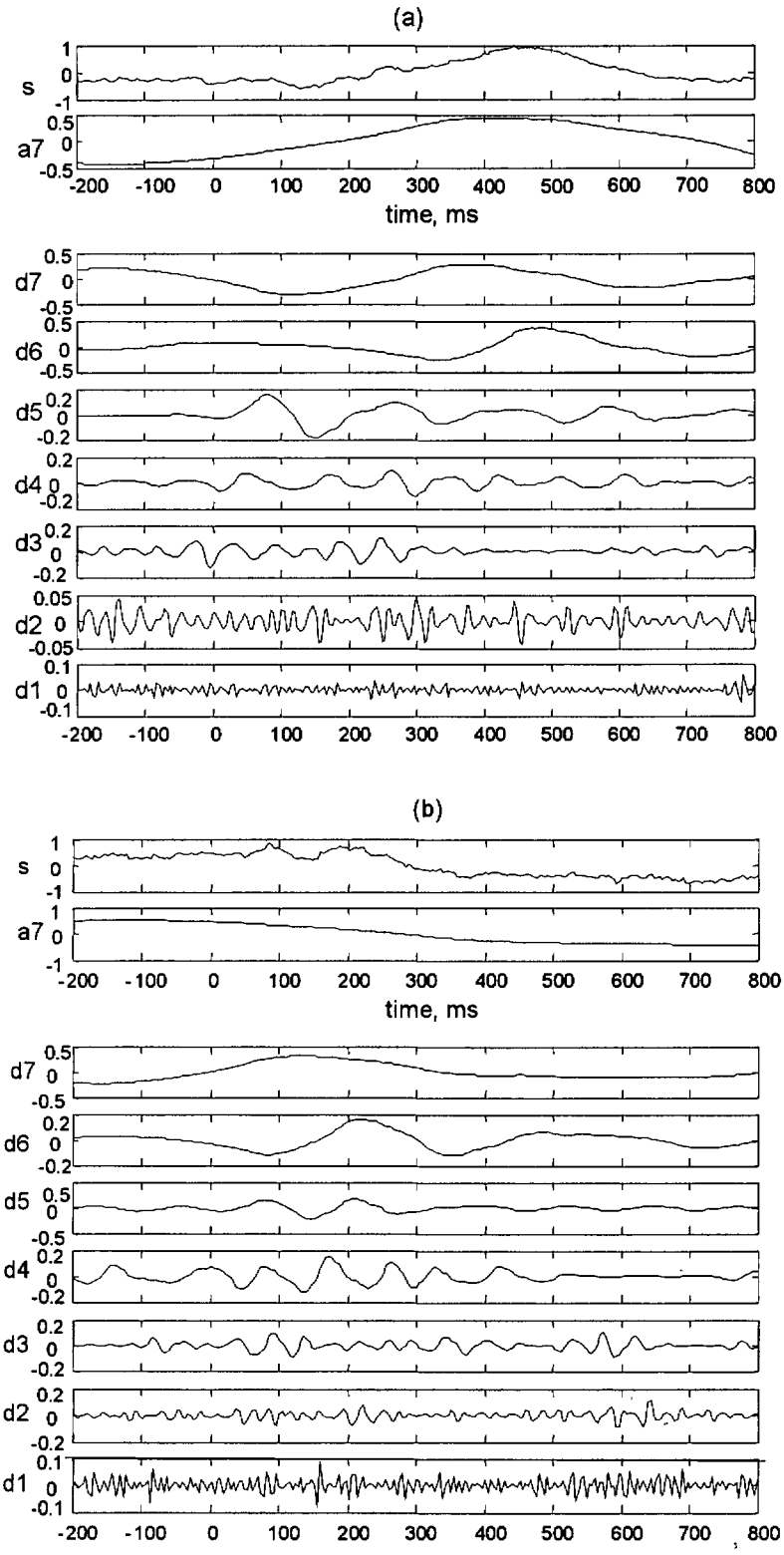


Figure 5. 3: Approximation of Daubechies 4 wavelet coefficients (+++) compared to original signal (solid line). Number of coefficients—(a) 40, (b) 62, (c) 100, and (d) 169.

The results presented in Table 5. 3 through Table 5.6 are the classification performance of the wavelet coefficients using an MLP neural network. The results are labeled for the level of cross validation and training/testing scenarios used. The data for training and testing was randomly selected in all cases with the multiple averages per patients kept together in either the training or testing dataset, never both. The number of AD and number of normal subjects were evenly split in the training or testing data as well. The network parameters, specifically the number of hidden layer nodes were varied from 20 to 35 and the error goal was varied from 0.005 to 0.1 in increments of 0.005. This was done to test the different variations of the network to determine the architecture that yields the best performance values given the training/testing scheme. The results for



KEY	
decomposition level	associated frequencies
s (original signal)	0~128 Hz
d1 (detail level 1)	64~128 Hz
d2 (detail level 2)	32~64 Hz
d3 (detail level 3)	16~32 Hz
d4 (detail level 4)	8~16 Hz
d5 (detail level 5)	4~8 Hz
d6 (detail level 6)	2~4 Hz
d7 (detail level 7)	1~2 Hz
a7 (approx level 7	0~1 Hz

Figure 5. 4: Wavelet decomposition, (a) normal subject, (b) AD subject.

training with half the data and testing with the other half of the data, 3, 5, and 10-fold cross validation with training/testing schemes mentioned in the captions are shown in the following tables. (These parameters are the same for the results given in Section 6.3.2 for the quadratic B-spline wavelet). The performance for training/testing with half the data was calculated using the best 10 out of 30 trials. The leave-one-out algorithm (limiting case for cross validation) is also included to give a better estimate of the performance of the algorithm.

5.4.1a MLP RESULTS FOR DAUBECHIES 4

Consistently, regardless of the level of cross validation, it appears that training and testing with the overall averages yields better results. Averaging the signals together smoothes the curves slightly and removes some of the noise, so this outcome makes sense.

The confidence intervals are large in most cases due to the small amount of data for this study. There are 28 patients total, so for 3-fold cross validation there are 9-10 subjects in the test set at a given time and for 10-fold cross validation there are only 2-3 subjects in the test set for each of the ten iterations. The results from training with half and testing with half of the data is the average over ten trials. The confidence intervals are smaller in this case because there is less variation between each different trial.

Table 5. 3: Train ½, test ½ --result is the overall average over ten trials (32 hidden layer nodes, 0.035 error goal).

DB4-MLP, 0.05 error goal, 20 hidden layer nodes	
SENSITIVITY	80.0+/-13.1
SPECIFICITY	87.8+/-6.0
+ PRED. VALUE	81.6+/-7.2
OVERALL PERF.	85.0+/-3.3

Table 5. 4: 3-fold cross validation—(a) train multiple/ test overall averages, (b)train/test overall averages, (c)train/test multiple averages, and (d)train multiple/test multiple + overall averages.

(a) 0.075 error goal, 28 hidden layer nodes	
SENSITIVITY	77.8+/-21.8%
SPECIFICITY	94.4+/-10.9%
+ PRED. VALUE	88.9+/-21.8%
OVERALL PERF.	88.9+/-12.6%

(b) 0.02 error goal, 22 hidden layer nodes	
SENSITIVITY	77.8+/-43.5%
SPECIFICITY	100.0+/-0.0%
+ PRED. VALUE	100.0+/-0.0%
OVERALL PERF.	92.6+/-14.5%

(c) 0.06 error goal, 33 hidden layer nodes and 0.075 error goal, 30 hidden layer nodes	
SENSITIVITY	52.1+/-27.9%
SPECIFICITY	89.6+/-4.8%
+ PRED. VALUE	71.7+/-11.8%
OVERALL PERF.	76.7+/-8.8%

(d) 0.065 error goal, 23 hidden layer nodes	
SENSITIVITY	59.6+/-10.6%
SPECIFICITY	88.0+/-15.8%
+ PRED. VALUE	73.5+/-30.9%
OVERALL PERF.	78.1+/-12.5%

Table 5. 5: 5-fold cross validation—(a) train multiple/ test overall averages, (b)train/test overall averages, (c)train/test multiple averages, and (d)train multiple/test multiple + overall averages.

(a) 0.1 error goal, 27 hidden layer nodes, 0.03 error goal, 20 hidden layer nodes, and 0.075 error goal, 21 hidden layer nodes	
SENSITIVITY	70.0+/-24%
SPECIFICITY	95.0+/-9.8 %
+ PRED. VALUE	93.3+/-13.1%
OVERALL PERF.	85.3+/-7.3%

(b) 0.04 error goal, 32 hidden layer nodes	
SENSITIVITY	80.0+/-24%
SPECIFICITY	95.0+/-9.8%
+ PRED. VALUE	93.3+/-13.1%
OVERALL PERF.	90.0+/-8.0%

(c) 0.095 error goal, 28 hidden layer nodes	
SENSITIVITY	60.0+/-23.9%
SPECIFICITY	90.1+/-9.8%
+ PRED. VALUE	77.5+/-19.6%
OVERALL PERF.	79.2+/-10.5%

(d) 0.095 error goal, 35 hidden layer nodes	
SENSITIVITY	52.5+/-31.4%
SPECIFICITY	93.6+/-9.5%
+ PRED. VALUE	67.1+/-36.6%
OVERALL PERF.	78.7+/-12.6%

Table 5. 6: 10-fold cross validation—(a)train multiple/ test overall averages, (b)train/test overall averages, (c)train/test multiple averages, and (d)train multiple/test multiple + overall averages.

