Rowan University

Rowan Digital Works

Theses and Dissertations

12-11-2009

Decision-based data fusion of complementary features for the early diagnosis of Alzheimer's disease

Metin Ahiskali

Follow this and additional works at: https://rdw.rowan.edu/etd

 Part of the Electrical and Computer Engineering Commons

Recommended Citation

Ahiskali, Metin, "Decision-based data fusion of complementary features for the early diagnosis of Alzheimer's disease" (2009). *Theses and Dissertations*. 21.
https://rdw.rowan.edu/etd/21

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact graduateresearch@rowan.edu.

www.manaraa.com

DECISION-BASED DATA FUSION OF COMPLEMENTARY FEATURES FOR THE EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE

by
Metin B. Ahiskali

A Thesis

Submitted in partial fulfillment of the requirements of the
Master of Science in Engineering Degree
of
The Graduate School
at
Rowan University
December 11, 2009

Thesis Chair: Robi Polikar, Ph.D.

© 2009 Metin B. Ahiskali

Dedicated to:

... my mom and dad, your support and guidance
through the years has brought me to where I am today.

and to Courtney, without you I never would have had the motivation to
finish. Thank you for keeping me grounded and for all the love
and support you have given me through everything.

ABSTRACT

Metin B. Ahiskali
DECISION BASED DATA FUSION OF COMPLEMENTARY FEATURES
FOR THE EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE
2008/09
Robi Polikar, Ph.D.
Master of Science in Engineering

As the average life expectancy increases, particularly in developing countries, the prevalence of Alzheimer's disease (AD), which is the most common form of dementia worldwide, has increased dramatically. As there is no cure to stop or reverse the effects of AD, the early diagnosis and detection is of utmost concern. Recent pharmacological advances have shown the ability to slow the progression of AD; however, the efficacy of these treatments is dependent on the ability to detect the disease at the earliest stage possible. Many patients are limited to small community clinics, by geographic and/or financial constraints. Making diagnosis possible at these clinics through an accurate, inexpensive, and noninvasive tool is of great interest. Many tools have been shown to be effective at the early diagnosis of AD. Three in particular are focused upon in this study: event-related potentials (ERPs) in electroencephalogram (EEG) recordings, magnetic resonance imaging (MRI), as well as positron emission tomography (PET). These biomarkers have been shown to contain diagnostically useful information regarding the development of AD in an individual. The combination of these biomarkers, if they provide complementary information, can boost overall diagnostic accuracy of an automated system.

EEG data acquired from an auditory oddball paradigm, along with volumetric T2 weighted MRI data and PET imagery representative of metabolic glucose activity in the brain was collected from a cohort of 447 patients, along with other biomarkers and metrics relating to neurodegenerative disease. This study in particular focuses on AD versus control diagnostic ability from the cohort, in addition to AD severity analysis. An assortment of feature extraction methods were employed to extract diagnostically relevant information from raw data. EEG signals were decomposed into frequency bands of interest through the discrete wavelet transform (DWT). MRI images were preprocessed to provide volumetric representations of specific regions of interest in the cranium. The PET imagery was segmented into regions of interest representing glucose metabolic rates within the brain. Multi-layer perceptron neural networks were used as the base classifier for the augmented stacked generalization algorithm, creating three overall biomarker experts for AD diagnosis. The features extracted from each biomarker were used to train classifiers on various subsets of the cohort data; the decisions from these classifiers were then combined to achieve decision-based data fusion.

This study found that EEG, MRI and PET data each hold complementary information for the diagnosis of AD. The use of all three in tandem provides greater diagnostic accuracy than using any single biomarker alone. The highest accuracy obtained through the EEG expert was 86.1±3.2%, with MRI and PET reaching 91.1±3.2% and 91.2±3.9%, respectively. The maximum diagnostic accuracy of these systems averaged 95.0±3.1% when all three biomarkers were combined through the decision fusion algorithm described in this study. The severity analysis for AD showed similar results, with combination performance exceeding that of any biomarker expert alone.

ACKNOWLEDGEMENTS

To begin with, I would like to thank Tejash Patel for his initial work on MRI and EEG data fusion techniques with this lab, which ultimately led to this study. His guidance and assistance in the initial stages of my research were paramount, and helped set the foundation for the findings in this study. I would like to thank Michael Russell for his help in automating a great deal of the EEG data analysis, specifically the directory and file parsing techniques we hashed out in MATLAB together. Additionally, I owe William Weaver a debt of gratitude for his work and assistance in the actual data fusion process which involved the automated selecting, parsing, and processing of the hundreds of subjects necessary to create the various datasets used for analysis in this study. I would like to thank Ryan Elwell, Greg Ditzler, and James Ethridge for their help in brainstorming and discussion of all the minutiae involved in pattern recognition algorithms and MATLAB coding/syntax nuances.

I wish to extend my gratitude to the entire Rowan University engineering faculty for providing me with an incredible and rewarding undergraduate (and graduate) experience that has helped shape me into the engineer I am today. Specifically, I would like to thank Dr. Robi Polikar for all his guidance and granting me the opportunity to work alongside him on this enriching and rewarding research. Furthermore, I would like to thank thesis committee members Dr. Maria Tahamont and Dr. Shreekanth Mandayam for taking time out of their busy schedules to be an integral part of this thesis, as well as for their guidance throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Modern medicine has increased the average life expectancy of individuals, particularly those in developing countries. In the United States alone, the average life expectancy of an individual has risen from 47.3 years to 77.8 years over the last century [1]. Society has arguably benefitted from this increase in life expectancy—more individuals are living healthier and longer lives. However, there are other medical consequences to such advances. As the average life expectancy increases, the prevalence of neurodegenerative diseases also increases. Within the most populous generation in American history (commonly referred to as the "baby boomers" and consisting of over 78 million individuals), the frequency of elderly-targeted diseases is starting to increase dramatically. Every minute, approximately 7 people from this generation are turning 50, with almost 330 turning 60 years of age every hour [2,3]. As the average life expectancy increases, research must focus on the prevention, detection, and curing of late-life neurodegenerative diseases.

The number of individuals effected by Alzheimer's disease (AD) has been steadily rising each year. In the United States, over 5.3 million people are diagnosed with AD, with a new case diagnosed every 70 seconds [4]. AD constitutes the 6^{th} leading cause of death and also was the only cause that saw an increase in cases every year, from 2000 to 2006, the change in cause of death due to AD jumped 47.1%. In comparison, the top four

causes of death, heart disease, breast cancer, prostate cancer, and stroke on average dropped 11.13% over the same time period. This trend is particularly alarming, as there is currently no treatment to stop or reverse the effects of AD. It is predicted that by the year 2050 if the current trends continue, that over 15 million Americans will be diagnosed with AD [4-6].

1.1 ALZHEIMER'S DISEASE RESEARCH

Worldwide, there are over 26 million people diagnosed with Alzheimer's disease. This number is expected to grow to more than 106 million individuals by 2050. As the most common form of dementia, AD generally begins after the age of 60, with the risk of developing the disease increasing every year thereafter. Over 50% of individuals over 85 years of age suffer from AD, and this age group constitutes the fastest growing segment of the American population. Only 5% of all AD cases are considered to be hereditary; however, in these cases early-onset AD is more common, and can effect individuals as young as 30 years [5]. The affliction rate of AD increases dramatically past the age of 65. Of those between 65 and 74, 2% are affected by AD; between 75 and 84 the number increases to almost 20%, and over 50% of those above the age of 85 are thought to have AD [4,5]. With the prevalence of this disease increasing rapidly, AD could soon become a leading cause of death in the US as well as worldwide in the coming decades.

Aside from the emotional and physical burden imposed by AD on both patients and their loved ones, the disease creates a tremendous economic burden. In 2005, the global cost of all dementia related care and treatment was estimated to be $315.4 billion. In the United States, AD ranks as the third most expensive disease, with the cost of care at approximately $180,000 on average from diagnosis to death The estimated annual cost

for AD in the US by 2050 is projected to rise exponentially, with over 15 million individuals diagnosed with this disease [7].

1.1.1 DIAGNOSIS AND BIOMARKERS

Such startling statistics indicate a pressing need for the early diagnosis and treatment of AD. One of the major hindrances behind the successful treatment of Alzheimer's disease is the lack of clinical testing to accurately diagnose the disease. If screened by expert clinicians specialized in memory disorders and neurodegenerative diseases, individuals can expect an 80-90% diagnostic accuracy [8]. The clinicians in these cases make their diagnosis based on a multitude of factors; these elements include interviews with patients and caregivers, medical history, clinical observation, and memory tests. This type of longitudinal study occurs over the course of months to years—the combination of these factors is what helps the neurologist make a final determination on the subject's condition [8,9].

A significant issue with the diagnosis of AD is the ability to rule out other contributing diseases/conditions, such as Mild Cognitive Impairment which is hard to differentiate from AD at an early stage. While tests exist to rule out other such diseases, many patients essentially have "waited too long" for a diagnosis. By the time family and caregivers begin to notice the symptoms of the disease, it is generally too late to utilize any currently available treatments effectively [10]. Many AD patients do not have access, either financially or geographically, to expert neurologists; furthermore, general health practitioners in community clinics cannot match the diagnostic accuracy[1] of expert

[1] Diagnostic accuracy refers to the ability to accurately diagnose a given disease properly (correct vs. incorrect), also measuring metrics regarding true/false positives (see section 3.9 for detailed explanation).

neurologists, generally falling to a diagnostic accuracy of 75% with a sensitivity of 83%, and a specificity of 55% [8]. Currently, the only definitive diagnosis for AD can only be determined post-mortem. In this case, specific protein markers for AD are identified from a biopsy of brain tissue under a microscope. To biopsy a brain in a living patient is extremely invasive and high-risk, and is rarely, if ever used. Brain tissue biopsy is reserved for a definitive diagnosis through autopsy [6].

For this study, metabolic, anatomical, and physiological biomarkers for AD were analyzed for their effectiveness in the early diagnosis of the disease. Each of these three biomarkers was obtained through clinical study and evaluation of subjects with the hope that each biomarker contained relevant and pertinent information about the patient's condition than a written clinical evaluation based on medical history and various memory and cognitive tests.

The metabolic marker examined for this study relates to the blood glucose levels in the brain, obtained through positron emission tomography (PET). These levels have been shown to contain information with regards to AD diagnosis, and can be used to aid in an automated system for subject classification [11].

Magnetic resonance imaging (MRI) of the brain is used to study the anatomical biomarker for AD. Specifically, this method is used to quantify the shrinkage of gray matter in the brain, generally attributed to neuronal death caused by AD [12,13].

Finally, the use of electroencephalography (EEG) to record event-related potentials in subjects was explored to determine the effectiveness of the electrophysiological biomarker for AD as a means of early diagnosis [14]. The detection

and measurement of event-related potentials (ERPs) in EEG data was used to determine the integrity of neuronal systems in subjects relating to AD diagnosis [9,15]. Due to the fundamental differences in their modalities, metabolic, anatomical and electrophysiological, PET, MRI, and EEG provide complementary information.

1.1.2 ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

A project worth noting is the 'Alzheimer's Disease Neuroimaging Initiative' (ADNI). Principal investigator of ADNI as well as the director of the Center for Imaging of Neurodegenerative Diseases, Dr. Michael W. Weiner, MD is a professor of Medicine, Radiology, Psychiatry, and Neurology at the University of California. The primary goal of ANDI is "to define the rate of progress of mild cognitive impairment and Alzheimer's disease, to develop improved methods for clinical trials in this area, and to provide a large database which will improve design of treatment trials" [16].

A five year research project which began October 2004, the ADNI project studies the rate of change of cognition, brain structure (atrophy), and biomarkers in a large cohort of patients. Two hundred elderly controls, 400 MCI, and 200 AD probable patients make up the total cohort—with all the data from this research study becoming available publicly. Funding for the project comes from multiple sources, totaling $60 million. The National Institute of Aging and National Institute of Bioimaging and Bioengineering contributed $40 million, with the remaining $20 from donations by multiple foundations as well as the pharmaceutical industry, making this initiative one of the highest funded studies for AD.

In addition to the raw cohort data, ADNI provides a wealth of information regarding recording protocols and procedures for both MRI and PET imaging. Focusing

specifically on MRI and PET imaging data, the ADNI study aimed to set uniform recording procedures for both techniques as well quantification methods for analysis of the subsequent data. The grant proposal, as well as program announcements, MRI/Pet methods for non-ADNI studies and overall ADNI protocol summaries can be found on the ADNI-info website [17].

The research and results obtained in the work for this thesis are similar and related directly to the ADNI study. Much of the work done on EEG, MRI and PET data followed many of the protocols and procedures outlined in ADNI. The work done in this thesis parallels the research conducted in ADNI.

1.2 OBJECTIVES OF THIS STUDY

This thesis is part of an ongoing study, and subsequently builds and improves upon previous work completed. The overall goal of this study and of the previous work was to create a clinically available, non-invasive, and automated system to aid in the early diagnosis of Alzheimer's disease by using a fusion of complementary data sources. For this specific study, three biomarkers are used for the development of an automated system for the early diagnosis of AD. Specifically, ERPs are acquired from EEG data along with volumetric MRI and normalized blood glucose levels from PET scans from cohorts of both control normal and AD probable patients.

For this study, only AD probable and control normal (CN) patients were examined. Mild cognitive impairment (MCI) and Parkinson's disease (with and without dementia) were not investigated; this study focused on the accuracy of the early diagnosis and detection of AD vs. CN. The first series of tests performed focused solely on the

diagnostic accuracy when only ERPs from EEG data were used. Several smaller cohorts were generated from the main overall EEG dataset of 210 patients, comprised of 114 AD and 96 CN subjects. The algorithm for these EEG tests was a modified approach to Augmented Stacked Generalization, implemented with wavelet processing of the EEG signals. Similarly, two more initial tests were run using only volumetric MRI data and blood glucose data from PET scans. A total of 136 subjects were available in the MRI-only cohort, with 79 AD and 57 CN used for this work. For the PET-only cohort, 80 total subjects were available, with 37 AD and 43 CN also utilized for the work described within this thesis.

For each biomarker, an individual ensemble of classifier systems was created to determine classification performance. Each of these ensembles was then used for a decision-based data fusion approach. Various combinations of complimentary data from EEG, MRI, and PET were used to generate final evaluations and performance metrics. Specifically, EEG+MRI, EEG+PET, MRI+PET, and EEG+MRI+PET fusion combinations were all evaluated on specific subject cohorts. Since many subjects in the overall cohort would not necessarily all have EEG, MRI, and PET data, each fusion combination had a varying number of subjects. Various combinations of these subjects generated a wide range of results with regards to overall ensemble system classification performance on diagnostic accuracy.

1.3 ORGANIZATION OF THESIS

Chapter 2 provides a detailed literature review, covering the biomarkers used in this study for AD as well as other neurodegenerative disease and other current diagnosis techniques used for the classification of these diseases. The methodologies employed in this study

are explained in Chapter 3. Specifically, all aspects of ERP data (acquisition, preprocessing, and feature extraction) and MRI/PET data acquisition and preprocessing are discussed. Biomarker specific ensemble systems for the classification of AD versus CN are also detailed, along with the decision-based data fusion approach for the various biomarker combination systems. Chapter 4 is presents all results from this current study, including individual and combined biomarker expert diagnostic accuracies. Severity analysis is detailed, showing diagnostic accuracies of individual and combined biomarker experts. Finally, Chapter 5 provides discussion with respect to sources of error, conclusions and suggestions for future work.

CHAPTER II

BACKGROUND

2.1 ALZHEIMER'S DISEASE

Currently, the primary method to diagnosis Alzheimer's disease (AD) concentrates on longitudinal studies, based on interviews of the patient and caregivers. However, since small community clinics rarely achieve an 80% diagnostic accuracy, with even expert neurologists seldom surpassing 90%, the use of reliable biomarkers to aid in the diagnosis of this disease has become paramount [8]. Generally, the biomarkers used for this purpose can be sorted into one of four categories: biochemical, anatomical, metabolic, and physiological. For the purposes of this study, the anatomical, metabolic, and physiological biomarkers are investigated for diagnostic accuracy analysis.

2.1.1 DIAGNOSIS OF AD

Alzheimer's disease is considered an extremely difficult disease to accurately diagnose at the early stage [6]. Generally, when an individual is suspected to have some form of neurodegenerative disease, the initial step is to administer a series of standardized neurological tests, to determine baseline information about the subject's condition. The simplest and most commonly used test is the Mini-Mental State Exam (MMSE), which attempts to assess cognitive ability by asking the subject to answer several simple questions and perform trivial tasks. The Clinical Dementia Rating (CDR) and Severe

Impairment Battery (SIB) are then used as further examinations on a patient's cognitive state and degree of dementia.

2.1.1a MINI-MENTAL STATE EXAM

The concept behind the MMSE is to test a subject's short-term memory, as well as assess their overall cognitive state [18,19]. The questions generally focus on short-term memory recognition, requiring the subject to determine the current time, date, and location among other things. The subject's overall cognitive ability is evaluated through tests such as counting backwards in steps of seven, spelling common words backwards, copying and drawing a diagram, or writing sentences.

The test is scored from zero to thirty, with zero representing a vegetative state and thirty indicating normal cognitive ability. It is important to take into account each subject's educational background, as varying levels of education can impact outcome of the test. As such, most MMSE scores are accompanied by a number relating to the subject's highest level of education attained. As shown in Table 1, there are many ways to interpret the raw score from the MMSE test. No single test can be relied upon for a comprehensive and accurate diagnosis of AD; it is common to have the MMSE performed along with several other cognitive assessments.

Table 1 – MMSE Score Interpretations [20]

Method	Score	Interpretation
Single Cutoff	**< 24**	Abnormal
Range	**< 21**	Increased odds of dementia
	> 25	Decreased odds of dementia
Education	**21**	Abnormal for 8th grade education
	< 23	Abnormal for high school education
	< 24	Abnormal for college education
Severity	**24-30**	No cognitive impairment
	18-23	Mild cognitive impairment
	0-17	Severe cognitive impairment

2.1.1b CLINICAL DEMENTIA RATING

The CDR is a 5-point scale that is used to focus upon six areas of cognitive and functional performance applicable to AD as well as other dementias: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. The information to determine each rating is collected through a structured interview process of the patient and a primary caregiver, generally a family member [21]. The CDR table used for this interview process gives descriptive anchors to help the clinician make the most accurate assessment possible. Since this rating is highly subjective, it is important to have some form of unity throughout the interview process. Additionally, a simple algorithm to determine the CDR of a subject can be used to calculate the appropriate value, based on the clinical evaluations [22]. The scale used for the CDR is shown in Table 2.

Table 2 – The CDR Scale and Meanings [21]

0	Normal	**0.5**	Very Mild Dementia	**2**	Moderate Dementia
		1	Mild Dementia	**3**	Severe Dementia

It is important to note that, in general, CDR scores are integer numbers, aside from the 'very mild dementia' category, and do not fall between classifications. For this purpose, the aforementioned algorithm chooses the appropriate CDR based on the input from all aspects of the clinicians evaluation table for the subject. Various statistical methods can be used to determine a patient's average CDR score. Algorithms average, then round to the closest score value for most CDR score analyses. In addition, the use of median and mode of a set of patient's respective CDR scores are commonly used to represent the overall 'average' score of a specific group of subjects.

2.1.1c SEVERE IMPAIRMENT BATTERY (SIB)

Unlike the MMSE and CDR, the Severe Impairment Battery can be used as a means of assessing a severely demented patient, where standard neuropsychological tests can no longer be applied. This test focuses solely on those patients who have an average MMSE lower than 11, falling into the severe cognitive impairment range.

Subgroups of patients in this range (MMSE of 0-5 and 6-11) showed significant differences in scores on the SIB, while subgroups from 11 and higher did not show any truly statistical difference in scoring [23]. Since this test is primarily focused on the assessment of severely demented patients, it was not used for the assessment of any subjects for this work. SIB scores are available for all subjects used in the cohorts developed for this study.

2.1.1d GLOBAL DETERIORATION SCALE

While there are many methods for the classification of the stages of neurodegenerative diseases, the Global Deterioration Scale (GDS) has been used extensively in clinical environments to give a stage-based representation of the progression of a disease such as AD in any patient.

Initially developed by Reisberg, *et al.* (1982), this scale is used to measure the progression of AD into seven distinct stages. While this scale can be used to diagnose other dementias, the symptoms within the scale can vary between the different diseases. For the purposes of this study, the stages in Table 3 are a generally accepted as the method to categorize AD.

Table 3 – The Global Deterioration Scale, AD specific [24,25]

	Stage	Typical Symptoms
Stage 1	No cognitive decline (normal function)	- No memory or cognitive problems
Stage 2	Very mild cognitive decline (may be normal age related changes or earliest signs of AD)	- Memory lapses - Forgetting familiar names and locations of objects - Lapses are not typically obvious to others
Stage 3	Mild cognitive decline (early stage AD can be diagnosed in some, but not all, individuals with these symptoms)	- Mild forgetfulness - Difficulty learning new things - Difficulty concentrating or limited attention span - Problem with orientation, such as getting lost - Communication difficulties, such as finding right word - Loss or misplacing of valuable objects - Difficulty handling problems at work - Issues are noticeable to family, friend, or co-workers
Stage 4	Moderate cognitive decline (mild or early stage AD)	- Some memory loss of one's personal history - Difficulty with complex tasks (e.g. managing finances, travel, shopping) - Decreased knowledge of current events - Impaired ability to perform challenging mental arithmetic (example: counting backwards from 75 by 7)
Stage 5	Moderately severe cognitive decline (moderate/mid-stage AD)	- Major gaps in memory (e.g. phone number or names of close family members) - Help needed for day to day tasks
Stage 6	Severe cognitive decline (moderately severe/mid-stage AD)	- Continued memory loss (e.g. occasionally forgetting the name of a spouse or primary caregiver) - Loss of awareness of recent events and experiences in their lives - Assistance is needed with activated of daily living (e.g. getting dressed or bathing) - Difficulties counting - Personality and emotional changed such as confusion, anxiety, suspiciousness, anger, sadness/depression, hostility, apprehension, delusions and agitation - Obsessions such as petition of simple activities - Disruption of normal sleep/waking cycle - Increasing episodes of incontinence
Stage 7	Very severe cognitive decline (severe or late-stage AD)	- Vocabulary becomes limited and verbal abilities eventually disappear - Loss of ability to walk independently and sit without support - Help is needed with eating and using the toilet (incontinence issues)

As Table 3 details, the GDS breaks down the progression of AD into seven, distinct stages. In many cases, stages one and two tend to go unnoticed by most, even by the patient, and any respective memory/cognitive declines are generally associated with old age. As with any of the aforementioned cognitive testing, no one metric appears better than another

Section 2.1.2 details the pathology of the disease, tracking the effect on the brain and other cognitive abilities throughout the progression of AD from diagnosis to death.

2.1.2 ALZHEIMER'S DISEASE PATHOLOGY

The human brain is regarded as one of the most complex and intriguing organs in the body. While composed essentially of the same structure as most other mammals, the human brain sets itself apart with an expanded cerebral cortex, most notably in the frontal lobes. Much of what makes us human comes from the expansion of these regions; self-control, planning, reasoning, and abstract thought all originate from these areas [26]. The physiological changed caused by AD target the complex network of the brain, specifically damaging and eventually destroying neurons. This damage caused by AD amounts to the complex communication network in the brain shutting down and misfiring, leading to the decline cognitive ability overtime. In particular, protein misfolding causes plaques and tangles, that have been shown as a root cause to the destruction of neurons, and are hallmark indicators of AD [27].

2.1.2a BETA-AMYLOID PROTEIN

Protein misfolding has been shown to be a key culprit in the progression and development of AD. While the cause for this phenomenon has not been discovered to date, protein misfolding has been linked to the primary cause of neuronal death, and

consequently the deterioration of cognitive ability. Two proteins have been linked to the development of AD; one of these proteins, beta-amyloid, originates from segmented fragments from a larger protein called amyloid precursor protein (APP). In a healthy brain, proteins such as the β-amyloid are normally metabolized and removed. While APP has been shown to be important in the growth and development of neurons, the β-amyloid fragment from APP begins to develop "plaques" in the brain, eventually leading to neuronal death [28].

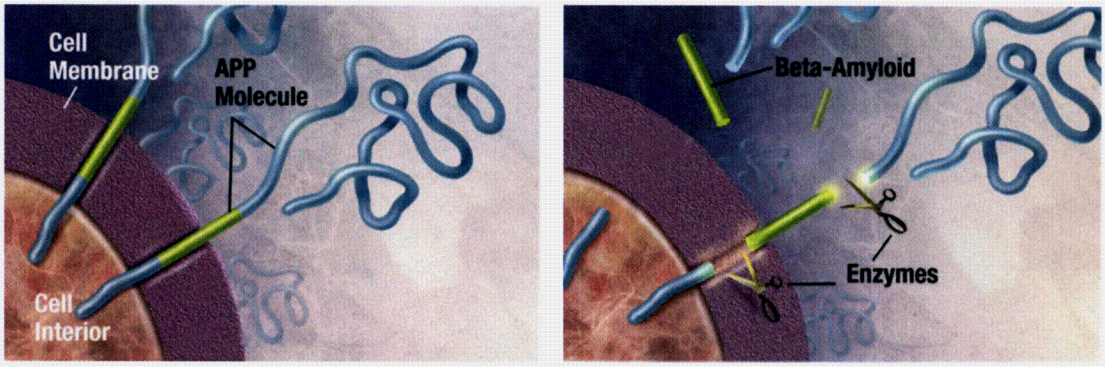

Figure 1 – From APP to Beta-Amyloid [29]

Enzymes break apart the APP molecule, creating β-amyloid fragments which later combine to form β-amyloid plaques that cling to the outer membrane of neurons.

As shown in Figure 1, a normal APP molecule is fragmented by enzymes, with these segments becoming β-amyloid proteins. These β-amyloid protein fragments then begin to "clump" with other molecules, forming insoluble amyloid plaques which begin to stick to the membranes of neurons. These plaques have a tendency to develop within the hippocampus, which is considered to play the largest role in short term memory [28].

Additionally, plaque build-up has been shown to occur in parts of the cerebral cortex, generally associated with decision making and higher-level cognitive abilities, such as reasoning and thinking. It is important to note that it is currently unknown whether these β-amyloid plaques cause AD, or if they are a byproduct of another

neurophysiological process in the development of the disease [27,28,30]. A rare, inherited form of AD, known as early-onset AD has been shown to develop due to alterations in the structure of APP. Large concentrations and accumulations of the insoluble β-amyloid plaques on the membranes of neurons can block the synaptic pathways, which eventually leads to neuronal cell death [28,31].

2.1.2b HYPERPHOSPHORYLATED TAU PROTEIN

The second protein linked to AD is actually part of the internal structure of a neuron. The hyperphosphorylated-τ protein is essential for the structural support and cohesion of a microtubule assembly in the neuron; additionally, these microtubules are crucial in guiding nutrients and other molecules towards far away axons [32]. The tau molecules stabilize these microtubules, and are routinely replaced over time as the tubules assemble. Figure 2 illustrates a healthy neuron along with a microtubule and the stabilizing τ molecules within it.

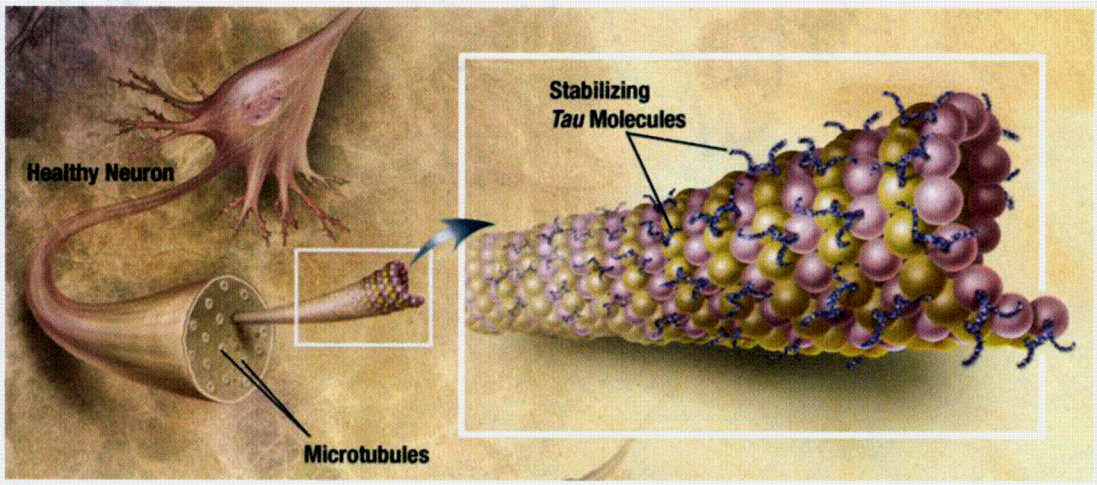

Figure 2 – A healthy neuron and the stabilizing tau molecules for microtubules [29]

In AD, the hyperphosphorylated-τ protein does not follow the conventional renewal cycle. Instead, these molecules begin to break off from the microtubules and

start to pair off with other threads of the τ proteins. Eventually, enough of these molecules clump together, forming neurofibrillary tangles [31,32]. The loss of the τ protein causes the microtubules to disintegrate, losing structural integrity as they fall apart. With the breakdown of the microtubules, vital nutrients and molecules are prevented from being transported. The inability to transport material within the neurons ultimately results in cell destruction [27,28,30]. Figure 3 illustrates a diseased neuron, and the breakdown of a microtubule caused by the clumping of τ proteins.

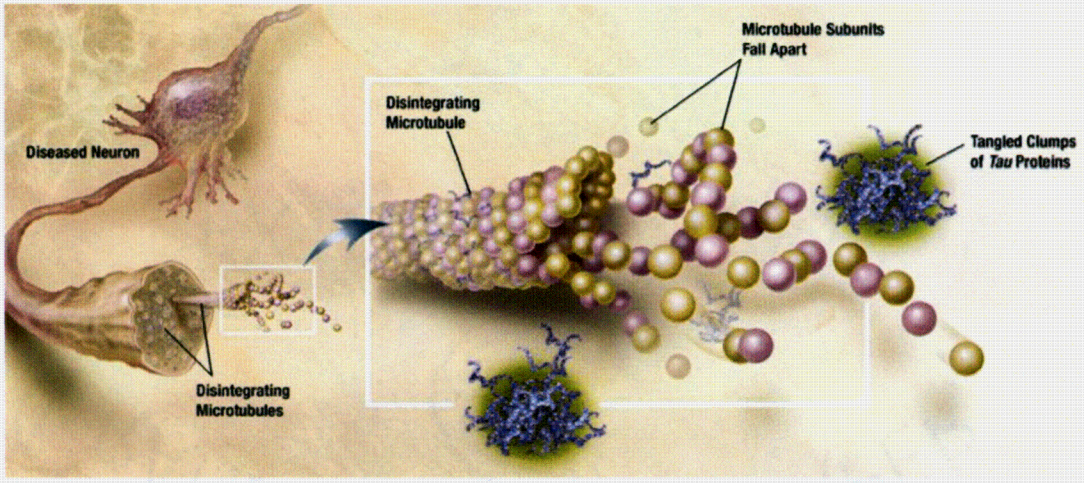

Figure 3 – The tau protein in the disintegration of neuronal microtubules [29]

An unknown mechanism causes hyperphosphorylated-τ to breakdown improperly, leading to the disintegration of microtubules in the neuron. These tau proteins then begin to clump, forming neurofibrillary tangles.

2.1.2c ATROPHY OF THE BRAIN

Neuronal death occurs due to the amyloid plaques and neurofibrillary tangles, caused by the buildup of the β-amyloid and hyperphosphorylated-τ proteins, respectively. As the microtubules break apart, material can no longer be transported within neurons, ultimately leading to the destruction of the cell. With β-amyloid fragments, the resulting plaques begin to stick to the neuron's outer membranes, eventually causing damage and

disintegration of the cell itself [28]. The death of these cells eventually leads to brain atrophy, where significant portions of the gray matter dissolve. There are four primary stages to AD: preclinical, mild, moderate, and severe. Each stage effects the overall atrophy of the brain in a progressive manner. While the course of AD may vary from patient to patient, the overall track of symptoms develop over these same general stages [33,34].

In the preclinical stages of AD, the disease begins to effect the areas of the hippocampus most related to the creation of memories. As these areas begin to atrophy, subjects will start to experience short-term memory problems. The onset of AD at this stage can go unnoticed for up to twenty years before the first signs of memory loss are noticed [4,5]. Some patients with preclinical AD actually have a condition called amnesic mild cognitive impairment (MCI). The differentiating factor between MCI and AD patients is that MCI patients generally have more memory problems than others at their age, but do not exhibit the severity of memory and cognitive problems that plague AD patients. It has been hypothesized that MCI is a transitional phase between the normal brain and AD; however, while more MCI patients go on to develop AD than people without MCI, only about 30% of those individuals with MCI actually develop AD .

Upon the initial diagnosis of AD, life expectancy for most patients ranges between 8 to 10 years. Of course, this number varies among patients depending on the current stage of the disease, along with other risk factors, such as family history and age for example [35]. In most cases, preclinical AD has very little brain atrophy, as illustrated in Figure 4. The blue shading in Figure 4b denotes the region of the brain affected by preclinical AD.

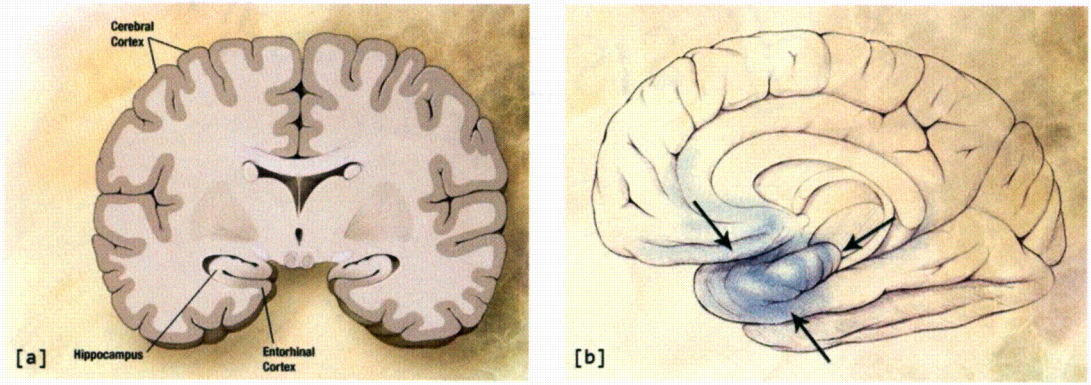

Figure 4 – Preclinical AD: Frontal and side illustrations of the brain [29]

The blue shading and arrows in Figure 4b indicate the region of the brain beginning to be effected by the preclinical stage of AD. Little to no brain atrophy is generally visible at this point, as illustrated in Figure 4a.

As the disease progresses into the mild AD stage, overall brain atrophy increases. This in turn results in noticeable memory loss as the damage to the cerebral cortex increases, also negatively affecting the patient's cognitive abilities. It is at this point that symptoms such as mood/personality changes, loss of initiative, poor judgment, and increased anxiety begin to manifest themselves. For most patients, this is when AD is first diagnosed, as both the patients and their caregivers start to accept that the memory and cognitive declines are more than what would be considered normal through aging.

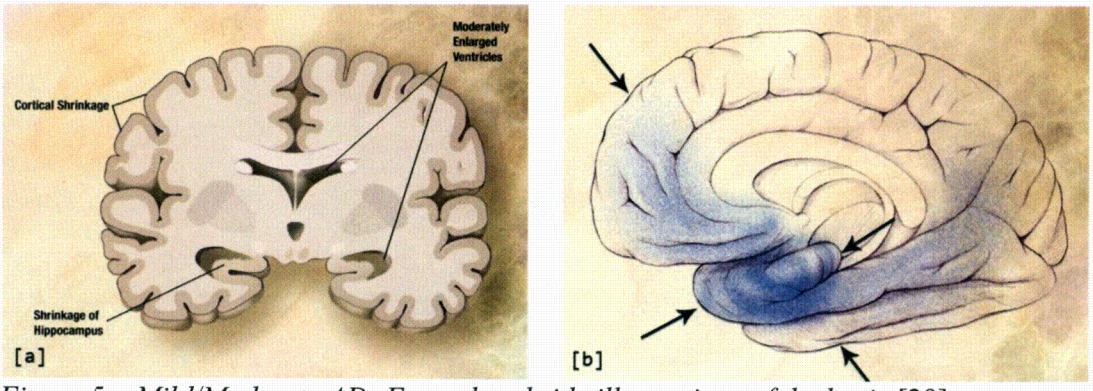

Figure 5 – Mild/Moderate AD: Frontal and side illustrations of the brain [29]

In this stage of AD, the regions of the effected brain matter grows—indicated in Figure 5b by blue/dark shading and arrows. The hippocampus/cortical regions begin to show shrinkage, as brain matter starts to die—moderate enlarging of the ventricles occurs at this stage.

As the overall atrophy of the brain increases in the moderate stage of AD, regions of the brain that control language, reasoning, and perception—specifically the cerebral cortex and frontal/parietal regions—are effected. Figure 5 shows the effect on the brain in the mild to moderate stages of AD. At this point, patients begin to lose their ability to organize thoughts properly and start to lose the ability to recognize friends and family. Memory and other cognitive abilities deteriorate. Some patients may still have an awareness of their condition at this point. Generally, at this stage patients begin to require assistance with many daily tasks, increasing the burden on caregivers [34]. In the severe stage of AD, the β-amyloid plaques and hyperphosphorylated-τ tangles have lead to a significant overall atrophy of the brain. Figure 6 shows an illustration of the brain in this stage, indicating a significant loss of brain matter. It is at this point that many patients lose the ability completely to recognize loved ones and to communicate coherently—if at all. From a physical standpoint, significant weight loss occurs, as some of the most basic of motor skills are lost. Complete assistance is required at this point, as patients with severe AD cannot swallow or control their bowel movements [33].