(a) 0.065 error goal, 22 hidden layer nodes	
SENSITIVITY	70.0+/-29.9%
SPECIFICITY	95.0+/-9.8%
+ PRED. VALUE	65.0+/-29.4%
OVERALL PERF.	86.7+/-10.7%

(b) 0.025 error goal, 25 hidden layer	
SENSITIVITY	80.0+/-26.1%
SPECIFICITY	95.0+/-9.8%
+ PRED. VALUE	35.0+/-29.4%
OVERALL PERF.	86.7+/-13.6%

(c) 0.07 error goal, 25 hidden layer nodes	
SENSITIVITY	63.3+/-26.6%
SPECIFICITY	86.3+/-10.7%
+ PRED. VALUE	60.0+/-25.5%
OVERALL PERF.	74.4+/-16.6%

(d) 0.075 error goal, 33 hidden layer nodes	
SENSITIVITY	62.5+/-24.5%
SPECIFICITY	92.0+/-7.0%
+ PRED. VALUE	74.2+/-21.3%
OVERALL PERF.	81.1+/-9.3%

The performance values for all networks are better than the clinical evaluation performance of 75% except for the performance from 10-fold cross validation while training/testing with the multiple averages. Considering the confidence intervals for each of the cases, positive predictive value, sensitivity and specificity, most of these ranges include the values for a clinical evaluation. In summary, the performance of the algorithm is about par with the performance values given for a clinical evaluation.

5.4.1b LEARN++ RESULTS FOR DAUBECHIES 4

Results with Learn++ were achieved using weak MLP networks. The parameters for the neural networks were varied: number of hidden layer nodes (15, 20, 25, 30) and error goals (0.15, 0.1, and 0.2 for some cases). The number of classifiers was also varied. Training/testing with half the data over ten trials was implemented with Learn++ and the results are shown in Table 5. 7. The performance for this is slightly better than those achieved with the same scenario with the MLP. Five-fold cross validation was performed with two variations in the data–training/testing with overall averages and training/testing with the multiple averages. Results are then compared with those of the MLP.

Table 5. 7: Learn++ results for training/testing with half the data using the overall averages.

DB4-LEARN++-0.1 ERROR GOAL, 30 HIDDEN LAYER NODES, 10 CLASSIFIERS	
SENSITIVITY	70.0+/-10.5
SPECIFICITY	96.7+/-4.6
+ PRED. VALUE	95.1+/-6.5
OVERALL PERF.	87.1+/-2.8

Table 5. 8: Learn++ Results from training/testing with multiple averages.

LEARN ++ RESULTS WITH MULTIPLE AVERAGES						
# CLAS-SIFIERS	ERROR GOAL	# HIDDEN LAYER NODES	OVERALL PERF.	+ PRED VALUE	SPECIFICITY	SENSITIVITY
20	0.1	30	71.3+/-8.4%	51.3+/-32.1%	83.2+/-17.7%	56. 7+/-35.2%
15	0.1	25	72.0+/-9.6%	60.8+/-15.2%	84.3+/-5.3%	53.3+/-28.1%
10	0.15	30	77.7+/-5.0%	92.0+/-15.7%	96.4+/-7.1%	45.0+/-19.0%
5	0.2	30	73.8+/-8.8%	71.4+/-23.6%	81.0+/-17.0%	60.0+/-13.1%
3	0.1	30	74.7+/-10.0%	57.0+/-32.4%	90.7+/-5.7%	46.7+/-31.3%

Table 5. 9: Learn++ Results from training/testing with overall averages.

LEARN ++ RESULTS WITH OVERALL AVERAGES						
# CLAS-SIFIERS	ERROR GOAL	# HIDDEN LAYER NODES	OVERALL PERF.	+ PRED VALUE	SPECIFICITY	SENSITIVITY
20	0.15,0.1	25	82.0+/-12.5%	73.3+/-38.1%	95.0+/-9.8%	60.0+/-36.7%
15	0.15	15	81.3+/-12.5%	90.0+/-19.6%	86.7+/-26.1%	70.0+/-24.0%
10	0.1	15	85.3+/-7.3%	93.3+/-13.1%	95.0+/-9.8%	70.0+/-24.0%
5	0.1	30	90.0+/-13.1%	80.0+/-39.2%	100.0+/-0.0%	70.0+/-39.2%
3	0.1	30	82.0-10.5%	83.3+/-20.7%	83.3+/-20.7%	80.0+/4.0%

In this case the overall averages produce better classification performance compared to the multiple averages as they did with the MLP. Learn++ also yields a performance that in overall is comparable to that of a clinical evaluation. These results also indicate that Learn++ performs as well as a single strong classifier for classifying the Daubechies 4 wavelet coefficients.

5.4.1c LEAVE-ONE-OUT RESULTS FOR DAUBECHIES 4

The leave-one-out algorithm is the limiting case of cross validation and therefore gives the best estimate of the generalization performance of the classifier for the particular dataset in use. Results are shown in Table 5. 10 and Table 5. 11 for a single MLP and the Learn ++ algorithm, respectively. The results between the two algorithms are identical. The low values for the positive predictive value, sensitivity and specificity are explained

Table 5. 10: Leave-one-out results-MLP, 0.05 error goal, 20 hidden layer nodes.

	AVERAGE	CI
OVERALL PERF.	82.1429	14.4466
+ PRED. VALUE	21.4286	15.4776
SENSITIVITY	21.4286	15.4776
SPECIFICITY	60.7143	18.422

Table 5. 11: Leave-one-out results-Learn++ 0.15 error goal, 20 hidden layer nodes, 3 classifiers.

	AVERAGE	CI
OVERALL PERF.	82.1429	14.4466
+ PRED. VALUE	21.4286	15.4776
SENSITIVITY	21.4286	15.4776
SPECIFICITY	60.7143	18.422

as follows. Since there is only one instance being tested each of the 28 iterations, one of the values for A, B, C and D (refer to Appendix A) will be 1 while the rest will be 0. Whenever A is 1, the positive predictive value and sensitivity are 100% and whenever D is 1, the specificity is 100%. If B or C is 1 at any time then all values are 0 because these are only 1 if an instance was misclassified. This means that the positive predictive value and sensitivity will be zero when a normal subject is classified and can only have a value when an AD patient is classified correctly (the highest these values could be for this data-set would be 35.7% or 10/28. The reverse is the case for specificity, it will only have a value when a normal patient is classified correctly (the highest possible value would be 64.3% or 18/28). Also for each case, the overall performance can be calculated by (A+D)/(A+B+C+D).

Since only one instance is classified at a time, the overall performance turns out to be the sum of the specificity and positive predictive values which is actually the sum of

the patients classified correctly. The overall performance values are the only values that are meaningful in this case.

5.4.2 QUADRATIC B-SPLINE

The implementation of the quadratic B-spline wavelet was performed using filter coefficients given in [52] and the sub-band coding scheme described in section 3.2.6. The number of coefficients used was determined and are shown in Table 5. 12. The coefficients for the levels a7, d7, d6, d5, d4 were used which correspond to the frequencies below 16 Hz giving a total of 121 coefficients.

Just as with the db4, the coefficients were plotted to determine if there are significant differences in altering the number of coefficients used. This type of wavelet performs best with 121 coefficients, which incidentally corresponds to the 0~16 Hz frequency band and does poorly when more coefficients are added. The reconstruction of the signals in Figure 5. 5 using the specified number of wavelet coefficients seems to make this more obvious.

In Figure 5. 6 the decomposition of a normal control and that of an AD patient are shown. As with the Daubechies 4 wavelet in the previous section, differences can be seen between the two signals, however this is not always the case. The results for cross validation with an MLP are provided in Table 5. 13 through Table 5. 16 for the quadratic B-spline wavelet.

Table 5. 12: Number of coefficients in each decomposition level for Quadratic B-spline wavelet.

approximation level 7	detail level 7	detail level 6	detail level 5	detail level 4	detail level 3	detail level 2	detail level 1
20	20	22	26	33	48	78	138

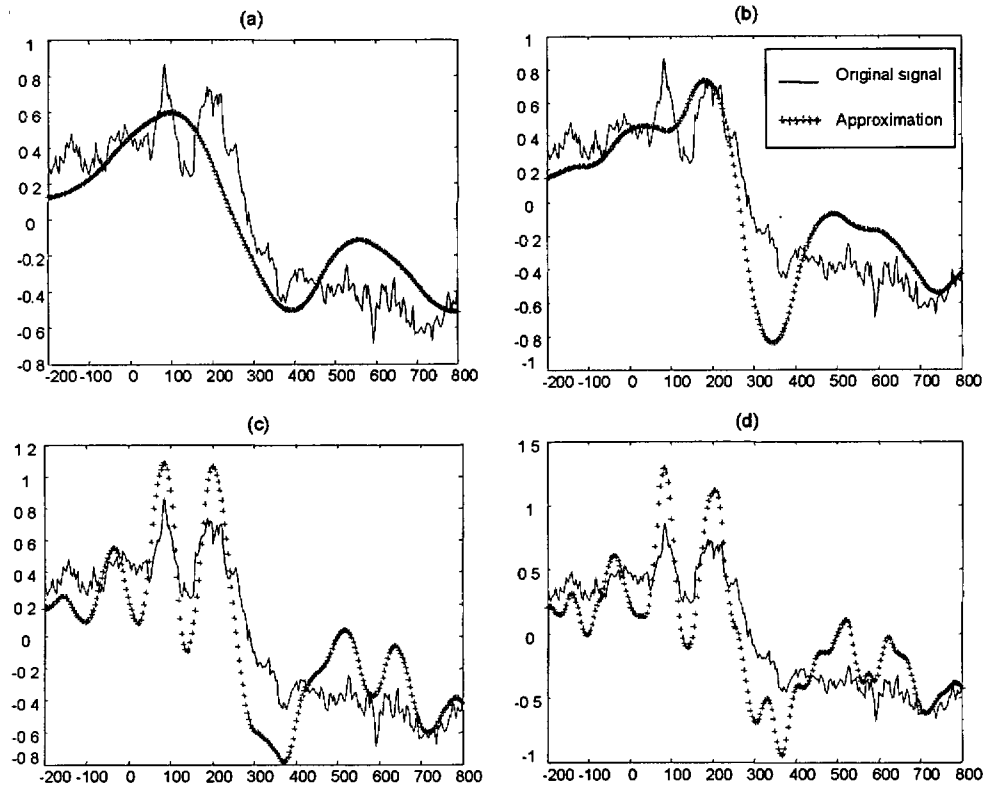
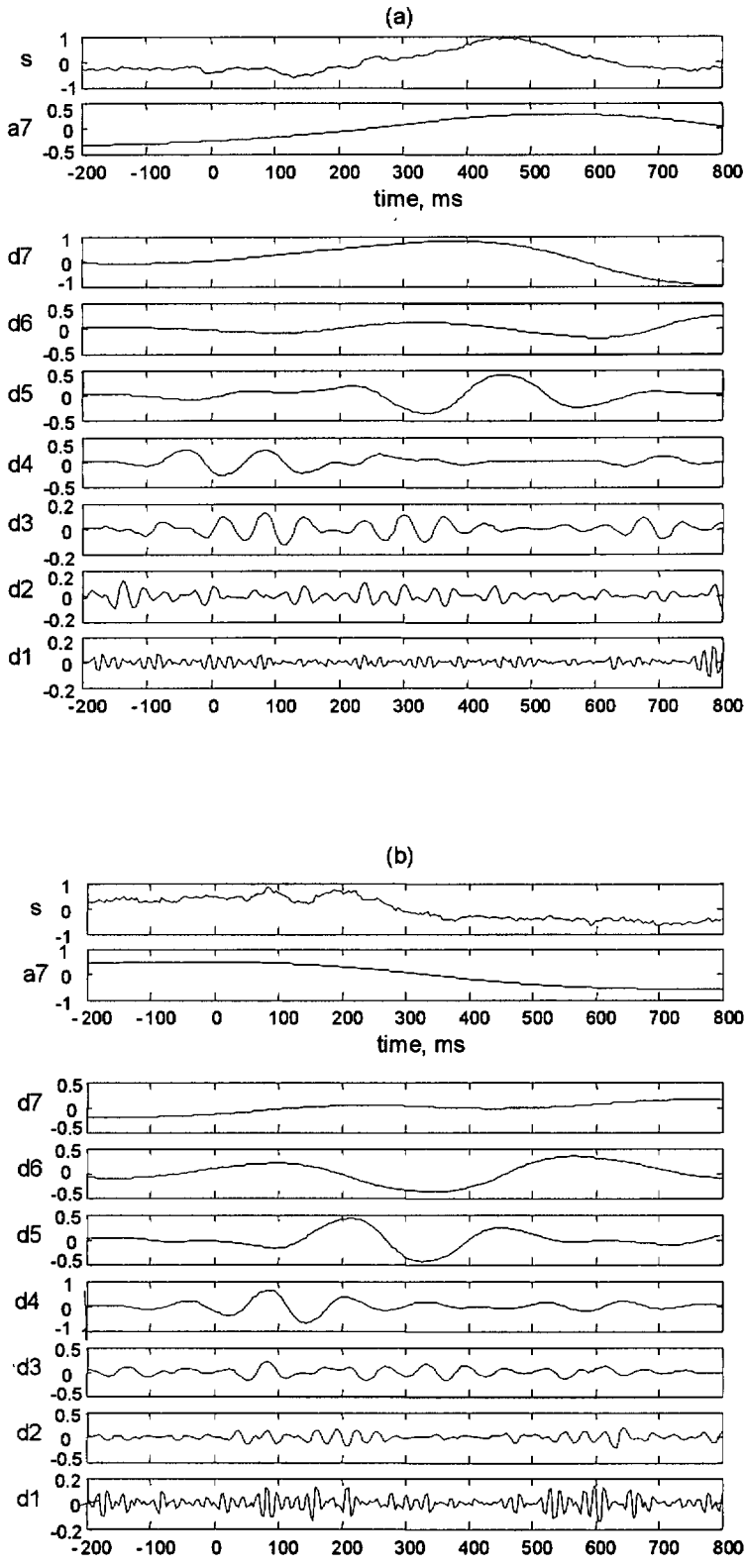


Figure 5. 5 Approximation of Quadratic B-Spline wavelet coefficients (red) compared to original signal (blue). Number of coefficients—(a) 88, (b) 121, (c) 169 and (d) 247.



KEY	
decomposition level	associated frequencies
s (original signal)	0~128Hz
d1 (detail level 1)	64~128Hz
d2 (detail level 2)	32~64Hz
d3 (detail level 3)	16~32Hz
d4 (detail level 4)	8~16Hz
d5 (detail level 5)	4~8Hz
d6 (detail level 6)	2~4Hz
d7 (detail level 7)	1~2Hz
a7 (approx level 7	0~1Hz

Figure 5. 6: QBS Wavelet Decomposition, (a) normal subject, (b) AD subject.

5.4.2a MLP RESULTS FOR QUADRATIC B-SPLINE

Table 5. 13: Train ½, test ½ --result is the overall average over ten trials.

QBS-MLP-0.085 error goal, 34 hidden layer nodes	
SENSITIVITY	64.0+/-9.8
SPECIFICITY	91.1+/-5.4
+ PRED. VALUE	84.2+/-9.3
OVERALL PERF.	81.4+/-3.1

Table 5. 14: 3-fold cross validation—(a)train multiple/test overall averages, (b)train/test overall averages, (c)train/test multiple averages, and (d)train multiple/test multiple + overall averages.

(a) 0.065 error goal, 32 hidden layer nodes			(b) 0.08 error goal, 20 hidden layer nodes	
SENSITIVITY	77.8+/-43.6%		SENSITIVITY	80.6+/-19.6%
SPECIFICITY	88.9+/-21.8%		SPECIFICITY	94.4+/-10.9%
+ PRED. VALUE	86.7+/-26.1%		+ PRED. VALUE	91.7+/-16.3%
OVERALL PERF.	85.2 +/-14.5%		OVERALL PERF.	89.7+/-11.3%
(c) 0.05 error goal, 27 hidden layer nodes			(d) 0.05 error goal, 28 hidden layer nodes	
SENSITIVITY	51.5+/-8.9%		SENSITIVITY	65.6+/-12.7%
SPECIFICITY	93.9+/-0.5%		SPECIFICITY	94.2+/-11.4%
+ PRED. VALUE	80.2+/-6.1%		+ PRED. VALUE	91.1+/-17.4%
OVERALL PERF.	79.1+/-7.2%		OVERALL PERF.	84.7+/-3.6%

Table 5. 15: 5-fold cross validation—(a)train multiple/test overall averages, (b)train/test overall averages, (c)train/test multiple averages, and (d)train multiple/test multiple + overall averages.