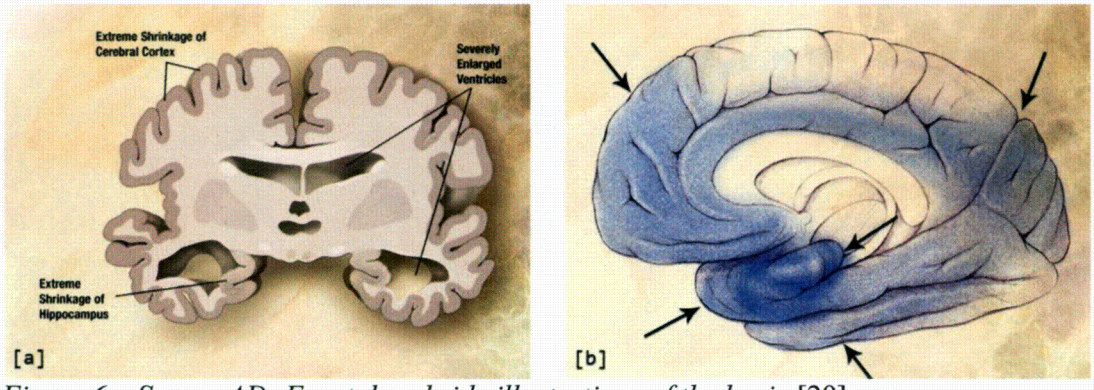

Figure 6 – Severe AD: Frontal and side illustrations of the brain [29]

By the severe stage of AD, a large portion of brain matter has atrophied, leading to extreme shrinkage of both the Cerebral Cortex and the Hippocampus. The ventricles are also shown to be enlarged greatly, a direct result of the brain matter loss. In Figure 6a, the blue/dark shading and arrows once again represents the main regions affected by this atrophy.

2.1.3 TREATMENT FOR ALZHEIMER'S DISEASE

Before treatment can begin, an accurate diagnosis of AD is of the utmost importance. Other causes of dementia must be ruled out, especially conditions that can be treated, and even reversed. If diagnosed early enough, the progress of AD may be slowed through the use of medications and other care, greatly improving the overall quality of life for the patient [10,35]. The use of medications to slow the progression of AD is a heavily researched area of interest for major pharmaceutical companies. In the United States, all drugs must pass rigorous clinical trials before acquiring FDA approval for prescription use. According to the National Institute of Health, there are currently four prescription medications available to slow the effects and progression of AD.

Table 4 – Medications to Treat Alzheimer's Disease[2] [36]

Drug Name	Usage	How it Works	Common Side Effects
Namenda® (memantine)	- Moderate to severe AD	- Regulates glutamate activation	- Dizziness - Headache - Constipation - Confusion
Razadyne® (galantamine)	- Mild to moderate AD	- Prevents breakdown of acetylcholine - Stimulates production of acetylcholine	- Nausea - Vomiting - Diarrhea - Weight loss - Loss of appetite
Exelon® (rivastigmine)	- Mild to moderate AD	- Prevents breakdown of acetylcholine and butyrylcholine	- Nausea - Vomiting - Diarrhea - Loss of appetite - Muscle weakness
Aricept® (donepezil)	- Mild to moderate - Moderate to severe AD	- Prevents breakdown of acetylcholine	- Nausea - Vomiting - Diarrhea

The primary goal of these medications is to reduce the symptoms of AD, by serving as a means to slow memory loss, control behavior problems, or even improve

[2] New medications are being tested all the time, and clinical trial volunteers are necessary for eventual FDA approval. More information regarding these medications and current clinical trials can be found at the NIA's website, cited in source [36].

sleep. These medications are prescribed after consultation with expert physicians in the field of neurodegenerative diseases [10,36].

Other methods for the treatment of AD have also been proposed. One such method is antioxidant treatment. The idea is that the generation of "free radicals" leads to the damaging and death of normal cells—a suggested cause of aging [37]. Free radicals from a chemical standpoint are essentially chemicals whose molecular or ionic structure includes an unpaired (or free) electron. Biologically, most free radicals contain oxygen. These "free radicals" are suspected causes for tissue damage caused by a range of conditions such as radiation and aging. Many physicians and scientists debate whether the use of antioxidants (such as Vitamin E) has any positive effect on AD patients [10,37].

Referring back to Table 4, it is clear that every drug (with the exception of Namenda®) attempts to prevent the breakdown of acetylcholine. This chemical has been identified as a neurotransmitter that is produced by neurons in the brain; specifically, acetylcholine is believed to be involved in cognitive abilities, such as learning, memory, and mood [38]. The enzyme acetyl-cholinesterase rapidly breaks down acetylcholine—the three medications described previously each act as a cholinesterase inhibitor, which has been postulated to help improve both cognitive and neuropsychiatric symptoms [10,38]. Another method to treat AD focuses instead on another chemical in the brain that acts on NMDA receptors. NMDA are glutamate receptors, which are the primary molecular mechanism that controls synaptic response and memory function [39]. The medication Namenda® works to regulate this chemical, glutamate. Research has shown that overstimulation of NMDA receptors may be detrimental to neuronal health [10,36].

2.2 BIOMARKERS FOR AD

In the current study, metabolic, anatomical, and physiological biomarkers for AD were all analyzed for their effectiveness in the early diagnosis of the disease. Each of these three biomarkers was obtained through clinical study and evaluation of subjects with the hope that each biomarker contained relevant and pertinent information about a given patient's condition rather than a standard medical evaluation.

The datasets used in this study were created through various biomarker cohorts, collected at the University of Pennsylvania. The collection methods for each biomarker are explained in sections 2.2.1 through 2.2.3, as well as relevant pre and post processing techniques. Literature reviews regarding previous works done on these biomarkers are also explored. Section 3.3 covers the CERND study in greater detail.

2.2.1 ELECTROENCEPHALOGRAPHY

The use of electroencephalography (EEG) to record event-related potentials in subjects has been rigorously explored in previous studies to determine its effectiveness as a physiological biomarker for early diagnosis of AD.

Electrodes were placed at specific intervals along the scalp to measure brain activity. Minute voltage changes were recorded from a large number of neurons firing in proximity to these electrodes. A summation of these voltages collected at each electrode constituted the EEG signal. The international 10/20 system was the electrode placement method used for this study, and is shown in Figure 7.

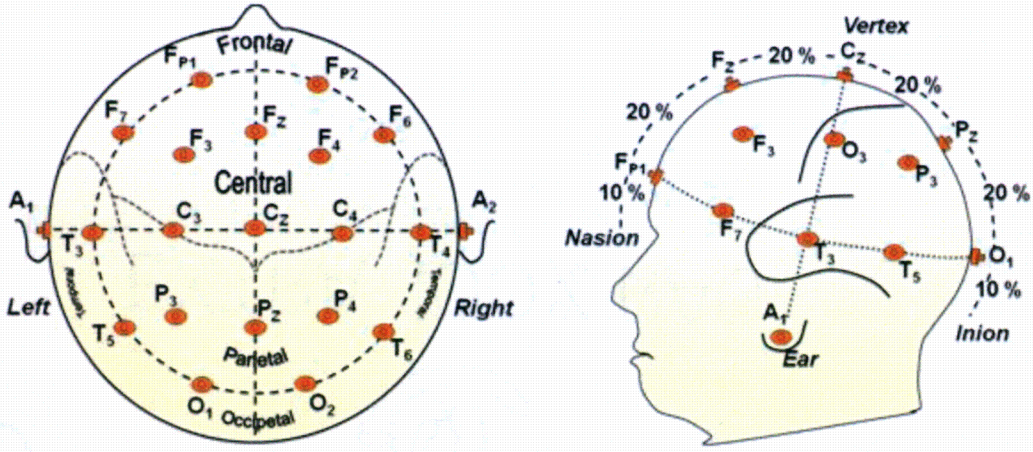

Figure 7 – The International 10/20 System for EEG Electrode Placement [40]

Based on the relationship between the location of an electrode and the underlying area of the cerebral cortex, this system is the most widely used and accepted method for the placement of electrodes in an EEG setup. A conductive gel or paste was applied between the electrode and the scalp to minimize the impedance caused by dead skin and hair. Each site has a corresponding letter and number to describe the lobe and hemisphere location, respectively. Even numbered electrodes referred to the right hemisphere, where odd numbers represented the left hemisphere. In some EEG system setups, the reference point is located near the eyes—in others it consists of two electrodes attached to the subject's ear lobes (electrodes A1 and A2, as seen in Figure 7). Slight modifications to this system are not uncommon, such as the replacement of the O1 and O2 electrodes with the OZ electrode, located on the mid-line in the occipital region [41].

The lettering of the system indicates the region of the brain where the electrode is located. Specifically, this lettering system is defined by an "F" for the frontal lobe, "T" for the temporal lobe, "C" for the central lobe, "P" for the parietal lobe, and "O" for the occipital lobe. While physically there is no central lobe in the cerebral cortex, these

electrodes are given this designation to indicate their location on the midline between the ears with the "FP" electrodes standing for 'front polar.' Nasion indicates the point between the forehead and nose, while inion is the "bump" at the bottom-rear of the skull. This method allows quick and easy configuration of electrode placement across all subjects of varying skull sizes. The name of the system stems from a reference to the 10% and 20% inter-electrode distances [41-45].

EEG signals provide measurement of the electrical activity of the brain in millisecond time-resolution. Because of the ease of collection, and relatively inexpensive setup, EEG has become the biomarker of choice to analyze brain activity in many disorders and disease, such as epilepsy, dyslexia, and sleep disorders [44,45].

Considerations on the accuracy of EEG measurement must be made. While these signals have superb time resolution, the spatial resolution suffers due to the inherently poor localization properties of EEG. Since signals are generally measured from the scalp, the electrodes will pickup post-synaptic events from every neuron firing during some neurological episode. While these electrodes can detect these events in real time, it is impossible to determine the locale of the actual event since the EEG acquired at any given electrode is a summation of signals from neighboring areas of the brain. Consequently, the specific synaptic origin of an EEG signal cannot be determined through the use of these scalp electrodes.

In an effort to increase spatial resolution in EEG, subdural electrode techniques have been employed in various studies. While subdural electrodes have helped improved the localization of the EEG signal, this procedure is highly invasive and can result in infection or even brain hemorrhaging [46,47]. A particular field of interest is the study of

source localization of EEG. Essentially, this focus attempts to determine the localization source in the brain of various EEG signals. There are many methods, such as low resolution brain electromagnetic tomography (LOERTA); however, source localization reaches beyond the scope of this thesis. Interested readers may reference the LORETA homepage for more information on this field [48]. Another consideration to be made for EEG signal acquisition is its inherent susceptibility to artifacts and noise. Artifacts can be introduced into the signal through poor electrode contact to the scalp, impeded by hair and dead skin cells, improper application of the gel/paste, and sweating. A main source of artifacts in EEG is caused by other electrical activity like impulses generated during muscle movement (EMG) such as eye blinking, cardiac rhythm, or even simple muscle movement on or near the scalp. An example of this typical artifact is shown in Figure 8.

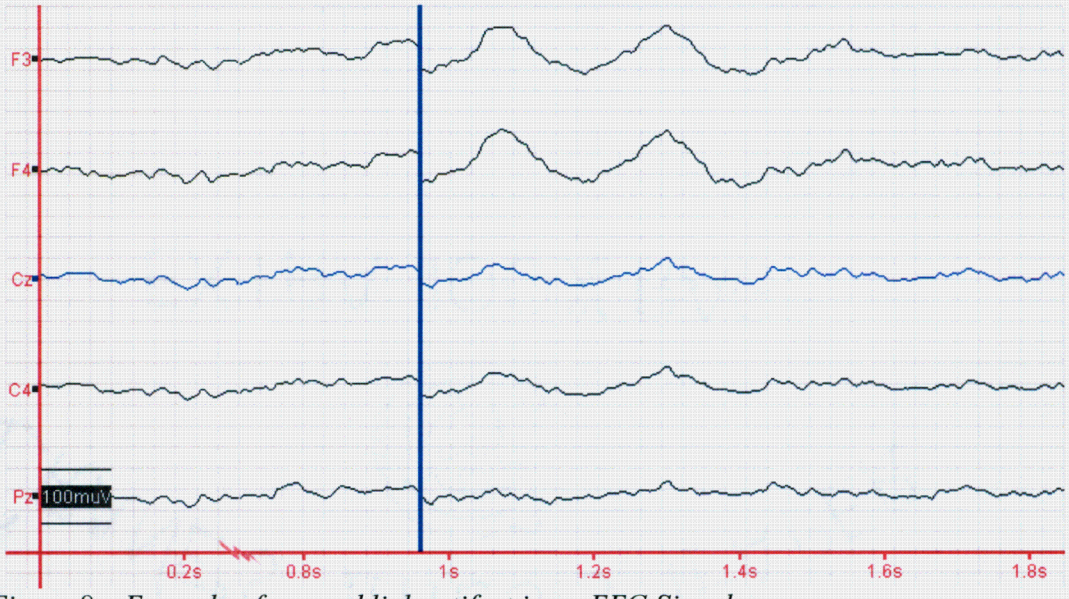

Figure 8 – Example of an eye blink artifact in an EEG Signal

The artifact can be clearly seen in the F3 and F4 channels, starting at the blue line (~1s). Since the F3/F4 electrodes are very close to the eye, it exhibits the greatest distortion—as the data is viewed from electrodes further away from the eye, the distortion lessens in magnitude (such is the case at CZ, C4, and PZ in this example, acquired from our EEG setup).

The electrical potential created in a muscle movement is much greater than that of an EEG, and easily skews the entire signal readout. Fortunately, such artifacts have been greatly researched and most artifact-rejection techniques can handle these errors [45]. In even the most controlled acquisition environments, artifacts are still introduced to the EEG signal. Other than the typical EMG artifacts, other sources of error such as electrical noise can be presented. The 60 Hz noise[3] introduced to EEG from AC power is commonly removed through the use of a notch or low pass filter, as most information in an EEG resides at lower frequencies [45]. It is imperative that such sources of noise, be it physiological or external, are removed prior to analysis by either a trained EEG technician or an automated artifact rejection process.

2.2.1a SPECTRAL CONTENT

Early research on the EEG was first described by Richard Caton in 1875 through his experiments on rabbits and monkeys. This early work involved the placement of electrodes on the exposed brains of the animals while visual stimuli were presented to track changed in the EEG—Canton showed that flashes of light created a change in the electrical activity, and his was the earliest known experiment in ERPs [6,26].

By 1902, Han Berger started work on EEG waves in canines and in 1920 he began studies on human subjects. In 1929, Berger published his work, which included experiments performed on numerous subjects, including him and his son, in a paper titled "On the Electroencephalogram of Man" in which he identified two basic brain wave patterns [26].

---

[3] For this study, this refers to the United States—worldwide AC power specifications can be either 60 Hz or 50 Hz depending on locale. In such instances these high-frequency components are filtered on a location basis. A complete list of per country power standards is available in source [49].

One pattern was described by Berger was a "large, regular wave, that occurred 10 to 11 times per second" and the second as a "smaller, irregular one, at a frequency of 20 to 30 cycles per second" [50]. He named these two waves "alpha" and "beta," relating to the larger/slower and smaller/faster kinds respectively. Soon after, researchers identified more types of brain waves, and named them using the same Greek letter convention as Berger, calling them gamma, delta, theta, kappa, lambda, and mu. The characteristics of these EEG waves (frequency, amplitude, patterns) in the same individual are generally rather consistent [50,51].

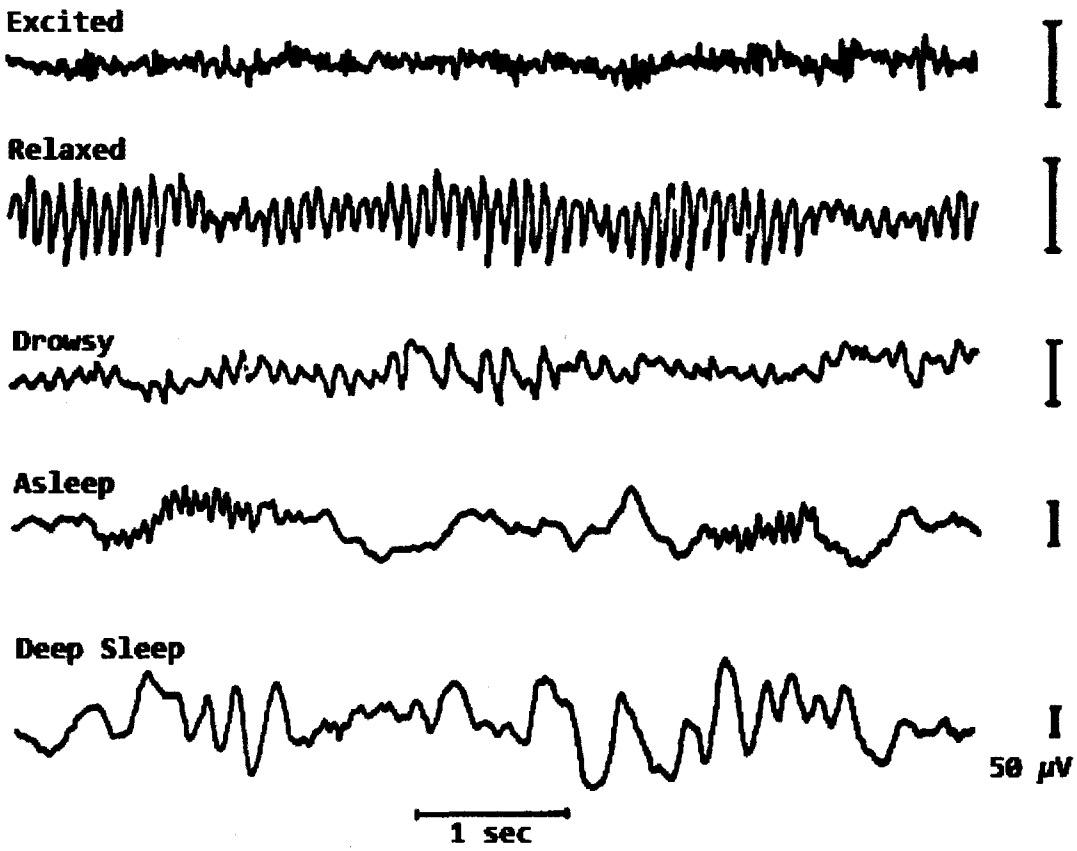

Figure 9 – EEG signal variation between different stages

Excitement is characterized by a rapid frequency, low amplitude signal. Varying levels of sleep states alternatively are distinguished by an increased irregularity and the manifestation of "slow waves" [50].

EEG has been used in many applications for the diagnosis of different sleeping disorders and dementia. Figure 9 illustrates the differences in EEG signals based on various states of awareness. There are seven primary EEG wave types, each generally associated with certain brain activities. These wave types occur below the 50/60Hz range so high frequency noise is easily filtered and avoided through pre/post processing techniques without disturbing these waves.

Delta Waves (0.5-3.5 Hz): The delta wave is a low frequency signal, with high amplitudes ranging from 20 to 200μV. In most cases, this wave only appears during deep sleep in normal individuals—when observed in an alert person, delta waves could suggest an irregularity in the brain caused by cerebral damage or a tumor. The delta wave is shown at the bottom most trace labeled "Deep Sleep" in Figure 9. Dominant in infants, this wave is thought to be the result of unconscious processes within the mind [50,52].

Theta Waves (4-7 Hz): The theta wave is rarely found in adults and is more common in children. Similar to delta waves, the amplitudes of theta waves are higher than most, ranging from 20 to 100μV. In adults, it has been shown to occur during states of fantasy, both displeasure and pleasure, and drowsiness. The appearance of this rhythm in normal waking adults can be a sign of high stress or disease [50].

Alpha Waves (8-13 Hz): These rhythmic oscillations usually have amplitudes between 20 and 60μV. Most people can generate alpha waves in a resting position, with eyes closed in a meditative state. As soon as the individual begins any type of mental or physical activity, these waves generally decrease in amplitude or disappear. This wave is illustrated in Figure 9 by the "relaxed" state. The established view on alpha waves associates them with relaxation; however there have been published studies by Shaw and

Mulholland that illustrate several exceptions to alpha wave association with relaxation [53,54]. In up to one third of individuals tasked with performing mental arithmetic there was either no change in amplitude or an enhancement of alpha waves was observed. More evidence shows alpha wave augmentation before performing skilled actions involving aiming, such as archery, shooting, and in putting golf. This wave originates in the occipital and frontal cortex regions, and is the most common type of EEG activity in a healthy awake adult brain [50,51,53-55].

Kappa Waves (~10 Hz): While technically within the alpha band range, Kappa waves at about 10 Hz have been observed during thinking. Kennedy, *et al.* (1948) reported that this phenomena occurred in about 30% of their subjects [50,56].

Mu Waves (8-13 Hz): Even though the mu rhythm resides entirely within the alpha wave definition, it has been shown that these waves are independent phenomena due to differences between the two in source generation. Appearing in the normal EEG of about 7-8% of the population, this rhythm has been attributed to movement or the intent to move. The mu wave is not effected by eye blinks—rather, this wave is more connected to the motor cortex, relating specifically to the movement of the extremities [50].

Beta Waves (14-30 Hz): An irregular wave that has an amplitude generally between 2 and 20μV, the beta wave are most common when a person in engaging in mental or physical activity. This is illustrated in Figure 9 by the excited state, shown in the top most trace. Generally seen symmetrically in the brain, it primarily originates from the frontal regions of the cortex [50].

Gamma Waves (~40 Hz): First reported in humans by Galambos, *et al.* (1981), the gamma wave is described as a rhythmic EEG activity occurring in response to sensory stimuli—either auditory or visual. This rhythm represents the highest frequency in human brain activity. Usually correlated to simultaneous processing and communication between different regions of the brain, the gamma wave has a resting frequency of 40 Hz, and ranges between 36 and 44 Hz in most cases. Strong and well-regulated gamma activity has been attributed to good memory; conversely, a weak gamma rhythm is indicative of learning disabilities [50,57]. This band may be of interest in future studies for AD-based diagnosis through EEG; however, it has not been extensively researched at this point.

2.2.1b ACQUISITION PROTOCOLS

In most cases, EEG recording is done while using specific protocols to generate stimuli in an attempt to evoke certain responses. These stimuli can be visual, auditory, somatosensory, or olfactory. Generally, a visual stimulus consists of patterns, colors, or words presented at random intervals to the subject. A somatosensory stimulus commonly is presented in the form of a targeted electrical pulse to some part of the patient [58,59]. An olfactory stimulus is an odor presented to the patient to generate some response for the EEG recording. In most cases, the auditory stimulus is a series of tones that vary in frequency, but also can contain everyday sounds for the same purpose. Interestingly, a study by Finney, *et al.* (2003) demonstrated that when presented with visual stimuli, a response was observed in the auditory cortex in deaf subjects [60]. Irrespective of their originating source, the responses generated in an EEG from these stimuli are called event related potentials (ERP) [61]. Many protocols can be created to generate these

responses; in some cases multiple stimuli can be combined with cognitive tasks to produce ERPs.

2.2.1c ERP AND THE ODDBALL PARADIGM

One such protocol that is widely used in ERP generation is called the auditory oddball paradigm. In the traditional setup, two tones are presented to the subject—a standard tone, generally at 1 kHz, is played in short segments for the majority of the EEG recording. Scattered sporadically throughout the procedure, an oddball tone is presented, usually at 2 kHz. The two stimuli in this situation are presented randomly throughout the recording session, with a random amount of inter-stimulus time. In order to amplify the possible ERP from the oddball tone, the subject is usually instructed to push a button or count the total number of oddball tones [62,63].

A modification to the auditory oddball paradigm was presented by Yamaguchi, *et al.* (2000) which has been implemented in this study [64]. In addition to the standard and oddball (target) tones from the original protocol, the proposed protocol by Yamaguchi's group added a third stimulus type, consisting of novel sounds. These sounds were never repeated throughout the EEG recording and had slight variations in length. Each sound was a common everyday noise, such as a door slamming or a dog barking. As with the original protocol, patients were instructed to press a button upon hearing the 2 kHz target tone. The overall breakdown of stimuli presentation throughout this modified oddball paradigm was as follows: 65% standard, 20% target, and 15% novel. The inter-stimulus interval varied randomly between 1 to 1.3 seconds. In this setup, the subjects were not warned of novel sounds; nor were they instructed to respond in any manner to them. By evoking a response in an entirely different manner with novel tones, the ERPs generated

from both novel and target stimuli would represent activity from different regions of the brain, possibly providing discriminatory information between different types of neurodegenerative dementias [64].

Once this process is complete, the subject recordings are segments with respect to stimulus type and are averaged together to obtain the overall ERPs. Cohen and Polich (1997) say the minimum number of trials (epochs) necessary to have a statistically significant average of the ERP is twenty [65]. However, in this study and many others like it, a much larger number of ERPs are averaged to aid in the reduction of noise in individual epochs. Since the overall amplitude of EEG is small to begin with, and the amplitude of ERPs in the EEG can be minute as well, it is very possible that characterizing information could be lost in noise introduced in the recording. Once averaged over many trials, the features of an ERP become more distinguished as the noise is minimized.

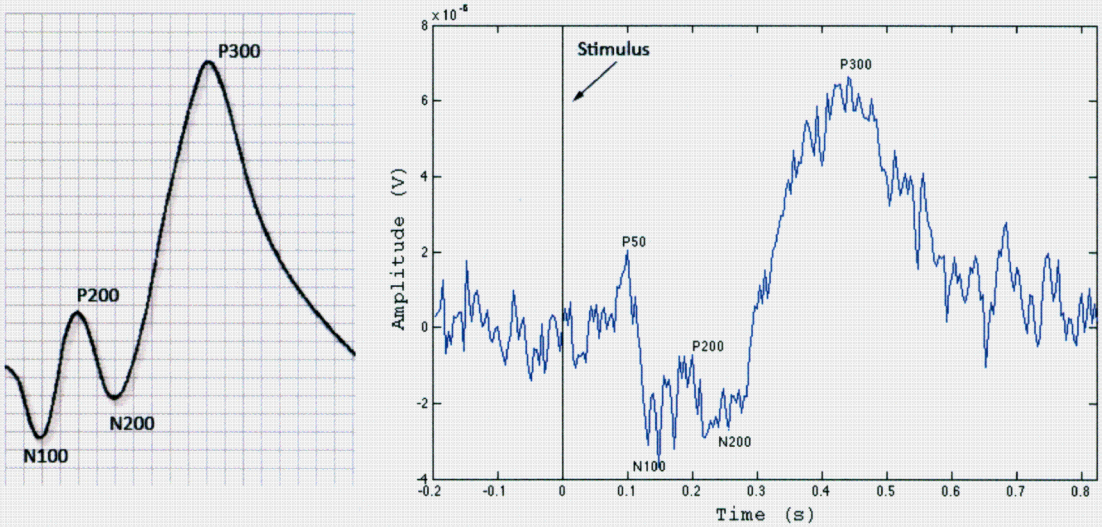

Figure 10 – Typical ERPs, labeled with features of interest

While ERPs vary from patient to patient, the general form of an ERP can be described by the illustrations in Figure 10. In the illustration on the left, an idealized case is shown, labeled with the four main components that generally make up the ERP. The letter refers to a peak (positive) or trough (negative), with the number indicating the milliseconds post-stimulus. While peaks and latency can vary, the order does not change. This is a general case as the right illustration in Figure 10 depicts, these peaks and troughs do not always perfectly line up over their respective post-stimulus times [63].

2.2.1d THE P300 AS A BIOMARKER

The feature of most interest in the ERP, and subsequently the most studied, is the P300. The presence and amplitude of this peak in ERP data has been related to cognitive ability, and generally occurs 250-500ms after the stimulus is presented to the patient [66,67]. Two and three stimuli oddball paradigms tend to generate the P300; however, this peak has been shown to differ between stimuli. In novel stimuli, the P300 tends to arrive sooner (lower latency, P300a) than when elicited by a target stimuli (P300b) [59,67].

It is important to note that a visual interpretation of the P300 has not been considered an effective diagnostic tool. This becomes apparent across various patients. In diagnosing AD, one would expect a general pattern of late P300 latency and diminished amplitude across AD patients. For normal control patients, one would expect a stronger and higher amplitude P300 response. In some publications and research, the P300 is referred to as P3. Similarly, other peaks and troughs are identified in this format, such as N100 referenced as N1 or P200 as P2. Other publications even flip the vertical axis, running from negative to positive, top to bottom. Figure 11 illustrates an example

comparison of AD probable and normal control patient's averaged ERP plots using the oddball paradigm.

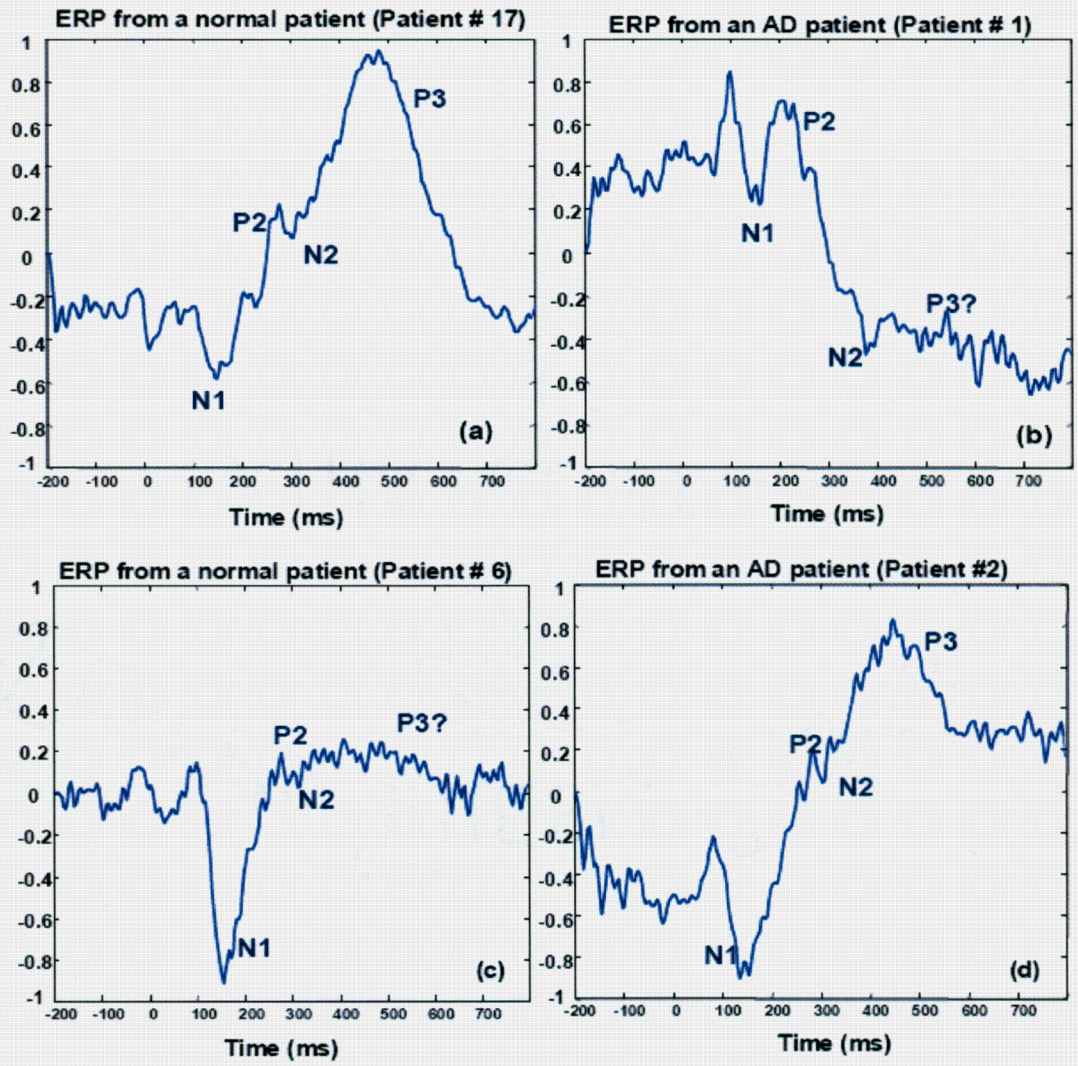

Figure 11 – ERP Comparisons: The P300 in CN and AD subjects

Plots (a) and (b) in Figure 11 show expected patterns in AD and CN ERPs, with the CN patient having a fairly typical ERP and the AD patient with a poor P300 response, possibly indicating poor cognitive ability. However, as plots (c) and (d) indicate, this is not always the case. While a strong P300 may indicate good cognitive processes, it is not uncommon to find strong P300 amplitude in AD patients and a weak P300 in normal

patients. These visual anomalies demonstrate the problem with the use of P300 as a diagnostic method.

The P300 itself has been studied at length since it was first observed in 1965. Over the past 45 years, the P300's relation to cognitive ability, its origin within the brain, and its many research applications has been comprehensively studied. Originally, the P300 was thought to stem from deep within the brain, in the hippocampal region; however, Polich and Kok in 1995 determined a possibility for multiple sources for the P300 within the brain. They postulated that the P300 most likely originated in the temporo-parietal junction [67]. Interestingly, this region of the brain has been found to play a role in self-other distinction processes, theory of mind,[4] as well as attributed to out of body experiences[5] [69,70].

In 1997, Polich and Margala described a single stimulus protocol to compare directly with a two stimulus procedure. They determined that while a P300 generated through a two stimuli, auditory oddball paradigm had slightly stronger amplitude and more defined P300, the differences between the two were statistically insignificant. The report did show that irrespective to the stimuli, the source of the P300 appeared to remain the same [62].

A 1999 study determined that both auditory and visual stimuli presented with an "active task condition" created a much stronger P300 than when no task was assigned. The active task condition refers to a test in which a subject was instructed to press a

[4] Theory of mind relates to the ability to attribute specific mental states (belief, intent, desire, imagination, knowledge) to oneself and others. This is also concerned with the ability to understand that others mental states are different from one's own [68].

[5] These out of body experiences (OBE), described in [69] have been shown to occur in part due to damage to the temporoparietal junction. OBEs can be induced by electrically stimulating this region of the brain.

button on target stimuli. Furthermore, the auditory protocol generated a strong response in both tasking situations as opposed to the visual stimuli [71]. Strüber and Polich (2002) compared the oddball paradigm to the single stimulus protocol once again, this time with a focus on how the P300 would be affected by altering the inter-stimulus interval (ISI). They showed that the oddball paradigm produced a stronger P300 overall than the single stimulus protocol with an ISI of only 2.5 seconds. When the ISI was increased to 30 seconds, both protocol produced similar P300s. The only discerning difference between the two protocols in this case was that the P300 in the oddball paradigm was of a longer duration [72].

The P300 is now considered to form from multiple sources simultaneously within the brain. Polich (2007) detailed two subcomponents of the P300, the earlier P3a and the later P3b, which he postulated are created depending on the difficulty of the task at hand, and the attention required by the subject to standard, target, and novel stimuli. The P3a appeared to originate primarily from the frontal regions of the brain during periods where active tasking and attention was required, while the P3b seems to originate from the temporal and parietal regions more associated with memory processing. Because of this, it is believed that the P3a and P3b carry different information regarding cognition from each other. The P3a is generally only found in 10-15% of cognitively active, normal young adults [73].

2.2.1e WAVELET ANALYSIS OF THE P300

As discussed in section 2.2.1a, the EEG contains many different frequency bands, some overlapping, that represent various fundamental processes occurring in the brain. Analysis of this information requires the ability to access specific frequency bands within

this signal. On the other hand, since ERPs are one-second signals that consist of averages taken from EEG fragments, the latency of the individual components of the ERP are just as crucial for analysis as isolating their respective frequency bands. This is related to the non-stationary nature of EEG signals, for which the wavelet transform is considered to be one of the most appropriate analysis methods.

Başar, *et al.* (2001) compared three different protocols: the oddball paradigm, the single stimulus (varying ISI), and the single stimulus (constant ISI) with every fourth stimulus removed [74]. The ERPs from each case were decomposed using the wavelet transform, and the resulting signals were compared across various frequency bands. Their key observation was that the P300 component was visible across all electrode locations when using the oddball paradigm with delta peaks around 2 Hz [74]. Previously, in another paper published by Demiralp *et al.* (1999), it was shown that viewing ERPs only in the time domain left the analysis susceptible to the diverse disposition of the P300 as well as other components in the signal. They hypothesized this susceptibility to be caused by concurrent brain processes that initially created the signal. By using the wavelet transform to decompose the signal into its respective time and frequency components, they were able to record and accurately measure components from events in the brain. Since frequency components in different bands of EEG (delta, theta, and alpha for example) indicate different facets of cognition, the process of wavelet decomposition provided the capacity to witness new mechanisms involved in the simultaneous processing of cognitive ability that could not be seen in the time domain alone [75].

Başar *et al.* (2001) then compared the wavelet transform for ERP data to other signal processing techniques, such as the Fourier transform. Their primary goal was to separate the standard EEG frequency bands from the original data (namely the alpha, theta, delta and gamma bands). In using the wavelet decomposition technique, they found that more characterizing information could be extracted from the delta and alpha band that was not possible through standard Fourier spectral analysis. The main advantages of the wavelet transform over other pre-existing signal processing techniques included the ability to resolve in both frequency and time, significant data compression ability, as well as no need for a fixed time window to analyze the signal (as required in Fourier-based techniques) [74]. Demiralp *et al.* (2001) used a five-octave wavelet transform to analyze both single-trial and averaged ERP. The main emphasis of their study was to determine whether this transform can be used to compare the components of ERP from subjects of different age and cognitive ability. They confirmed that the delta band contained the most pertinent and distinguishing feature in an ERP after wavelet decomposition [75].