(a) 0.06 error goal, 34 hidden layer nodes			(b) 0.065 error goal, 21 hidden layer nodes, 0.06 error goal, 32 hidden layer nodes	
SENSITIVITY	80.0+/-24.0%		SENSITIVITY	80.0+/-24.0%
SPECIFICITY	95.0+/-9.8%		SPECIFICITY	86.7+/-16.0%
+ PRED. VALUE	93.3+/-13.1%		+ PRED. VALUE	83.3+/-20.7%
OVERALL PERF.	89.3+/-8.6%		OVERALL PERF.	84.7+/-14.6%
(c) 0.07 error goal, 20 hidden layer nodes			(d) 0.06 error goal, 32 hidden layer nodes	
SENSITIVITY	51.7+/-13.1%		SENSITIVITY	76.7+/-27.0%
SPECIFICITY	87.1+/-7.7%		SPECIFICITY	88.0+/-20.4%
+ PRED. VALUE	72.0+/-16.6%		+ PRED. VALUE	63.3+/-20.7%
OVERALL PERF.	74.8+/-5.4%		OVERALL PERF.	83.9+/-13.6%

Table 5. 16: 10-fold cross validation—(a) train multiple/test overall averages, (b)train/test overall averages, (c)train/test multiple averages, and (d)train multiple/test multiple + overall averages.

(a) 0.085 error goal, 20 hidden layer nodes			(b) 0.005 error goal, 29 hidden layer nodes, 0.065 error goal, 33 hidden layer nodes	
SENSITIVITY	70.0+/-29.9%		SENSITIVITY	90.0+/-19.6%
SPECIFICITY	90.0+/-13.1%		SPECIFICITY	90.0+/-13.1%
+ PRED. VALUE	65.0+/-29.4%		+ PRED. VALUE	80.0+/-21.7%
OVERALL PERF.	83.3+/-14.6%		OVERALL PERF.	90.0+/-10.0%
(c) 0.09 error goal, 29 hidden layer nodes			(d) 0.06 error goal, 25 hidden layer nodes	
SENSITIVITY	63.3+/-22.7%		SENSITIVITY	65.0+/-22.2%
SPECIFICITY	87.5+/-14.5%		SPECIFICITY	90.4+/-7.6%
+ PRED. VALUE	71.7+/-21.3%		+ PRED. VALUE	79.7+/-16.2%
OVERALL PERF.	79.0+/-10.3%		OVERALL PERF.	81.3+/-9.6%

The best performance for both 3-fold and 10-fold cross validation was achieved with the overall averages, however, with the 5-fold cross validation the best performance is given through training with the multiple averages and testing with the overall averages. For all test cases it is observed that the training and testing with the multiple averages yields a poorer performance for all levels of cross validation when compared with the other test scenarios.

Overall, in this case as with the Daubechies wavelet, the performance values are similar to the performance of a clinical evaluation and the confidence intervals in most cases include the range of the clinical evaluation. This is significant because it suggests that similar performances can in fact be obtained without requiring the expensive university hospital setting clinical evaluation.

5.4.2b LEARN++ RESULTS FOR QUADRATIC B-SPLINE

Results for Learn++ are shown in Table 5. 17 to Table 5. 19. For training/testing with half the data, Learn++ performs slightly lower than the MLP, however, when the

Table 5. 17: Learn++- overall averages, train/test with half the data.

QBS-LEARN++-0.1 ERROR GOAL, 25 HIDDEN LAYER NODES, 15 CLASSIFIERS	
SENSITIVITY	60.0+/-10.1
SPECIFICITY	91.1+/-4.4
+ PRED. VALUE	81.0+/-8.7
OVERALL PERF.	80.0+/-3.5

Table 5. 18: Learn++ results from training/testing with the multiple averages.

LEARN ++ RESULTS WITH MULTIPLE AVERAGES						
# CLAS-SIFIERS	ERROR GOAL	# HIDDEN LAYER NODES	OVERALL PERF.	+ PRED VALUE	SPECIFICITY	SENSITIVITY
20	0.1	15	83.3+/-8.3%	87.1+/-17.2%	90.9+/-13.8%	70.0-+/-11.1%
15	0.15	20	76.7+/-9.1%	61. 7+/- 32.5%	92.2+/-7.0%	48.3+/-25.5%
10	0.1	25	72.1+/-14.5%	66.0+/-21.1%	84.8+/-11.9%	50.0+/-27.3%
5	0.1	20	71.1+/-6.6%	49.3+/-26.6%	89.9+/-5.7%	35.0+/-22.2%
3	0.15	20	70.4+/-13.4%	45.6+/-25.0%	83.1+/-10.1%	50.0+/-35.8%

Table 5. 19: Learn++ results from training/testing with the overall averages.

LEARN ++ RESULTS WITH OVERALL AVERAGES						
# CLAS-SIFIERS	ERROR GOAL	# HIDDEN LAYER NODES	OVERALL PERF.	+ PRED VALUE	SPECIFICITY	SENSITIVITY
20	0.1	30	78. 7+/-6.1%	73. 3+/-38.1%	95.0+/- 9.8%	50.0+/-31.0%
15	0.15	30	85.3+/-7.3%	100.0+/-0.0%	100.0+/-0.0%	60.0+/-19.6%
10	0.1	30	82.0+/-10. 5%	83. 3+/-20.7%	88. 3+/-14.2%	70.0+/-24.0%
5	0.1	25	75.3+/-7.0%	70.0+/-39.2%	95.0+/-9.8%	40.0+/-19.6%
3	0.1	25	78.0+/- 8.9%	66.7+/-35.8%	90.0+/-12.0%	60.0+/-36.7%

confidence intervals are considered, the differences are negligible. Overall the results are comparable to those of the MLP and a clinical evaluation.

5.4.2c LEAVE-ONE-OUT RESULTS FOR QUADRATIC B-SPLINE

The discussion in Section 5.5.1c for the leave-one-out algorithm applies to these results as well. The comparison between Learn++ and an MLP for the quadratic B-spine wavelet show that the MLP performs slightly better in the limiting case. Since the other levels of cross validation for this wavelet performed better than the MLP while this case

Table 5. 20: Leave-one-out results-MLP, 0.05 error goal, 25 hidden layer nodes.

	AVERAGE	CI
OVERALL PERF.	75	16.3333
+ PRED. VALUE	25	16.3333
SENSITIVITY	25	16.3333
SPECIFICITY	50	18.8601

Table 5. 21: Leave-one-out results-Learn++, 0.05 error goal, 25 hidden layer nodes, 3 classifiers.

	AVERAGE	CI
OVERALL PERF.	71.4286	17.0403
+ PRED. VALUE	14.2857	13.1993
SENSITIVITY	14.2857	13.1993
SPECIFICITY	57.1429	18.6667

performed worse, it can be said that these results indicate that Learn++ performs as well as a single strong classifier.

5.6 RESULTS USING DATA FUSION

Results for data fusion have been obtained using the two sets of wavelet coefficients as features. Results are presented in Table 5. 22 and Table 5. 23 for the overall and for the multiple averages respectively. Five-fold cross validation was performed for comparison with previous results using the Learn++ algorithm. The number of classifiers was kept the same for each dataset to avoid any bias toward a particular set.

The performance for fusion with the overall averages is the same as the better of the two datasets. The performance for fusion with the multiple averages is slightly improved from each of the individual datasets but not significantly. The values for the fusion performance for the overall and for the multiple averages are similar to the results from each type of wavelet with Learn++. The results for this data fusion imply that no

additional information is gained by using both types of wavelets together since there is no significant improvement in the performance.

Table 5. 22: Data fusion results from five-fold cross validation using overall averages.

	Feature set 1 Quadratic B-spline coefficients (0.1error goal, 35 hidden layer nodes, 10 classifiers)	Feature set 2 Daubechies 4 coef-ficients (0.1error goal, 30 hidden layer nodes, 10 classifiers)	Data fusion of both sets
OVERALL PERFORM-ANCE AVERAGE	78.0+/-13.6	82.0+/-1.6	82.0+/-1.6
POSITIVE PREDICTIVE VALUE AVERAGE	73.3+/-22.2	93.3+/-13.1	93.3+/-13.1
SENSITIVITY AVERAGE	71.7+/-19.6	95.0+/-19.6	95.0+/-19.6
SPECIFICITY AVERAGE	90.0+/-26.1	60.0+/-9.8	60.0+/-9.8

Table 5. 23: Data fusion results from five-fold cross validation using multiple averages.

	Feature set 1 Quadratic B-spline coefficients (0.1error goal, 30 hidden layer nodes, 5 classifiers)	Feature set 2 Daubechies 4 coef-ficients (0.1error goal, 30 hidden layer nodes, 5 classifiers)	Data fusion of both sets
OVERALL PERFORM-ANCE AVERAGE	70.6+/-11.4	72.2+/-10.1	73.0+/-12.8
POSITIVE PREDICTIVE VALUE AVERAGE	51.4+/-31.7	70.0+/-24.0	55.0+/-36.0
SENSITIVITY AVERAGE	55.0+/-27.1	43.3+/-20.7	41.7+/-28.3
SPECIFICITY AVERAGE	81.2+/-18.4	89.0+/-8.0	91.5+/-9.0

CHAPTER 6

CONCLUSIONS

6.1 SUMMARY OF ACCOMPLISHMENTS

It has been shown that different types of wavelets are suitable for identifying features of an EEG signal that are characteristic of AD. The two types of wavelets used in this study have each performed as well as a clinical evaluation, in most cases, signifying that they are each capable of isolating the features of interest within the relevant frequency range. In comparing the two types of wavelets and their performance using the MLP with different training/testing schemes, it appears that the both wavelets perform better in classifying the overall averages. The overall results are similar for both types of wavelets and these performance values with their confidence interval include the clinical evaluation values. This suggests that similar performances to a clinical evaluation can be obtained through the automated algorithm using Learn++ or an MLP.

Learn++ had not previously been used for this classification problem and has demonstrated performances similar to that of a single strong classifier. The results from Learn++ were better than the results for the MLP in most cases for the Quadratic B-Spline wavelet however for classifying the Daubechies 4 wavelet coefficients, Learn++ performance was similar to that of a single strong classifier. The use of Learn++ in the problem allows for more analysis when additional data becomes available. The incremental learning feature of Learn++ could allow several additional classes to be intro-

duced into the problem such as the various degrees of severity of AD or classes including other forms of dementia.

It was expected that if the two types of wavelets were extracting different information from the signals, data fusion of this relevant information would provide a more informed classification decision, thereby improving upon previous methods. For the multiple averages with data fusion, the performance is slightly better than each of the individual cases but not by an amount significant enough to suggest the wavelets extract complimentary information from the signals. The results for both the overall and multiple averages indicate that the wavelets are not likely extracting complementary information from the signals since the combination of these two wavelets does not consistently produce a more informed decision than either individually.

The use of data fusion for combining features is a feasible approach to this classification problem. Data fusion should improve the classification performance if the features are providing different relevant information. In summary, the data fusion method using Learn++ provides a building block for future work in this area. The performance with two feature sets consisting of the wavelet coefficients did not improve upon the individual feature results however with the right features, this method has the potential to provide a more informed classification decision.

The use of data fusion, as well as the use of a single MLP or Learn++ as an automated program is a feasible approach for diagnosing AD in a community health clinic. The accuracy of each is similar to that of a clinical evaluation but is not limited to a research hospital setting.

6.2 SOURCES OF ERROR

The patients recruited for this study were diagnosed using a clinical evaluation. As mentioned previously this type of evaluation has an accuracy of only 75% however the classification algorithms were trained as if this were 100% (the 'gold standard'). The original misdiagnosis of a test subject is a potential source of error in this study. The only way to obtain a diagnosis with 100% accuracy is through an autopsy. Inclusion of a post mortem analysis of the test subjects to obtain the true 'gold standard' would prevent this error.

6.3 RECOMMENDATIONS FOR FUTURE WORK

The use of Learn++ allows for the introduction of new classes which could allow for a more involved study of these signals, perhaps introducing the degree of severity of AD indicated in the clinical evaluation. The patients recruited for this study were primarily in the earliest stages of the disease however there are variations in their clinical scores which could be partitioned to indicate severity.

The number of wavelet coefficients used for each type of wavelet was limited to the 16 Hz and below range to be consistent in comparisons between the two types of wavelets. There may be a more suitable number of coefficients for each wavelet and should be further investigated. Also, other types of wavelets should also be analyzed.

Additional electrodes should be explored as well as the use of the novel responses since these have both given interesting results in other studies. Investigation of the coherence across different electrodes has given promising results in previous studies discussed in Chapter 2 and may provide additional information relevant to this study.

Different features can also be combined with the data fusion classification method which may serve to increase the performance, provided the feature sets carry complementary information. Data fusion with the right features should yield a more informed decision for classification. Additional data would be useful for determining the true performance of the networks. For this project the additional data from the remaining patients is crucial for obtaining a better estimate of the true performance of any algorithm developed.

For any of the algorithms it may be a good idea to determine the cost of a misdiagnosis. Would it be worse for a person who is normal to be misdiagnosed as having AD or for a person with AD to be misdiagnosed as normal? In any of the algorithms, the weights of the networks could be modified so that a misdiagnosis would be more likely to occur for the least costly case.

REFERENCES

[1] Lim, A., et al. "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series ". *Journal of the American Geriatrics Society.* (1999) Vol. 47. Iss. 5. pp. 564-569.

[2] http://www.alz.org/

[3] Vellas, B. and Fitten, L.J. *Research and Practice in Alzheimer's Disease vol.3.* Springer Publishing Company. New York: 2000. pp. 27-32.

[4] Soukup, James E. *Alzheimer's Disease: A Guide to Diagnosis, Treatment, and Management.* Praeger Publishers. Connecticut: 1996. pp. 11-15.

[5] http://www.cnn.com/HEALTH/library/DS/00161.html

[6] Shenk, David. *The Forgetting, Alzheimer's: Protrait of an Epidemic.* Doubleday. New York: 2001.

[8] Polich, John and Kok, Albert. "Cognitive and biological determinants of P300: an integrative review". *Biological Psychology.* Volume 41. (1995) Pages 103-146.

[8] Lopes de Silva, Fernando *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* 4th edition. Lippincott, Williams and Wilkins, Philadelphia: 1999.

[10] Quiroga, Rodrigo Quian. 'Quantitative analysis of EEG signals: Time-frequency methods and Chaos theory'. *Thesis:*Institute of Physiology and Institute of Signal Processing - Medical University Lubeck. 1998.

[11] Karakas, Sirel; Erzengin, Omer Utku; and Basar, Erol. "The genesis of human event-related responses explained through the theory of oscillatory neural assemblies". *Neuroscience Letters.* Vol. 285. (2000) pp. 45-48.

[12] Basar, Erol, et al. "Gamma, alpha, delta, and theta oscillations govern cognitive processes". *International Journal of Psychophysiology.* vol. 39. (2001) pp. 241-248.

[13] Basar, Erol, et al. "Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG?". *Neuroscience Letters.* Vol. 259. (1999) pp. 165-168.

[14] Basar, Erol, et al. "Event-related oscillations are 'real brain responses'—wavelet analysis and new strategies". *International Journal of Psychophysiology.* Volume 39. (2001) Pages 91-127.

[15] Maltseva, Irina,; Geissler, Hans-Gearg and Basar, Erol. "Alpha oscillations as an indicator of dynamic memory operations—anticipation of omitted stimuli". *International Journal of Psychophysiology.* vol. 36. (2000) pp. 185-197.

[16] Schurmann, Martin, and Erol Basar. "Functional aspects if alpha oscillations in the EEG". *International Journal of Psychophysiology.* Vol. 39. (2001) pp.151-158.

[17] Basar-Eroglu, Canan, et al. "Gamma-band responses in the brain: a short review of physiological correlates and functional significance". *International Journal of Psychophysiology.* vol. 24. (1996) pp. 101-112.

[18] Karakas, S.and Basar, E. "Early Gamma response is sensory in origin: a conclusion based on cross-comparison of results from multiple experimental paradigms". *International Journal of Psychophysiology.* vol. 31. No. 1. (1998) pp. 13-31.

[19] Slobounov, S., et al. " Human oscillatory brain activity within gamma band (30-50Hz) induced by visual recognition of non-stable postures". *Cognitive Brain Research.* Vol. 9. (2000) pp. 177-192.

[20] Polikar, Robi, et al. "Multiresolution Wavelet Analysis of ERPs for the Detection of Alzheimer's Disease". *Proceedings-19th International Conference IEEE.* Oct. 30-Nov. 2, 1997.

[21] Polich, J. "Neurophysiology of P3a and P3b: A theoretical overview". From *Advances in Electrophysiology in Clinical Practice and Research.* Kjellberg, Inc.: Wheaton, IL. 2002.

[22] Yamaguchi, S., et al. "Event-related brain potentials in response to novel sounds in dementia". *Clinical Neurophysiology.* Volume 111. (2000) Pages 195-203.

[23] Polich, John, et al. "P300 Topography of Amplitude/Latency Correlations". *Brain Topography.* 1997. Vol. 9, Iss. 4:275-282.

[24] Golob, E. J. and Starr, A.. "Age-related qualitative differences in auditory cortical responses during short-term memory". *Clinical Neurophysiology.* Volume 111. (2000) Pages 2234-2244.

[25] Basar-Eroglu, Canan, et al. "Topological distribution of oddball 'P300'responses". *International Journal of Psychophysiology.* Volume 39. (2001) Pages 213-220.

[26] Katayama, Jun'ichi, and John Polich. "Auditory and visual P300 topography from a 3 stimulus paradigm". *Clinical Neurophysiology.* Vol. 110. (1999) pp.463-468.

[27] Polich, John and Herbst, Kathryn L. "P300 as a clinical assay: rationale, evaluation, and findings". *International Journal of Psychophysiology.* Volume 38. (2000) Pages 3-19.

[28] Penntila, M., et al. "Quantitative analysis of occipital EEG in different stages of Alzheimer's disease". *Electroencephalography and clinical Neurophysiology.* Vol. 60. Iss. 1. (1985) pp. 1-6.

[29] Babiloni, Claudio, et al. "Mapping distributed sources of cortical rhythms in mild Alzheimer's disease. A multicentric EEG study". *NeuroImage.* Article in Press (2004).