2.2.1f ERPS IN ALZHEIMER'S DISEASE RESEARCH

Using electroencephalography to capture event related potentials has become relatively affordable and feasible for use in even the smallest of community clinics; subsequently, the viability of ERPs for AD diagnosis has been researched heavily over the past decade. Much of the research in this area has focused on merging and expanding upon key information from previous works [41,44]. While there is a plethora of information regarding the use of ERPs in medical diagnosis, the following section summarizes the work most relevant to the current study.

The most conventional method to analyze ERPs in a patient is through visual inspection. In such an analysis, a technician determines the amplitudes of latencies of standard ERP components such as the P300, either in a single trial, or through an averaged set of trials [14]. The latency of the P300 peak has been shown to gradually increase proportionally to the age of a cognitively normal individual. This latency also increases in patients with AD as compared to the P300 of cognitively normal subjects of the same age range. It is important to note that while true for AD, this latency is also apparent in other neurodegenerative diseases and dementias, and therefore is not a reliable marker for AD [64,76]. A similar parallel can be drawn between cognitively normal subjects and AD patients of the same age group with respect to the amplitude of the P300. It was initially reported that patients with AD had decreased P300 amplitude in comparison to normal subjects. These data were later contested by several other studies [62]. While parallels and trends can be drawn when analyzing entire cohorts of data, it is not generally statistically relevant to apply the same convention to an individual patient, due to the high variation of ERPs between subjects.

A recent paper concerned specifically with the development of a better method to distinguish various dementias through the use of EEGs was published by Hidasi *et al.* (2009). They performed a coherence analysis on a cohort consisting of 14 AD probable patients along with 10 control subjects. Relative frequency band changes were analyzed between the two groups, and coherence analysis was performed between these bands. Their findings reported a much higher theta band power in AD patients as compared to controls; furthermore, they noted that the alpha-2 band, described in the paper to cover 9.5-11 Hz, had an increase in the AD group as compared to the normal group. Even more

telling, they reported that the alpha-1 band (8-9 Hz) had a visible increase in control subjects when compared to the AD patients [77].

Alzheimer's disease and its links to genetic susceptibility have also been extensively researched. It has been shown that while a small number of cases of have come from known genetic causes, the majority of AD cases do not. A study by Boutros *et al.* (1995) studied ERPs in patients at risk for AD. The study cohort consisted of 19 healthy subjects and 33 subjects known to have a family history of AD. The auditory oddball paradigm was used for ERP generation. An overall increase in amplitude in the P50, P300, and N100 responses from subjects with familial AD as compared to the normal subjects demonstrated a correlation between the P300 and early cognitive decline in patients with first-degree familial history of AD [78].

A similar study was conducted by Ally *et al.* (2006) with the biological children of patients with AD to examine the feasibility of the P300 as a preclinical maker for the disease. Once again, the oddball paradigm was used, the study consisted of the AD group, their children, as well as age and gender matched control groups for each. Their results confirmed the findings of others: the amplitudes of the P300 were significantly lower in the AD group as compared to the control group. However, the more fascinating result from their study shows that both the amplitude and latency of the P300 in the children of the AD group were lower in comparison to the control group of subjects. This study showed that the use of the P300 as a diagnostic measure could be extended for preclinical use in a high-risk patient group [79].

2.2.1g DIAGNOSIS WITH EEG

Certain traits within the ERP can be used as a biomarker for AD; either through wavelet processing or visual inspection, distinguishing characteristics can be identified in the ERP to aid in the classification of AD and normal subjects. Many different classification methods and algorithms have been created in an attempt to construct an accurate system for automated diagnosis of AD.

An early method for automated classification using artificial neural networks was proposed by Polikar *et al* (1997). A 28 patient cohort, of 14 AD and 14 normal controls, was analyzed and ERPs generated through the oddball paradigm were extracted from the corresponding EEG data. For this study, the ISI was 1.5 seconds with each tone 200 ms in length. Eighty six percent of the presented stimuli were standard 1 kHz signals with 14% constituting the 2 kHz target tones. The initial classification was done on the raw time domain signals, yielding 64% accuracy at the best trial. When wavelet coefficients were used in place of the raw signals, accuracy improved within a range of 79% to 93%. While the results from this study proved promising, such a small cohort created an unfavorably large confidence interval; therefore, a larger patient cohort was necessary to properly validate these results [80].

In a study aimed at validating the plausibility of using a recurrent neural network (RRN) to diagnose AD, Petrosian *et al.* (2001) used EEG collected from patients with AD and an age-matched control group. Instead of using an oddball paradigm to elicit ERPs, they collected EEG recordings for two minutes at a resting state from the parietal and occipital chancels from 10 early AD and 10 control subjects. Preprocessing was done with the wavelet transform, using the Daubechies 4 wavelet (db4). They trained the

RNN using 3 AD and 3 control subjects and utilized the remaining 14 for testing. All control subjects were diagnosed accurately, with only 2 misdiagnoses for the AD test subjects, leading to 80% sensitivity and 100% specificity. Again, while this study appeared promising, the same algorithm needed to be applied to a larger patient cohort for statistical significance [81].

In 2002, Yagneswaran *et al.* studied signal power frequency and wavelet characteristics in EEG for AD analysis. Their cohort consisted of 9 AD probable and 10 control subjects; nine-channel EEG data was recorded, with signal power frequency and wavelet coefficients used to make the diagnosis. Overall, the information from the power spectra proved inconsequential, so a band pass FIR filter utilizing the Hamming windows was used to separate the delta, theta, alpha, and beta bands. These filtered bands were used alongside relative power and slow wave ratio to train and test an artificial neural network. They also trained and tested another neural network using data from the same EEG recordings, this time utilizing the Daubechies 5 (db5) wavelet coefficients as a means of comparison. Interestingly, the classification accuracy of the network trained on the power spectrum features reached 94.7%, with the network trained on the db5 wavelet coefficients only reaching 89.4% [82]. This study gave interesting insight to the use of a raw signal filtered into bands as opposed to wavelet decomposition for diagnosis of AD. Again, the small cohort, only 19 subjects, indicates it would be advisable to repeat this experiment on a much larger pool of subjects.

A different method of analysis, called approximate entropy (ApEn) was examined in a 2003 study by Abasolo *et al.* ApEn is a "regularity statistic" that attempts to represent the randomness of fluctuations in time series data, such as EEG. ApEn is

essentially a representation of the complexity of the given data [61]. They collected EEG data from the P3 electrode only, on a cohort of 7 AD and 7 control subjects. The signals were then divided at 5 second intervals at which time the ApEn analysis was performed. They showed through ANOVA testing that the higher ApEn complexity scores in the AD subjects in comparison to the controls were statistically significant [83].

Tao and Tian (2005) tested the diagnostic accuracy for AD versus MCI diagnosis. In their cohort, 12 AD and 18 MCI patients had their EEG signals recorded from 21 electrodes. The protocol included five minutes of resting state EEG recorded with the subject's eyes open, followed by five minutes with the eyes closed. They then generated 21 random numbers, and displayed them to the subject in 30 seconds; each subject was asked to make a note of the number of odd digits shown as the cognitive task to elicit ERPs. The subsequent EEG signals were then decomposed using the Mexican hat wavelet, primarily to extract the gamma band from the data. Coherence analysis[6] was then performed on pairs of electrode channels. For the eyes closed resting state, AD patients showed a much lower coherence in comparison to the MCI and control groups. However, in the cognitive task both the AD and MCI groups shared a decline in coherence when compared to the normal control subjects [84].

Clearly, one of the main challenges in working with EEG data is feature extraction. Chapman *et al.* (2007) used principal component analysis (PCA) to extract eight features from a cohort of 12 AD and 12 normal subjects. They implemented the 'number-letter' paradigm, which elicits ERPs through visual stimulation. A discriminant function based classification technique was then tested on the data. When trained and

[6] Coherence analysis allows for the detection and measurement of linear dependencies in multichannel time-series data. Essentially, this analysis finds variations in signals with similar spectral properties .

tested with a 50-50 validation technique, the performance reached 92% with a sensitivity and specificity of 100% and 83%, respectively. When trained and tested with a leave one out cross validation method, the classification performance reached 79%, with corresponding sensitivity and specificity at 83% and 75% [85]. While the 50-50 validation had a higher classification accuracy, it is not truly representative of the technique—the leave one out configuration provides a more accurate depiction of the system's diagnostic capability.

Henderson, *et al.* (2006) collected EEG data from 30 AD probable, 6 vascular dementia, 3 mixed dementia, and 42 cognitively normal subjects. For their classification method, they implemented a fractal-dimension based approach as well as an approach based on a probability density function with zero-crossing intervals that used the raw EEG as data. Achieving a sensitivity of 67%, with a specificity of 99.9%, the fractal based approach did fairly well in identifying the control subjects. The probability density approach obtained 78% sensitivity, improving upon the fractal based approach. Both methods need to be researched more thoroughly for multi-class situations [86].

2.2.2 MAGNETIC RESONANCE IMAGING (MRI)

The use of magnetic resonance imaging (MRI) as an anatomical marker for AD has gained in popularity with the declining costs of these machines. The ADNI study also helped in the proliferation of this technique, as new standards and modalities were created to use MRI as a diagnostic tool for AD classifications [16].

MRI is as a medical imaging technique, most commonly used to visualize the internal structure of the body. Powerful magnetic fields are used to align the nuclear magnetization of hydrogen atoms within the body, and radio frequency fields are then

used to systematically change the alignment of this magnetization, which causes the hydrogen nuclei to produce a rotating magnetic field. This field is subsequently detected by the scanner, and with enough manipulation through additional magnetic fields, a visual representation of the scan be can constructed [87]. In methods similar to those used for CT, three-dimensional models can be constructed from multiple MRI scans.

As discussed previously, the stages of AD can also be characterized by the atrophy of the brain. This change in overall brain volume can be detected by various volumetric MRI imaging techniques. MRI enhances the differences in tissue matter based on the ratio of bound to unbound water molecules, which gives it a far greater contrast than that of computed tomography (CT)[7]. This binding of water molecules is what makes the MRI useful in imaging the brain; specifically, brain matter has more bound water molecules in comparison to the surrounding cerebral spinal fluid (CSF) and bone structures. For example, the T2 weighted MRI technique boosts contrast in the image, showing water and fluid containing tissues as bright regions of the brain. This method accentuates the CSF and allows for greater contrast in viewing the brain tissue and surrounding fluid [89].

2.2.2a MRI MODALITIES

Targeted neuroimaging software can be used to isolate regions of interest (ROI) in an MRI of the brain. These regions can then be quantized through volumetric image processing, isolating the low contrast brain matter from the bright CSF in the MRI. Using this information, slices of the brain can be compiled to determine overall inter-cranial volume (ICV), or specific regions using various segmentation methods. However,

[7] CT scans utilize tomography (digital geometry processing) to generate a three-dimensional image of an object's internals from a series of two-dimensional X-ray images taken from a single axis of rotation [88].

while a single-visit MRI as a diagnostic tool is convenient to a point, longitudinal studies generally give a better representation of overall brain atrophy [90].

There are two primary types of MRI that can be obtained, referred to as T1 and T2 imaging. Referred to as the "longitudinal" relaxation time, the T1 type indicates the time required for a substance to become magnetized after being subjected to a strong magnetic field. This also can refer to the time required to regain a longitudinal magnetization after the RF pulse. Essentially, T1 relaxation is concerned with the water molecule's magnetization vector returning to its initial state after rotation caused by the RF pulse. This relaxation time reflects the relationship between the frequency of these movements as well as the resonance frequency. When these two components are comparable, T1 is short, which means magnetization recovery occurs quickly; conversely, when different, the T1 is long, corresponding to a long magnetization recovery time. Biologically, the water molecule in tissue is small, and therefore moves too quick for T1 relaxation, whereas large proteins in body tissue move slow. Generally, tissues and fluids in the brain that have a large ratio of *bound* to *unbound* water molecules have short T1 times (fast magnetization recovery), whereas tissues with a large ratio of *unbound* to *bound* water molecules have longer T1 times, corresponding to a long magnetization recovery time [91].

T2 is known as the "transverse" relaxation time, which is a measure of how long transverse magnetization would persist in a perfectly uniform magnetic field. It can also be considered a measure of how long the resonating protons remain in phase, given an orthogonal radio frequency (RF) pulse. Generally speaking, T2 relaxation refers to the rotation of water molecule's magnetization vector away from the transverse axis

following the RF pulse in MRI imaging. The decay in T2 imaging is caused by magnetic relations that occur between spinning protons. Unlike T1, T2 interactions do not involve any transfer of energy; rather, a change in phase occurs, which leads to a loss of coherence. Therefore, we would expect an opposite reaction to that of T1 relaxation; specifically, in T2 relaxation tissues and fluids in the brain that have a large ratio of *unbound* to *bound* water molecules have short T2 times, whereas tissues with *bound* to *unbound* ratios have long T2 times [91].

T2 relaxation becomes less efficient and T2 time increases as the natural frequency of the protons heightens. Rapid irregular movements average out in T2 so internal fields become insignificant. This occurs in liquids and creates a more uniform internal magnetic surrounding. In the brain, the hydration-layer has a shorter T1 than 'bulk-phase' water (e.g. cerebral spinal fluid or CSF). The motions of these protons are not too slow to create efficient T2 relaxation. Therefore, the T2 time stays long. T1 and T2 imaging modalities for MRI have advantages and disadvantages, each visually representing tissue and fluid differently in the brain [87,91].

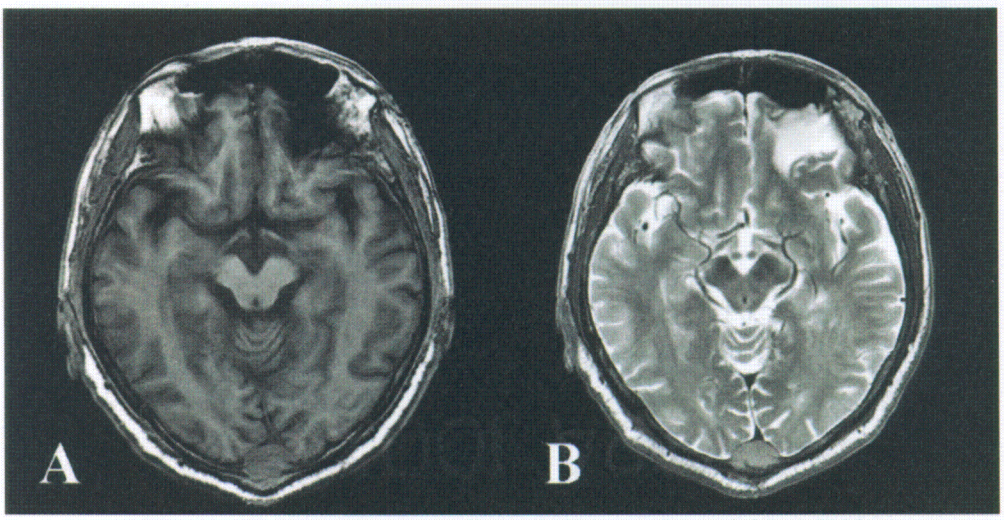

Figure 12 – MRI T1 (A) and T2 (B) Imaging Modalities [92]

Figure 12 illustrates the two modalities on a sample brain image of a patient. These two methods result in significantly different contrasts, depending on the type of tissue present [89]. In T1 weighting, the image shows excellent tissue contrast, with CSF appearing much darker. T2 weighting shows less spatial resolution for tissue matter, whereas CSF appears bright and high in contrast.

2.2.2b DIAGNOSIS WITH MRI

The use of MRI as a practical method for the diagnosis of AD has also been extensively researched. A 2006 study by Fritzsche *et al.* utilized a 68 subject cohort consisting of 27 normal controls, 16 MCI, and 25 AD subjects. They performed an automated volumetric analysis of the cerebral spinal fluid distribution, and MRI images were then spatially oriented and normalized. They implemented an image segmentation algorithm to classify regions consisting of CSF, gray matter and white matter. They then used the distribution of the various tissue type counts determined from this algorithm as well as overall brain atrophy to create six features. These features were used in Fisher Linear Discriminant (FLD) analysis as well as a neural network for classification. Overall, they achieved classification accuracies of 80% for AD and 85% for controls. When testing between MCI and controls, they achieved a specificity and sensitivity of 80% and 81%, respectively; however when comparing MCI and AD patients, their overall classification accuracy dropped to 59%. While they showed that automated classification of AD and normal subjects is possible through image segmentation of MRI data, the distinction between MCI and AD through the same process was poor [12].

MRI imaging is not just limited to two-dimensional slices, converted into volumetric regions of interest. Torabi *et al.* (2006) used MRI images from a cohort of 50

control and 25 AD subjects. These images were separated into simple 60/40 training and testing datasets. The scans from each patient were preprocessed and fitted to the same overall brain image map, where texture feature processing was then computed. These features are created by considering 16 "landmark" pixels and then analyzing neighboring pixels and the similarity to the landmarks. Overall, this process generated 336 features, which were then reduced in size through principal component analysis (PCA). The resulting reduced dimensionality data was used to train and test a radial basis function neural network. Overall, they were able to distinguish between AD and control subjects with approximately 95% accuracy [93].

Longitudinal tracking of the course and progression of AD is commonplace in many studies, and MRI can provide meaningful diagnostic information for researchers in this respect. A study performed by Fox *et al.* in 2001 utilized a cohort of 20 control, 20 AD, and 4 patients that did not exhibit any symptoms but came from families with genetic risk factors for the development of early-onset AD. Subjects in the trial had MRI scans taken over the course of five to eight years, with the first used as a baseline for brain mass. They compared subsequent scans to the baseline with a non-linear fluid matching algorithm. Their results showed progressive atrophy in the parietal and medial temporal lobes of individuals at risk for early-onset AD. For patients with AD, they essentially verified all clinical assumptions about brain loss; all regions generally associated with AD showed significant atrophy over time. These results showed that using series of MRI along with other classic clinical trials through longitudinal studies can be used as an early detection method for AD—provided a baseline MRI is taken early enough in the progression of AD [94].

Alone, MRI has proven to be a viable diagnostic tool for AD; however, other readily available or easily attainable patient data could be used to complement the diagnosis from MRI. In 2006, a study by van der Hiele, *et al.* aimed to find whether cognitive function in normal and AD patients is represented better in MRI, EEG or a combination of both. A cohort of 33 total subjects was used, 10 AD, 11 MCI, and 12 controls. EEG data were collected from each patient while memory tasks were administered to generate ERPs and help evaluate cognitive ability. The MRI data collected from each patient was processed to quantize the frequency of gray/white matter and CSF within the brain. They performed several linear regression analyses between both sets of data. They determined that complementary information could be found between EEG and MRI data, with EEG measuring brain function and MRI determining overall brain atrophy. While each is associated with a different measure of cognitive decline in the progression of AD, the information in both sets supplemented the other to aid in a better distinction between the diagnostic classes in the groups [95].

2.2.3 POSITRON EMISSION TOMOGRAPHY (PET)

First described in the late 1950s by David E. Kuhl and Roy Edwards, this method of tomography measures the metabolic activity of cells in the body, producing a visual representation of the body's biochemistry. Their work led to further refinements in the imaging technique by Ter-Pogossian *et al.* (1975) paper submitted to *Radiology* [96]. Unlike MRI which displays tissue and fluid response to a strong magnetic field and radio frequency, PET shows the underlying chemical processes within the body.

To perform a PET scan, the patient is first injected with a radiopharmaceutical[8] or "radioactive tracer". After injection, the scan is delayed anywhere from a few seconds to a few minutes to allow the radio-isotope to be transported throughout the area under test. As the radio-isotope decays, it emits a positron which travels a small distance before annihilation with an electron. The annihilation of a positron by an electron emits two high-energy photons (approximately 511 keV) that propagate in almost opposite directions [99,100]. The photons emitted are gamma rays which can be detected by the scanning device that surrounds the patient. A computer then analyzes the collection of gamma rays to create a map of the area under test. The amount of radiopharmaceutical collected in the tissue reflects how brightly the tissue will appear on the computer generated image, indicating the level of tissue function. Unfortunately, the cost of PET systems is extremely prohibitive to be installed anywhere other than high-tech clinics or university hospitals and research centers. Also, the preparation of the various doses of radiopharmaceuticals is complex, and many of these begin to decay almost immediately, necessitating immediate scanning after injection. For purposes of this study, the imagery produced provides a representation of glucose activity in the brain, considered a metabolic biomarker for AD [100,101].

2.2.3a NORMALIZATION TECHNIQUES

In MRI, normalization is generally not as prudent as in PET imaging. Through the various types of MRI (T1, T2), contrast levels vary enough to provide distinguishable characteristics and pertinent information. However, in PET the use of normalization

[8] In higher doses, these "tracers" are used for cancer treatment. The radiation absorbed by the body in the small doses for PET is extremely low, and considered safe. The most common emitter used is Flourine-18 (^{18}F), with F18-FDG (Flurodeoxyglucose) used in brain scans for cognitive impairment studies [97,98].

becomes important to isolate regions of specific interest. A common technique for PET is to normalize with respect to the pons, as this provides a baseline for metabolic activity normalization on a per patient basis. Biologically, the pons is located on the brain stem, anterior to the cerebrum, as illustrated in Figure 13. In PET scans of the brain, the glucose levels of specific regions are determined; these representations of metabolic activity are indicators for AD and other neurodegenerative diseases [101]. When performing pons normalization, the glucose readings from the pons in the patient are used as a baseline for the rest of the brain. In this sense, a patient's glucose metabolic rates in the brain map relative to the pons, and allows for uniform visual look between patients.

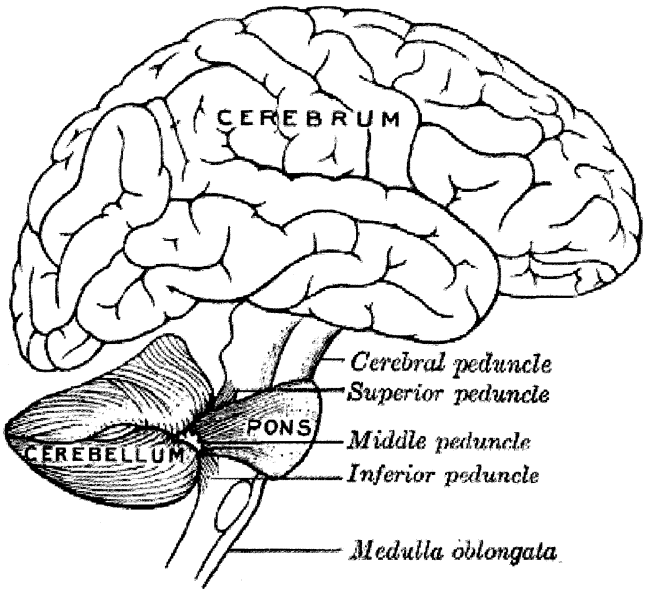

Figure 13 – Location of the pons for PET normalization [102]

Another method commonly used in PET image normalization involves the statistical process called z-score or standard score. The concept of z-scoring is simple: if given the mean and standard deviation of a population, it is possible to derive a relationship between a given score and the population [103]. This scoring method

indicates how many standard deviations an observation is above or below the mean. In a mathematical sense, the formula for z-scoring is as follows, with x representing the raw score, μ as the population mean, and σ as the standard deviation of the population:

$$z = \frac{x - \mu}{\sigma} \tag{2.1}$$

However, this method assumes knowledge of the entire population mean and standard deviation. In most cases, only a sample of the population data is available, making this method analogous to the t-test/t-statistic. For the use of PET normalization, it is common practice to use z-scoring to create statistical maps of hypometabolism, relative to elderly normal control subjects. Specifically, this requires the use of either one or an average of many control subjects as the mean and standard deviation parameters for this normalization technique.

The PET data specific to this thesis was processed through a program called Neurostat[9], which is a software library for neurological image processing and analysis. Developed by Satoshi Minoshima, the program performs baseline alignment, error detection/correction, as well as both normalization parameters described previously. This also allowed researchers to select specific ROI to normalize individual for greater accuracy in specific region metabolic activity visualization [103].

Figure 14, shows an example of a PET scan with the various stages of processing from subject with AD. The first row is a reference map, indicating the various regions of the brain of specific interest. The second row shows the elderly control glucose metabolism, averaged across 27 control subject PET scans. In the third row, the subject's

[9] Neurostat is freely available online through registration at the developer's website. 3D-SSP (used for display of processed images) is also available online for further analysis and research purposes [104].

individual glucose metabolism map is displayed, normalized to pons. The final row shows the result of z-scoring against the elderly control map, clearly accentuating the parietal lobe where hypometabolism is evident.

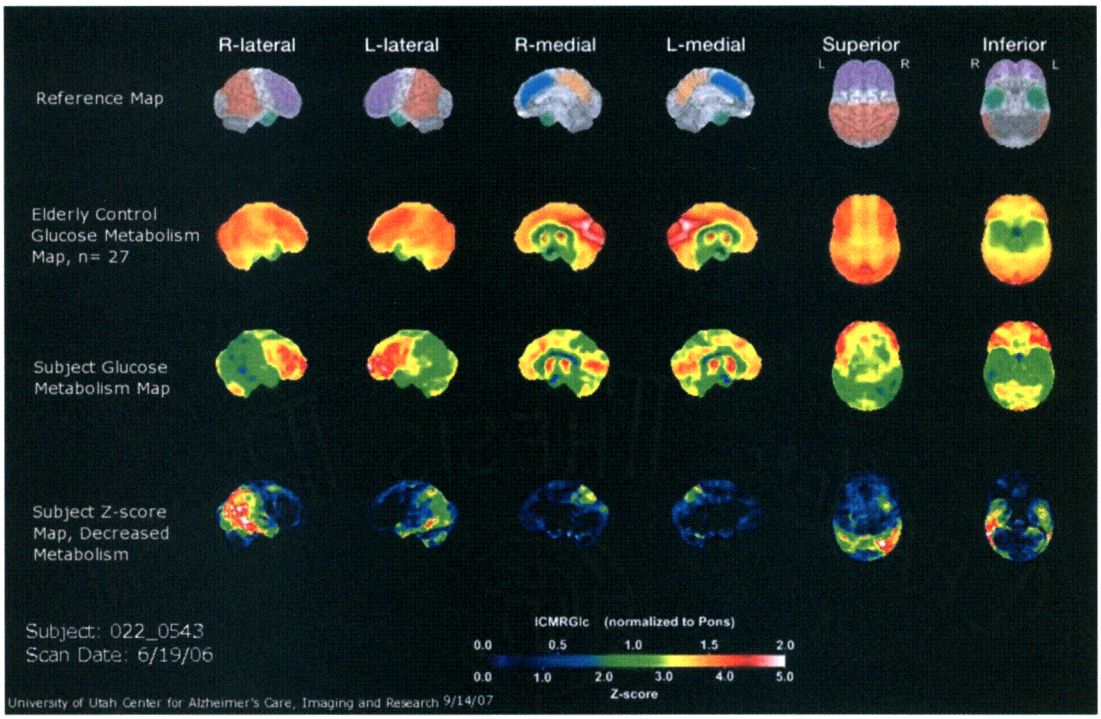

Figure 14 – Neurostat Image showing pons and Z-score normalization [103]

2.2.3b DIAGNOSIS WITH PET

Various manual (visual inspection) and automated (computer-aided classification) methods utilizing PET data for AD diagnosis have been developed and tested. In an early 1994 study performed by Salmon *et al.* differential diagnosis of AD with PET was attempted. Based on previous work that demonstrated hypometabolism in AD probable and possible patients, the study evaluated the diagnostic power of cerebral metabolic patterns for neurodegenerative diseases. PET scans were obtained from 129 subjects overall, 65 with probable AD, and the remainder of the cohort representing several other dementias. These subjects were referred for differential diagnosis, as a final decision had

not yet been determined. Using visual analysis of PET images from each patient, as well as prior clinical evaluations, the group determined that the use of PET as a diagnostic tool for AD is relevant. Of the 65 patients with a final clinical diagnosis of AD, they determined that 97% exhibited abnormal metabolic scans, with 94% showing a pattern of hypometabolism [11].

Kazunari *et al.* (2001) described a method for statistical brain mapping of ^{18}F-FDG PET in AD. Specifically, they investigated the use of statistical parametric mapping (SPM)[10] as well as the aforementioned Neurostat software. The PET cohort included 10 AD and 10 age-matched controls, where each subject underwent ^{18}F-FDG PET imaging. The resulting images were standardized using a stereotactic brain template, using either SPM or Neurostat. The data was normalized to the overall global metabolic activity. They performed within-group comparison of image sets by both the SPM and Neurostat methods, with between group comparisons of the AD and control using SPM statistical routines. They also created simulated PET image sets, generated from segmented MRI sets of 5 controls and 5 AD. Their results showed that while both SPM and Neurostat are viable methods for statistical routines in PET image processing, the SPM technique had 20% more gray matter mismatches—which resulted in an improper representation of metabolic activity. Furthermore, they concluded that the extent, severity, and location of metabolic changes in the brain from AD were inconsistent, regardless of the statistical mapping method employed. Specifically, deformation accuracies become an issue in heavily atrophied brains [105].

[10] SPM is a statistical technique for analyzing differences in brain activity captured through various imaging techniques (such as PET). This process focuses on voxel analysis, with each voxel of an image generally representing the activity of a specific region.

In a recent study by Foster, *et al* (2007), ^{18}F-FDG PET was used to improve the accuracy of diagnosis between AD and frontotemporal dementia (FTD), which can be confused in clinical diagnoses. The experimental cohort consisted of 45 patients, 31 of whom were pathologically confirmed with AD and 14 with FTD. Six dementia experts with experience in ^{18}F-FDG PET analysis made independent and forced choice diagnoses for each subject based on five separate methods: review of clinical information only, a diagnostic checklist, both the clinical information and checklist, transaxial ^{18}F-FDG PET scans, and ^{18}F-FDG PET stereotactic surface project (SSP) metabolic and statistical maps (an example of which previously shown in Figure 14). Their results proved promising for the use of ^{18}F-FDG PET in distinguishing AD and FTD. Through the visual interpretation of the 3D-SSP images alone, they were able to achieve an overall diagnostic accuracy of 89.6%, with specificity and sensitivity at 97.6% and 86%, respectively. The researchers noted that the use of the ^{18}F-FDG PET images was more helpful when the experts were not certain in the initial diagnosis. This study alone showed that visual interpretation of ^{18}F-FDG PET after simple training for doctors and clinicians is actually more reliable overall when distinguishing FTD and AD than clinical methods alone. Traditionally, these two diseases are often confused through clinical diagnosis, and ^{18}F-FDG PET gives additional information that increased diagnostic confidence, even with dementia experts [106].

CHAPTER III

METHODS

Much of the work presented within this thesis is an extension of many years of previous work and research accomplished by earlier studies. Section 3.1 details all previous work completed relevant to the current study, as well as how the earlier work contributed to this thesis. Sections 3.2 through 3.9 cover all background and methods specific to the most current study.

3.1 PREVIOUS WORK SPECIFIC TO THIS STUDY

The efforts of the earlier studies focused on only ERP EEG data collected from a previous cohort (Cohort A) 71 AD and cognitively normal subjects. This data was used to evaluate the feasibility of ERP based analysis for automated diagnosis of AD. The primary approach taken in the early stages essentially involved decomposition of the signal using the discrete wavelet transform of individual ERPs for each patient, which were then used to train and test a neural network classifier.

The first stage of the efforts prior to this thesis started with a cohort of 32 subjects. EEG signals obtained at the PZ electrode were processed through a wavelet feature extraction technique, using both the Daubechies 4 (db4) and quadratic B-spline wavelets to obtain the related signal coefficients. These coefficients were used to train an ensemble of classifiers system using the multi-layer perceptron as the base classifier to

distinguish between AD and normal control subjects. With this setup, a performance averaging around of 80% accuracy was obtained [107].

The next phase of the project expanded the size of the cohort to its final size of 71 subjects, 34 AD and 37 control normal. Once again, the ensemble approach was evaluated on the cohort, this time utilizing ERP data collected from the PZ, CZ, and FZ electrodes. These signals were subsequently decomposed through wavelet processing using the Daubechies 4 wavelet to obtain their respective coefficients. A single trial run of the algorithm utilizing the MLP as the base classifier yielded an accuracy of approximately 83.1%. A five trial average of the algorithm exhibited an accuracy of 79.2% [108].

The 71 patient cohort data contained ERPs from 19 overall electrodes; however only three had been used by the prior studies. The following study aimed to investigate additional electrode locations, with particular emphasis on the parietal region. Once again, the Daubechies 4 wavelet was used to decompose the ERPs into specific frequency bands. The three lowest bands were used as features for this analysis, generally considered to contain the most pertinent information for ERPs [64,108]. The ensemble based approach was expanded to allow further combination of classifiers; in this case, multiple classifiers were trained on different feature sets. These feature sets consisted of a specific set of wavelet coefficients relating to a given frequency band acquired from various electrodes. Furthermore, these feature sets were separated by the stimulus type that generated the ERP. The sum, product, weighted majority voting, and decision templates were utilized for decision level fusion to combine the various feature set classifiers. Classification accuracy reached 83.1% for the entire cohort [109].

In a later stage of the studies prior to this thesis, the investigation shifted using a different type of ensemble system. This investigation used the stacked generalization algorithm for AD/CN classification. Classification accuracy in distinguishing normal from AD reached 85.65% over averaged trials. Severity analysis was computed for this cohort, with the overall AD group being split into mild and moderate subjects based on their clinical diagnoses. Creating a three class problem (mild AD vs. moderate AD vs. normal) produced a classification accuracy of 71.34% [110].

In 2005, a new study started to investigate various biomarkers of AD, with a new cohort (Cohort B). In the initial stages of this second study 16-channel EEG data were recorded from 62 subjects, also based on the oddball paradigm for ERP generation. An ensemble of classifiers based classification algorithm was used. Once again, all channels from the raw EEG signals were processed through wavelet decomposition with the Daubechies 4 wavelet, and the three lowest frequency bands were used for analysis. Each band was analyzed individually in an attempt to obtain the most complimentary feature sets to use for classification to boost overall accuracy. The best mixture of classifiers were then combined using decision level fusion, once again utilizing the sum rule, product rule, and weighted majority voting with the addition of the Dempster Safer rule. The system was tested on both the old (Cohort A) and new (Cohort B) patient cohorts, yielding accuracies ranging from the mid 80% range to the low 90% with the optimal combinations [111].

The most recent phase of the study expanded the new cohort to 98, with 49 normal and 40 AD subjects overall. This data was used to reassess the work previously done on ERP-based diagnostic accuracy between AD and normal subjects. The cohort

was then expanded further with the addition of 39 Parkinson's disease and 39 MCI, bringing the total overall cohort to 161 subjects. Additional biomarker data was made available for each subject, including (but not limited to) EEG, volumetric MRI T2, and biochemical marker data. Similar to previous processes with multiple classifiers as experts, this method utilized the different biomarkers available as experts in an ensemble based classification algorithm. The decision level fusion in this setup was performed through simple majority voting, weighted majority voting, and the sum rule in a cross-validation setup. The classification performance of MRI and ERP combined was compared with that of cerebral spinal fluid (CSF)[11] analysis, the single most accurate diagnostic test currently available for the diagnosis of AD in a living patient. ERP alone classified with an accuracy of 80.61%, with MRI alone classifying 85.54% in AD vs. control only. When combined, the ERP+MRI fusion system achieved a diagnostic accuracy of 93.99%, exceeding the CSF biochemical analysis performance of 92.77% [112].

3.2 CURRENT RESEARCH

Present work now includes the compete cohort from the CERND research initiative (see section 3.3). In the two studies prior to this, a portion of the final subject set was used as data was acquired. With the recruitment process over, data from a total of 447 individual patients was collected, with EEG, MRI, PET, as well as various clinical evaluations and dementia scores. Specific to this study, EEG, MRI and PET data for AD and normal patients are analyzed for the first time, based on a modification to the stacked

[11] CSF analysis consists of a spinal tap and chemical evaluation of specific proteins (in the case of AD diagnosis both β-amyloid and hyperphosphorylated-τ are evident). Elevated levels of these proteins in the spinal fluid have proven to be reliable in AD diagnosis [112,113].

generalization algorithm. Each biomarker constitutes an overall ensemble system expert, with the experts comprised of augmented stacked generalization subsystems. The classification performances of EEG, MRI, as well as PET were all compared individually, and in all possible fusion combinations. Basic severity analysis was also performed, splitting the AD group into both mild and moderate groups, and computing performance metrics with the control group. For the purposes of this study, the cutoffs for the mild and moderate AD severity analysis groups were based solely on the respective MMSE scores of the subjects. Those AD patients with scores lower than 23 were categorized as the moderate AD group and the remaining subjects as mild AD. Figure 15 shows the basic flow diagram of information from data collection to final decision, based on the ERP, MRI, and PET classification experts.

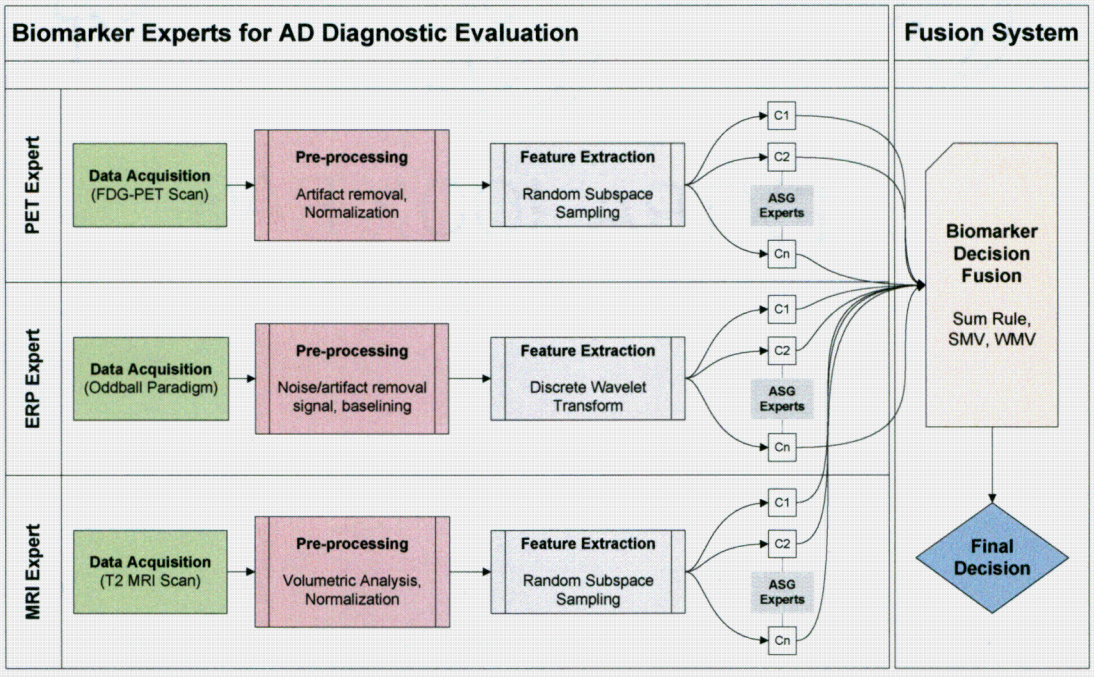

Figure 15 – Overall fusion system process diagram

3.3 CERND STUDY COHORT

The primary data used in this study (hereafter referred to as 'Cohort B') were collected at the University of Pennsylvania and Drexel University in the Center of Excellence for Research in Neurodegenerative Disease (CERND) study [114]. The cohort included 447 patients; the average age of the entire cohort was 72.02 (standard deviation 8.62) years. This cohort included patients diagnosed with AD, MCI, Parkinson's disease (PD) and Parkinson's disease with dementia (PDD), as well as 63 total normal controls. Each patient underwent a battery of neurological and physiological tests to obtain a diagnosis, as well as create a large set of data for future analysis. Some patients had as many as three visits, so multiple sets of data have been generated from one patient, increasing the overall usable cohort size. The CERND study generated vast amounts of data for each patient that can be used for automated diagnoses of each neurological disease. EEG, MRI, and PET are three of the datasets that are used in this thesis study. Proprietary image segmentation software was used to perform volumetric analysis of MRI data for each patient at the University of Pennsylvania prior to our use at Rowan University. The PET data presented was in the form of raw numerical quantization of various ROI in the brain. Both pons and z-score normalization was done at Rowan on the dataset provided. EEG signals were also provided in their raw form. Baseline correction, artifact rejection, filtering and signal averaging were all performed at Rowan.

Each patient had an accompanying MMSE, CDR, DSRS, and GDS scoring results. Each of these tests was designed to evaluate the overall cognitive and mental state of the patients in order to provide more information to aid in the final diagnosis. The MMSE was the simplest of the tests used to diagnose dementia and is scored from 0

to 30. Depending on the level of education for the individual, a score below 19-20 indicates cognitive impairment. The CDR was scored incrementally as 0, 0.5, 1, 2, or 3, with zero indicating normal cognition, and three indicating severe dementia. The DSRS was administered with multiple-choice questions that incorporate both cognition and daily living tasks. Higher scores indicate higher levels of dementia. The GDS was a stage-based scale used for dementia. Higher scores in the GDS indicate higher levels of dementia and overall poor cognitive ability. Furthermore, a numerical value is assigned to each patient regarding their level of education, which had an impact on their expected abilities on several dementia tests. This number indicated years of formal education attained.

3.3.1 INCLUSION CRITERIA

There were several key factors that determined the overall inclusion criteria for the CERND study. Listed below in Table 5 are the inclusion and exclusion criteria for normal controls as well as AD/MCI patients relevant to this specific study.

Table 5 – Inclusion/Exclusion Criteria for CN, AD, and MCI Patients [114,115]

	Cognitively Normal	**AD/MCI**
Inclusion	• Age > 55 • CDR = 0 • MMSE > 26 • No indication of functional or cognitive decline for two years prior to enrollment	• Age > 55 • CDR ≥ 0.5 • MMSE ≤ 26 • Presence of functional and cognitive decline of the previous 12 months • Satisfaction of National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association Criteria for probable AD
Exclusion	• Evidence of any central nervous system neurological disease (e.g. stroke, multiple sclerosis, PD, etc) by history or exam • Use of sedative, anxiolytic or anti-depressant medications 48 hours prior to evaluation, testing or data collection	

Table 6 lists relevant statistical information for all patients in Cohort B, along with their respective diagnostic class. The respective subject count per in each category, average ages and MMSE scores are provided, along with relevant standard deviations. The second half of the table documents the average DSRS scores along with the standard deviation for each class. CDR scores were computed, with the overall average and standard deviation displayed, as well as the group median. Since the CDR score scales only as 0, 0.5, 1, 2 or 3 the median was displayed for the best representation of the group. A similar process was done for the level of education attained for each group, with the mean, standard deviation and median provided. Education level is indicated in the table by the total years of education attained.

Table 6 – Cohort B Statistics (entire 447 patients set, all classes)

	# of Patients	Age	Std. Dev.	MMSE (avg)	Std. Dev.
Normal	63	70.10	9.93	28.43	2.34
AD	121	75.00	8.67	18.86	5.75
MCI	111	70.88	8.99	24.88	3.84
PD	126	70.31	6.70	27.50	2.47
PDD	26	75.96	7.12	21.79	3.70

	DSRS (avg)	Std. Dev	CDR (avg)	Std. Dev.	Median
Normal	1.27	1.97	0.07	0.17	0
AD	16.27	8.39	1.11	0.58	1
MCI	8.01	5.77	0.53	0.18	0.5
PD	4.89	5.04	0.26	0.30	0
PDD	18.08	8.73	1.31	0.92	1

	GDS (avg)	Std. Dev	Education	Std. Dev.	Median
Normal	1.66	2.58	14.48	4.93	16
AD	2.17	2.65	12.41	4.45	12
MCI	3.21	3.21	13.09	5.26	14
PD	1.77	2.31	--	--	--
PDD	2.42	2.90	15.23	3.02	16

3.3.2 *COHORTS SPECIFIC TO THIS STUDY*

Many datasets were generated for this study, each of which contained different information and patients. Furthermore, many patients had multiple visits, so any one patient could have three unique EEG trials, increasing the total number of effective subjects. Henceforth, all usable trials (multiple visits from each patient or not) will be referred to as *subjects*, with *patients* referring to individual patients.

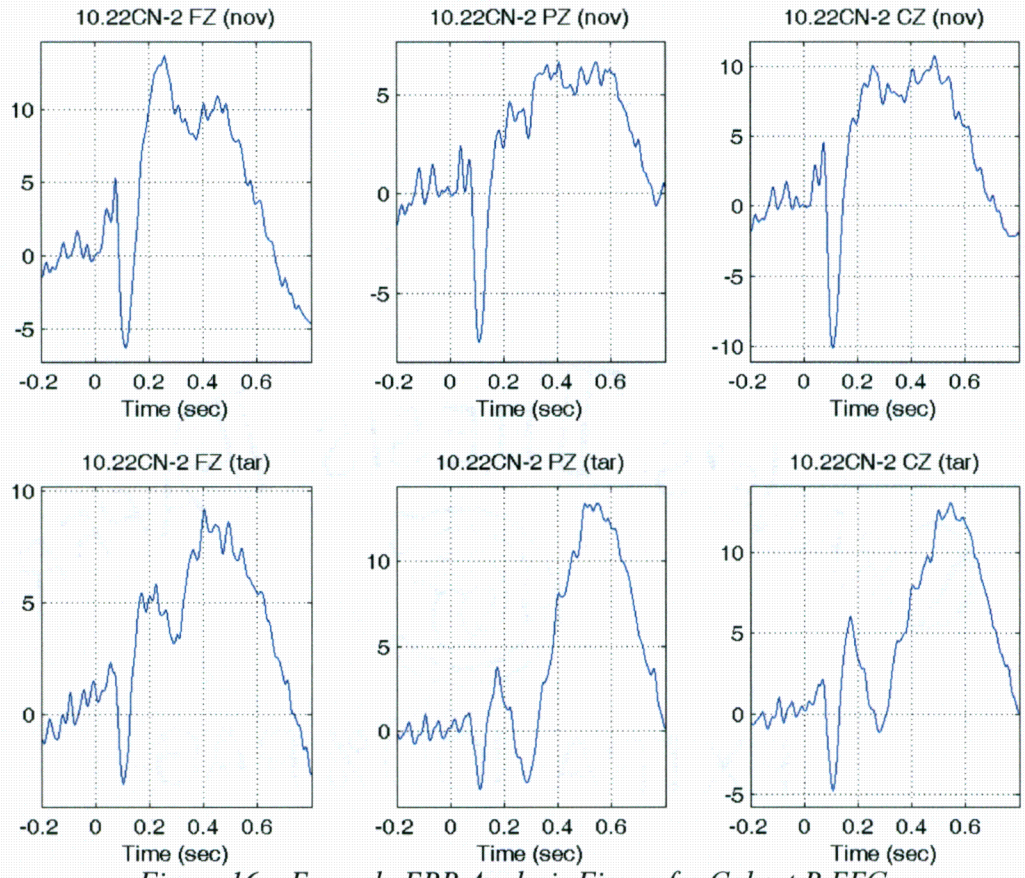

Figure 16 – Example ERP Analysis Figure for Cohort B EEG

Multiple datasets for EEG were generated, but only three were primarily used for standalone EEG analysis. Each of these sets was generated based on visual inspection of subject ERPs generated from the FZ, PZ and CZ electrodes for both novel and target

stimuli. Standard stimuli were not utilized for any part of this research project, as the most notable ERP generation in the oddball task comes from target and novel stimuli. Figure 16 is an example from this analysis for the second visit of patient 10.22, which was determined to be satisfactory for both novel and target ERP generation. This analysis was subjective, and was determined by looking for signals with low overall noise and distortion, and strong peaks and troughs. In Figure 16, a strong N100 and P300 can be seen, indicative of a desirable P300 response.

3.3.2a EEG COHORTS

Satisfactory EEG Dataset: The following sub-cohort was generated from the overall EEG dataset of 491 available subjects (all classes, CN/AD/MCI/PD/PDD) to create a 107 subject cohort of AD and normal controls. The subjects were selected based on visual ERP analysis where all ERPs appeared clean and satisfactory overall. Both target and novel responses for the subject had to pass this test, otherwise all data for that subject was excluded from this subset. Table 7 details the breakdown of the sub-cohort information, specifically the total number of males/females, average age as well as the respective average MMSE scores.

Table 7 – Satisfactory EEG Dataset Statistics

	Male	Female	Total	Avg. Age	Avg. MMSE
AD	21	33	54	76.44	20.60
CN	13	40	53	70.28	28.94
Totals	34	73	**107**		

Satisfactory + Marginal EEG Dataset: This dataset incorporated all the previous subjects in the 'satisfactory only' set and added those subjects with ERPs considered "marginal." These marginal subjects were determined again through visual inspection—if only one

stimuli (target or novel) was deemed marginal and other satisfactory, it still fell into this sub-cohort. This cohort was generated to encompass a larger swath of the available subjects to determine if these marginal ERPs in addition to the satisfactory ones would cause a drop in diagnostic accuracy.

Table 8 – Satisfactory + Marginal EEG Dataset Statistics

	Male	Female	Total	Avg. Age	Avg. MMSE
AD	39	46	85	76.02	19.20
CN	24	55	79	72.16	29.05
Totals	63	101	**164**		

First Visit Satisfactory + Marginal EEG Dataset: This dataset is similar to the previous, however only first visit subjects were included. Essentially, this signifies that only data from satisfactory and marginal subjects that was generated from the first visit diagnosis was used. While this reduces the size of the cohort, it was created to determine what effect, if any, a dataset including only first visit patients would have on the ensemble system performance. Table 9 gives the statistical breakdown for this sub-cohort.

Table 9 – First Visit Satisfactory + Marginal EEG Dataset Statistics

	Male	Female	Total	Avg. Age	Avg. MMSE
AD	22	35	57	76.39	20.49
CN	13	31	44	70.19	28.79
Totals	35	66	**101**		

3.3.2b MRI COHORT

The cohort used for MRI analysis is a subset of the overall 447 patient cohort for this study. All AD/CN patients with respective MRI data was used to generate this dataset. Similar to the EEG procedures, each unique trial is considered a subject, with each patient possibly having multiple trials. This cohort was used for MRI-only ensemble analysis. Table 10 details the relevant statistics for this dataset.

Table 10 – MRI Dataset Statistics

	Male	Female	Total	Avg. Age	Avg. MMSE
AD	30	49	79	75.02	18.79
CN	15	75	57	71.01	28.96
Totals	45	91	**136**		

3.3.2c PET COHORT

Similar to the MRI dataset, the PET cohort is a subset of the overall 447 patient cohort used in this study. Once again, only AD/CN patients were used, some with multiple trials giving more usable data. This dataset was used for PET-only ensemble analysis. Table 11 gives the relevant subject statistics for this dataset.

Table 11 – PET Dataset Statistics

	Male	Female	Total	Avg. Age	Avg. MMSE
AD	9	28	37	74.15	18.98
CN	14	29	43	70.79	28.88
Totals	23	57	**80**		

3.3.3 ERP ACQUISITION

All ERPs were generated and acquired through the auditory oddball paradigm. Each patient was seated and made comfortable in a relatively quiet and distraction-free room while wearing headphones. Before data collection occurred, a 1 kHz tone was presented to the subject in order to adjust the volume in accordance to the patient's minimum hearing threshold.

For each patient, the stimulus was provided at 60dB above their hearing threshold, in order to remove any bias between patients and their respective hearing levels. Both standard and target tones were presented for 100 ms each, with standard tones playing at 1 kHz and target (oddball) at 2 kHz. Novel environmental sounds were at least 200 ms in

duration, and were not repeated throughout the duration of the recording session. In each session, a total of 1000 stimuli were presented, with standard, target, and novel tones played 65%, 20%, and 15% of the total time respectively. Patients were instructed to press a button upon hearing the target tone, and were instructed to not respond for any other stimuli. Each stimulus and the associated EEG data are considered one epoch, with many epochs recorded in one session. For this study, the inter-stimulus interval (ISI) varied between 1.0 to 1.3 seconds randomly. For most patients, the recording sessions ran for 30 minutes which included of three minutes of rest every five minutes.

Relevant EEG data were collected from 16 tin electrodes applied directly to the scalp of the patient, along with two reference electrodes located on the mastoids (protrusion of the temporal bone behind the ear). All recording was continuous—that is, recording was not stopped during inter-stimulus times. Technicians administering the recording sessions ensured impedances for all electrodes did not exceed 20 kΩ, through the use of electrically conductive sterile gel applied between the electrode and the patient scalp. As hair and dry skin can contribute to poor impedance for EEG recording, each patient was carefully connected to the recording setup to minimize noise, distortion, and other unwanted artifacts. All signals were amplified and sampled at 256Hz per electrode channel. This constituted the raw data presented to us for this study.

The processing technique employed in this study involves low pass filtering, artifact rejection based on an automated derivative analysis process, separation into stimulus type, and subsequent epoch averaging to obtain averaged ERPs for each patient, separated into stimulus type. Signals were first filtered with a 20^{th} order FIR low pass filter, with the passband and stopband set at 20Hz and 40Hz respectively. Upon

completion of low pass filtering, the raw EEG signals for each patient were separated based on stimulus type. Analysis was conducted 200ms prior to the stimulus and recorded for another 800ms post-stimulus, creating a 1 second window for ERP analysis. A 20th order derivative was then used for artifact rejection. In this setup, any one second segment (in this case, any individual epoch) whose derivative was above a threshold was removed from the averaging process. Equation 3.1 describes this approach.

$$f(x) = \begin{cases} f(x) & \text{where } f(x+20) - f(x) < threshold \\ \text{remove} & \text{where } f(x+20) - f(x) \geq \text{threshold} \end{cases} \tag{3.1}$$

Upon removal of the epochs deemed unwanted artifacts, the remaining segments were averaged for each stimulus type and patient. This provided three overall ERPs for every subject, and patients with multiple visits for EEG recordings were included.

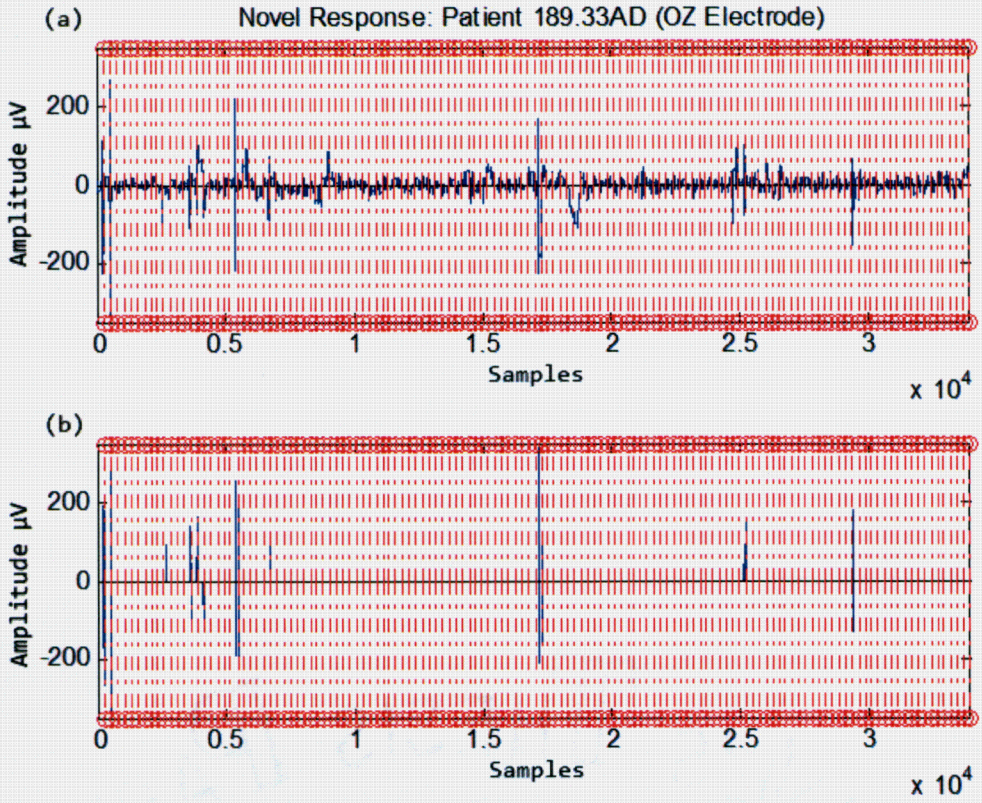

Figure 17 – Visual representation of EEG artifact removal process

The vertical red lines in Figure 17 indicate the one-second epochs for each stimulus. Any epoch that exceeds a threshold based on this derivative analysis was removed prior to averaging. Figure 17a shows the raw EEG signal with artifacts; Figure 17b indicates the location of artifacts identified, which were subsequently removed. All remaining epochs were then averaged to create the overall response at a given electrode, for a given stimulus, for each subject. Therefore, each subject had 3 overall stimuli ERPs, at 16 electrodes. The data length of the digitized vector for each stimulus, and each electrode was 257; as a result, the overall matrix for a given subject was 257×16.

3.3.4 EEG FEATURE EXTRACTION

In the volumetrically processed MRI and normalized glucose region-based analysis with PET data, single numerical values were given for each feature. No other information could truly be gleaned from such data—only varying normalization techniques could be performed. In EEG data however, a vast and generally hidden amount of information lies within these signals, pertaining to overall cognitive ability. Unfortunately, when these signals are collected at a scalp electrode, they are essentially a combination of information originating from various other sources within the brain. While generally not visible in the raw EEG signal, various frequency bands within the original can be used to identify specific traits and characteristics, as described previously in Section 2.2.1a.

Isolation of these bands can be achieved in several ways. Techniques to separate these information sources in the frequency domain allow for a visualization of the frequency content of the signals . The most common method to achieve basic frequency domain representation is the Fourier transform (FT). The continuous Fourier transform

(CFT) is shown in equation 3.2, with the inverse CFT shown in equation in 3.3. $X(f)$ constitutes the frequency domain representation of the time domain signal, $x(t)$.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-2j\pi ft}\,dt \tag{3.2}$$

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{2j\pi ft}\,dt \tag{3.3}$$

The primary disadvantage to using the Fourier transform is the loss of time resolution. Any signal passed through this transform will have its frequency components identified; though the points at which these frequencies occurred are lost. Such information can be extremely important in time-varying signals, such as the EEG. While the time information remains in the phase response of the FT, this is not generally useful or relatable in analysis. Hence, the Fourier transform is not an effective approach to frequency domain analysis of time-varying signals such as the EEG.

One modification to the FT, the short-time Fourier transform (STFT) uses a windowing approach to overcome the temporal limitation of the FT. This provides a time-frequency representation of a signal in various time windows. In this approach, a window is used to segment portions of a time-varying signal, which are then processed sequentially with the FT to determine the frequency content of these segments .

Equation 3.4 details the STFT, with $x(t)$ as the original time-varying signal, $w(t)$ as the windowing function, and τ is the translation of the window. The fixed width of the windowing function $w(t)$ in the STFT determines the tradeoff between time and frequency resolution. The Heisenberg Uncertainty Principle, generally applied to the uncertainty in measuring the momentum and location of particles, can also be related to

signal analysis. Essentially, the same axiom holds true for time-frequency analysis: exact frequency and exact time cannot be known simultaneously, thus limiting the trade-off in resolution .

$$\mathrm{STFT}_x^{\omega}(\tau, f) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j2\pi t}\,dt \tag{3.4}$$

3.4 THE WAVELET TRANSFORM

The tradeoff between time and frequency resolution in the STFT analysis becomes an issue in non-stationary signals such as the EEG. When the window size is fixed, certain frequency bands can be lost in the processing. The Wavelet Transform (WT) addresses this issue by varying the window size based on the frequency band being prolonged. Several versions of the WT exist, such as the Continuous Wavelet Transform (CWT), a discretized version of the CWT called the wavelet series, and the Discrete Wavelet Transform (DWT) [116]. The DWT is used for the isolation of various frequency bands in the EEG data used in this study for feature extraction purposes.

3.4.1 CONTINUOUS WAVELET TRANSFORM

In CWT, the time-frequency representation of a signal is calculated using basis (windowing) functions. The supports for these functions vary, based on scale. Instead of implementing the FT within each window, the CWT utilizes a correlation metric with the wavelet function. The width varies based on the scale (or frequency band) in question. Mathematically, the CWT is described in Equation 3.5, where $x(t)$ represents the original time-varying signal, $*$ means complex conjugation, τ and s represent the translation and

scale variables for the mother wavelet $\psi(t)$. The mother wavelet is translated and scaled (according to τ and s) as shown in Equation 3.6

$$\Psi_x^{\psi}(\tau, s) = \int x(t)\psi^*(t)dt \tag{3.5}$$

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{3.6}$$

The Fourier transform uses complex exponentials and fixed basis functions. In wavelet transform, any wavelet function that abides by the following constraints can be used as the basis function. The wavelet must be a wave with an area of zero (Equation 3.7). Furthermore, these functions must also be of finite duration and finite energy (Equation 3.8).

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{3.7}$$

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \tag{3.8}$$

In order to calculate the CWT, the mother wavelet's scale is kept constant and translated across the signal. The scale is then changed, and the translation is then repeated. This process of compression and expansion is repeated for all desired values of scale and translation. Unfortunately, the CWT uses a continuously scalable wavelet function—that is, it is not practical for most real world applications reliant on discrete signal processing [117,118]. Instead of using an infinite number of scaling and translation variables, the Discrete Wavelet Transform (DWT) steps these variables in discrete amounts.

3.4.2 DISCRETE WAVELET TRANSFORM

CWT is not a discrete transform (e.g. it is represented by an infinite integration for the scaling and translation parameters), and therefore cannot be implemented on a computer. Furthermore, the CWT gives redundant information in the signal after transformation, translating to a high memory requirement. The DWT solves both these issues by providing a discrete method to transform a signal without the unwanted redundancy for frequency decomposition and reconstruction of the original signal. Because of this, it is easier to digitally implement, and also is less computationally and memory expensive in comparison to the CWT. The entire basis of the DWT comes from the concepts of multiresolution analysis and subband coding [119].

3.4.2a MULTIRESOLUTION ANALYSIS

Multiresolution analysis (MRA) allows a signal to be analyzed at various resolutions, or levels of approximation. This process essentially allows any complex function (signal) to be divided into multiple smaller and simpler functions (frequency bands). Time resolution is possible with this technique as well, allowing for the time localization of spectral components. MRA also provides the ability to dynamically change the sampling rate at each processing level, allowing for a significant reduction in the overall signal length at each of these levels.

The raw signal is sampled with each consecutive level, at half the prior frequency. This process removed redundant information, as signal coefficients that pertain specifically to each band are only retained, thus eliminating overhead. This procedure is repeated for a given number of iterations (levels), which results in the approximation of the signal. This approximation accounts for the low frequency representation of the

original signal. The information removed between samplings corresponds to the high frequency detail coefficients for each level. At each of these levels, the approximation coefficients are added to the approximation subspace (A_s) with the detail information added to the Wavelet subspace (W_s). W_s is complimentary with A_s; therefore, when the detail coefficients from W_s are combined with the approximation information from A_s, it forms A_{s+1}, which represents the high frequency (level) approximation. This process is shown in equation 3.9.

$$A_s \oplus W_s = A_{s+1} \tag{3.9}$$

In Equation 3.9, the symbol $\oplus$ represents a direct summation. These subspaces are directly related to the approximation and detail coefficients of the signal, where each of which can be related to $x(t)$ and $y(t)$ such that $x_s(t) \in A_s$ and $y_s(t) \in W_s$. Equations 3.10 and 3.11 indicate the calculations used to obtain $x_s(t)$ and $y_s(t)$, respectively [120].

$$x_s(t) = \sum_k a_{k,s}\phi(2^s t - k) \tag{3.10}$$

$$y_s(t) = \sum_k w_{k,s}\psi(2^2 t - k) \tag{3.11}$$

3.4.2b SUBBAND CODING

In the most general sense, subband coding is a method that splits a signal into various frequency bands, with encoding performed on each independently. This type of decomposition is commonly used in the first step of compression for digital audio and video signals. A common codec used for compression that implements subband coding is MPEG audio [121]. The discrete wavelet transform implements MRA through subband coding, utilizing digital filters. Specifically, quadrature mirror filters (QMF) are used,

which are half band lowpass and half-band highpass filters. These filters are used to create a filter bank that splits the input signal into two bands, resulting from the high and lowpass filters. The resulting signals are then decimated by 2, yielding a critically sampled two-channel representation of the original signal. Equation 3.18 shows the general form of the QMF, where Ω is the frequency and the sampling rate is normalized to 2π.

$$\left|H_0(e^{j\Omega})\right|^2 + \left|H_1(e^{j\Omega})\right|^2 = 1 \tag{3.18}$$

Essentially, equation 3.18 states that the power sum of the highpass and lowpass filters equals one, with their responses symmetric about $\Omega = \pi/2$. In terms of the DWT, the wavelet being implemented determines the parameters of the lowpass filter for decomposition, given by the impulse response $h[n]$. To generate the mirrored highpass version of the filter, a conversion is performed on the impulse response of the LPF. Equation 3.19 show this process, where L is the length of the filter, n as samples, and $g[n]$ representing the highpass filter.

$$g[L-1-n] = (-1)^n h[n] \tag{3.19}$$

The actual decomposition of the signal occurs when the original signal is filtered by the lowpass filter $h[n]$ or the highpass filter $g[n]$. Simple convolution is performed for the filtering process, done in discrete time. The lowpass filtering method for this is shown in Equation 3.20 where $*$ represents convolution.

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k] \tag{3.20}$$

Upon completion of filtering, the two signals now hold the approximation and detail coefficients of the original signal. For arguments sake, assume that π is the highest frequency in the original signal. The signal is therefore sampled at 2π to adhere to the Nyquist criterion; however, post-filtering, the highest frequency in the signals is reduced to $\pi/2$, so the signal need only be sampled at π. This process effectively downsamples by 2, denoted by $2\downarrow$, and removes the redundant information at the same time.

After this first step, there are two signals—one from the lowpass filtering and the other from the highpass filter. The signal from the lowpass filter gives the approximation of the original signal, with a frequency band of 0 to $\pi/2$. Conversely, the signal from the highpass filter yields a frequency band of $\pi/2$ to π. As stated previously, both the signals are downsampled by 2. The overall filtering and downsampling process for both the high and lowpass filters is shown in Equations 3.21 and 3.22, respectively.

$$y_{\text{high}} = \sum_{n} x[n] \cdot g[2k - n] \tag{3.21}$$

$$y_{\text{low}} = \sum_{n} x[n] \cdot h[2k - n] \tag{3.22}$$

Equations 3.18 to 3.22, done for the first iteration, constitute the first level of approximation and detail coefficients for the signal. The detail coefficients are then set aside, with the remaining approximation coefficients used for the next iteration of filtering and downsampling. This process is repeated until the final level, or frequency band, is reached. The detail coefficients from all levels constitute the DWT of the original signal. With each successive filtering and downsampling, the time resolution is reduced by half. Because of this, the halved bandwidth causes the frequency resolution to double.

Figure 18 shows an illustration of this process, with $g[n]$ as a HPF and $h[n]$ as the LPF defined by the wavelet being used. This process is illustrated in Figure 18.

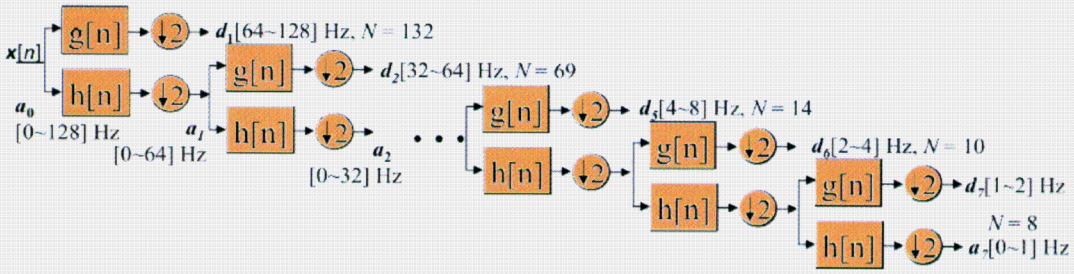

Figure 18 – Seven-level DWT decomposition example [122]

This shows the process for the decomposition of an ERP signal with 128 Hz max frequency content.

3.4.2c SIGNAL RECONSTRUCTION

As with any decomposition and frequency analysis technique, it is necessary for a fully reversible procedure that allows for reconstruction of the original signal. Wavelet decomposition meets this criterion, allowing for perfect reconstruction of the original signal provided all levels of approximation and detail coefficients are available [118]. An estimate of the signal at any level can be obtained by summing the detail and approximation coefficients for that level. This process is outlined in Equation 3.23, where s represents the approximation level, a and w are the approximation and detail signals, respectively, with ϕ and ψ as the scale and wavelet functions.

$$x_s(t) + y_s(t) = \sum_k a_{k,s}\phi_{k,s}(t) + \sum_s w_{k,s,}\psi_{k,s,}(t) = x_{s-1}(t) \tag{3.23}$$

A similar process is followed in the discrete case. The high and low frequency reconstruction filters can be obtained from the original decomposition filters used in the initial process. Similar to Equation 3.23, a summation process for the discrete signal portions is used for reconstruction. Equation 3.24 illustrates this process, which is repeated at each level and summed to obtain the original signal.

$$x[n] = \sum_{k=-\infty}^{\infty} y_{\text{high}}[k] \cdot g[-n + 2k] + y_{\text{low}}[k] \cdot h[-n + 2k] \tag{3.24}$$

3.4.2d DAUBECHIES WAVELET

There are several wavelet functions available, each with varying characteristics and generally chosen to fit the specific situation in which it is being implemented. Specifically, the wavelet must be chosen to match the frequency characteristics of the signal in question. The Daubechies wavelet allows for perfect reconstruction of the

signal and was chosen for use in this study. In particular, the Daubechies wavelet with 4 vanishing moments (db4) was implemented due to its optimal frequency similarities with the ERP. The smoothness of the wavelet function adds to its optimality for this situation. Since the db4 is only an 8-point filter, it is ideal for use in the 256 point ERP signals for this study. The Wavelet and scaling functions, along with the decomposition and reconstruction filters are shown in Figure 19.

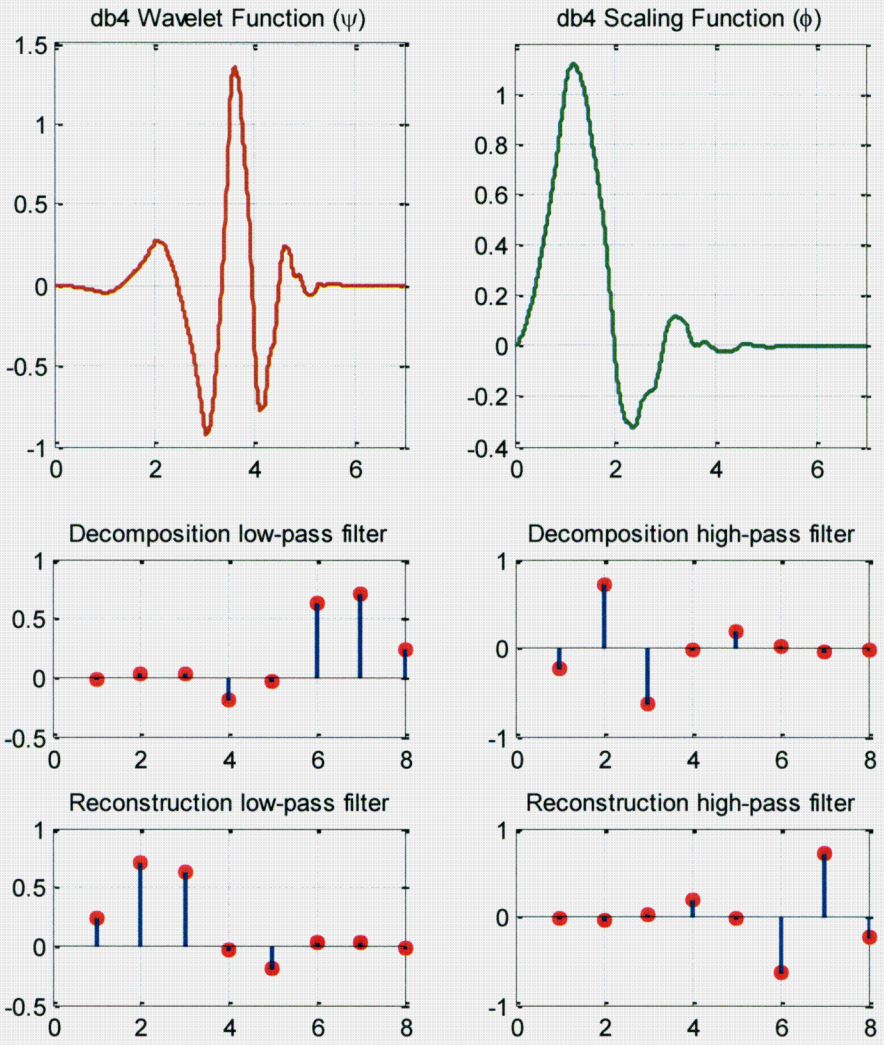

Figure 19 – db4 Wavelet, Scaling function, decomposition/reconstruction

Figure 20 shows a full wavelet decomposition of a sample ERP signal, taken from an AD patient. The top plot denotes the original signal (red) followed by the 6th level approximation coefficients (blue). The remaining plots show the six levels of detail coefficients for the signal (green). The approximation and following decomposition layers each represent a different frequency band of interest within the ERP.