[30] Schreiter-Gasser, Ursula; Gasser, Theo; and Zeigler, Peter. "Quantitative EEG analysis in early onset Alzheimer's disease: correlations with severity, clinical characteristics, visual EEG and CCT". *Electroencephalography and clinical Neurophysiology.* Vol. 90. (1994) pp.267-272.

[31] Grunwald, M., et al. "Correlation Between Cortical θ Activity and Hippocampal Volumes in Health, Mild Cognitive Impairment, and Mild Dementia". *Journal of Clinical Neurophysiology.* Vol. 18. No.2. (2001) pp.178-184.

[32] Claus, Jules, J., et al. "Quantitative Spectral Electroencephalography in Predicting Survival in Patients With Early Alzheimer Disease". *Archives of Neurology.* Vol. 55. August 1998. pp.1105.

[33] Buchwald, J. S., et al. "Midlatency auditory evoked responses : differential abnormality of P1 in Alzheimer's disease". *Electroencephalography and clinical Neurophysiology.* Vol. 74. (1989) pp.378-384.

[34] Green, J.B., et al. " The middle latency auditory evoked potential may be abnormal in dementia". *Neurology.* Vol. 42. Iss. 5. (1992) pp. 1034-1036.

[35] Fein, George; Biggins, Christie, and Van Dyke, Craig. "The auditory P50 response is normal in Alzheimer's disease when measured via a paired click paradigm". *Electroencephalography and clinical Neurophysiology.* Vol. 92. (1994) pp.536-545.

[36] Polich, John and Lisa D. Hoffman. "Alzheimer's Disease and P300: evaluation of modality and task difficulty". *Brain Topography Today*. 1998: pp. 527-536.

[37] Frodl, Thomas, et al. "Value of event-related P300 subcomponents in the clinical diagnosis of mild cognitive impairment and Alzheimer's Disease". *Psychophysiology*. Volume 39. (2002) Pages 175-181.

[38] Polich, J. "EEG, ERPs, and Aging: Data and Measurement Issues". *Cognitive Changes Due to Aging and Fatigue as Revealed in the Electrical Brain Activity*. Germany: 1999. Pages 146-171.

[39] Pucci, E., et al. "EEG spectral analysis in Alzheimer's disease and different degenerative dementias". *Archives of Gerontology and Geriatrics*. Volume 26. (1998) Pages 283-297.

[40] Pucci, E., et al. "EEG power spectrum differences in early and late onset forms of Alzheimer's disease". *Clinical Neurophysiology*. Volume 110. (1999) Pages 621-631.

[41] Rodriguez, Guido, et al. "EEG spectral profile to stage Alzheimer's disease". *Clinical Neurophysiology*. Volume 110. (1999) Pages 1831-1837.

[42] Hibino, Shin, et al. "Fuzzy Neural Network Model for Assessment of Alzheimer-Type Dementia". *Journal of Chemical Engineering of Japan*. Vol. 34. No. 7. (2001) pp. 936-942.

[43] Besthorn, C., et al. "EEG coherence in Alzheimer's disease". *Electroencephalograchy and clinical Neurophysiology*. Vol. 90. (1994) pp.242-245.

[44] Locatelli, T., et al. "EEG coherence in Alzheimer's Disease". *Electroencephalography and clinical Neurophysiology*. Vol. 106. (1998) pp. 229-237.

[45] Jelic, V. "Quantitative electroencephalography in mild cognitive impairment: longitudinal changes and possible prediction of Alzheimer's disease". *Neurobiology of Aging*. Volume 21. (2000) Pages 533-540.

[46] Hogan, Michael J., et al. "Memory-related EEG power and coherence reductions in mild Alzheimer's disease". *International Journal of Psychophysiology*. Volume 49. (2003) pages 147-163.

[47] Brunovsky, Martin, et al. "Objective Assessment of the Degree of Dementia by Means of EEG". *Neuropsychobiology*. Vol. 48. (2003) pp.19-26.

[48] Musha, Toshimitsu, et al. "A new method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease". *Clinical Neurophysiology*. Volume 113. (2002) Pages 1052-1058.

[49] Jeong, Jaeseung, Gore, John C., and Peterson, Bradley S.. "Mutual information analysis of the EEG in patients with Alzheimer's disease". *Clinical Neurophysiology*. Volume 112. (2001) Pages 827-835.

[50] Petrosian, A.A., et al. "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG". *Clinical Neurophysiology*. Vol. 112 (2001) pp1378-1387.

[51]Demiralp, T., et al. "Time–Frequency Analysis of Single-Sweep Event-Related Potentials by Means of Fast Wavelet Transform". *Brain and Language*. Vol. 66. pp. 129–145. (1999).

[52] Demiralp, T., et al. "Detection of P300 Waves in Single Trials by the Wavelet Transform (WT)". *Brain and Language*. Vol. 66. (1999) pp108–128.

[53] Quiroga, R., et al. "Wavelet Transform in the analysis of the frequency composition of evoked potentials". *Brain Research Protocols*. Vol. 8. Iss. 1. August 2001. pp. 16-24.

[54] Bose, Tamal. *Digital Signal and Image Processing.* John Wiley and Sons, Inc. 2004.

[55] Mitra,Sanjit K. *Digital Signal Processing : A Computer-Based Approach.* McGraw Hill. Boston: 2001.

[56] Nunez, Paul L., et al. "EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales". *Electroencephalography and clinical Neurophysiology.* Vol. 103. (1997) pp.499-515.

[57] Ademoglu, Ahmet; Micheli-Tzanakou, Evangelia; and Istefanopulos,Yorgo. "Analysis of Pattern Reversal Visual Evoked Potentials (PRVEP's) by Spline Wavelets". *IEEE Transactions on Biomedical Engineering.* 1997: pp. 881-890.

[58] Chui, Charles K. Wavelet *Analysis and Its Applications Volume 1: An Introduction to Wavelets.* Academic Press, San Diego, 1992.

[59] Polikar, Robi. *Wavelet Tutorial* http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html

[60] wavelet class website-- http://galaxy.rowan.edu/~polikar/CLASSES/ECE554/page2.html

[61] Matlab wavelet toolbox-
http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/wavelet.html

[62] Rao, Raghuveer M., and Ajit S. Bopardikar. *Wavelet Transforms: Introduction to Theory and Applications.* Addison-Wesley. Reading: 1998.

[63]Quiroga, R. Quian, et al. "Wavelet Transform in the analysis of the frequency composition of evoked potentials". *Brain Research Protocols.* 2001. Vol. 1: 1-9.

[64] Jacques, G., et al, "Multiresolution wavelet analysis for early diagnosis of Alzheimer's Disease". *Proc. of 26th Int. Conf. of IEEE Eng. in Med. and Biology Soc.*, San Francisco, CA, 2004.

[65] Duda, Richard O., Hart. Peter E., and David G. Stork. *Pattern Classification.* John Wiley and Sons, Inc. New York: 2001.

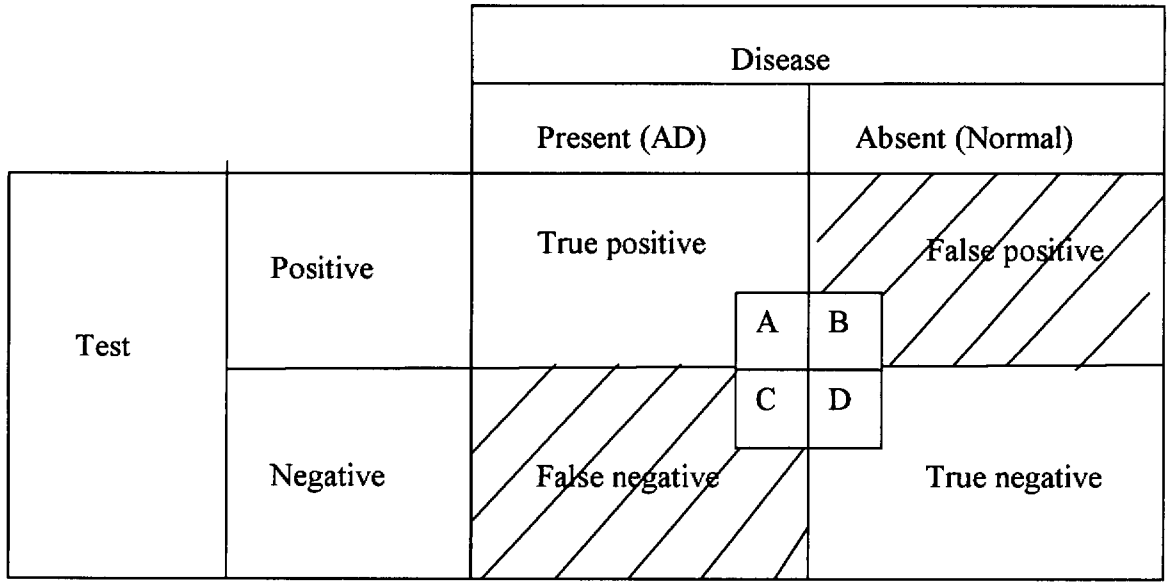
[66]PR website http://galaxy.rowan.edu/~polikar/CLASSES/ECE504_04/

[67] Polikar, Robi, et al. "Learn++: A Classifier Independent Incremental Learning Algorithm for Supervised Neural Networks". *Proc. Of Int. Joint Conference on Neural Networks (IJCNN2002),* vol. 2, pp. 1742-1747, Honolulu, HI, 12-17 May 2002.