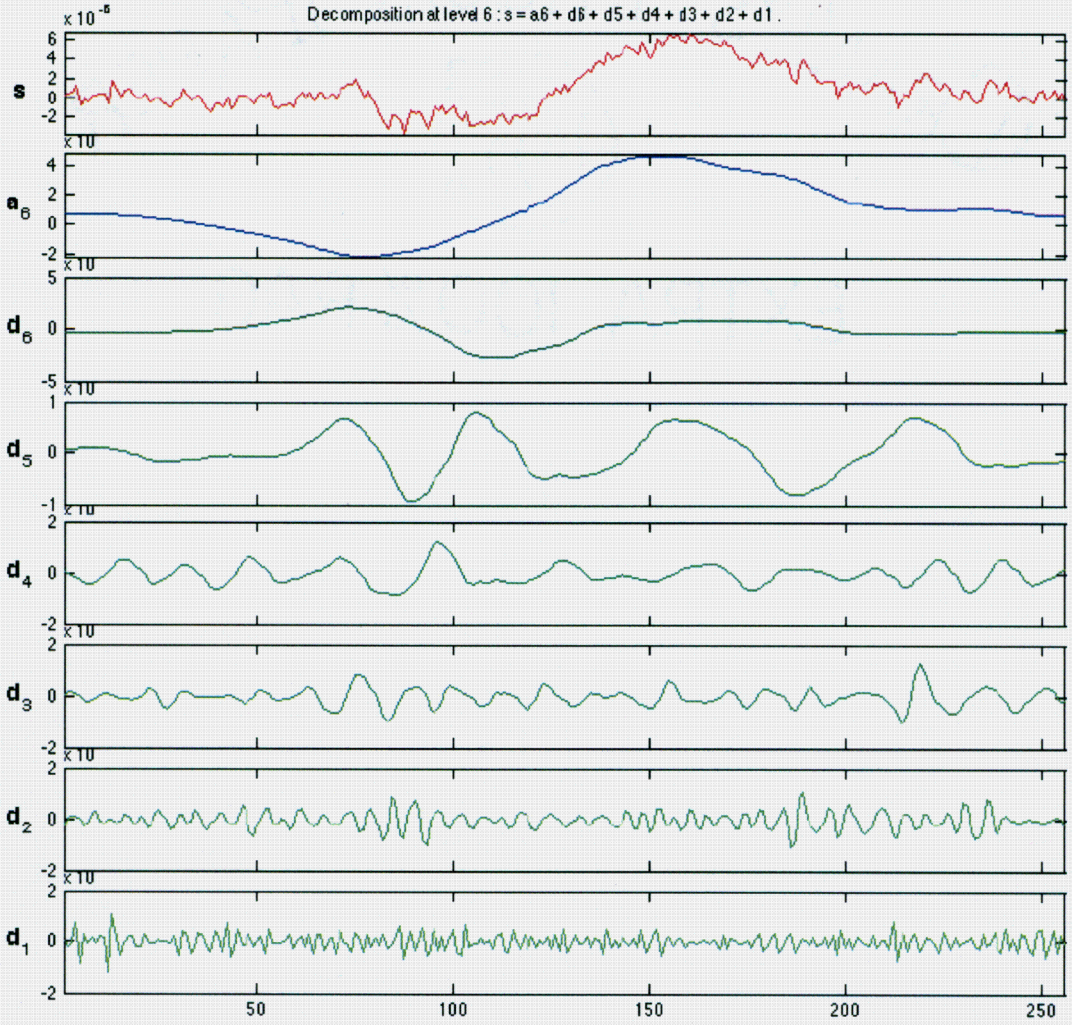

Figure 20 – Wavelet decomposition of an example ERP signal [112]

3.5 RANDOM SUBSPACE METHOD

In the realm of pattern recognition, it is generally accepted that "more is better" with respect to the overall ability to adequately train a system. A growing number of situations incorporate vast amounts of data, with as many features possible collected. However, this "data overload" poses a real challenge to traditional pattern recognition techniques. Specifically, the number of samples for any given dataset is generally still limited with respect to the number of features available. The most classic of pattern recognition techniques will perform poorly on small sample size, large feature count datasets; therefore a more meaningful method of making the best use of data is necessary.

Classically, the more features for a given problem, the better—as there is inherently more information available for use. However, the addition of "too many" features will lead to a higher probability of error, known as the curse of dimensionality [123]. This phenomenon is inherently caused by the limited number of samples available for any given dataset. Even with a large set of features, the system is still limited by the total available samples for each feature [124]. In relation to this thesis for example, the MRI T2 dataset contains 29 overall features, with each feature represented by a numerical value indicting the volume of the specific region of interest. The PET dataset is even larger, consisting of 43 features with a single numeric representation for each region of interest as well. Furthermore, we wish to generate several "feature sets" from the available data, such that these feature sets are different from one another and contain different combinations of all usable features. This process will be further explained in section 3.6. Be it through supervised removal of features deemed "noisy" or irrelevant, or through some automated, unsupervised process, all feature selection strategies all share

a common goal of reducing the overall dimensionality for a given problem. First introduced in 1998 by Tin Kam Ho, a random subspace method (RSM) was developed to reduce a problem's overall dimensionality. Specifically, Ho focused on avoiding overfitting on a set of training data while attaining high accuracy with decision trees.

A pseudorandom process for selecting subsets of components in a dataset's feature space to be used in a decision tree method for classification was implemented. Specifically, this process is determinant on various parameters. The more important parameter to choose is the number of feature to be selected in the splitting process. These sampling or splitting processes are determined via various functions and thresholds, which are detailed in [125]. Ho showed significant classification accuracy improvements with this method when compared to utilizing all available features for training the decision tree. Furthermore, it was shown that was more features were introduced to the problem, the classification accuracy improved when using this method. Conversely, when a smaller amount of features were present, RSM did not show a statistically significant advantage over using all available features [125].

This method was expanded and tested for use in this thesis. Specifically, the MRI T2 and PET datasets were analyzed as candidates for this algorithm. The random subspace sampling method was employed to aid in feature set diversity, such that each set trained for a respective biomarker would be "diverse enough" to be used adequately in an ensemble system (see section 3.6 for detailed explanation of the ensemble systems employed in this study for each biomarker). By using RSM, the overall dimensionality of the problem is reduced and more diverse feature sets are generated that can be used in later ensemble system evaluation [124].

3.6 DATA PROCESSING TECHNIQUES

The algorithm developed in this thesis utilizes features extracted from the EEG, MRI, and PET data for all subjects available in the cohort. As discussed previously, the EEG data features are the wavelet coefficients of specific frequency bands of interest, ranging from 0Hz to 8Hz. For MRI, the raw data consists of topographic images of the brain, taken in series perpendicular to the coronal and parallel to the transverse planes. Since the overall cranium size of patients can vary, an elastic warping algorithm was implemented to transform images into a standardized space, while preserving the overall morphological characteristics of the brain.

The raw MRI image was then segmented into white matter, gray matter, CSF, and ventricles. An overall density map of the image was calculated, and the total inter-cranial volume (ICV) for the patient was determined. Through these density maps, conclusions were drawn regarding the overall atrophy of various segments of the brain. The overall density map was split into various overlapping portions, based on an automated region of interest (ROI) analysis. This removed cranium and bone regions visible in the MRI, leaving only volumetric representations of specific regions in the brain. This process was repeated for every image slice, with each consecutive slice compiled and merged. Voxel analysis is performed to quantize the volume of each ROI [112]. The overall process used for this analysis shown in Figure 21.

With this voxel analysis complete, a total of 14 unique, anatomically defined regions were created in each hemisphere of the brain, giving a total of 28 features for use in the MRI data. Different regions of interest were then analyzed across different patient

groups and classifications. Each patient had an inter-cranial volume value associated, which allowed for normalization between patients.

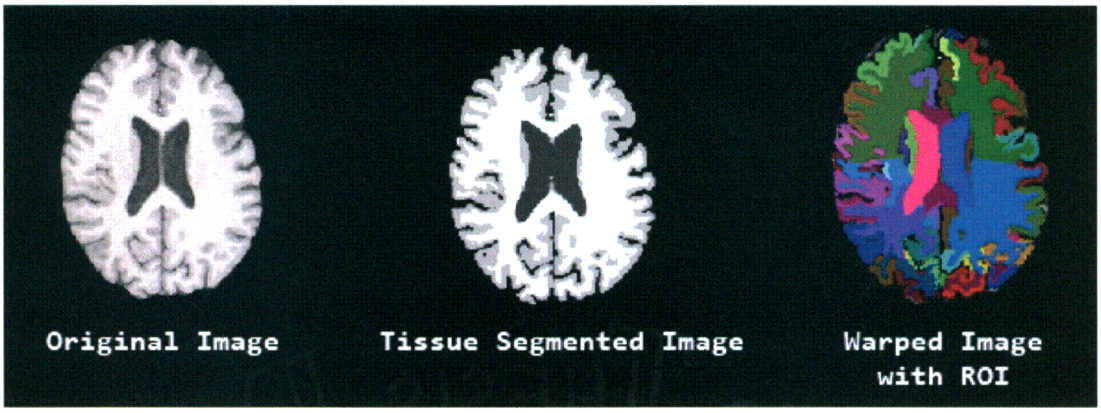

Figure 21 – Sequence of image processing for T2 MRI data [112]

One slice of processing is illustrated, showing the original image on the left, the tissue segmented image in the middle, and the warped image with the ROIs highlighted. ROIs are used to define regions for mathematic quantization of the volumetric data within the MRI.

The region-based features generated through this ROI analysis were as follows, with each region consisting of both a left and right hemisphere component as well: anterior/posterior cingulate cortex, frontal lobe gray/white matter, parietal lobe gray/white matter, occipital lobe gray/white matter, temporal lobe gray/white matter, insular cortex, lateral ventricle, medial temporal lobe, and the hippocampus [126-129]. For training and testing, a random subspace sampling method was used to create diverse classifiers from these 28 overall features. Random subsets ranging from 16 to 28 features were taken from the overall set to generate distinct classification sets in order to match the number of ERP and PET experts for the ensemble system, which helps to ensure similar voting weights and expert counts.

After preprocessing (filtering, baselining, artifact rejection), the ERP data was decomposed using the discrete wavelet transform with the Daubechies 4 wavelet.

Computing the DWT on a 1 second ERP signal sampled at 256 Hz yields 7 detail coefficients. The ERP consisted of a 200ms pre-stimulus recording, and a 800ms post-stimulus recording. However, since the majority of pertinent information in the ERP is found between 0-600ms post stimuli, the 200ms pre-stimulus and last 200ms of the overall signal are removed. Four frequency bands of interest are generated through the wavelet decomposition of the ERP, shown in Table 12.

Table 12 – DWT Coefficients and Frequency bands for ERP

Frequency Band	**0-1Hz**	**1-2Hz**	**2-4Hz**	**4-8Hz**
# of coefficients used	7	6	7	11

Figure 22 shows a breakdown of the seven detail and one approximation coefficients generated through the wavelet decomposition of the ERP. Highlighted are the four frequency bands of interest. This is similar to Figure 20, with the exception of the addition of the 7th level decomposition.

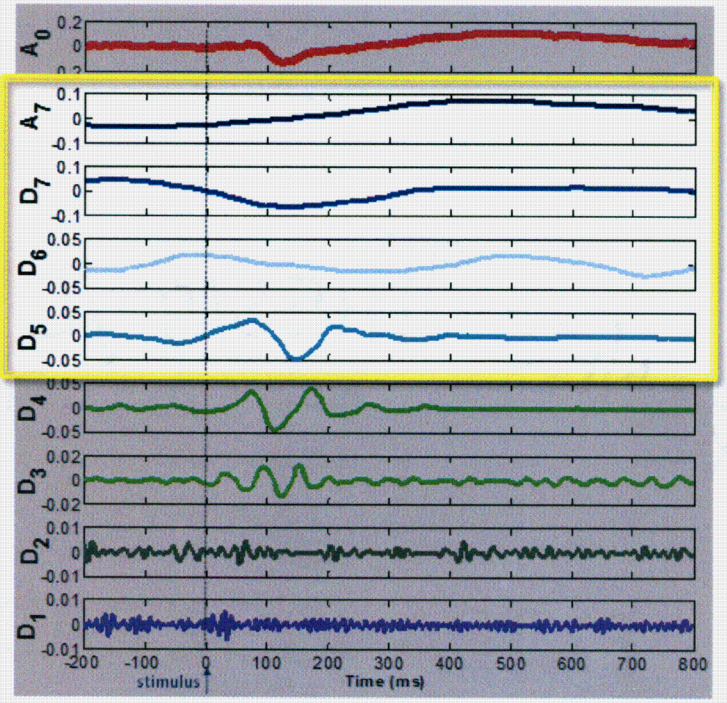

Figure 22 – DWT of ERP with frequency bands of interest highlighted [130]

These varying frequency bands were used to create multiple sets of features, referred to in this study as "feature sets." These ERP feature sets were combinations of differing electrodes, stimulus types (novel or target), and frequency bands. Table 13 lists several example feature sets used in this study.

Table 13 – ERP Feature Set Examples

Combination Type			Abbreviation
Novel Tone	Electrode C_Z	2-4Hz Band	NC_Z24
Novel Tone	Electrode F_4	1-2Hz Band	NF_412
Target Tone	Electrode P_3	0-1Hz Band	TP_301
Target Tone	Electrode P_z	1-2Hz Band	TP_z12

For this study, only the first three frequency bands were analyzed (0-1Hz, 1-2Hz, and 2-4Hz). Earlier studies have shown better performances with these frequency bands, as more complementary information appears to lie within these ranges . The final ERP expert system utilized multiple feature sets to evaluate and generate a decision for any given subject. These "experts" were combined with various classifier combination rules, which will be discussed later in this document.

The PET data was processed in a similar fashion to the MRI data. In this case, multiple features were created from over 40 regions of interest in the brain. Instead of measuring volumetric data relating to brain mass, the PET data consisted of images representative of the metabolic activity from glucose of various regions of the brain. These readings have been shown to provide pertinent information in the deterioration and decline in brain function [11]. The image slices were taken into Neurostat software, which is specific to PET scan analysis. The data presented to us comes after all image processing, and is similar to that of the MRI data. 43 total features were available, along with the data from each subject's pons, which can be used for normalization.

3.7 CLASSIFICATION METHODS

There are many methods for automated classification on a computerized system. These methods can be supervised (training with labeled class data) or unsupervised (training with unlabeled examples). For the purposes of this study, all classification was done in a supervised manner, as the classifier was presented with labeled training data for learning purposes. The entire basis of an automated classification algorithm is that it can learn hidden patterns within data to identify and classify previously unseen new data, based on these underlying distributions.

Generally, the first step prior to any classification system implementation is the extraction of relevant features to be used in the training and evaluation of the system. The classifier is then trained on this data, which is done by presenting correctly labeled instances of various classes. Training for the majority of these classifiers is essentially variations on a method to find a function that maps a given set of features to their respective correct class labels. As such, many classifiers are also referred to as "function approximators" [131].

One of the most commonly researched and implemented classification algorithms originates from artificial neural networks. These systems were initially modeled from a biological viewpoint on how neurons within the brain communicate and "learn." These algorithms work by adjusting internal parameters based on provided training data. Upon training, these algorithms then make a classification on previously unknown test or field data, based on their weights adapted to learn the training data. One of the most common types of artificial neural networks is the multilayer perceptron (MLP).

3.7.1 MULTILAYER PERCEPTRON

The multilayer perceptron is a 'feed-forward' artificial neural network that attempts to map input data onto a set of appropriate outputs, based on classification labels. It is a modification to the single-layer perceptron (SLP). The overall concept of the perceptron comes from the biological model of the neuron, shown in Figure 23.

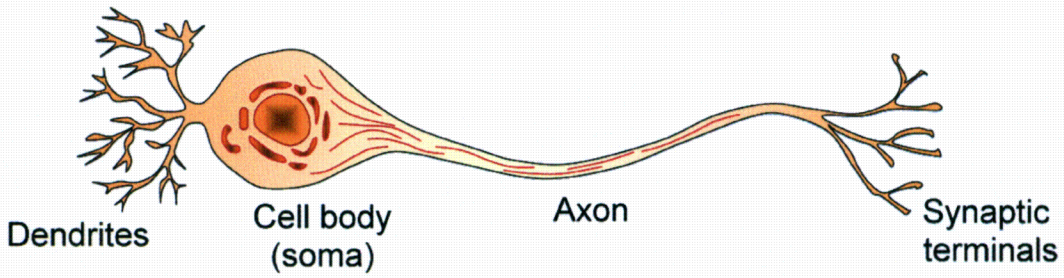

Figure 23 – Diagram for the biological representation of the neuron [132]

One of the greatest strengths of the MLP over the SLP is its ability to classify data that are not linearly separable. This strength is created by the addition of hidden layers of neurons with non-linear activation functions, allowing for greater complexity and the ability to generate non-linear hyperplane decision boundaries. Figure 24 shows a typical network structure for the MLP, with three features for a two class problem.

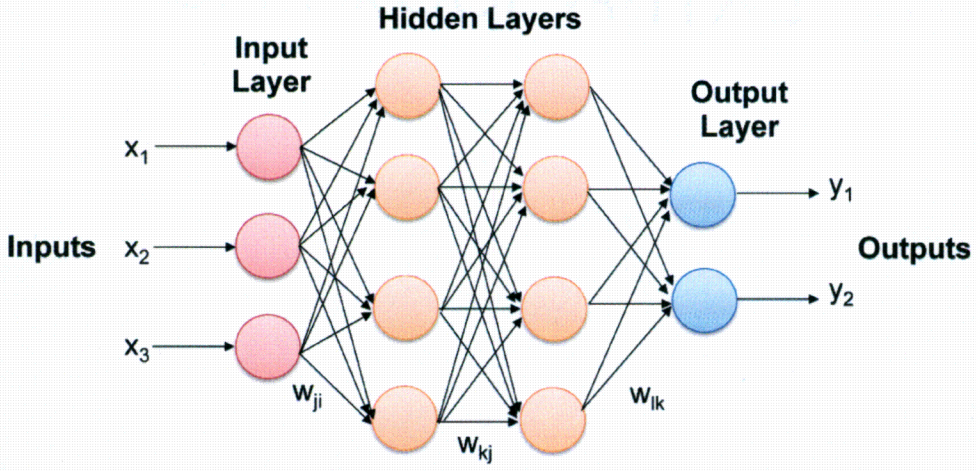

Figure 24 – Structure of the MLP with two Hidden Layers

The first set of nodes corresponding to the input layer, which accepts relevant input data (features) and forwards this information to the 'hidden' layer nodes. The total number of input nodes corresponds to the number of features in the data being used for training. The number of hidden layers and corresponding nodes is specified by the user prior to training, and varies based on the dataset, features, and other factors. The final layer constitutes the output nodes, each of which represents a different class or label. These outputs are encoded in a binary sequence. In the MLP, the input layer only serves as a method to accept data into the network; the hidden layer and output layer nodes perform computations from data based on specific weights assigned to each during training. Specifically, these weights are generated and optimized through the back propagation algorithm [112,133].

The back propagation algorithm was developed to optimize the synaptic weights assigned to hidden layer and output nodes in the multilayer perceptron. The process for back propagation is a four-tiered procedure: initialization of weights, presentation of training data, forward computation, and backward computation. Pairs of nodes are "linked" together with a synaptic weight attached to the coupling. These weights constitute the "knowledge" of the neural network. In the initialization stage, the weights are randomly initialized from a zero-mean standard distribution. The variance is chosen such that the standard deviation of the induced local field of a node lies in the transition area between the linear and saturated region of the sigmoid activation function [134]. For this study, the hidden layer nodes were used with a tangential sigmoid activation function; this activation function created restrictions on the output of a node given some input characteristics.

Upon creation of the neural network, which included initialization of the synaptic weights for all nodes, training data were presented at the input, along with correct corresponding class labels. A process of forward and backward computation was performed for each instance of the training data, until various stop conditions were reached. The two most common of these restrictions was a maximum number of epochs (iterations) or a minimum error goal. There are numerous activation functions that can be used in the MLP. Two of the most commonly implemented are the *logarithmic* and *tangential* sigmoid shown in equations 3.25 and 3.26. Figure 25 shows a graphical representation of these functions [135].

$$\varphi(n) = \frac{1}{1 + e^{-n}} \tag{3.25}$$

$$\varphi(n) = \frac{2}{1 + e^{-2n}} - 1 \tag{3.26}$$

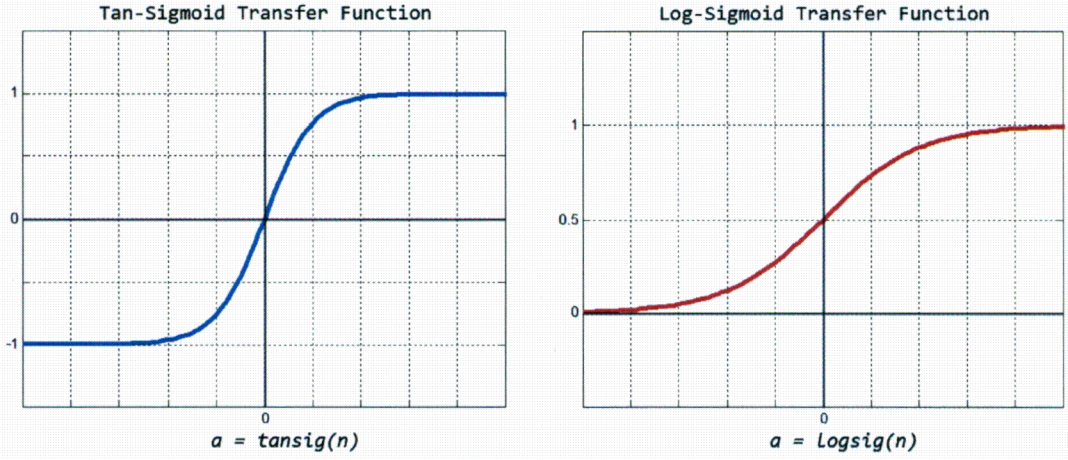

Figure 25 – Tangential and Logarithmic Sigmoid Activation Functions

Regardless of the activation function used, the overall back propagation operates the same. This process was then repeated by presenting new sets of training data and corresponding labels until the stop conditions were met.

3.7.2 K-FOLD CROSS VALIDATION TECHNIQUE

In real-world situations, only one dataset is available for training and testing. With 50/50 validation scenarios, where half the data is used for training and the other half for testing, a true generalization performance cannot be determined. The use of a cross-validation method is the best overall technique to estimate the true generalization performance of a given classification system. In K-fold cross validation, the data is split into K blocks, where $K-1$ blocks are used for training and the K^{th} block for testing. This process is repeated K times, with each iteration using a different block for testing. Each run performance is then averaged for the best estimation on the system performance [136]. Figure 26 illustrates this process.

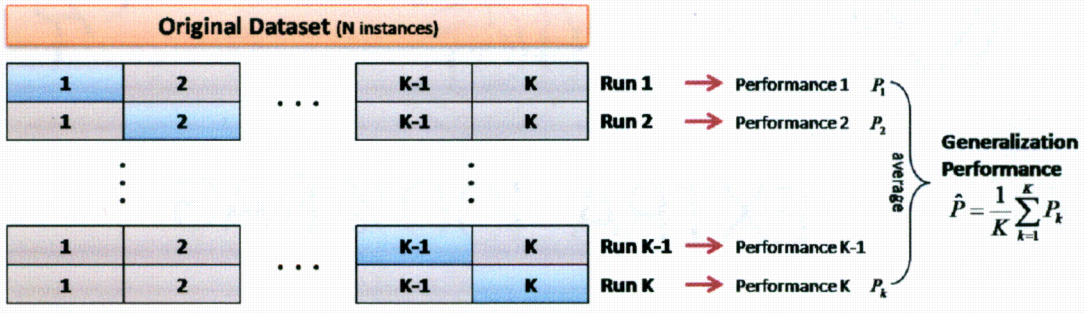

Figure 26 – K-Fold Cross validation for generalization performance

Blue blocks for indicate the test set used. Gray blocks comprise the training data for the system. The K total runs are averaged to give the overall generalization performance.

In this setup, a small K value will reduce the amount of data available for training, which is of particular concern for small datasets. Conversely, a large K value will result in a larger amount of training data available and possibly a more accurately trained classifier; however, this comes at the expense of greater computational time and a more varied overall performance. If a dataset contains N instances, and K-fold cross validation is performed with $K = N$, a specific type of validation is created, termed "leave one out"

(LOO). The initial trials done in this study involved LOO at one step of the classification process; however, this has since been removed and replaced with a 10-fold cross validation loop instead. It was shown that the LOO method was overly computational expensive and did not provide a better estimate in comparison to 10-fold CV.

3.8 ENSEMBLE OF CLASSIFIERS BASED DECISION MAKING

The concept of an ensemble of classifiers based decision making system was to combine information from multiple sources in the attempt to achieve a greater generalization performance. In this study, a decision-based data fusion approach was taken to combine EEG, MRI, and PET data to make the most informed decision possible in the classification of AD versus normal control patients. The decision based data fusion in this study was done by generated an ensemble of classifiers in a mixture of experts structure, based on the relevant EEG, MRI, and PET data available. Each biomarker for AD constituted an overall expert, with each expert containing a complex ensemble system that has multiple classifiers. These multiple classifiers can be related to a panel of experts in decision making, which attempts to increase reliability and accuracy in making the final decision for any given patient. Since each classifier in this system generated its own distinct decision boundaries, each classifier subsequently have its own error. The combination of these decisions from multiple classifiers helps to reduce overall error in such a system. However, it is crucial that such an ensemble system contain sufficiently diverse classifiers, each trained on complementary data. Therefore, the overall EEG, MRI, and PET experts all contained sub-experts, trained on complementary feature sets that promote classifier diversity to boost overall system generalization performance [137].

3.8.1 *DECISION-BASED DATA FUSION*

Decision fusion based data fusion refers to combination of decisions made by individual classifiers, as opposed to feature based data fusion where individual features are usually concatenated. In this study, we used a modified (augmented) stacked generalization approach to obtain feature level decision fusion for each biomarker (EEG, MRI, PET), followed by a combination of these decisions to create the final, across-features decision fusion for the specific marker. Each biomarker was itself an expert, which was combined using a classic classifier combination method or another stacked generalization routine.

3.8.2 *STACKED GENERALIZATION*

The primary goal of stacked generalization is to confirm or correct what has been learned by a group of preliminary (Tier-1) classifiers with the use of a meta-classifier. Any instance in a certain region of the feature space (e.g., near the decision boundary), may be more likely to be misclassified by certain classifiers than others. This trend can be learned by mapping the outputs of an ensemble of classifiers to their true labels (Figure 27) thereby allowing the system to adapt to classifiers with varying performances [138].

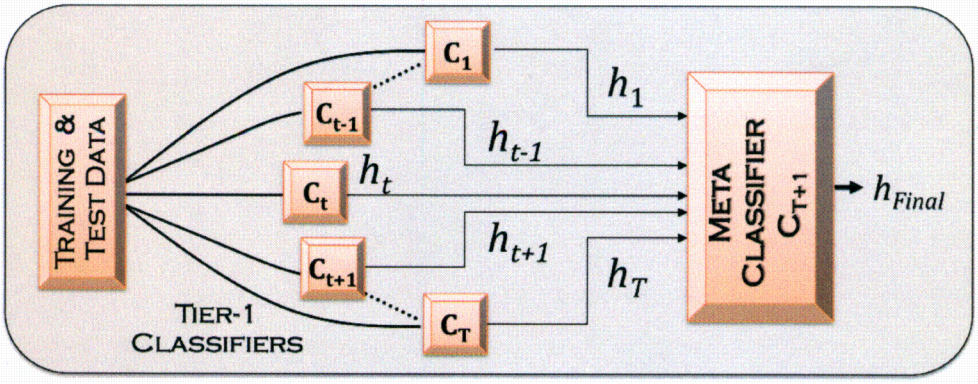

Figure 27 – Stacked Generalization diagram [139]

3.8.3 *AUGMENTED STACKED GENERALIZATION*

We use a modified version of stacked generalization, called the augmented stacked generalization (ASG) by augmenting the Tier-1 classifier outputs with the *original data* used to train them, before training the meta-classifier. Such a process enriches the intermediate feature space used by the meta-classifier, aiding in overall system performance. This process is shown in Figure 28.

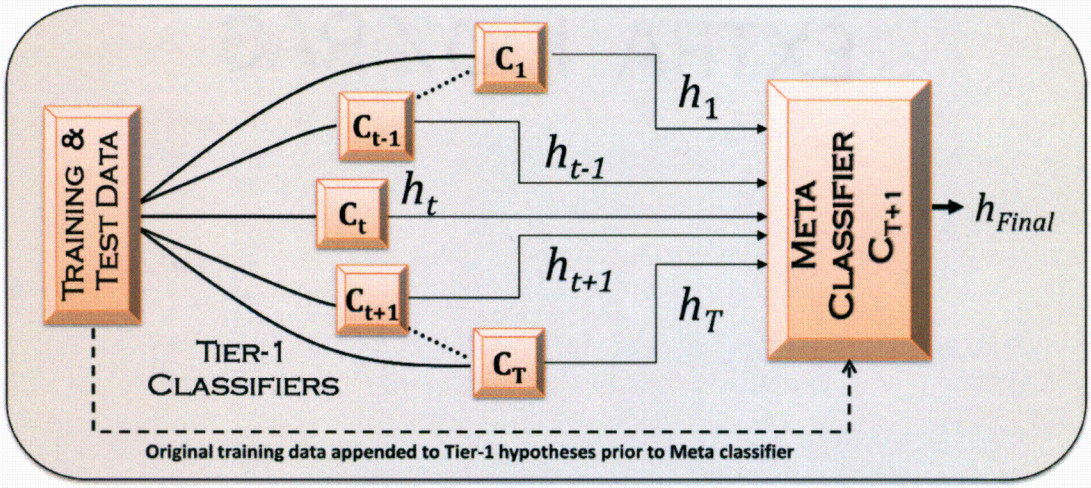

Figure 28 – Augmented Stacked Generalization diagram [139]

The actual training process is rigorous, to ensure a true generalization performance metric. The initial data of length K was first segmented into six blocks of approximate length $K/6$. One block was set aside to be used for a final evaluation of all feature sets generated in the ASG algorithm; the remaining five blocks were fed into the ASG algorithm as training data. This training data was split once again through 5-fold cross validation, with one block set aside to for evaluation. The remaining data was sent to the internal/tier-1 classifiers for training and testing. This data was split through 10-fold cross validation, with the test set used to evaluate the tier-class classifiers. The respective classifier outputs were placed into matrix form. These classifier hypotheses

were augmented with the test data used in that fold, and the process is repeated until all ten folds have run. Upon completion of these ten folds, a matrix sized $(18 + d) \times N$ is created where d represents the length of the appended test data, N as the overall ASG training data, and 18 being the support outputs from the nine tier-1 base classifiers. This entire matrix was used to train meta-classifier with the correct respective class labels, thus completing one internal fold of ASG. A diagram of this process is illustrated in Figure 29.

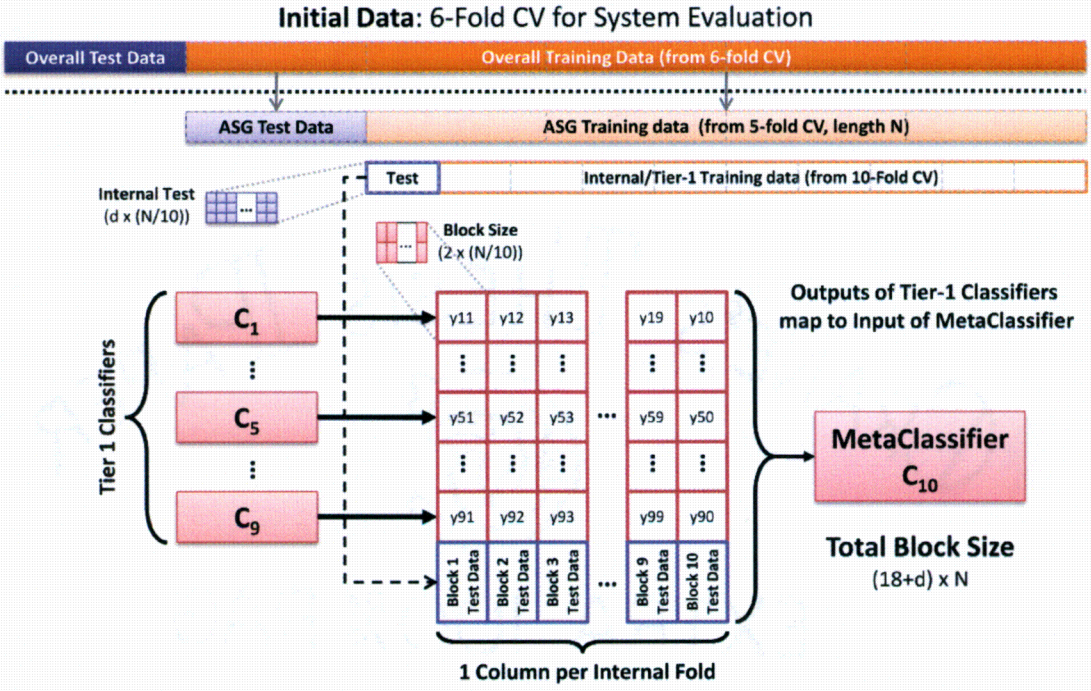

Figure 29 – Illustration of the three-tiered cross validation setup for ASG

The original training labels are used in the training process, allowing the meta-classifier to determine – and correct – poorly performing classifiers. In the intermediate training stage of ASG, the original Tier-1 classifiers are discarded, and all instances of the training data are collected. The Tier-1 classifiers are then retrained on the entire training data subset. During the testing stage, a given test instance is sent to the Tier-1

classifiers. The output from these classifiers (augmented with the original feature vector) is then sent as an input to the meta-classifier. This output from the meta-classifier then constitutes the final decision and output of the system.

Inputs for Algorithm Augmented Stacked Generalization (ASG)

- Training data $\boldsymbol{S} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, $\mathbf{x}_i \in R^d$ with correct labels $y_i \in \Omega$, $\Omega = \{\omega_1, \cdots, \omega_C\}$
- Supervised classifiers for **BaseClassifier** and a **MetaClassifier**, which can be of the same type.
- Number of classifiers T to be generated

Initial Training: ***Train Tier-1 classifiers:***

- Divide $\boldsymbol{S}$ into K blocks of size N/K, i.e. $\boldsymbol{S}_1$, $\boldsymbol{S}_2$,...,$\boldsymbol{S}_k$, (extreme case $K = N$, LOO Validation)
- **Do for** k= 1,2,...,K
  - Define $\boldsymbol{S}^{(-k)} = \boldsymbol{S} - \boldsymbol{S}_k$ be the concatenation of k-1 blocks of data, not including the k^{th} block $\boldsymbol{S}_k$
  - Train classifiers C_1,...,C_T on $\boldsymbol{S}^{(-k)}$
  - Test classifiers C_1,...,C_T on $\boldsymbol{S}_k$
  - Let $y_{t,k} \in [0\ 1]^{C \times N/K}$, t = 1,...,T (each of size $C \times [N/K]$) be the continuous-valued output of C_t

  End Do loop
- Combine K blocks of $y_{t,k}$ from all T classifiers to obtain new instances $\mathbf{Y}_n$, n=1,...,N each of length $C \cdot T$

Train meta-classifier C_{T+1}:

- Form the augmented (intermediate feature-space) vector $\mathbf{Y}^n$ by concatenating $\mathbf{Y}_n$ with the original $\mathbf{x}_n$ used to obtain $\mathbf{Y}_n$ from Tier-1 classifier outputs: $\mathbf{Y}^n$=[$\mathbf{Y}_n$ $\mathbf{x}_n$], n=1,...,N, each of size $C \cdot T+d$
- Train **MetaClassifier** C_{T+1} using N instances of $\mathbf{Y}^n$ along with their corresponding class labels ω_1,...,ω_C;

Intermediate training

- Discard all Tier-1 classifiers, and pool all training data to obtain $\boldsymbol{S}$.
- Retrain Tier-1 classifiers C_1,...,C_T on the entire training data $\boldsymbol{S}$

Testing: Given an unlabeled instance $\mathbf{z}$

- Obtain and concatenate the outputs of C_1,...,C_T, to form the intermediate space feature vector $\mathbf{Z}$
- Augment $\mathbf{Z}$ with the original $\mathbf{z}$ to obtain $\mathbf{Z}$ =[$\mathbf{Z}$ $\mathbf{z}$]
- Obtain the output of C_{T+1} for $\mathbf{Z}$ as the predicted label for $\mathbf{z}$

Figure 30 – Pseudocode for Augmented Stacked Generalization [139-141]

The output from the ASG algorithm is a decision-fusion based expert for each data source. Training an ensemble of such experts (each with data from different sources) creates a decision fusion based data fusion approach. There are many methods available for combining ensemble of classifiers' outputs, and the methods implemented in this study will be discussed in the following sections. The final combination of all these

experts is shown in Figure 31. Each expert is essentially a feature set for a given set of biomarker data. For example, with EEG data a feature set would be the ASG algorithm trained on data derived from a novel response at the PZ electrode with a frequency band of 1-2Hz (NP_Z12).

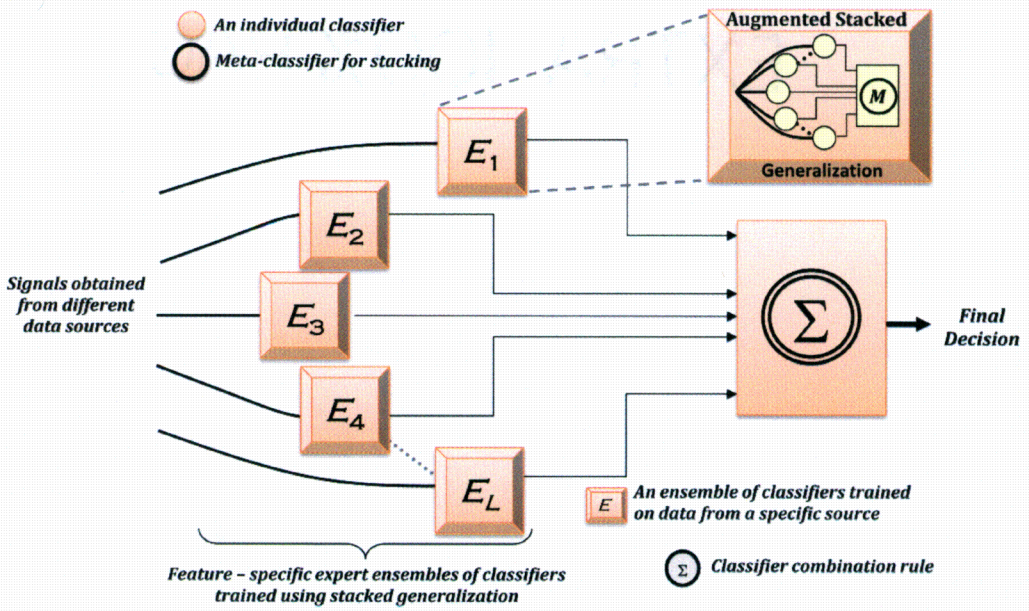

Figure 31 – ASG ensemble system, showing expert decision fusion [139]

This figure illustrates an example of one biomarker, with several experts. For individual biomarker classification, these experts are combined with some classifier combination rule. For multiple biomarker combination, the experts from all biomarkers are combined.

The system is modified only slightly for each biomarker. As stated previously, each feature set in EEG constitutes a different frequency band, electrode, and stimuli. For MRI and PET, each feature set used in ASG is created through random subspace sampling of the available features in order to create a diverse set of experts. Furthermore, the number of feature sets between all biomarker data sets was chosen to be the same in order to facilitate a fair and balanced decision fusion system. Irrespective of the biomarker used, the ensemble system generates a user-defined number of total feature sets, each containing an ASG expert to be used for final decision fusion.

3.8.4 COMBINATION RULES

There are numerous classifier combination rules that can be used for ensemble system decision fusion. For this study, three of the most common rules were analyzed as each has consistently proven to prove the best overall method for expert combination.

Sum Rule

The sum rule works on the principle of support based or probabilistic classifier outputs. For example, the MLP gives a continuous output, $d_{i,j} \in [0,1]$ with a support given by a classifier C_i to class j, where $i = 1, \ldots, T$ and $j = 1, \ldots, c$ with T as the total number of classifiers and c as the total number of classes. The support for each class from a given ensemble system are summed, and the class with the largest sum is selected as the overall decision. Equation 3.42 shows this process mathematically.

$$\mu_j(x) = \frac{1}{T}\sum_{j=1}^{T} d_{i,j}(x) \qquad d_{i,j}(x) \in [0,1] \tag{3.42}$$

Simple Majority Voting

Simple majority voting works on the principle of class decisions based on the principle of majority rule. Essentially, a final class decision is chosen based on which class the majority of classifiers in the system selected. The support structure for each class is shown in Equation 3.43.

$$\mu_j(x) = \sum_{j=1}^{T} d_{i,j}(x) \qquad d_{i,j}(x) \in [0,1] \tag{3.43}$$

Weighted Majority Voting

Weighted majority voting is similar to simple majority voting in that total votes are simply tallied. The difference here is the ability to "weight" the decisions of each expert/classifier. Generally, these weights are assigned based on validation performance. Equation 3.44 shows the support structure for each class based on this method.

$$\mu_j(x) = \sum_{j=1}^{T} W(C_i) d_{i,j}(x) \qquad d_{i,j}(x) \in [0,1] \tag{3.44}$$

Figure 32 illustrates a sample of all three combination methods, based on a two class problem with five classifiers. Respective weights are given for each classifier, along with their support outputs for each class.

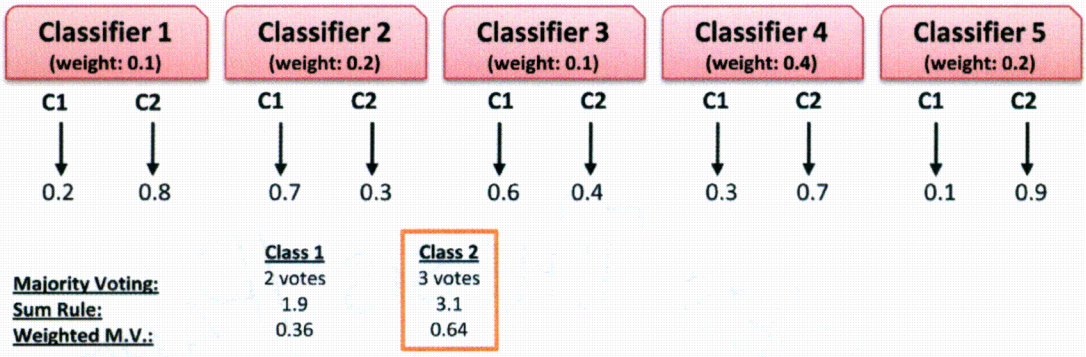

Figure 32 – Example of classifier/ensemble system combination rules

3.9 CLINICAL DIAGNOSITIC MEASURES

Arguably more important than merely overall system accuracy, clinical diagnostic measures allow a more in-depth view at how well a given system performs on various classification scenarios. Specifically, these measures include positive/negative predictive value, sensitivity, and specificity. Each of these measures is calculated based on the total

number of true positive/negative and false positive/negative diagnoses achieved from a classification system. The breakdown of how each is calculated is shown in Table 14.

Table 14 – Breakdown of diagnostic performance metrics

		Diagnostic Condition	
		Positive	**Negative**
Test Condition	**Positive**	True Positive **(TP)**	False Positive **(FP)**
	Negative	False Negative **(FN)**	True Negative **(TN)**

To calculate positive/negative predictive value, sensitivity, and specificity the diagnostic performance metrics of a system must first be known. Positive predictive value (PPV) is the probability that a subject has the disease, given that they tested positive. Similarly, negative predictive value (NPV) is the probability that a subject does not have the disease in question, given they tested negative. Sensitivity is a measure of how well the system can recognize all true positives; in other words, those individuals that truly have the disease. Likewise, specificity measures how well the system can correctly identify those patients that do not truly have the disease in question. Based on the information from Table 14, the following relationships for each measure can be determined:

$$\textbf{PPV} \qquad \frac{TP}{TP + FP} \tag{3.45}$$

$$\textbf{NPV} \qquad \frac{TN}{TN + FN} \tag{3.46}$$

$$\textbf{Sensitivity} \qquad \frac{TP}{TP + FN} \tag{3.47}$$

$$\textbf{Specificity} \qquad \frac{TN}{TN + FP} \tag{3.48}$$

CHAPTER IV

RESULTS

The results presented in this section are divided into three main segments: single dataset performance, ensemble system combination performance, and preliminary disease severity analysis. All three biomarkers were tested alone, and with each other in all possible combinations. ERP from EEG processed with the wavelet transform were used, along with quantized MRI data and PET data. Normalization methods for MRI and PET data were also tested, with respective results tabulated.

For this study, only AD versus CN was analyzed; however, in the severity analysis, mild/moderate AD versus severe AD versus normal control subjects was compared. All accuracies within this section are based upon the diagnosis of expert neurologists, and may not be representative of the true patient diagnosis. For example, a performance of 95% would indicate a diagnostic accuracy of 95% correct with respect to the expert diagnoses provided for each cohort.

4.1 SINGLE DATA SET PERFORMANCES

Different subject data and normalization techniques were explored for each biomarker to determine the best overall method to be used for ensemble system combination. Basic demographics and MMSE scores for each cohort used for their respective biomarker are tabulated and detailed in each of the following sections. A final visual comparison of all individual performances along with confidence intervals is provided in section 4.1.4.

4.1.1 EEG (P300 ERP ANALYSIS)

In the EEG P300 analysis, there were a total of five different cohorts and electrode feature sets tested. Each table lists the cohort used, as well as relevant statistics for all subjects. Furthermore, various feature sets were analyzed and their combination performances for the ASG algorithm are detailed. There were two primary sub-cohorts generated for this section based on the visual ERP analysis performed for each subject. Table 15 lists basic information for the first of such sub-cohorts, based on the satisfactory only ERPs from the visual analysis.

Table 15 – Satisfactory Dataset Cohort Statistics

	Male	Female		MMSE	Age
AD	21	33	**54**	20.60	76.44
CN	13	40	**53**	28.94	70.28
Totals	34	73	**106**		

Table 16 lists basic information for the second sub-cohort generated, this time based upon both the satisfactory as well as marginal ERPs from the visual analysis.

Table 16 – Satisfactory + Marginal Dataset Cohort Statistics

	Male	Female		MMSE	Age
AD	39	46	**85**	19.20	76.02
CN	24	55	**79**	29.05	72.16
Totals	63	101	**164**		

These two sub-cohorts make up the primary data used for analysis of the EEG biomarker. A subset of the satisfactory only cohort was used to perform analysis on first visit only subjects, detailed in section 4.1.1a.

4.1.1a FIRST VISIT SUBJECT ANALYSIS (18 FEATURE SETS)

This dataset consists of 101 subjects (44CN, 57AD) chosen for first visit subjects only from the satisfactory only dataset. The ideology behind this choice was to determine whether using date that consisted of subjects with AD from first diagnosis had any effect

on system performance with respect to later tests run with multiple visit subjects. Basic demographic and MMSE scores for this subset are shown in Table 17. The accuracies of the individual feature sets within this cohort are shown in Table 18.

Table 17 – Satisfactory Dataset (1st Visit) Cohort Statistics

	Male	Female		MMSE	Age
AD	22	35	**57**	20.49	76.39
CN	13	31	**44**	28.79	70.19
Totals	35	66	**101**		

Table 18 – Satisfactory Dataset (1st Visit) Cohort Performances

'FS'	'OGP'	'GP5'	'GP'	'SN'	'SP'	'PPV'	'NPV'
'TP301'	67.43±1.9%	74.25%	83.15%	58.50%	52.43%	62.69%	58.06%
'TP312'	67.84±2.0%	75.46%	84.25%	59.66%	51.63%	62.73%	58.44%
'TP324'	71.30±1.6%	77.87%	88.17%	61.60%	57.48%	66.34%	62.66%
'TPZ01'	67.11±1.8%	73.25%	81.13%	57.96%	51.98%	62.34%	57.91%
'TPZ12'	67.47±1.6%	74.08%	83.21%	58.87%	51.29%	62.11%	58.22%
'TPZ24'	68.14±1.8%	75.24%	87.13%	58.51%	53.41%	63.05%	58.96%
'TCZ01'	65.02±1.1%	72.10%	81.19%	56.67%	48.67%	59.98%	55.28%
'TCZ12'	64.73±2.0%	71.91%	80.33%	56.03%	49.47%	60.10%	54.97%
'TCZ24'	69.30±1.6%	75.86%	83.21%	60.55%	53.48%	63.83%	60.52%
'NP301'	65.92±0.8%	72.89%	83.15%	56.87%	51.29%	61.16%	56.55%
'NP312'	67.34±1.0%	75.85%	87.13%	57.76%	53.45%	62.80%	57.76%
'NP324'	69.49±1.0%	76.28%	85.23%	60.31%	54.96%	64.50%	60.85%
'NPZ01'	66.42±1.3%	72.87%	81.19%	57.50%	51.13%	61.30%	57.19%
'NPZ12'	66.45±1.4%	73.14%	79.29%	58.80%	49.68%	61.61%	56.91%
'NPZ24'	70.66±1.0%	77.63%	86.15%	61.98%	55.50%	65.66%	62.47%
'NCZ01'	64.43±1.1%	72.36%	83.27%	56.54%	47.94%	59.51%	55.24%
'NCZ12'	65.09±1.6%	72.50%	80.27%	57.91%	47.75%	60.20%	56.03%
'NCZ24'	65.92±1.5%	72.70%	81.25%	58.38%	49.56%	61.24%	56.92%

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

In Table 18, it is clear that the P3 electrodes tend to perform the highest, with PZ close behind, with higher accuracies for target tone analysis. The first visit cohort used 18 total feature sets including the P3, PZ, and CZ electrodes only, with frequency bands from 0-4Hz. Table 19 shows the feature set combinations used and respective

performance for sum rule and simple majority voting decision fusion. The individual feature set combination lists (FS1 – FS4) are listed below.

FS1: TP324, NPZ24, NP324, TCZ24, TP312, TPZ24
FS2: NCZ01, NCZ12, NCZ24, TCZ01, TCZ12, NP301
FS3: TP301, TP312, TP324, NP301, NP312, NP324
FS4: NPZ24, TCZ24, TPZ24, NP312, TPZ01, TPZ12, TP324

Table 19 – 1st Visit Data Fusion Performances

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	81.3±2.9	90.5	81.3±2.1	80.0±2.9	83.5±3.5	81.6±2.4
FS2	76.9±3.2	83.8	78.2±2.3	81.2±2.2	81.5±2.5	80.4±1.9
FS3	79.2±2.8	88.3	76.5±2.9	79.2±2.7	78.5±2.0	78.6±2.6
FS4	**81.8±2.1**	**91.3**	**81.6±2.8**	**83.5±2.4**	**82.4±1.9**	**81.5±3.1**
SMV	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	**79.2±2.7**	89.2	80.2±2.5	81.2±2.4	79.8±2.2	80.2±2.5
FS2	75.7±2.3	86.2	77.3±2.4	80.5±2.3	79.4±2.6	80.3±3.0
FS3	75.4±1.9	85.6	78.2±3.1	78.5±3.1	77.4±2.1	79.4±2.8
FS4	77.0±2.9	90.2	79.3±2.8	77.4±2.7	78.3±3.2	78.5±2.6

In this test, the highest overall diagnostic accuracy reached was through a sum rule combination of the FS4 combination list, achieving 81.8±2.1%, shown in Table 19. From Table 18, it is clear that the parietal electrodes tend to have a higher overall diagnostic accuracy when compared to the CZ electrode. Target stimuli also proved to show a slight performance advantage when compared to novel stimuli.

4.1.1b NEURONETRIX ELECTRODE ANALYSIS (36 FEATURE SETS)

For the next two cohorts tested, 36 total feature sets were generated, corresponding to the P3, P4, C3, C4, F3, and F4 electrodes, once again only using frequency bands between 0-4Hz. Neuronetrix is a company based out of Louisville, KY for which we perform EEG analysis. The electrodes in this test represent the electrodes used in Neuronetrix's prototype COGNISION system for AD diagnosis currently under FDA review [142]. The first test for this analysis utilized the satisfactory only EEG dataset, comprised of 107 total subjects (54CN, 53AD). Individual feature set accuracies are shown in Table 20.

Table 20 – 36 FS: Satisfactory Dataset Cohort Performances

'FS'	'OGP'	'GP5'	'GP'	'SN'	'SP'	'PPV'	'NPV'
'TP301'	65.94±1.5%	75.36%	81.54%	64.32%	66.67%	66.35%	64.85%
'TP312'	70.99±1.1%	78.18%	84.32%	71.52%	69.36%	71.38%	70.46%
'TP324'	65.43±1.2%	73.14%	77.78%	63.89%	65.60%	65.64%	64.22%
'TP401'	66.50±2.1%	74.63%	84.32%	65.73%	65.83%	66.18%	65.56%
'TP412'	65.96±2.1%	74.48%	80.61%	66.42%	64.13%	65.87%	65.07%
'TP424'	65.83±2.1%	73.25%	80.78%	63.75%	66.41%	65.80%	64.74%
'TC301'	63.78±1.0%	70.79%	78.81%	64.28%	62.19%	64.14%	62.40%
'TC312'	65.74±2.1%	74.49%	83.50%	66.46%	64.36%	65.70%	65.14%
'TC324'	64.60±1.0%	73.16%	83.44%	63.71%	64.01%	64.33%	63.50%
'TC401'	63.55±1.5%	71.30%	77.83%	63.86%	61.50%	63.49%	62.54%
'TC412'	65.94±1.2%	72.76%	78.76%	65.24%	65.48%	66.33%	65.05%
'TC424'	65.41±0.9%	73.91%	81.59%	64.85%	63.61%	65.04%	63.72%
'TF301'	61.29±1.7%	69.57%	77.78%	61.67%	59.70%	60.91%	60.40%
'TF312'	63.11±1.4%	71.66%	79.74%	63.08%	62.60%	63.50%	62.48%
'TF324'	62.71±1.6%	70.95%	82.52%	61.76%	61.95%	62.21%	61.42%
'TF401'	59.98±1.9%	68.53%	77.02%	59.24%	59.18%	59.71%	58.77%
'TF412'	61.45±0.6%	71.28%	78.71%	61.76%	59.56%	61.41%	59.98%
'TF424'	63.41±1.7%	70.73%	79.74%	63.57%	61.97%	63.72%	62.16%
'NP301'	62.74±1.6%	70.21%	75.11%	63.43%	60.94%	62.64%	61.69%
'NP312'	65.50±1.9%	72.76%	81.48%	64.79%	64.85%	65.55%	64.47%
'NP324'	64.61±2.2%	72.28%	80.67%	63.92%	63.76%	64.63%	63.53%
'NP401'	66.08±1.4%	72.77%	77.78%	66.01%	64.88%	65.69%	65.18%
'NP412'	62.39±0.7%	70.17%	79.69%	62.01%	61.09%	61.86%	61.30%
'NP424'	66.94±1.0%	75.48%	82.57%	65.81%	66.80%	67.04%	65.50%
'NC301'	60.94±1.1%	67.74%	78.76%	59.35%	60.74%	60.42%	59.66%
'NC312'	61.67±1.0%	69.27%	77.89%	60.03%	61.65%	61.33%	60.40%
'NC324'	63.87±1.8%	72.79%	79.69%	63.05%	63.55%	63.39%	63.07%
'NC401'	61.74±1.8%	70.22%	76.96%	61.91%	60.39%	61.88%	60.41%
'NC412'	65.50±1.8%	74.27%	79.74%	66.35%	62.79%	65.22%	64.05%
'NC424'	63.23±0.6%	71.29%	79.69%	62.01%	63.44%	63.24%	62.43%
'NF301'	63.25±1.3%	70.93%	78.81%	62.18%	63.09%	63.14%	62.33%
'NF312'	66.12±1.6%	73.15%	79.74%	63.76%	67.02%	66.10%	64.77%
'NF324'	60.46±1.6%	66.69%	73.20%	58.96%	60.12%	60.43%	58.89%
'NF401'	62.38±2.1%	69.63%	75.06%	61.67%	61.48%	61.85%	61.18%
'NF412'	62.45±0.9%	70.55%	79.69%	61.66%	62.06%	61.99%	61.64%
'NF424'	60.01±1.4%	67.23%	78.87%	58.09%	60.71%	59.72%	58.93%

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

Once again, we notice a trend with parietal electrodes tending to have higher accuracies, with target tone analysis yielding the highest results. The frontal electrodes

tend to have the lowest overall accuracies of the test, regardless of stimuli. The combinations used for this test are listed below.

FS1: TP301, TP312, TP324, TP401, TP412, TP424
FS2: TP301, TP312, TP324, TP401, TP412, TP424, TC301, TC312, TC324, TC401, TC412, TC424
FS3: NP301, NP312, NP324, NP401, NP412, NP424, NC301, NC312, NC324, NC401, NC412, NC424
FS4: TF401, NF424, NF324, NC301, TF301, TF412
FS5: TP301, TP312, NP424, NP401, TC312, TP401, NF312, NC412

Table 21 – 36 FS: Satisfactory Data Fusion Performances

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	82.0±3.7	91.5	82.4±2.9	81.5±3.0	83.6±2.6	81.4±2.8
FS2	82.8±3.9	95.4	83.9±2.9	84.5±3.2	82.4±1.9	83.4±2.5
FS3	77.1±4.1	83.5	80.1±2.1	81.2±3.1	79.1±2.0	80.3±2.1
FS4	59.8±4.1	71.2	64.1±3.0	66.1±2.9	68.5±2.7	66.4±2.7
FS5	**82.8±2.9**	**90.2**	**83.1±2.4**	**84.5±2.2**	**86.5±2.3**	**84.7±3.2**
SMV	**Avg (%)**	**Best (%)**	**SN (%)**	**SP (%)**	**PPV (%)**	**NPV (%)**
FS1	70.7±3.1	80.8	75.2±2.2	74.1±3.7	76.2±4.1	74.3±2.9
FS2	**79.8±2.8**	96.3	78.3±3.4	77.4±3.0	76.9±3.0	77.3±3.2
FS3	69.5±3.3	74.2	73.5±3.6	70.2±3.1	70.0±2.8	71.1±3.4
FS4	63.0±3.3	73.8	68.1±2.1	65.2±2.9	64.5±3.1	66.5±3.6
FS5	75.1±2.8	85.6	77.2±3.0	78.1±2.4	79.0±2.0	77.3±2.9

The FS5 combination list achieved the highest accuracy at 82.8±2.9%. The lowest performance came from the FS4 combination list, which was comprised of the lowest accuracy individual feature sets. These sets were all frontal electrodes, with the exception of the C3 electrode.

The second test utilizing this 36FS combination makes use of the Satisfactory + Marginal EEG dataset containing 164 overall subjects (79CN, 85AD), shown in Table 16. The feature set combinations used for this test were the same as in the satisfactory only dataset cohort, listed previously. The performance metrics for the individual sets shown in Table 22 with combination accuracies shown in Table 23.

Table 22 – 36 FS: Satisfactory + Marginal Dataset Cohort Performances

'FS'	'OGP'	'GP5'	'GP'	'SN'	'SP'	'PPV'	'NPV'
'TP301'	69.74±1.0%	74.16%	80.34%	64.13%	66.48%	66.16%	64.66%
'TP312'	71.59±1.4%	76.98%	83.12%	71.33%	69.17%	71.19%	70.27%
'TP324'	64.23±0.6%	71.94%	76.58%	63.70%	65.41%	65.45%	64.03%
'TP401'	68.90±1.1%	73.43%	83.12%	65.54%	65.64%	65.99%	65.37%
'TP412'	64.76±0.7%	73.28%	79.41%	66.23%	63.94%	65.68%	64.88%
'TP424'	64.63±0.9%	72.05%	79.58%	63.56%	66.22%	65.61%	64.55%
'TC301'	62.58±1.4%	69.59%	77.61%	64.09%	62.00%	63.95%	62.21%
'TC312'	64.54±2.4%	73.29%	82.30%	66.27%	64.17%	65.51%	64.95%
'TC324'	63.40±1.3%	71.96%	82.24%	63.52%	63.82%	64.14%	63.31%
'TC401'	62.35±0.9%	70.10%	76.63%	63.67%	61.31%	63.30%	62.35%
'TC412'	64.74±0.9%	71.56%	77.56%	65.05%	65.29%	66.14%	64.86%
'TC424'	64.21±0.5%	72.71%	80.39%	64.66%	63.42%	64.85%	63.53%
'TF301'	60.09±1.5%	68.37%	76.58%	61.48%	59.51%	60.72%	60.21%
'TF312'	61.91±1.2%	70.46%	78.54%	62.89%	62.41%	63.31%	62.29%
'TF324'	61.51±1.3%	69.75%	81.32%	61.57%	61.76%	62.02%	61.23%
'TF401'	58.78±1.2%	67.33%	75.82%	59.05%	58.99%	59.52%	58.58%
'TF412'	60.25±1.1%	70.08%	77.51%	61.57%	59.37%	61.22%	59.79%
'TF424'	62.21±1.1%	69.53%	78.54%	63.38%	61.78%	63.53%	61.97%
'NP301'	61.54±0.7%	69.01%	73.91%	63.24%	60.75%	62.45%	61.50%
'NP312'	64.30±1.8%	71.56%	80.28%	64.60%	64.66%	65.36%	64.28%
'NP324'	63.41±1.6%	71.08%	79.47%	63.73%	63.57%	64.44%	63.34%
'NP401'	64.88±1.7%	71.57%	76.58%	65.82%	64.69%	65.50%	64.99%
'NP412'	61.19±0.6%	68.97%	78.49%	61.82%	60.90%	61.67%	61.11%
'NP424'	65.74±1.3%	74.28%	81.37%	65.62%	66.61%	66.85%	65.31%
'NC301'	59.74±0.8%	66.54%	77.56%	59.16%	60.55%	60.23%	59.47%
'NC312'	60.47±0.6%	68.07%	76.69%	59.84%	61.46%	61.14%	60.21%
'NC324'	62.67±0.7%	71.59%	78.49%	62.86%	63.36%	63.20%	62.88%
'NC401'	60.54±1.8%	69.02%	75.76%	61.72%	60.20%	61.69%	60.22%
'NC412'	64.30±0.8%	73.07%	78.54%	66.16%	62.60%	65.03%	63.86%
'NC424'	62.03±1.8%	70.09%	78.49%	61.82%	63.25%	63.05%	62.24%
'NF301'	62.05±0.5%	69.73%	77.61%	61.99%	62.90%	62.95%	62.14%
'NF312'	64.92±2.3%	71.95%	78.54%	63.57%	66.83%	65.91%	64.58%
'NF324'	59.26±0.5%	65.49%	72.00%	58.77%	59.93%	60.24%	58.70%
'NF401'	61.18±1.6%	68.43%	73.86%	61.48%	61.29%	61.66%	60.99%
'NF412'	61.25±0.2%	69.35%	78.49%	61.47%	61.87%	61.80%	61.45%
'NF424'	58.81±2.0%	66.03%	77.67%	57.90%	60.52%	59.53%	58.74%

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

This test evaluated the same feature sets as the previous, this time utilizing the larger cohort containing 164 subjects. Similar parallels can be drawn between overall

electrode performances, with the parietal tending to be the highest and frontal electrodes again representing the lower accuracies of the test. Target stimuli performance is on average greater than that of novel stimuli accuracy, as with the previous test.

Table 23 – 36 FS: Satisfactory + Marginal Data Fusion Performances

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	**81.5±3.4**	**87.8**	**80.2±3.2**	**81.1±2.2**	**79.4±3.0**	**77.6±2.1**
FS2	80.2±2.9	95.2	79.4±3.6	78.9±3.9	81.5±2.7	80.0±3.0
FS3	79.7±3.5	87.8	81.3±2.9	79.9±4.0	81.0±3.0	80.3±2.8
FS4	69.3±3.3	76.5	66.4±3.9	69.4±4.1	62.0±3.5	63.3±3.2
FS5	80.8±3.7	86.0	82.4±3.5	81.4±2.8	79.5±2.6	79.9±3.1
SMV	**Avg (%)**	**Best (%)**	**SN (%)**	**SP (%)**	**PPV (%)**	**NPV (%)**
FS1	71.0±3.9	80.5	72.4±3.0	70.4±2.9	73.1±3.6	71.2±2.7
FS2	72.9±3.2	81.7	73.5±4.1	72.4±3.2	71.4±2.8	70.5±2.8
FS3	**75.5±3.6**	84.1	75.1±3.6	74.8±4.0	73.3±2.9	71.5±3.2
FS4	67.2±3.4	73.1	69.2±3.1	70.2±4.2	67.8±3.5	68.8±3.2
FS5	70.2±2.9	80.4	75.4±3.0	72.4±2.8	70.1±2.7	73.9±2.9

Both tests on the 36 feature set evaluation, using satisfactory only (4.1.1d) as well as satisfactory + marginal data (4.1.1c), achieved overall combination accuracies of over 80%. The highest overall performance for this feature set evaluation came to 82.8±2.9% using the satisfactory only dataset on the FS5 combination. The next test focuses on parietal only electrodes, detailed in section 4.1.1c.

4.1.1c PARIETAL ONLY ELECTRODE ANALYSIS (30 FEATURE SETS)

The next EEG test used both the satisfactory and satisfactory + marginal cohorts, to test a 30 FS, parietal electrode combination list. All bands between 0 Hz and 4 HZ were tested on all the parietal electrodes (P3, P4, P7, P8, and PZ), with both novel and target tone stimuli.

Table 24 – 30 FS: Satisfactory Dataset Cohort Performances

'FS'	'OGP'	'GP5'	'GP'	'SN'	'SP'	'PPV'	'NPV'
'TP301'	67.92±1.9%	74.54%	85.14%	66.86%	63.44%	65.72%	64.31%
'TP312'	72.75±2.1%	79.26%	87.04%	71.41%	68.53%	70.20%	70.19%
'TP324'	67.43±1.9%	76.85%	86.22%	63.83%	66.04%	66.05%	63.86%
'TP401'	67.84±1.8%	74.20%	80.61%	64.29%	65.95%	65.94%	64.40%
'TP412'	69.43±2.2%	76.24%	84.32%	67.68%	65.99%	67.45%	66.67%
'TP424'	67.57±1.7%	75.94%	81.54%	63.00%	67.13%	66.27%	64.22%
'TP701'	66.64±2.2%	75.16%	83.34%	64.51%	63.72%	64.66%	63.44%
'TP712'	67.89±2.2%	75.33%	81.54%	66.17%	64.17%	65.67%	64.87%
'TP724'	66.42±1.9%	74.81%	81.49%	62.97%	64.33%	64.48%	63.29%
'TP801'	64.95±2.0%	72.12%	82.36%	63.87%	60.79%	62.60%	62.11%
'TP812'	66.13±1.1%	74.54%	81.49%	63.23%	63.47%	63.81%	62.97%
'TP824'	68.03±1.5%	75.88%	83.34%	63.30%	67.35%	66.50%	64.58%
'TPZ01'	69.53±1.6%	77.98%	86.22%	66.50%	67.57%	67.28%	67.13%
'TPZ12'	69.78±1.8%	77.19%	84.32%	67.68%	66.31%	67.65%	66.78%
'TPZ24'	66.32±1.8%	74.80%	84.37%	61.96%	65.08%	64.53%	62.88%
'NP301'	63.74±2.0%	70.15%	77.84%	61.95%	60.12%	61.49%	60.50%
'NP312'	65.58±2.0%	72.71%	78.71%	62.08%	63.09%	63.59%	61.84%
'NP324'	66.30±2.1%	72.71%	78.71%	63.49%	64.36%	64.67%	63.32%
'NP401'	67.92±1.7%	77.03%	84.32%	64.72%	65.44%	65.41%	65.18%
'NP412'	64.62±1.7%	70.85%	77.73%	61.34%	62.24%	62.53%	61.23%
'NP424'	68.30±1.5%	75.19%	83.39%	65.44%	66.33%	66.73%	64.99%
'NP701'	64.86±1.9%	72.90%	79.63%	63.64%	60.80%	63.02%	61.58%
'NP712'	66.18±1.9%	75.35%	84.32%	63.45%	63.19%	63.70%	63.25%
'NP724'	66.65±2.0%	71.95%	80.51%	63.10%	65.10%	64.80%	63.67%
'NP801'	68.57±2.2%	78.28%	88.89%	66.51%	65.38%	66.23%	65.83%
'NP812'	66.47±2.1%	73.28%	79.63%	64.15%	63.69%	64.26%	63.74%
'NP824'	69.42±1.6%	77.00%	87.10%	64.53%	68.92%	67.70%	66.54%
'NPZ01'	66.03±2.8%	74.93%	84.26%	62.99%	63.88%	64.47%	62.74%
'NPZ12'	67.00±2.4%	75.90%	82.47%	65.83%	63.52%	65.07%	64.40%
'NPZ24'	67.61±2.0%	75.72%	82.41%	63.29%	65.77%	65.27%	64.05%

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

Table 24 lists all the diagnostic accuracies for each feature set in this test with the satisfactory dataset. A trend that has been noticed in previous tests reappears in this evaluation—target stimuli feature sets tend to perform higher than comparable novel stimuli sets. It is also clear that even the lowest performing individual feature set from this test only drops to 63.74±2.0%, with the highest reaching 72.75±2.1%. The feature set combinations used for this setup were as follows:

FS1: TP301, TP312, TP324, TP401, TP412, TP424
FS2: TP301, NP824, NP401, NP324, TP312, NP724, TP801
FS3: TP301, TP312, TP324, NP301, NP312, NP324
FS4: TP812, TP724, TP701, TPZ24, TP424, NP412
FS5: TP301, NP824, TP312, NPZ01, TP401, NP401, TP801, NP324

Table 25 – 30 FS: Satisfactory Data Fusion Performances

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	**85.7±3.1**	**95.4**	**83.2±3.8**	**81.4±2.9**	**81.8±3.1**	**82.0±3.0**
FS2	82.6±4.1	96.7	80.2±2.9	80.6±2.9	81.3±3.2	79.9±2.8
FS3	80.1±3.2	91.1	78.4±3.0	79.0±3.1	77.3±2.4	77.0±2.6
FS4	79.1±4.0	89.5	74.1±2.3	73.1±3.5	74.2±2.0	72.9±2.9
FS5	78.1±3.7	90.5	75.0±3.0	74.2±3.2	75.3±3.3	76.2±3.9
SMV	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	**81.3±2.9**	90.5	79.8±2.9	77.4±2.7	78.0±2.5	80.2±3.1
FS2	79.9±3.6	84.7	76.2±3.4	74.3±2.9	77.1±3.2	76.8±3.4
FS3	77.1±4.2	90.5	75.2±3.1	74.6±3.2	75.4±2.9	73.8±3.6
FS4	77.1±3.7	91.5	74.3±2.6	75.6±3.2	76.4±3.1	75.0±2.8
FS5	74.2±4.2	91.3	72.4±2.7	73.1±2.9	72.1±3.0	70.0±2.9

The highest overall combination accuracy for this test was achieved with the sum rule combination using the FS1 combination list, achieving 85.7±3.1% as shown in Table 25. Another trend visible for all EEG tests thus far is that the sum rule combination gives a higher overall accuracy than that of simple majority voting for the same combination lists. The next test used the same feature sets with the satisfactory + marginal cohort (164 total subjects, 79CN 85AD) for evaluation.

Table 26 – 30 FS: Satisfactory + Marginal Dataset Cohort Performances

'FS'	'OGP'	'GP5'	'GP'	'SN'	'SP'	'PPV'	'NPV'
'TP301'	69.70±1.5%	73.49%	78.31%	66.99%	66.83%	68.78%	65.05%
'TP312'	68.99±1.3%	73.20%	81.03%	66.67%	65.62%	68.04%	64.27%
'TP324'	67.75±0.9%	72.44%	82.01%	67.00%	62.84%	66.45%	63.62%
'TP401'	68.62±1.4%	73.16%	80.78%	66.35%	65.13%	67.54%	64.13%
'TP412'	67.44±0.6%	72.03%	79.54%	65.76%	63.47%	66.16%	63.23%
'TP424'	66.65±1.1%	70.43%	77.80%	64.20%	63.70%	65.92%	62.16%
'TP701'	66.23±1.3%	70.66%	76.18%	65.30%	61.74%	65.24%	61.86%
'TP712'	67.53±0.7%	72.89%	78.03%	66.30%	63.11%	66.51%	63.08%
'TP724'	65.97±1.0%	69.71%	76.18%	64.07%	62.00%	64.91%	61.42%
'TP801'	68.79±1.3%	73.60%	77.76%	66.35%	65.25%	67.65%	64.10%
'TP812'	65.20±1.2%	69.09%	74.29%	63.14%	61.99%	64.52%	60.63%
'TP824'	67.58±1.2%	71.74%	78.16%	64.77%	64.55%	66.49%	63.04%
'TPZ01'	68.24±1.0%	73.57%	79.88%	65.90%	65.13%	67.24%	64.15%
'TPZ12'	68.48±1.3%	74.09%	80.99%	67.11%	64.08%	67.37%	63.98%
'TPZ24'	66.26±0.7%	71.05%	75.33%	63.72%	63.82%	65.76%	61.65%
'NP301'	67.45±0.6%	73.10%	78.82%	64.65%	64.32%	66.47%	62.61%
'NP312'	68.80±1.7%	72.48%	79.48%	66.33%	65.53%	67.75%	64.29%
'NP324'	69.10±0.9%	72.76%	79.76%	65.88%	66.55%	68.46%	64.25%
'NP401'	69.19±1.2%	72.42%	80.78%	67.97%	64.79%	68.13%	65.02%
'NP412'	67.16±1.1%	71.80%	77.99%	64.39%	64.24%	66.44%	62.42%
'NP424'	67.21±0.4%	72.05%	78.93%	64.49%	64.14%	66.21%	62.44%
'NP701'	67.13±1.0%	71.55%	80.42%	64.99%	63.43%	66.13%	62.52%
'NP712'	67.58±1.1%	71.91%	78.90%	64.99%	64.56%	66.65%	63.03%
'NP724'	68.64±1.4%	71.75%	76.75%	66.75%	64.54%	67.25%	64.23%
'NP801'	67.46±1.0%	71.93%	78.56%	64.48%	64.91%	66.57%	63.15%
'NP812'	67.47±1.6%	70.94%	78.61%	66.79%	62.47%	66.48%	62.93%
'NP824'	70.12±0.7%	73.74%	78.31%	66.02%	68.71%	69.59%	65.41%
'NPZ01'	68.37±0.9%	73.67%	81.40%	66.38%	64.51%	67.36%	63.82%
'NPZ12'	67.43±1.7%	72.09%	78.59%	65.03%	64.23%	66.51%	62.96%
'NPZ24'	67.92±0.8%	71.66%	76.46%	65.26%	65.26%	67.29%	63.26%

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

Table 26 details results similar to the previous test, with the majority of the feature sets listed surpassing 66% diagnostic accuracy. While novel tone feature set performances appear to surpass the comparable target stimuli sets, the accuracies are close enough, with a few within confidence intervals, to make this a statistically insignificant difference. Table 27 shows the combination performances.

Table 27 – 30 FS: Satisfactory + Marginal Data Fusion Performances

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	85.7±3.3	92.8	83.2±2.9	81.8±3.1	80.3±3.0	81.2±2.7
FS2	85.9±3.2	96.5	81.4±3.1	82.5±3.0	83.1±3.1	80.8±3.6
FS3	85.3±2.4	92.8	83.0±2.7	81.2±3.9	82.0±2.7	80.9±2.5
FS4	81.5±2.8	92.8	79.8±2.9	78.5±3.0	80.1±3.9	77.9±3.1
FS5	**86.1±3.2**	**95.5**	**84.3±3.2**	**80.2±2.7**	**80.9±2.9**	**81.3±3.2**
SMV	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	71.2±3.2	80.5	70.5±2.7	70.9±3.2	69.9±3.0	70.0±3.6
FS2	**79.2±2.9**	85.4	74.3±3.5	73.6±2.7	73.5±3.2	75.1±3.7
FS3	78.6±2.9	85.4	75.0±3.6	70.2±3.0	73.2±3.1	72.8±2.5
FS4	72.3±2.8	77.8	69.5±2.8	69.9±3.1	70.1±3.3	68.5±2.9
FS5	78.6±3.4	85.4	74.2±2.9	73.6±4.0	72.9±3.6	70.5±2.5

This test constituted the highest overall accuracy attained for any of the EEG evaluations across all cohorts at 86.1±3.2% with the FS5 combination list as shown in Table 27. The sum rule continued to outperform the simple majority voting method for classifier combination in both tests.

4.1.1d RANDOM LABEL TESTING

The final EEG test focused on random label performance evaluations. Such performance metrics are important in determining the algorithm's susceptibility to overfitting[12] on training data. The ASG algorithm was tested with random labels assigned to the satisfactory cohort, in addition to the full Cohort A dataset to further evaluate the system's tolerance to overfitting. The Cohort A dataset contained 71 (34 AD, 37CN) subjects with 19 electrodes.