[68] Lewitt, Mike, and Robi Polikar. "An Ensemble Approach for Data Fusion with Learn++". *4th International Workshop on Multiple Classifier Systems* (MCS 2003), Surrey, England, June 11-13, 2003

APPENDIX A

SENSITIVITY, SPECIFICITY AND POSITIVE PREDICTIVE VALUE



Sensitivity = A / A+C
-the probability that a symptom is present given that the person has the disease.

Specificity = D / B+D
-the probability that a symptom is not present given that the person does not have the disease.

Positive Predictive Value (PPV) = A/A+B
-the probability that a person has the disease given a positive test result

APPENDIX B

HEISENBERG UNCERTAINTY PRINCIPLE

The more precisely the position is determined, the less precisely the momentum is known in this instant, and vice versa.

--Heisenberg, uncertainty paper, 1927

An electron moving all by itself through empty space where the particle is described by certain measured properties of the particle itself. The position of the electron, its momentum (the electron's mass times its velocity), its energy, and the time are the properties that appear as "variables" in the equations that describe the electron's motion.

$$\Delta p \Delta q \geq \hbar / 4\pi \qquad \text{(B. 1)}$$

$$\Delta E \Delta t \geq \hbar / 4\pi \qquad \text{(B. 2)}$$

where Δq is the uncertainty of the position measurement, Δp is the uncertainty of the momentum in the q direction at the same time the q was measured, ΔE is the uncertainty in the energy, Δt is the uncertainty in the time at the same time as the energy is measured, and $\hbar$ is Planck's constant from quantum theory.

If the position of a moving electron is measured with great accuracy such that Δq is very small, then Δp is very large. The precision of the position measurement gets so great that the uncertainty Δq gets so small that it approaches zero, then Δp gets so large that it approaches infinity or becomes completely undefined.

If the position is known with certainty then this gives uncertainty to the exact momentum.

For wavelets, the window length defines the time and frequency resolutions. According to the Heisenberg's inequality there cannot be arbitrarily good time and frequency resolutions at a given time. There can be one or the other at any one time but never both at once.

APPENDIX C

GLOSSARY OF TERMS

Ergodic signal - a positive recurrent aperiodic state of a stochastic system

Orthonormal basis- A subset $v_1, v_2, ..., v_k$ of the vector space V, is said to be orthonormal if the inner product, $< v_i, v_j >$ when $i \neq j$., meaning that the vectors are mutually perpendicular.

Orthogonal basis - Set of vectors $\{x_j\}$ that satisfy $x_j x_k = C_{jk}\delta_{jk}$ and $x^\mu x_\nu = C_\nu^\mu \delta_\nu^\mu$, where C_{jk}, C_ν^μ are constants (not necessarily equal to 1), δ_{jk} is the Kronecker delta. When the constants are all equal to 1, then the set of vectors is called an orthonormal basis.

Biorthogonal basis - two different bases which are orthogonal to each other, but each do not form an orthogonal set.

Frame- a system that is more or less tight depending on how well the energy of the signal is maintained in the decomposition.

APPENDIX D

DIPOLE SOURCE ANALYSIS

In normal brains, distributions of neurons are approximately equally activated by active current generators. This type of activation produces: (1) uniform electric current density at the cortical surface; (2) currents within the cortical sulci (grooves) that cancel each other; and (3) uniform distribution of the scalp potential as a result. When damage exists, wither cortically or subcortically, randomly oriented electric current sources arise becasue the sulcal currents no longer cancel out and therefore the scalp potential becomes less uniform.

To determine the loss of uniformity of an observed alpha EEG potential distribution, various distributions of source current dipole models are simulated to find the best fitting model. The best model is given as the smallest mean squared difference between the alpha dipole potential of the model and the subject at each electrode, summed across all electrodes. This leads to the dipolarity, D, of the subject defined as:

$$D = \sqrt{1 - \left\langle \left(u_{obs} - u_{dip}\right)^2 \right\rangle_{av} \Big/ \left\langle u_{obs}^2 \right\rangle_{av}} \qquad \text{(D. 1)}$$

APPENDIX E

MUTUAL INFORMATION ANALYSIS

Mutual information quantifies information from one system through the measurement of another. When applied to EEGs, it can be considered a measure of functional connectivity. The information in bits is defined as:

$$\log_2 \frac{1}{P_X(x_i)} = -\log_2 P_X(x_i) \tag{E. 1}$$

where x_i is a measurement drawn from a set $X=\{x_i\}$, $P_X(x_i)$ is the normalized histogram of the distribution of values observed for the measurement x and $P_X(x_i)$ is the probability that an isolated measurement will find the system in the i^{th} element of the bin. The probabilities are evaluated by constructing a histogram with the variations of the measurement x_i.

The entropy, H, of the system is the amount of information obtained from any observation of X.

$$H(X) = -\sum_{x_i} P_X(x_i)\log_2 P_X(x_i) \tag{E. 2}$$

Before X is measured, this is information is called uncertainty. For $Y=y_j$, $H(X)$ must be replaced by the conditional uncertainty on X:

$$H\left(X \mid Y = y_j\right) = -\sum_{x_i} \frac{P_{XY}(x_i, y_j)}{P_Y(y_j)} \log_2 \frac{P_{XY}(x_i, y_j)}{P_Y(y_j)} \tag{E. 3}$$

where $P_{XY}(x_i,y_j)$ is the joint probability density for the measurements X and Y that produce the values X and Y. $H\left(X \mid Y = y_j\right)$ is indicative of the uncertainty in a

measurement of x, given that y has been measured and found to be equal to y_j. The mean conditional uncertainty on X over y_j given Y is known is:

$$H(X \mid Y) = \sum_{y_j} P_Y(y_j) H(X \mid Y = y_j) = -P_{XY}(x_i, y_j) \log_2 [P_{XY}(x_i, y_j) / P_y(y_j)] \quad \text{(E. 4)}$$

$$= H(X,Y) - H(Y) \quad \text{(E. 5}$$

where

$$H(X,Y) = -\sum_{x_i, y_j} P_{XY}(x_i, y_j) \log_2 [P_{XY}(x_i, y_j)] \quad \text{(E. 6)}$$

The a posteriori uncertainty on X, given a measurement of y is $H(X \mid Y)$, while the a priori uncertainty on X is H(X). The amount that a measurement of y reduces the uncertainty of x is

$$I_{XY} = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X,Y) \quad \text{(E. 7)}$$

which can also be expressed as:

$$I_{XY} = \sum_{x_i, y_j} P_{XY}(x_i, y_j) \log_2 \frac{P_{XY}(x_i, y_j)}{P_X(x_i) P_Y(y_j)} \quad \text{(E. 8)}$$

I_{XY} is the cross mutual information (CMI), the average mutual information between the measurements of X and measurements of Y.

The auto mutual information (AMI) is the mutual information between x_i and $x_{i+\tau}$, given as:

$$I_{XX_\tau} = \sum_{x(t), x(t+\tau)} P_{XX_\tau}(x(t), x(t+\tau)) \log_2 \frac{P_{XX_\tau}(x(t), x(t+\tau))}{P_X(x(t)) P_{x_\tau}(x(t+\tau))} \quad \text{(E. 9)}$$

$I_{X(t)Y(t+\tau)}$ is the time-delayed CMI or the mutual information of the EEG between different electrodes as a function of time delay.

APPENDIX F

ADDITIONAL RESOURCES FOR INFORMATION ON TESTS USED IN CLINICAL EVALUATION

For information on Clinical Dementia Rating (CDR), a useful chart can be found at: http://www.adrc.wustl.edu/cdrGrid.html .

For information on Mini Mental Scores (MMS), a copy of the actual exam can be found at: http://www.merck.com/mrkshared/mm_geriatrics/figures/38f1.jsp

For information on National Institute of Neurological and Communicative Disorders and Stroke Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria: http://www.uni-koeln.de/med-fak/neurologie/nest-dd/english/info/clinical_criteria.htm#Alzheimer

APPENDIX G

BACKGROUND OF QUADRATIC B-SPLINE WAVELETS

"Polynomial spline functions with equally spaced simple knots" are referred to as "cardinal splines". The set Z of all integers can be considered the "knot sequence" [56]. B-splines of the n^{th} order are a basis of the subspace of all continuous piecewise polynomial functions of degree n with derivatives up to $n-1$. $\phi^n(x)$ of this space can be defined as:

$$\phi^n(x) = \sum_{i=-\infty}^{\infty} c(i)\beta^n(x-i) \qquad \text{(G. 1)}$$

where $\beta^n(x)$ represents the normalized B-spline function of order n with $n+2$ equally spaced knots. $\beta^n(x)$ is defined as:

$$\beta^n(x) = \beta^{(n-1)}(x) * \beta^0(x) = \beta^0(x) * \beta^0(x) * \ldots * \beta^0(x), \quad n+1 \text{ times} \qquad \text{(G. 2)}$$

The function $\phi^n(x)$ can be determined by its B-spline coefficients, $c(i)$. For Cardinal Spline Interpolation, the coefficients should be determined such that $\phi^n(x)$ matches the values of some discrete sequence, $f(k)$ at the knot points: $\phi^n(x)=f(k)$ for $\{k=-\infty \ldots +\infty\}$ [50,55].