The basic procedure for the random label testing began with a random assignment of class labels for all subjects. ASG training was performed on the data, with five total

[12] Overfitting occurs in a classification system when the classifier is over-trained on data, essentially learning noise. A random label test that produces performance around random chance is desirable as it indicates little to no occurrence of overfitting.

iterations per feature set, as opposed to the 10 in normal testing. This entire process was repeated six times for different training and testing folds of the data, creating a 6-fold cross validation setup. The individual feature set and combination performances for both datasets used in this testing procedure are shown in the following tables, similar to the presentation of previous cohorts and combination sets.

Table 28 – Random Label Performances (Cohort A – 71 subjects)

AVERAGED TRIALS							
'FS'	OGP	GP5	GP	SN	SP	PPV	NPV
'TF812'	50.24%	56.05%	67.38%	52.50%	49.43%	50.01%	51.58%
'TFP212'	47.08%	52.02%	61.03%	47.56%	47.38%	46.14%	48.29%
'TPZ01'	50.14%	52.48%	59.84%	48.53%	52.77%	50.27%	50.94%
'NPZ12'	52.02%	57.57%	71.43%	50.22%	55.42%	52.83%	52.59%
'NT812'	53.22%	58.57%	68.57%	52.10%	54.35%	52.84%	53.93%
'NPZ24'	48.87%	52.24%	60.24%	49.66%	50.78%	49.72%	50.78%
'NCZ24'	49.54%	52.92%	60.95%	50.04%	49.68%	48.80%	50.96%
'NCZ12'	53.86%	59.97%	68.81%	53.67%	54.94%	54.37%	54.86%
'NOZ12'	49.81%	56.21%	65.63%	49.37%	51.17%	49.63%	50.11%
'TCZ24'	49.20%	53.00%	63.81%	46.51%	52.24%	49.15%	49.67%
'TPZ12'	47.76%	54.33%	62.14%	46.97%	49.68%	46.91%	49.41%
'TP312'	48.62%	52.75%	61.43%	46.66%	50.86%	48.44%	49.24%
'NFZ24'	48.25%	53.27%	64.29%	47.87%	49.19%	47.67%	49.33%
'TP324'	47.23%	53.48%	65.24%	45.33%	50.44%	46.77%	49.11%
'TPZ24'	47.10%	50.73%	61.35%	46.19%	48.46%	46.50%	48.01%
'NPZ01'	50.72%	54.10%	67.62%	48.17%	53.53%	49.09%	52.60%

```
FS1: NCZ12, NCZ24, NPZ24, NT812, NPZ24, TFP212, TPZ01, TF812, TP324
FS2: NPZ24, NPZ24, NCZ24, NT812, NCZ12, TFP212
FS3: TF812, TPZ01, NCZ12, NPZ24, NT812, NPZ12, TFP212
```

Table 29 – Random Label Data Fusion Performances (Cohort A)

SUM	Avg (%)	Best (%)
FS1	53.09	60.11
FS2	43.17	53.11
FS3	48.91	59.03

SMV	Avg (%)	Best (%)
FS1	49.70	59.03
FS2	44.01	55.03
FS3	53.13	61.71

The process was repeated for the satisfactory dataset. Table 30 shows the individual feature set random label performances for this dataset, with Table 31 listing

the combination performances using both the sum rule and simple majority voting, as performed for the Cohort A random label tests.

Table 30 – Random Label Performances (Satisfactory Dataset – 107 subjects)

AVERAGED TRIALS							
FS	OGP	GP5	GP	SN	SP	PPV	NPV
'TP312'	50.87%	53.79%	61.04%	51.58%	49.85%	50.28%	51.20%
'NFZ24'	51.35%	52.86%	66.71%	50.57%	51.30%	50.45%	51.42%
'TP324'	51.55%	55.63%	63.82%	51.35%	49.76%	50.35%	51.01%
'TPZ24'	51.45%	55.11%	68.67%	50.01%	51.87%	50.71%	51.41%
'NPZ01'	53.23%	56.44%	64.86%	50.41%	54.42%	51.69%	52.77%
'TF312'	51.95%	55.27%	60.12%	50.58%	52.23%	51.04%	51.90%
'TFP212'	50.47%	54.02%	59.25%	46.88%	52.25%	49.49%	49.90%
'TPZ01'	52.20%	55.13%	63.00%	50.37%	52.25%	50.36%	51.81%
'NPZ12'	50.22%	54.70%	65.73%	48.52%	50.47%	48.94%	49.90%
'NT812'	50.05%	52.88%	61.04%	49.43%	49.25%	49.30%	48.98%
'NPZ24'	50.16%	53.57%	61.97%	48.44%	50.70%	48.76%	50.46%
'NCZ24'	52.39%	55.87%	65.78%	53.20%	50.57%	51.06%	52.43%
'NCZ12'	50.46%	54.54%	63.88%	50.63%	48.74%	48.94%	50.27%
'NOZ12'	49.91%	52.63%	60.12%	48.08%	50.37%	48.30%	50.23%
'TCZ24'	50.97%	55.48%	63.00%	51.59%	50.21%	50.53%	51.35%
'TPZ12'	53.29%	57.20%	66.87%	50.43%	54.57%	52.45%	53.23%

```
FS1: NCZ12, NCZ24, NPZ24, NT812, NPZ24, TFP212, TPZ01, TP312, TP324
FS2: NPZ24, NPZ24, NCZ24, NT812, NCZ12, TFP212
FS3: TP312, TPZ01, NCZ12, NPZ24, NT812, NPZ12, TFP212
```

Table 31 – Random Label Data Fusion Performances (Satisfactory Dataset)

SUM	Avg (%)	Best (%)
FS1	53.12	61.58
FS2	51.26	63.81
FS3	49.16	58.12

SMV	Avg (%)	Best (%)
FS1	49.15	55.79
FS2	51.22	59.46
FS3	47.55	53.47

It is clear from these results that the system is generally around random chance performance, indicating little to no overfitting on both cohorts of EEG data. Both individual and fusion performances for Cohort A and the satisfactory dataset are within an acceptable range of random chance.

4.1.2 MRI (VOLUMETRIC ANALYSIS)

The following performance figures are derived from the 136 subject MRI-only dataset, normalized to inter-cranial volume for each patient. This dataset includes all available volumetric MRI scans from AD and CN patients, some with multiple visits, constituting a total of 79 AD and 57 CN subjects. Table 32 lists the basic statistics for this dataset.

Table 32 – MRI Cohort Statistics

	Male	Female		MMSE	Age
AD	30	49	**79**	18.79	75.02
CN	15	75	**57**	28.96	71.01
Totals	45	91	**136**		

Data was normalized to inter-cranial volume for each patient to balance all features for this cohort. A random subspace sampling method was employed to randomly pick 18 features from the total available of 28 for the 16 total feature sets[13].

Table 33 – MRI Cohort Feature Set Performances

FS	OGP	GP5	GP	SN	SP	PPV	NPV
1	83.76±2.2%	88.91%	94.36%	79.31%	76.23%	84.39%	71.21%
2	86.80±2.1%	90.63%	97.10%	82.52%	78.84%	86.87%	74.14%
3	85.14±1.9%	88.94%	96.17%	80.59%	77.83%	85.51%	72.61%
4	88.65±2.3%	92.76%	98.50%	84.31%	80.49%	88.39%	76.25%
5	86.36±2.3%	89.62%	95.12%	82.16%	78.11%	86.20%	73.81%
6	84.51±2.0%	88.31%	94.36%	80.32%	75.70%	84.65%	71.75%
7	86.75±2.1%	91.72%	98.91%	82.71%	78.66%	86.46%	74.45%
8	81.85±2.0%	87.95%	93.08%	76.87%	75.22%	83.04%	68.79%
9	86.48±2.5%	92.03%	99.49%	81.54%	80.04%	87.11%	74.01%
10	86.76±2.6%	92.32%	97.10%	82.38%	79.60%	87.06%	74.40%
11	84.28±2.7%	88.64%	96.64%	78.50%	77.76%	84.94%	71.22%
12	82.59±2.7%	87.27%	91.68%	78.25%	74.46%	83.02%	69.47%
13	86.42±2.2%	90.25%	95.47%	82.33%	77.99%	86.68%	73.62%
14	84.10±2.1%	88.93%	94.36%	79.24%	77.12%	84.79%	71.30%
15	83.79±2.4%	89.41%	97.63%	79.25%	76.36%	84.23%	71.13%
16	87.31±2.5%	91.44%	97.39%	82.08%	80.88%	87.59%	74.97%
AVG	**85.35%**	**89.95%**	**96.09%**	**80.77%**	**77.83%**	**85.68%**	**72.70%**

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

[13] This number was chosen to match the number of feature sets used in EEG analysis, such that the ASG expert combination system would be similar across datasets to ensure fairness and equal voting.

In Table 33, the average performance reached for individual feature set accuracy was 85.35%. Three overall feature set combinations were created through a random combination of the 16 individual feature sets. Additionally, all 16 available feature sets were combined and analyzed.

```
FS1: 01, 04, 06, 02, 09, 10
FS2: 03, 11, 16, 15, 08, 05
FS3: 07, 14, 13, 12, 02, 04
```

Table 34 – 16 FS: MRI Cohort Data Fusion Performances

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	87.2±3.2	92.2	84.3±2.9	83.6±3.1	85.3±2.9	84.4±2.1
FS2	88.6±3.4	93.2	87.2±2.3	84.4±2.9	82.7±3.1	82.5±3.0
FS3	89.1±3.1	94.0	87.6±3.0	84.9±2.5	82.2±3.0	85.0±3.3
ALL	**89.5±3.0**	94.2	87.5±3.1	85.0±2.7	83.1±3.1	85.1±3.0
SMV	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	89.2±3.0	94.2	87.3±2.4	85.4±2.6	87.1±3.3	85.4±2.8
FS2	87.8±3.2	95.3	88.8±3.0	85.0±3.0	88.2±2.9	86.3±2.7
FS3	90.0±2.9	96.9	90.0±2.9	85.9±3.1	88.0±3.0	85.9±2.8
ALL	**91.1±3.2**	**97.0**	**90.5±3.0**	**86.1±3.2**	**88.0±2.8**	**86.2±2.9**

In Table 34, it is clear that the simple majority voting rule tended to perform the best in combining the feature sets, with all 16 proving to achieve the highest overall accuracy at 91.1±3.2%. Unlike EEG, simple majority voting achieved a slightly higher diagnostic accuracy in comparison to the sum rule combination method.

Much like the EEG dataset tests performed, random label testing was done to determine if overfitting was occurring with the MRI implementation of ASG. The method of random subspace sampling was used to create individual feature sets that were later combined for decision fusion. 16 feature sets were generated. Random class labels were used for all subjects and the ASG algorithm was then trained and tested with this information. Random feature set combinations were used to evaluate the ensemble. The process taken for the random label testing was the same as done in EEG—randomization

of class labels, five iterations of ASG, with this entire process repeated six times to form a six fold cross-validation setup. Table 35 and Table 36 show the individual feature set accuracy as well as the decision fusion accuracy.

Table 35 – Random Label Testing (MRI-only Cohort – 136 subjects)

	AVERAGED TRIALS						
FS	**OGP**	**GP5**	**GP**	**SN**	**SP**	**PPV**	**NPV**
1	49.73%	52.78%	55.46%	50.25%	52.46%	54.76%	46.77%
2	51.24%	53.45%	59.42%	49.56%	53.14%	48.24%	46.57%
3	50.39%	52.42%	60.11%	54.20%	48.44%	53.22%	53.24%
4	48.55%	50.32%	60.23%	45.66%	42.55%	58.64%	55.71%
5	58.42%	60.12%	62.54%	53.21%	51.44%	55.14%	54.63%
6	42.98%	46.80%	58.46%	51.79%	53.26%	52.23%	50.19%
7	51.24%	53.12%	60.24%	49.54%	58.97%	44.68%	49.55%
8	55.10%	57.81%	60.71%	48.66%	55.46%	49.68%	49.68%
9	43.12%	46.21%	56.13%	43.25%	42.15%	53.24%	48.20%
10	48.55%	50.19%	57.21%	50.24%	48.59%	53.16%	47.54%
11	50.10%	52.17%	56.23%	53.66%	49.97%	54.55%	58.21%
12	46.70%	49.52%	52.13%	58.46%	52.46%	54.14%	43.06%
13	59.40%	60.21%	63.44%	51.24%	55.79%	43.25%	42.46%
14	60.10%	62.11%	65.79%	56.87%	48.57%	47.54%	55.79%
15	45.28%	48.56%	55.61%	57.77%	41.23%	40.36%	57.45%
16	54.50%	55.60%	60.54%	50.11%	56.24%	42.36%	59.60%
AVG	50.96%	53.21%	59.02%	51.53%	50.67%	50.32%	51.17%

Table 36 – Random Label Data Fusion Performances (MRI-only cohort)

SUM	Avg (%)	Best (%)
FS1	54.11	65.97
FS2	50.12	66.15
FS3	49.58	63.24
ALL	51.00	61.02

SMV	Avg (%)	Best (%)
FS1	51.24	63.12
FS2	53.15	67.45
FS3	50.33	62.10
ALL	54.10	66.79

Individual feature set accuracy averaged 50.96%, with fusion performances generally between 49% and 54%, indicating little to no overfitting as these numbers fall within the random chance range.

4.1.3 PET (METABOLIC ANALYSIS)

The performance figures listed in the following tables come from the 80 subject PET-only dataset. All available PET scans from AD and CN patients are used, again with some patients having multiple scans, creating a total of 37 AD and 43 CN subjects. Table 37 lists relevant subject demographics for this cohort.

Table 37 – PET Cohort Statistics

	Male	Female		MMSE	Age
AD	9	28	**37**	18.98	74.15
CN	14	29	**43**	28.88	70.79
Totals	23	57	**80**		

Data was normalized to pons as well as by z-scoring. Both sets of data were used in testing procedures to determine which normalization method provided the best overall generalization performance with the ASG algorithm. Once again, a random subspace sampling method was used to choose 36 features from the total 43 for diversity.

Table 38 – PET Cohort Feature Set Performances (pons)

'FS'	'OGP'	'GP5'	'GP'	'SN'	'SP'	'PPV'	'NPV'
1	84.39±4.1%	92.20%	98.49%	82.92%	94.19%	89.51%	82.48%
2	81.20±3.9%	87.14%	96.59%	78.54%	86.58%	80.77%	84.79%
3	80.45±3.7%	86.93%	96.30%	76.32%	88.07%	80.86%	84.24%
4	83.90±4.1%	92.13%	98.81%	84.91%	92.54%	88.80%	89.93%
5	83.31±4.2%	91.82%	97.58%	84.15%	90.96%	88.91%	88.48%
6	82.50±4.0%	95.00%	97.56%	80.69%	91.08%	85.63%	88.32%
7	81.23±3.8%	89.50%	95.20%	80.57%	90.79%	84.66%	87.19%
8	84.61±3.9%	91.62%	98.20%	83.44%	93.83%	88.14%	90.16%
9	81.22±3.7%	87.41%	95.96%	76.67%	86.49%	79.60%	82.93%
10	80.51±4.1%	88.16%	96.64%	80.36%	89.53%	83.72%	86.53%
11	82.48±4.0%	91.73%	97.90%	81.84%	90.85%	86.19%	87.85%
12	82.71±4.1%	89.94%	95.40%	82.33%	91.41%	86.22%	88.60%
13	80.94±3.8%	87.65%	93.74%	80.33%	89.39%	84.85%	86.47%
14	79.97±3.9%	89.19%	94.20%	79.97%	84.13%	78.54%	84.81%
15	80.93±3.5%	90.59%	95.19%	79.63%	89.85%	85.32%	86.49%
16	83.76±4.0%	96.54%	98.18%	84.33%	82.48%	88.00%	89.24%
AVG	**82.13%**	**90.47%**	**96.63%**	**81.07%**	**89.52%**	**84.97%**	**86.73%**

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

As Table 38 illustrates, the average performance for individual feature sets was 82.13%. The confidence intervals for this dataset however show a larger range than that of the MRI results. Three overall feature set combinations were created through a random combination of the 16 individual feature sets. Additionally, all 16 available feature sets were combined and analyzed.

```
FS1: 04, 07, 09, 11, 15, 16
FS2: 01, 05, 06, 10, 13, 14
FS3: 02, 03, 05, 08, 12, 16
```

Table 39 – 16 FS: PET Cohort Data Fusion Performances (pons)

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	86.3±3.9	91.3	83.2±3.9	80.1±3.0	85.4±3.5	86.1±3.9
FS2	88.9±3.1	93.6	85.3±4.1	82.4±4.0	84.0±3.7	83.9±3.6
FS3	88.2±3.7	93.9	86.7±2.9	85.1±3.9	81.2±3.6	82.9±3.5
ALL	88.5±3.8	94.0	86.4±2.8	85.2±3.8	81.0±3.4	83.0±3.2
SMV	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	87.4±2.8	92.8	86.0±2.7	84.6±3.9	83.5±2.9	83.7±3.0
FS2	89.3±3.6	95.6	88.3±4.0	83.0±2.7	86.7±2.7	87.0±3.1
FS3	90.0±4.0	96.8	88.6±3.7	81.9±3.8	85.0±3.1	84.4±2.9
ALL	90.0±3.5	96.6	89.0±3.8	82.0±3.5	84.9±3.2	84.2±3.0

The highest overall accuracy attained through pons normalization of the PET data was 90.0±3.5% as shown in Table 39. Simple majority voting once again was the better combination method, as was the case in the MRI dataset. Next, the same dataset was normalized through z-scoring with respect to the normal control's data and evaluated through the same process as before. The individual feature set runs are shown in Table 40 with data fusion combination accuracies listed in Table 41.

Table 40 – PET Cohort Feature Set Performances (z-scoring)

FS	OGP	GP5	GP	SN	SP	PPB	NPV
1	85.29±2.2%	87.92%	96.77%	80.46%	77.38%	85.54%	72.36%
2	84.22±2.3%	86.57%	96.20%	83.67%	79.99%	88.02%	75.29%
3	87.26±2.9%	90.78%	98.10%	81.74%	78.98%	86.66%	73.76%
4	80.36±3.1%	84.26%	93.58%	85.46%	81.64%	89.54%	77.40%
5	82.24±2.8%	86.57%	93.45%	83.31%	79.26%	87.35%	74.96%
6	83.29±2.9%	86.79%	93.25%	81.47%	76.85%	85.80%	72.90%
7	80.48±3.1%	83.55%	90.28%	83.86%	79.81%	87.61%	75.60%
8	80.58±4.0%	86.79%	92.68%	78.02%	76.37%	84.19%	69.94%
9	83.47±4.3%	93.64%	98.79%	82.69%	81.19%	88.26%	75.16%
10	86.99±3.0%	94.22%	99.85%	83.53%	80.75%	88.21%	75.55%
11	85.48±3.6%	90.28%	97.25%	79.65%	78.91%	86.09%	72.37%
12	85.39±3.4%	96.00%	98.99%	79.40%	75.61%	84.17%	70.62%
13	86.24±3.8%	95.20%	97.60%	83.48%	79.14%	87.83%	74.77%
14	84.22±3.9%	90.00%	95.69%	80.39%	78.27%	85.94%	72.45%
15	85.00±3.9%	92.10%	97.52%	80.40%	77.51%	85.38%	72.28%
16	86.70±3.8%	93.99%	98.30%	83.23%	82.03%	88.74%	76.12%
AVG	**84.20%**	**89.92%**	**96.14%**	**81.92%**	**78.98%**	**86.83%**	**73.85%**

Acronyms: ***OGP*** *(overall group performance),* ***GP5*** *(best of 5 trial group performance)*
GP *(best group performance),* ***SN*** *(sensitivity),* ***SP*** *(specificity)*
PPV *(positive predictive value),* ***NPV*** *(negative predictive value)*

A comparable average accuracy to pons normalization was obtained through z-scoring, averaging 84.20% as shown in Table 40. Once again, three overall feature set combinations were created through a random combination of the 16 individual feature sets. All 16 available feature sets were combined and analyzed.

```
FS1: 01, 04, 08, 10, 11, 16
FS2: 02, 06, 07, 09, 14, 15
FS3: 03, 05, 07, 08, 12, 13
```

Table 41 – 16 FS: PET Cohort Data Fusion Performances (z-scoring)

SUM	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
FS1	87.3±3.9	92.5	82.7±2.7	81.0±3.5	84.7±3.0	87.9±2.9
FS2	89.0±3.6	94.0	84.9±3.1	80.9±3.9	85.3±2.7	86.8±3.1
FS3	90.7±3.9	95.0	86.2±4.0	83.1±3.7	84.9±2.8	85.7±3.6
ALL	**90.9±3.6**	95.0	86.3±2.7	83.5±2.9	85.0±3.0	86.0±3.2
SMV	**Avg (%)**	**Best (%)**	**SN (%)**	**SP (%)**	**PPV (%)**	**NPV (%)**
FS1	88.6±3.5	93.6	85.7±3.1	85.7±4.0	84.6±2.9	85.7±3.1
FS2	91.0±3.8	95.8	87.9±3.9	86.7±3.6	88.8±3.0	87.9±3.5
FS3	90.8±3.6	96.0	88.1±3.6	86.2±3.7	87.9±3.4	88.0±3.6
ALL	**91.2±3.9**	**96.2**	**88.0±3.5**	**86.9±3.1**	**88.5±3.1**	**87.9±3.7**

In Table 41, we can see that both the sum rule and simple majority voting perform within each other's confidence intervals, each reaching around 91% diagnostic accuracy. Through this analysis, it is shown that z-scoring showed slight improvement over pons normalization, albeit by a statistically insignificant margin. Once again, random label testing was performed to determine whether the system was overfitting on the PET data. The procedure followed was done exactly as for EEG and MRI, which is again outlined below for reference. Random subspace sampling was again used to create individual feature sets for later combination, with 16 in total generated. Random class labels were assigned for all subjects, and the ASG algorithm was trained and tested on this setup. Random feature set combinations were then used to evaluate the ensemble performance.

Table 42 – Random Label Testing (PET-only Cohort – 83 subjects)

	AVERAGED TRIALS						
FS	**OGP**	**GP5**	**GP**	**SN**	**SP**	**PPV**	**NPV**
1	59.46%	62.54%	69.50%	49.63%	49.78%	59.66%	66.57%
2	58.45%	62.99%	66.37%	50.25%	59.22%	57.45%	60.24%
3	60.29%	63.45%	64.25%	59.46%	42.89%	60.12%	58.22%
4	61.25%	65.88%	70.52%	43.25%	56.12%	48.45%	62.14%
5	55.69%	63.25%	75.25%	46.25%	57.56%	60.00%	40.22%
6	63.57%	68.45%	69.22%	56.78%	49.00%	48.99%	46.55%
7	61.25%	63.56%	66.39%	56.00%	50.12%	50.00%	52.55%
8	68.77%	69.00%	75.47%	59.00%	50.00%	52.46%	53.87%
9	53.98%	60.28%	63.25%	63.10%	43.44%	44.50%	59.44%
10	56.78%	63.14%	66.34%	57.48%	49.55%	46.87%	47.54%
11	59.45%	60.22%	65.87%	56.25%	50.24%	60.25%	59.45%
12	52.15%	59.78%	63.24%	51.02%	56.24%	50.21%	44.44%
13	50.23%	56.45%	66.12%	50.28%	57.11%	52.32%	50.29%
14	56.21%	59.45%	66.75%	55.47%	60.20%	56.33%	50.77%
15	53.25%	59.90%	68.11%	48.69%	66.10%	57.24%	49.97%
16	50.29%	58.00%	63.25%	50.12%	59.55%	49.45%	55.55%
AVG	57.57%	62.27%	67.49%	53.31%	53.57%	53.39%	53.61%

Table 43 – Random Label Data Fusion Performances (PET-only cohort)

SUM	Avg (%)	Best (%)
FS1	56.47	63.55
FS2	59.45	69.77
FS3	57.20	68.44
ALL	58.11	70.00

SMV	Avg (%)	Best (%)
FS1	54.87	66.78
FS2	56.79	65.79
FS3	59.99	67.80
ALL	58.79	69.12

The random label testing for PET data showed slight overfitting, with averages reaching the upper 50 percentile. However, this case of overfitting is still slight, and close to random chance.

4.1.4 BIOMARKER PERFORMANCE COMPARISONS

Figure 33 illustrates all the individual biomarker performances with confidence intervals shown as horizontal lines projecting from the group mean. The EEG tests all overlap with each other's confidence intervals at one point. MRI and PET cohorts all overlap and constitute the top end of accuracy in this system, setting these two biomarkers apart from EEG. Only two EEG tests overlapped with the accuracies from PET and MRI, both of which utilized the parietal only, 30FS evaluation. The PET and MRI datasets outperform three of the EEG tests completely, and two EEG tests overlap with PET and MRI.

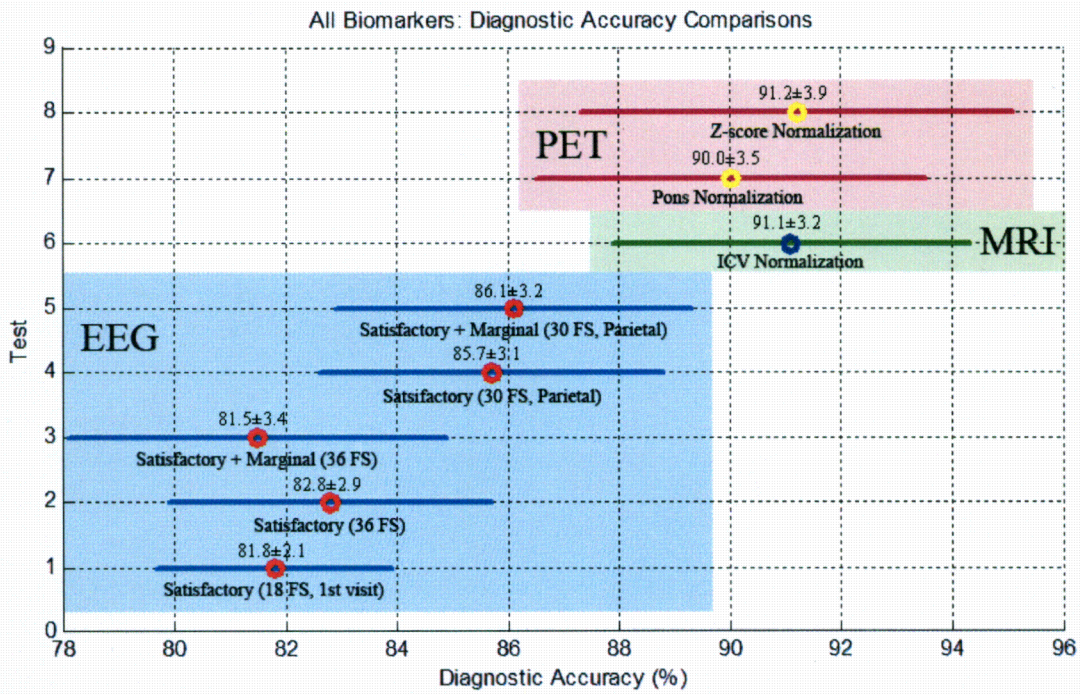

Figure 33 – Individual biomarker performances with confidence intervals

4.2 ENSEMBLE SYSTEM COMBINATION PERFORMANCES

The following section outlines the datasets used as well as the diagnostic accuracies of the different combinations of biomarker experts. Two methods were used to fuse the decisions of each biomarker expert—one employing the usual classifier combination schema (sum rule, simple majority voting, weighted majority voting) and the other utilizing a modified stacked generalization approach, where each expert from each biomarker is used to train another meta-classifier, acting as a final decision expert. This is done between the overall six-fold CV, in order to evaluate on patients the system has not yet seen. A diagram of this process (Figure 34) illustrates the system flow.

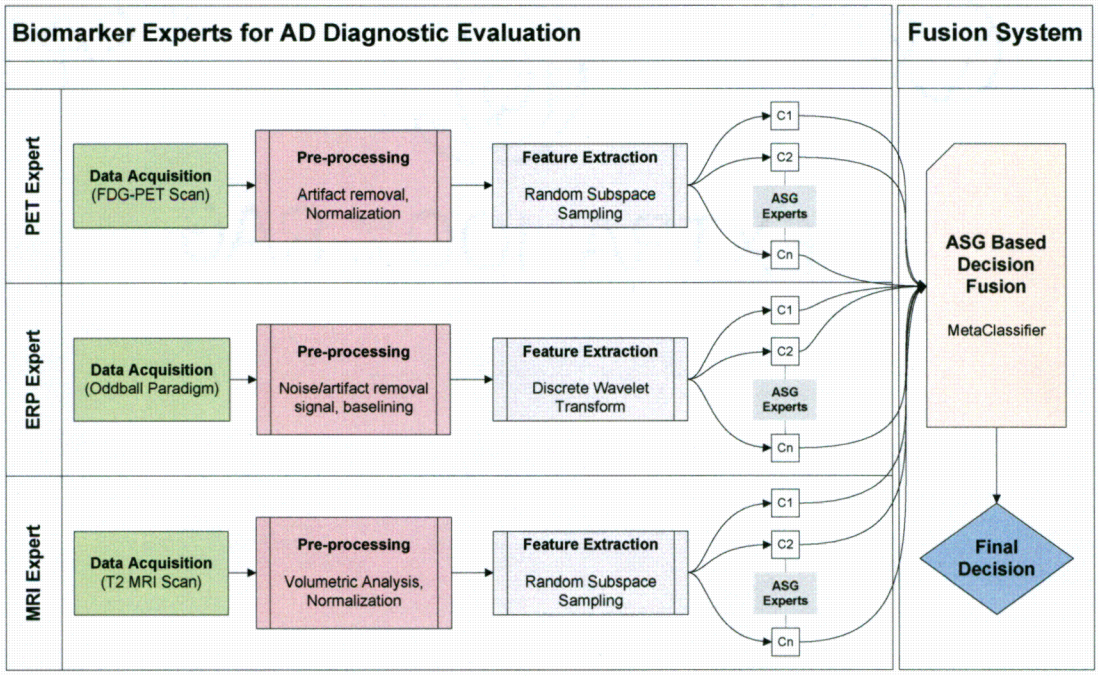

Figure 34 – Meta-classifier as decision fusion for biomarker expert combination

This implementation attempted to learn from the expert ASG outputs, creating a layer of stacked generalization for decision fusion on the final ensemble system's final output.

Each dataset presented shows statistics across all classes for informational purposes; however only AD and CN were compared for this study. For each combination

test, three overall methods were tested: sum rule combination, simple majority voting, weighted majority voting (with weights based on prior performances obtained in individual biomarker evaluation runs), and the use of the stacked generalization meta-classifier for learning and decision making (denoted by 'SG' in the results tables).

An overall average is shown of the combination of several cross-validation runs to ensure the best representation of generalization performance given with a 95% confidence interval. The best overall fold is also presented, along with diagnostic accuracy performance metrics. Furthermore, the 16 feature sets chosen for EEG were selected for diversity[14] and best overall generalization performance, determined through the previous individual EEG biomarker evaluation runs. In general, parietal electrodes tended to perform the best overall individually, and several are chosen because of this. The feature sets used for combination were:

```
Target: TP301, TP312, TP324, TP401, TPZ01, TPZ12, TPZ24, TC412
 Novel: NP324, NP401, NP424, NPZ01, NPZ24, NP801, NC412, NF312
```

4.2.1 EEG+MRI

Table 44 lists the basic cohort statistics (average age and MMSE scores, as well as the corresponding standard deviation) for EEG+MRI combination. All diagnostic classes are shown for reference; however, only CN versus AD diagnostic accuracy was tested.

Table 44 – EEG+MRI Cohort Statistics

		CN	AD	MCI	PD	PDD
Subjects		45	50	64	89	15
Age	Avg.	69.73	74.62	69.80	68.93	75.00
	Std.	9.63	8.50	8.61	6.09	5.53
MMSE	Avg.	29.04	19.39	24.97	27.79	22.38
	Std.	1.04	6.14	3.74	2.96	5.66

[14] Diversity with respect to the number of unique features between each feature set.

Table 45 – EEG+MRI Combination Performance

	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
SUM	91.3±3.1	95.8	88.6±3.0	90.1±2.9	92.7±2.7	91.6±3.0
SMV	92.8±3.0	96.4	89.7±3.0	91.0±3.0	92.9±2.8	90.8±3.2
WMV	93.5±3.2	97.0	92.7±2.7	93.1±3.1	94.0±3.0	93.7±2.9
SG	**94.7±3.1**	**97.4**	**93.5±2.5**	**93.6±3.2**	**93.7±3.1**	**94.0±3.0**

In Table 45, it is clear that the stacked generalization (SG) based method for decision fusion excels; however the weighted majority voting method is statistically comparable as both performances are within each other's confidence intervals.

4.2.2 EEG+PET

Again, Table 46 lists basic cohort statistics, in this case for the EEG+PET only combination cohort. Table 47 details the overall accuracies as well as the various diagnostic performance metrics for each test.

Table 46 – EEG+PET Cohort Statistics

		CN	AD	MCI	PD	PDD
Subjects		33	37	33	42	6
Age	Avg.	70.94	73.84	68.88	69.98	72.83
	Std.	9.02	8.43	8.21	6.51	5.08
MMSE	Avg.	29.00	19.86	25.03	27.26	22.20
	Std.	1.15	5.65	3.70	3.64	6.98

In Table 47, a similar trend is noted from the EEG+MRI cohort. Once again, the SG method outperforms all other decision-fusion techniques, outperforming WMV by a statistically small margin with the sum rule performing the poorest of all.

Table 47 – EEG+PET Combination Performance

	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
SUM	89.7±3.0	93.1	87.0±2.9	86.4±2.7	86.0±3.0	87.3±3.1
SMV	90.0±2.9	93.6	88.5±2.9	87.0±3.0	89.3±3.2	85.4±3.0
WMV	91.2±3.2	94.0	89.6±2.7	90.0±3.1	90.5±3.1	87.9±2.8
SG	**92.3±2.7**	**94.8**	**90.3±3.0**	**91.0±2.8**	**92.4±3.0**	**93.0±2.7**

4.2.3 MRI+PET

This combination dealt only with the MRI+PET combination cohort, with relevant statistics listed in Table 48. This cohort yielded the second most available subjects, with only EEG+MRI containing more. Overall accuracies and diagnostic performance metrics are shown in Table 48.

Table 48 – MRI+PET Cohort Statistics

		CN	AD	MCI	PD	PDD
Subjects		37	41	37	37	4
Age	Avg.	69.73	74.34	70.14	69.22	74.25
	Std.	9.61	8.40	9.03	5.90	3.30
MMSE	Avg.	28.51	19.15	24.73	27.24	23.00
	Std.	1.88	6.27	3.93	3.72	9.54

In Table 49, it can be seen that all performances across each fusion method remain approximately the same, with any small differences statistically insignificant due to confidence intervals. It is interesting to see this, as all previous combinations clearly showed a trend with sum rule performing lowest and SG attaining the highest diagnostic accuracy. However, this is not the case with the MRI+PET cohort.

Table 49 – MRI+PET Combination Performance

	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
SUM	90.7±2.5	93.4	87.2±3.4	85.4±2.8	85.7±2.7	84.2±2.9
SMV	90.9±3.0	93.8	88.5±3.0	87.2±2.9	88.0±2.9	86.9±3.1
WMV	90.6±2.9	93.7	89.4±3.1	90.0±3.5	87.4±2.6	87.0±3.2
SG	90.4±2.7	93.5	90.0±2.7	89.7±3.0	86.4±2.7	88.1±3.0

4.2.4 EEG+MRI+PET

In this final test, the combination of all three biomarkers used in this study was evaluated. The overall cohort yielded 32 normal control and 33 AD-probable subjects. Basic information for the cohort is shown in Table 50.

Table 50 – EEG+MRI+PET Cohort Statistics

		CN	AD	MCI	PD	PDD
Subjects		32	33	29	37	4
Age	Avg.	70.97	73.82	68.83	69.22	74.25
	Std.	9.17	8.67	8.35	5.90	3.30
MMSE	Avg.	28.97	19.56	25.03	27.24	23.00
	Std.	1.15	5.79	3.87	3.72	9.54

Once again, the pattern shown in the EEG+MRI and EEG+PET tests appears in this analysis, as shown in Table 51. It is clear that both SG and WMV tend to perform the best in this study for decision fusion. SG performed the highest overall, with WMV within the confidence interval range to make it just as viable as the best decision method for use in the ASG-based expert combination system.

Table 51 – EEG+MRI+PET Combination Performances

	Avg (%)	Best (%)	SN (%)	SP (%)	PPV (%)	NPV (%)
SUM	91.8±2.8	95.4	90.4±2.7	88.9±2.9	87.6±3.0	87.9±3.1
SMV	92.9±3.0	95.8	91.2±3.1	91.5±2.8	88.9±3.1	89.0±2.7
WMV	94.9±2.9	96.2	92.2±3.0	92.4±2.7	90.0±2.8	89.9±2.9
SG	**95.0±3.1**	**97.2**	**94.5±3.0**	**94.4±2.9**	**92.7±2.7**	**91.3±3.2**

Figure 35 illustrates the overall fusion accuracies for the various combinations, along with confidence intervals for a visual representation of system performance.