The B-spline wavelet transform is utilized for creating a sequence of embedded polynomial spline function spaces $\{S_{(i)}^n i \in Z\}$ of order n. This is performed such that $S_{(i)}^n \supset S_{(i+1)}^n$ *for* $i \in Z$, where Z is a set of all integers. $S_{(i)}^n$ is a subset of functions in $L_2(\Re)$ that are continuous with continuous derivatives up to the order of

n-1 and are equal to a polynomial of the n[th] degree on intervals

$[k2^i, (k+1)2^i]$ *with* $k \in Z$ [55, 56]

Then

$$S_{(i)}^n = \phi_i^n(x) = \sum_{k=-\infty}^{\infty} c_{(i)}(k)\beta_{2^i}^n(x - 2^i k) \qquad \text{(G. 3)}$$

where

$$\beta_{2^i}^n(x) = \beta^n\left(\frac{x}{2^i}\right) \qquad \text{(G. 4)}$$

$\beta^n(x)$ is the scaling function for a B-spline of the n^{th} order. The wavelet sequence q^n is defined as

$$q^n(k-1) = (-1)^k b^{2n+1}(k) * p^n(k) \qquad \text{(G. 5)}$$

where $b^{2n+1}(k) = \beta^{2n+1}(k)$ and $p^n(k)$ is the binomial kernel given by

$$p^n(k) = \frac{1}{2^n}\binom{n+1}{k}, \quad k = 0,\ldots,n+1 \qquad \text{(G. 6)}$$

The wavelet function, $\boldsymbol{\beta}_2^n(x)$, can be calculated once the wavelet and scaling sequences are determined using the dilation equation:

$$\boldsymbol{\beta}_2^n(x) = \boldsymbol{\beta}^n\left(\frac{x}{2^i}\right) = \sum_{k=-\infty}^{\infty} q^n(k)\beta^n(x-k) \qquad \text{(G. 7)}$$

The finest resolution level for $\phi^n(x)$ is given by

$$\phi_{(0)}(x) = \sum_{k=-\infty}^{\infty} c_{(0)}(k)\beta^n(x-k) \qquad \text{(G. 8)}$$

The B-spline wavelet is a polynomial with compact support satisfying the property: $\boldsymbol{\beta}_2^n \perp \boldsymbol{\beta}_2^n(x-2k)k \in Z$, such that for Equation (E.9), the first term on the right is the projection of $\phi_{(0)}^n$ on $S_{(1)}^n$ and the second term denotes the residual error.

$$\phi_{(0)}^{n}(x) = \sum_{k=-\infty}^{\infty} d_{(1)}(k)\beta_2^n(x-2k) + \sum_{k=-\infty}^{\infty} c_{(1)}(k)\beta_2^n(x-2k) \qquad \text{(G. 9)}$$

where $d_{(1)}$ are the details at level 1 and $c_{(1)}$ is the approximation of the signal at level 1. The decomposition can be implemented iteratively up to level I using the wavelet representation

$$\phi_{(0)}^{n}(x) = \sum_{i-1}^{\infty}\sum_{k=-\infty}^{\infty} d_{(i)}(k)\beta_{2^i}^{n}(x-2^i k) + \sum_{k=-\infty}^{\infty} c_{(I)}(k)\beta_{2^I}^{n}(x-2^I k) \qquad \text{(G. 10)}$$

where

$$\beta_{2^i}^{n}(x) = \beta_2^n\left(\frac{x}{2^{i-1}}\right) \qquad \text{(G. 11)}$$

These coefficients $\{d_{(i)}, \ldots, d_{(I)}\}$ are the detail coefficients and the sequence $\{c_{(I)}\}$ and the approximation coefficients at level I [55]. The lowpass and highpass filter kernels for the quadratic spline wavelets are:

$$h(k) = \frac{1}{2}[b^5]^{-1}(k) \uparrow 2 * b^5(k) * p^2(k) \qquad \text{(G. 12)}$$

$$g(k+1) = \frac{1}{2}[b^5]^{-1}(k) \uparrow 2 * (-1)^k p^2(k) \qquad \text{(G. 13)}$$

where

$$[(b^5)^{-1}](k) = Z^{-1}\left\{\frac{120}{z^2 + 26z + 66 + 26z^{-1} + z^{-2}}\right\} \qquad \text{(G. 14)}$$

and

$$p^2(k) = \frac{1}{2^2} Z^{-1}[1 + 3z^{-1} + 3z^{-2} + z^{-3}] \qquad \text{(G. 15)}$$

factoring $[b^5(k)]^{-1}$ gives

$$[(b^5)^{-1}](k) = \frac{\alpha_1\alpha_2}{(1-\alpha_1^2)(1-\alpha_2^2)(1-\alpha_1\alpha_2)(\alpha_1-\alpha_2)}\left[\alpha_1(1-\alpha_2^2)\alpha_1^{|k|} - \alpha_2(1-\alpha_1^2)\alpha_2^{|k|}\right] \tag{G.16}$$

where α_1=*-0.04309* and α_1=*-0.43057.*

The discrete B-splines are obtained by sampling the corresponding continuous functions $b^n(k) = \beta^n(k)$ with the starting condition:

$$b^0(k) = \begin{cases} 1 & for\ 0 \le k \le 1 \\ 0 & otherwise \end{cases} \tag{G. 17}$$

The discrete case has a similar convolution relation to the continuous case:

$$b(k) = b^0(k) * b^0(k) * \ldots * b^0(k) * b^n(k), \quad n+1\ times \tag{G. 18}$$

The interpolating function $\phi^n(x)$ of the form:

$$f(k) = \phi^n(k) = \sum_{i=-\infty}^{\infty} c(i) b^n(k-i) \tag{G. 19}$$

and can also be described by the convolution

$$f(k) = b^n(k) * c(k) \tag{G. 20}$$

$b^n(k)$ is the FIR operator and referred to as the indirect spline filter of order n. The additional convolution in (E.20) by $\{b^n(k)\}$ guarantees that the discrete B-spline provides the same values as the continuous basis functions at the node points *[k = -* $\infty \ldots +\infty$*].* The Z-transform of (E.20) gives

$$F(z) = B^n(z)C(z) \tag{G. 21}$$

where

$$B^n(z) = \sum_{k=-\infty}^{\infty} b^n(k) z^{-k} \tag{G. 22}$$

The spline coefficients, *c(k)*, are determined by inverse filtering:

$$S^n(z) = [B^n(z)]^{-1} \tag{G. 23}$$

A fast algorithm may be used to calculate the B-spline filter coefficients, $c(k)$ for the quadratic case (where $n = 2$) using:

$$c^+(k) = f(k) + b_1 c^+(k-1) \quad (k = 2,\ldots,K) \tag{G. 24}$$

$$c^-(k) = f(k) + b_1 c^-(k+1) \quad (k = K-1,\ldots,1) \tag{G. 25}$$

$$c(k) = b_0 (c^+(k) + c^-(k) - f(k)) \tag{G. 26}$$

where

$$b_0 = -8\alpha / (1-\alpha^2) \tag{G. 27}$$

$$b_1 = \alpha = \sqrt{8} - 3 \tag{G. 28}$$

The wavelet coefficients are then calculated from this for I levels by filtering and decimating by 2:

$$c_{(i+1)}(k) = [h(k) * c_{(i)}(k)] \downarrow 2 \tag{G. 29}$$

$$d_{(i+1)}(k) = [g(k) * c_{(i)}(k)] \downarrow 2 \quad \textit{for } i = 0,1,2,\ldots,I-1 \tag{G. 30}$$

where h and g are the lowpass and highpass filter kernels for decomposition [50, 55]. The z-transform of the sampled quadratic B-spline is

$$B^2(z) = \frac{z + 6 + z^{-1}}{8} \tag{G. 31}$$

The transfer function of the inverse (stable symmetric IIR) filter is given as:

$$S^2(z) = \frac{-8\alpha}{1-\alpha^2}\left(\frac{1}{1-\alpha z^{-1}} + \frac{1}{1-\alpha z} - 1\right) \tag{G. 32}$$

where $\alpha = \sqrt{8} - 3$. The impulse response of the filter is:

$$s^2(k) = \frac{8\alpha}{1-\alpha^2}\alpha^{|k|} \tag{G. 33}$$

[55].