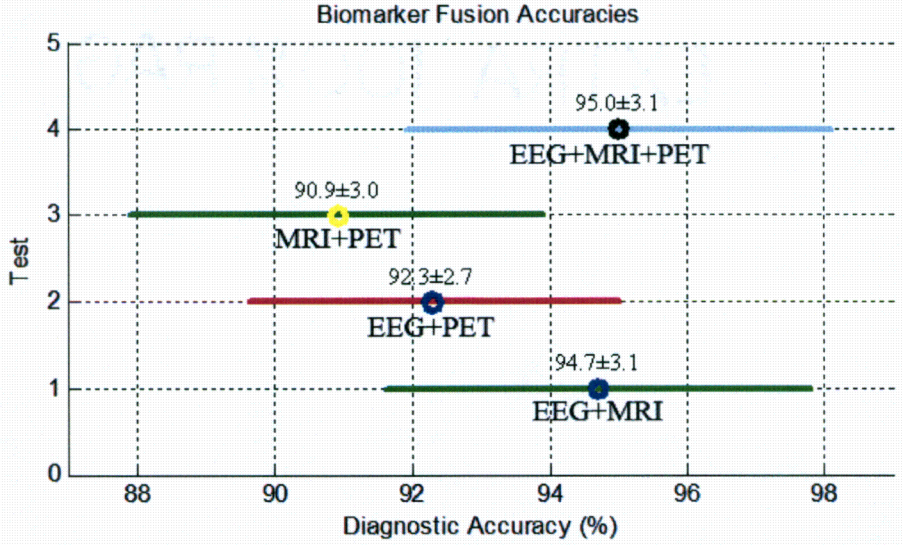

Figure 35 – Biomarker fusion performances with confidence intervals

In Figure 35, it is clear that all four tests have overlapping confidence intervals, with the EEG+MRI and EEG+MRI+PET sharing the greatest overlap. It is clear from these experiments that EEG tends to hold complementary information when combined with either MRI or PET. PET data on the other hand appears to do little when combined with other data types, possibly indicating less complementary information. For comparative purposes, all tests for both individual and data fusion evaluations are illustrated in Figure 36 with confidence intervals.

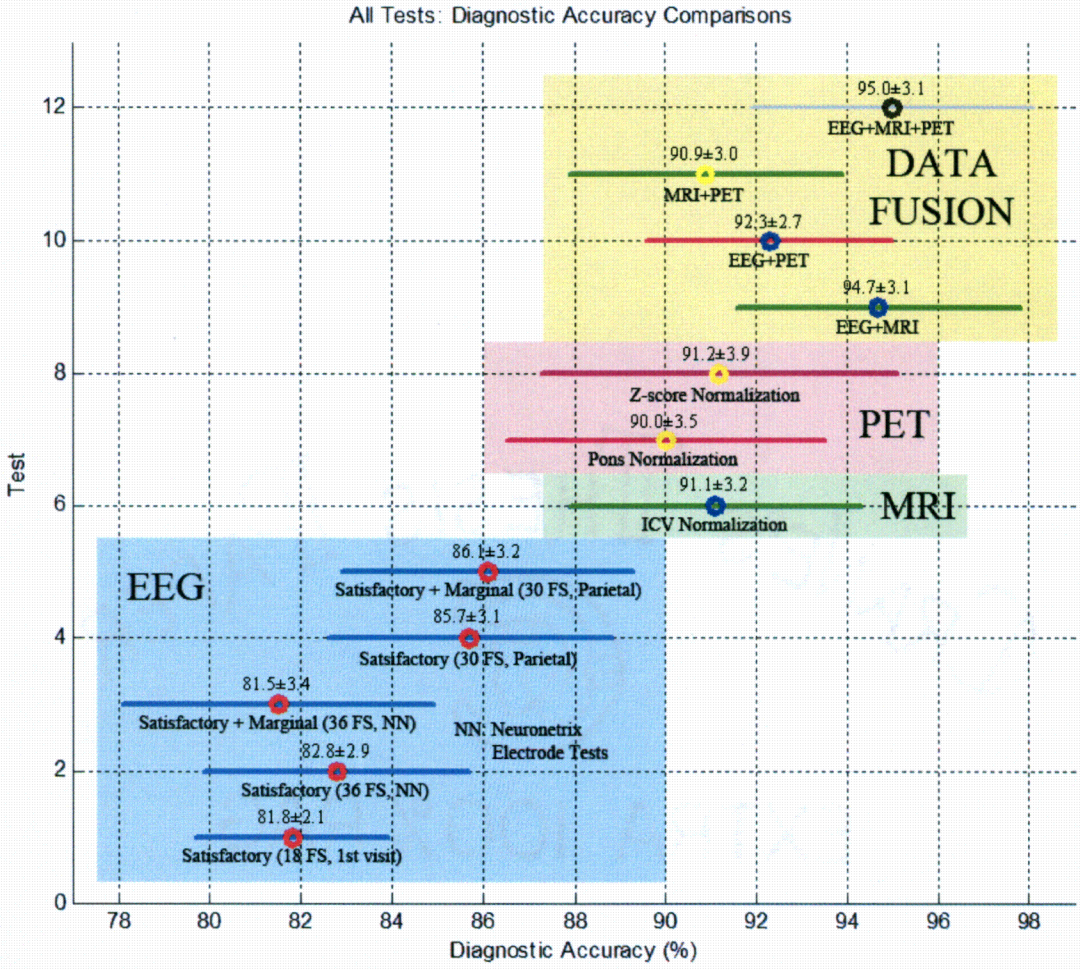

Figure 36 – Individual and data fusion performances with confidence intervals

EEG+MRI and EEG+MRI+PET are the only two data fusion methods that show a statistically significant improvement over parietal-only electrode based EEG evaluations.

4.3 DISEASE SEVERITY ANALYSIS

A simple disease severity analysis was performed to determine the efficacy of the method described in this study in separating mild and severe AD from normal control patients. EEG+MRI+PET data were combined, as this provided the highest overall accuracy in determining patient diagnosis. Individual biomarker performances were also evaluated for comparative use and completeness.

The basic metric used for AD severity distinction in subjects was the use of their respective MMSE scores as a cutoff point. Three cutoffs were used (20, 23, and 26)[15], where AD patients with a score above the cutoff were considered mild/moderate and all others below this value as severe. Modifications were made to the ASG and expert combination algorithm to allow for classification of three classes.

Table 52 shows the subject count breakdown for each cutoff level for the EEG+MRI+PET combination dataset. These subjects were also used in individual biomarker analysis, as opposed to the entire available subject set so a fair comparison could be made between biomarkers and the final combination accuracy. EEG feature set combinations were chosen identical to those used in the combination analysis from the previous section.

Table 52 – AD Severity Analysis Subject Counts

MMSE Cutoff	CN	Mild AD	Severe AD
20	32	22	11
23		12	21
26		7	26

[15] Recall: MMSE scores are a subjective scoring that rates a subject's cognitive ability. A perfect score is a 30. Scores lower than 26-28 generally indicate cognitive impairment, caused by diseases such as AD.

Three combination methods were tested for this setup: simple majority voting (SMV), weighted majority voting (WMV), as well as the stacked generalization based meta-classifier for decision making (SG).

Note that the SG decision analysis was only performed for the combination set, as the algorithm was not adapted to handle individual biomarker expert training. The performances shown in Table 53 are averaged from multiple cross validation runs, and represent the best estimate in system generalization performance for severity analysis.

Table 53 – AD Severity Analysis Performance

Cutoff: 20	EEG	MRI	PET	EEG+MRI+PET
SMV	57.9±2.1	53.0±2.7	56.3±2.9	65.1±2.7
WMV	60.3±3.2	56.9±3.1	56.4±2.4	66.7±3.0
SG	--	--	--	68.2±2.9

Cutoff: 23	EEG	MRI	PET	EEG+MRI+PET
SMV	63.4±2.9	60.5±2.7	60.8±2.8	66.1±3.0
WMV	66.7±2.7	63.1±260	62.9±3.0	69.4±2.7
SG	--	--	--	70.9±3.0

Cutoff: 26	EEG	MRI	PET	EEG+MRI+PET
SMV	60.2±2.7	58.4±3.0	56.9±3.2	62.5±2.6
WMV	59.4±2.5	57.0±3.1	56.7±2.6	63.4±3.2
SG	--	--	--	67.8±3.0

While performances are not as high as would be generally desired, they do exceed random chance by approximately double. In all cases of cutoff values, it is clear that the combination of all three available biomarkers yielded the highest diagnostic accuracy. Detailed discussion and interpretation of all results in this section, as well as preceding results, are provided in chapter 5 along with a summary of accomplishments in addition to recommendations for future work that can be performed on the data used for this study.

CHAPTER V

CONCLUSIONS

The main goal of this project was to expand and improve upon previous efforts in the development of an automated classification system for the diagnosis of neurodegenerative diseases. Specifically, this study focused on the diagnosis of AD-probable patients only. Previous studies relied upon primarily EEG data, with the exception of the most recent which implemented an ensemble system capable of fusing EEG and MRI data to aid in greater diagnostic accuracy. The current study incorporated the use of three biomarkers overall: EEG, MRI, and PET, which have complementary information for this diagnosis method. The use of these biomarkers as individual sources for classification information was evaluated, along with all possible combinations. Additionally, severity analysis for AD was performed, based on various cutoff values determined through patient MMSE scores. A decision-based data fusion approach was used by implementing variations on the augmented stacked generalization algorithm for ensemble system analysis.

5.1 SUMMARY OF ACCOMPLISHMENTS

Prior to the last study completed in this project, all previous research focused on the most optimal feature sets to diagnose AD using ERPs from EEG data only. Much of the prior work dealt with exhaustive searches for optimal electrode, stimuli, and frequency band combinations (from wavelet decomposition of the EEG signal) that yielded the highest

overall diagnostic accuracy. Additionally, work was introduced in the development and improvement of various ensemble systems to maximize the information available from EEG to realize even higher accuracy. The most recent work prior to this study added a heterogeneous data source to provide complementary information to aid in classification. The inclusion of MRI data proved to be highly beneficial to this process; used in conjunction with matching EEG data, classifications rates entered the low to mid 90 percentile, exceeding any single ERP analysis from previous studies.

As previous studies sought optimal electrode and frequency band combinations, along with the first attempt at data fusion with an additional heterogeneous data source, this present study built upon this foundation by expanding upon these optimal combinations, along with the addition of a third data source. The use of PET data allowed a third data source for classification, creating three expert biomarkers that could be used for automated diagnosis of AD. Specifically, the EEG expert was modified to allow for even greater flexibility in multi-input features, permitting a wide array of frequency band, electrode, and stimulus combinations to be created.

In general, it has been shown that the parietal lobe electrodes tend to provide the best information regarding AD diagnosis, as feature sets containing these electrodes consistently performed higher. The best overall combination in an EEG-only biomarker expert yielded a performance of 86.1±3.2%. The feature sets used in this combination were TP301, NP824, TP312, NPZ01, TP401, NP401, TP801, and NP324. This performance not only surpasses the accuracies reached in previous studies, but far outperforms the diagnostic accuracy rate of community clinics, generally accepted at 75% [8].

The MRI-only expert was implemented in a different manner than the previous study. As opposed to using a multi-tiered leave one out scenario for training, testing, and validation, a modified augmented stacked generalization algorithm was implemented, similar to that of the EEG expert. Random subspace sampling was used to choose 18 features from the available 28 in the MRI data, and 16 overall feature sets were created in this manner (to match the number of feature sets used in the EEG expert). The highest overall accuracy for the MRI expert was 91.1±3.2%, which exceeded the previous study highest MRI accuracy by a statistically irrelevant amount of percentage points. The previous study MRI accuracy peaked at 88.1±2.3%. The confidence interval of this current study's performance along with the previous overlaps; therefore a statistically significant performance increase is not realizable.

PET data from the CERND cohort was not previously analyzed; therefore, the results presented in this study are novel and have no previous work to compare to. A similar feature extraction method implemented for the MRI data was utilized for the PET data, in which a random subspace sampling technique chose 36 of the available 43 features to aid in feature set diversity. Again, 16 feature sets were created to match that of the EEG and MRI experts for later overall ensemble combination. The highest overall accuracy for this expert peaked at 91.2±3.9% with z-scoring used as the normalization method. This diagnostic accuracy puts this data source on par with the performance of the MRI expert.

It should be noted that random label testing was performed for all biomarker experts, in order to determine if overfitting was occurring. In all cases, overfitting did not appear to be an issue. All possible combination methods of EEG, MRI, and PET data

were then tested for this study. The highest overall performance was obtained through a combination of all three experts, using a meta-classifier as the final decision block reaching an accuracy of 95.0±3.1%. This performance metric exceeded all accuracies from previous studies, with the closest being 94.0±2.0, where EEG and MRI data was fused and implemented in a two-tiered leave one out algorithm [112]. The EEG+MRI combination performance for this study reached 94.7±3.1% with the meta-classifier from the stacked generalization implementation (SG) of decision used. This exceeded the previous work by a statistically insignificant margin. EEG+PET diagnostic accuracy reached a maximum of 92.3±2.7% where SG was the decision fusion method used. The MRI+PET attaining a maximum performance of 90.9±3.0% with simple majority voting (SMV) as the decision fusion block; however, it should be noted that all decision fusion methods (sum rule, SMV, WMV, and SG) were all within 0.5 percentage points of each other, falling well within each performance metric's confidence intervals.

Finally, a severity analysis was performed. The method used for the separating of AD groups into severe and mild was the same used in the previous work. MMSE scores were used to develop cutoff points to place subjects into different AD severity groups. This analysis was used to determine the severity of the disease in a given patient, not merely whether the patient indeed has the disease or not. The overall cohort contained the same amount of patients used for the three-biomarker evaluation set (32CN, 33AD). The three MMSE cutoffs used were 20, 23, and 26. This created the cohort shown in Table 52. Each biomarker was evaluated individually in this analysis, as well as a combination of all three for decision fusion. At an MMSE cutoff of 20, the best overall performance was reached in EEG+MRI+PET fusion, with the SG method for decision

fusion attaining 68.2±2.9%. For the cutoffs of 23 and 26, the maximum performance was again attained in the combination of all three with the SG for decision fusion, reaching 70.9±3.0% and 67.8±3.0% respectively. Overall we can note generally poor diagnostic performance from the system for the disease severity analysis. When the cutoff number was increased, the overall combination performance increased. This phenomenon can be explained in several ways. To begin with, as the cutoff increased, the severe AD group grew in number, which is to say that there were more patients that had a greater overall difference from normal patients. It may also be feasible that mild AD patients share more biological similarities across all biomarkers used at lower cutoff numbers, creating a harder problem for classification between mild AD and normal controls. Thus, this third class may have merely been contributing to overall performance loss, until the size of the class was reduced to a "small enough" number. The more likely explanation is that the severe and mild AD groups are hard to distinguish. Additionally, the use of the MMSE score alone as a cutoff for severity separation in AD patients may not paint a complete picture of AD progression in a given subject [115]. Supplementary information from other sources such as CDR and DRS data as well as greater analysis of other psychometrics could yield better class separation boundaries for automated analysis. Regardless, it is clear that as the MMSE cutoff increases, the performance of the severity analysis system increases [112].

5.2 POSSIBLE SOURCES OF ERROR

The preprocessing techniques used in this study generally do not effect the data in a negative manner. Therefore, the primary sources of error for all data used in this study can be primarily attributed to the data acquisition. EEG data in particular is susceptible to noise, mostly because the voltages from brain activity occur in the micro-volts range (μV), and are easily overshadowed by muscular electrical activity and external noise sources such as power-line and electromagnetic wave interference. Various EEG systems attempt to combat such errors by placing the amplifiers on the scalp electrode, as opposed to at the base unit. This method varies between each system, and any assumptions regarding the amplification technique cannot be made. Additionally, some AD patients of a low cognitive ability were not able to properly execute the auditory oddball paradigm. This meant their P300 responses to target and novel tones may have been greatly reduced, if apparent at all. We attempted to reduce such error by visually inspecting all AD and CN ERPs for all subjects used in this study, excluding those instances with a weak or non-existent P300.

However, the greatest source of error is also the largest assumption made in this study. All class labels for all patients in this study are determined from the decision of an expert neurologist, whom generally has a diagnostic accuracy of approximately 90% [8]. This diagnosis is accepted as the correct class information for all subjects, where in some cases this may not be correct. Some patients had inconclusive or contradictory class labels, and were subsequently excluded from all data sources for this study, as a conclusive class could not be determined with the information provided. Clearly, if any class labels were incorrect it would adversely effect the ability of our classification

algorithms to learn the data. While post-mortem analysis through autopsy is the only currently available method to accurately determine patient class, it cannot be immediately applied to our current datasets without further investigation to the stage and progression of the disease at time of data collection as well as at time of death. In the future, such correct diagnosis information could be applied to this dataset, with careful consideration placed in the progression of the diseases over time in comparison to the data collection date.

Finally, a contributing factor for error is that the data collected for this study, be it EEG, MRI, or PET, is a combination from various test sites and processing locales. Since the MRI and PET data was provided in a format processed with proprietary software, only specific features could be used. The raw EEG data was collected with an EEG technician on hand; however, expert EEG analysis was not performed and the data provided was merely sent through our automated filtering, baselining, and ratification rejection process. Error and poor EEG signals could be avoided with the use of professional EEG analysis software, or an expert electroencephalographer.

5.3 RECCOMMENDATIONS FOR FUTURE WORK

The CERND study has now reached the end of the data acquisition phase. With this important milestone, the available data across all biomarkers and tests constitutes the final patient cohort. This data is part of a vast project, and will be made available to researchers worldwide for further analysis in the hopes of developing methodology or drawing parallels for detection of neurodegenerative diseases. While the work presented in this study focused solely on the differentiation between AD and normal controls, the algorithms described herein can be adapted for multi-class problems with ease. The investigation into Parkinson's disease (with and without dementia) as well as mild cognitive impairment should be explored from this cohort.

Furthermore, the focus of this study as well as previous work dealt solely with biological markers for diseases—a large wealth of statistical information (psychometrics for example) is available. The inclusion of such statistical analysis along with automated biomarker based classification algorithms could yield even greater results. Severity analysis is also desirable; perhaps a system that could not only determine patient classification, but the stage of the disease would be of great use. Regardless of the approaches used in future work, there is a plethora of information readily available and should be used to its fullest extent. An extensive button press analysis of all subjects in the EEG dataset was conducted earlier in this study; however, this data was never fully utilized and could be considered as an additional exclusion criteria. Additionally, the removal of first-degree relatives to AD-probable patients in the control subject base should be explored. Previous studies have shown marked differences between those with and without first degree relation to AD subjects and should be considered.

REFERENCES

[1] National Center for Health Statistics, "Health, United States, 2008 with Special Feature on the Health of Young Adults,"2008.

[2] Baby Boomer HeadQuarters, "Generation Statistics," [Online] 2009: <http://www.bbhq.com/bomrstat.htm>, Last accessed: 9-21-2009

[3] US Census Bureau, "Facts for Features,"Jan.2006.

[4] Alzheimer's Association, "Alzheimer's Disease Facts and Figures, 2009," [Online] 2009: <http://www.alz.org/national/documents/report_alzfactsfigures2009.pdf>, Last accessed: 9-21-2009

[5] American Health Assistance Foundation, "The Facts on Alzheimer's Disease," [Online] 2009: <http://www.ahaf.org/alzheimers/about/understanding/facts.html>, Last accessed: 9-22-2009

[6] D. Shenk, *The Forgetting: Alzheimer's: Portrait of an Epidemic*. Harpswell: Anchor Publishing, 2003.

[7] A. Wimo, B. Winblad, and L. Jonsson, "An estimate of the total worldwide societal costs of dementia in 2005," *Alzheimer's and Dementia*, vol. 3, no. 2, pp. 81-91, Apr. 2007.

[8] A. Lim, D. Tsuang, W. Kukull, D. Nochlin, J. Leverenz, W. McCormick, J. Bowen, L. Teri, J. Thompson, E. R. Peskind, M. Raskind, and E. B. Larson, "Clinico-Neuropathological Correlation of Alzheimer's Disease in a Community-Based Case Series," *Journal of the American Geriatrics Society*, vol. 47, no. 5, pp. 564-569, May 1999.

[9] R. C. Sotero and N. J. Trujillo-Barreto, "Biophysical model for integrating neuronal activity, EEG, fMRI and metabolism," *Neuroimage*, vol. 39, pp. 290-309, Aug. 2007.

[10] Mayo Clinic, "Alzheimer's Disease Treatment," [Online] 2009: <http://www.mayoclinic.org/alzheimers-disease/treatment.html>, Last accessed: 10-12-2009

[11] E. Salmon, B. Sadzot, P. Maquet, C. Degueldre, C. Lemaire, P. Rigo, D. Comar, and G. Franck, "Differential Diagnosis of Alzheimer's Disease with PET," *Journal of Nuclear Medicine*, vol. 35, no. 3, pp. 391-398, Mar. 1994.

[12] K. Fritzsche, A. vonWangenheim, R. Dillmann, and R. Unterhinninghofen, "Automated MRI-Based Quantification of the Cerebral Atrophy Providing Diagnostic Information on Mild Cognitive Impairment and Alzheimer's Disease," in *Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems,* IEEE Computer Society, 2006, pp. 191-196.

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

[14] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, Jr. R. Johnson, G. A. Miller, W. Tirrer, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor, "Guideline for using human event-related potentials to study cognition: Recording standard and publication criteria," *Psychophysiology*, vol. 37, pp. 127-152, Mar. 2000.

[15] C. Babiloni, G. Frisoni, M. Steriade, L. Bresciani, G. Binetti, C. DelPercio, C. Geroldi, C. Miniussi, F. Nobili, G. Rodriguez, F. Zappasodi, T. Carfagna, and P. M. Rossini, "Frontal white matter volume and delta EEG sources negativelycorrelate in awake subjects with mild cognitiveimpairment and Alzheimer's disease," *Clinical Neurophysiology*, vol. 117, pp. 1113-1129, May 2006.

[16] M. Weiner, "ADNI-Info," [Online] 2009: <http://www.adni-info.org/>,

[17] M. Weiner, "ADNI Info - About ADNI," [Online] 2009: <http://www.adni-info.org/index.php?option=com_content&task=blogcategory&id=0&Itemid=43>,

[18] C. Hill, "What is the Mini Mental State Exam," [Online] 2008: <http://alzheimers.about.com/od/diagnosisofalzheimers/a/MMSE.htm>, Last accessed: 9-30-2009

[19] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'Mini-mental state': A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189-198, Nov. 1975.

[20] R. M. Crum, J. C. Anthony, S. S. Basset, and M. F. Folstein, "Population-Based Norm for the Mini-Mental State Examination by Age and Educational Level," *The Journal of the American Medical Association*, vol. 269, no. 18, pp. 2386-2391, May 1993.

[21] Washington University: Alzheimer's Disease Research Center, "What is the CDR?," [Online] 2006: <http://alzheimer.wustl.edu/cdr/AboutCDR/aboutcdr.htm>, Last accessed: 9-30-2009

[22] Washington University: Alzheimer's Disease Research Center, "Clinical Dementia Rating Assignment," [Online] 1999: <http://www.biostat.wustl.edu/~adrc/cdrpgm/index.html>, Last accessed: 9-30-2009

[23] M. Panisset, M. Roudier, J. Saxton, and F. Boiler, "Severe Impairment Battery: A Neuropsychological Test for Severely Demented Patients," *Archives of Neurology*, vol. 51, no. 1, pp. 41-45, Jan. 1994.

[24] B. Reisberg, S. H. Ferris, M. J. deLeon, and T. Crook, "The global deterioration scale for assessment of primary degenerative dementia," *American Journal of Psychiatry*, vol. 139, no. 9, pp. 1136-1139, Sept. 1982.

[25] Alzheimer Society, "Alzheimer's Disease: Progression of the Disease," [Online] 2009: <http://www.alzheimer.ca/english/disease/progression-intro.htm>, Last accessed: 10-12-2009

[26] D. Johanson and B. Edgar, *From Lucy to Language*. New York: Simon & Schuster, 1996.

[27] Q. Zhang, E. Powers, J. Nieva, M. E. Huff, M. A. Dendle, J. Bieschke, C. G. Glabe, A. Echenmoser, P. Wentworth, R. A. Lerner, and J. W. Kelly, "Metabolite-initiated protein misfolding may trigger Alzheimer's disease," *Proceedings of the National Academy of Sciences of the USA*, vol. 101, no. 14, pp. 4752-4757, Apr. 2004.

[28] D. J. Selkoe, "Cell biology of protein misfolding: The examples of Alzheimer's and Parkinson's diseases," *Nature: Cell Biology*, vol. 6, no. 11, pp. 1054-1061, Nov. 2004.

[29] National Institute on Aging, "High Resolution Images," [Online] 2009: <http://www.nia.nih.gov/Alzheimers/Resources/HighRes.htm>,

[30] G. Forloni, L. Terreni, I.Bertani, S. Fogliarino, R. Invernizzi, A. Assini, G. Ribizzi, A. Negro, E. Calabrese, M. Volonté, C. Mariani, M. Franceschi, M. Tabaton, and A. Bertoli, "Protein misfolding in Alzheimer's and Parkinson's disease: genetics and molecular mechanisms," *Neurobiology of Aging*, vol. 23, no. 5, pp. 957-976, Oct. 2002.

[31] S. A. Small and S. Gandy, "Sorting through the Cell Biology of Alzheimer's Disease: Intracellular Pathways to Pathogenesis," *Neuron*, vol. 52, no. 1, pp. 15-31, Oct. 2006.

[32] M. D. Weingarten, A. H. Lockwood, S. Hwo, and M. Kirschner, "A protein factor essential for microtubule assembly," *Proceedings of the National Academy of Sciences of the USA*, vol. 72, no. 5, pp. 1858-1862, May 1975.

[33] National Institute on Aging, "Symptoms," [Online] 2009: <http://www.nia.nih.gov/Alzheimers/AlzheimersInformation/Symptoms/>, Last accessed: 10-1-2009

[34] Mayo Clinic, "Alzheimer's Disease," [Online] 2009: <http://www.mayoclinic.com/health/alzheimers-disease/DS00161>, Last accessed: 9-28-2009

[35] J. Soukup, *Alzheimer's Disease: A Guide to Diagnosis, Treatment, and Management*. Westport, CT: Praeger, 1996.

[36] National Institute on Aging, "Alzheimer's Disease Medications Fact Sheet," [Online] 2009: <http://www.nia.nih.gov/Alzheimers/Publications/medicationsfs.htm>, Last accessed: 10-12-2009

[37] U.S. National Library of Medicine, "Free radicals," [Online] 2009: <http://ghr.nlm.nih.gov/glossary=freeradicals>, Last accessed: 10-12-2009

[38] J. Powers, "Acetylcholine," [Online] 2009: <http://www.chemistryexplained.com/A-Ar/Acetylcholine.html>, Last accessed: 10-12-2009

[39] F. Li and J. Z. Tsien, "Memory and the NMDA Receptors," *Clinical Implications of Basic Research*, vol. 361, no. 3, pp. 302-303, July 2009.

[40] R. Polikar, "10/20 System of Electrode Placement," [Graphic], 2006: <http://users.rowan.edu/~polikar>.

[41] J. Malmivue and R. Plonsey, *Bioelectromagnetism - principles and applications of bioelectric and biomagnetic fields*. New York: Oxford University Press, 1995.

[42] BrainMaster Technologies, "The 10/20 System," [Online] 2007: <http://www.brainmaster.com/generalinfo/1020/1020.html>, Last accessed: 10-12-2009

[43] Immrama Institute, "International 10-20 System of Electrode Placement," [Online] 2009: <http://www.immrama.org/eeg/electrode.html>, Last accessed: 10-12-2009

[44] L. Sornmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Boston: Elsevier Academic Press, 2005.

[45] E. Neidermeyer and F. Lopes da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5 ed. Philadelphia: Lippincoot Williams & Wilkins, 2004.

[46] M. C. Kushen and D. Frim, "Placement of Subdural Electrode Grids for Seizure Focus Localization in Patients With a Large Arachnoid Cyst," *Neurosurgical Focus*, vol. 22, no. 2, p. E5, June 2007.

[47] H. Blume, "Implanted EEG Electrodes," [Online] 2004: <http://professionals.epilepsy.com/page/surgery_electrodes.html>, Last accessed: 10-13-2009

[48] R. D. Pascual-Marqui, "LORETA: low resolution brain electromagnetic tomography," [Online] 2009: <http://www.uzh.ch/keyinst/loreta.htm>,

[49] Digital Tigers, "International Electrical Power Standards," [Online] 2006: <http://www.digitaltigers.com/international-power.asp>, Last accessed: 10-13-2009

[50] J. Andreassi, *Psychophysiology: Human Behavior and Physiological Response*, 5 ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2006.

[51] H. H. Stassen, D. T. Lykken, and G. Bombent, "The within-pair EEG similarity of twins reared apart," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 237, no. 4, pp. 244-252, June 1988.

[52] T. Patel, R. Polikar, C. Davatzikos, and C. M. Clark, "EEG and MRI Data Fusion for Early Diagnosis of Alzheimer's Disease," in *30th Annual International Conference of the IEEE EMBS*, 2008, pp. 1757-1760.

[53] T. B. Mulholland, *The concept of attention and the electroencephalographic alpha rhythm*. London: Butterworths, 1969, pp. 100-127.

[54] J. C. Shaw, "Intention as a component of the alpha-rhythm response to mental activity," *International Journal of Psychophysiology*, vol. 24, no. 1-2, pp. 7-23, Nov. 1996.

[55] S. Crawcour, A. Bowers, A. Harkrider, and T. Saltuklaroglu, "Mu wave suppression during the perception of meaningless syllables: EEG evidence of motor recruitment," *Neuropsychologia*, vol. 47, no. 12, pp. 2558-2563, Oct. 2009.

[56] J. L. Kennedy, R. M. Gottsdanker, J. C. Armington, and F. E. Gray, "A New Electroencephalogram Associated With Thinking," *Science*, vol. 108, no. 2811, pp. 57-529, Nov. 1948.

[57] R. Galambos, S. Makeig, and P. J. Talmachoff, "A 40-Hz auditory potential recorded from the human scalp," *Proceedings of the National Academy of Sciences of the USA*, vol. 78, no. 4, pp. 2643-2647, Mar. 1981.

[58] V. I. Khorevin, "Interaction between responses evoked by acoustic and somatosensory stimuli in neurons of the magnocellular part of the medial geniculate body," *Neurophysiology*, vol. 12, no. 4, pp. 241-245, July 1980.

[59] C. M. DeGiorgio, A. L. Rabinowicz, and P. S. Gott, "Predictive value of P300 event-related potentials compared with EEG and somatosensory evoked potentials in non-traumatic coma," *Acta Neurologica Scandinavica*, vol. 87, no. 5, pp. 423-427, Jan. 2009.

[60] E. M. Finney, B. A. Clementz, G. Hickok, and K. R. Dobkins, "Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG.," *Neuroreport*, vol. 14, no. 11, pp. 1425-1427, Aug. 2003.

[61] Y. Naganuma, T. Konishi, K. Hongou, T. Okada, J. Tohyama, and M. Uchiyama, "Event-related potentials (P300) and EEG activity in childhood partial epilepsy," *Brain and Development*, vol. 19, no. 2, pp. 117-121, Mar. 1997.

[62] J. Polich and C. Maragala, "P300 and probability: comparison of oddball and single-stimulus paradigms," *International Journal of Psychophysiology*, vol. 25, no. 2, pp. 169-176, 1997.

[63] S. Murali and V. Mulish, "Modeling of evoked potentials of electroencephalograms: An overview," *Digital Signal Processing*, vol. 17, no. 3, pp. 665-674, 2007.

[64] S. Yamaguchi, H. Tsuchiya, S. Yamagata, G. Toyoda, and S. Kobayashi, "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, vol. 111, no. 2, pp. 195-203, Feb. 2000.

[65] J. Cohen and J. Polich, "On the number of trials needed for P300," *International Journal of Psychophysiology*, vol. 25, no. 3, pp. 249-255, 1997.

[66] E. Kirino, "Correlation Between P300 and EEG Rhythm in Schizophrenia," *Clinical EEG and Neuroscience*, vol. 35, no. 3, pp. 137-146, July 2004.

[67] J. Polich and A. Kok, "Cognitive and biological determinant of P300: an integrative review," *Biological Psychology*, vol. 25, no. 3, pp. 249-255, 1995.

[68] S. Baron-Cohen, "Precursors to a theory of mind: understanding attention in others," 1991, pp. 233-251.

[69] O. Blanke and S. Arzy, "The out-of-body experience: disturbed self-processing at the temporo-parietal junction," *Neuroscientist*, vol. 11, no. 1, pp. 16-24, Feb. 2005.

[70] D. Samson, I. Apperly, C. Chiavarino, and G. Humphreys, "Left temporoparietal junction is necessary for representing someone else's belief," *Nature Neuroscience*, vol. 7, pp. 499-500, Apr. 2004.

[71] J. Bennington and J. Polich, "Comparison of P300 from passive and active tasks for auditory and visual stimuli," *International Journal of Psychophysiology*, vol. 34, no. 2, pp. 171-177, 1999.

[72] D. Strüber and J. Polich, "P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus interval effects," *International Journal of Psychophysiology*, vol. 45, no. 3, pp. 187-196, 2002.

[73] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128-2148, Oct. 2007.

[74] E. Basar, M. Schürmann, T. Demiralp, C. Basar-Eroglu, and A. Ademoglu, "Event-related oscillations are 'real brain responses' - wavelet analysis and new strategies," *International Journal of Psychophysiology*, vol. 39, no. 2-3, pp. 91-127, Jan. 2001.

[75] T. Demiralp, A. Ademoglu, Y. Istenfanopulos, C. Basar-Eroglu, and E. Basar, "Wavelet analysis of oddball P300," *International Journal of Psychophysiology*, vol. 39, no. 2-3, pp. 221-227, Jan. 2001.

[76] J. A. Pineda, "The functional significance of mu rhythms: translating 'seeing' and 'hearing' into 'doing'," *Brain Research Reviews*, vol. 50, no. 1, pp. 57-68, Jan. 2005.

[77] Z. Hidasi, B. Czigler, P. Salacz, E. Csibri, and M. Molnar, "Changes of EEG spectra and coherence following performance in a cognitive task in Alzheimer's disease," *Journal of the Neurological Sciences*, vol. 283, no. 1-2, p. 312, Aug. 2009.

[78] N. Boutros, M. Torello, E. Burns, S. Wu, and H. Hasrallah, "Evoked potentials in subjects at risk for Alzheimer's disease," *Psychiatry Resarch*, vol. 57, no. 1, pp. 57-63, June 1995.

[79] B. Ally, G. Jones, J. Cole, and A. Budson, "The P300 component in patients with Alzheimer's disease and their biological children," *Biological Psychology*, vol. 72, no. 2, pp. 180-187, May 2006.

[80] R. Polikar, M. H. Greer, L. Udpa, and F. Keinert, "Multiresolution wavelet analysis of ERPs for the detection of Alzheimer's disease," in *19th Annual International Conference of the IEEE EMBS*, 3 ed 1997, pp. 1301-1304.

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

[82] S. Yagneswaran, M. Baker, and A. A. Petrosian, "Power frequency and wavelet characteristics in differentiating between normal and Alzheimer EEG," in 1 ed 2003, pp. 46-47.

[83] D. Abasolo, R. Hornero, P. Espino, A. Alonso, and R. de la Rosa, "Electroencephalogram analysis with approximate entropy to help in the diagnosis of Alzheimer's disease," in *4th International IEEE EMBS Special Topic Conference on Information Technology applications in Biomedicine*, 2003, pp. 222-225.

[84] T. Hua-ying and X. Tian, "Coherence Characteristics of Gamma-band EEG during rest and cognitive task in MCI and AD," in *27th Annual International Conference of the IEEE EMBS*, 2006, pp. 2747-2750.

[85] R. Chapman, G. Nowlis, J. McCrary, J. Chapman, T. Sandoval, M. Guilily, M. Gardner, and L. Reilly, "Brain event-related potentials: Diagnosing early-stage Alzheimer's disease," *Neurobiology of Aging*, vol. 28, no. 2, pp. 194-201, 2007.

[86] G. Henderson, E. Ifeachor, N. Hudson, C. Goh, N. Outram, S. Wimalaratna, C. DelPercio, and F. Vecchio, "Development and assessment of methods for detecting dementia using the human

electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, pp. 1557-1568, Aug. 2006.

[87] R. A. Novelline, *Squire's Fundamentals of Radiology*, 6 ed. Boston: Harvard University Press, 2004.

[88] Merriam-Webster Online Dictionary, "Computed Tomography," [Online] 2008: <http://www.merriam-webster.com/dictionary/computed+tomography>, Last accessed: 10-20-2009

[89] J. Boone, J. Bushberg, E. Leidholdt, and A. Seibert, *The Essential Physics of Medical Imaging*, 2 ed. Philadelphia: Lippincott Williams & Wilkins, 2002.

[90] L. Mosconi, M. Brys, L. Glodzik-sobanska, S. DeSanti, H. Rusinek, and M. J. deLeon, "Early detection of Alzheimer's disease using neuroimaging," *Experimental Gerontology*, vol. 42, no. 1-2, pp. 129-138, 2007.

[91] W. G. Bradley, "Fundamentals of MRI: Part II," [Online] 2009: <http://www.e-radiography.net/mrict/fund%20mr2/fundmri%202.htm>, Last accessed: 10-20-2009

[92] Y. Shim, J. Ahn, J. Cho, and K. Lee, "Solitary skull metastasis as initial manifestation of heptocelluar carcinoma," *World Journal of Surgical Oncology*, vol. 6, no. 66 June 2008.

[93] M. Torabi, R. D. Ardekani, and E. Fatemizadeh, "Discrimination between alzheimer's disease and control group in MR-images based on texture analysis using artificial neural network," in *International Conference on Biomedical and Pharmaceutical Engineering,* 2006, pp. 79-83.

[94] N. Fox, W. Crum, R. Scahill, J. Stevens, J. Janssen, and M. Rossor, "Imaging of onset and progression of Alzheimer's disease with voxelcompression mapping of serial magnetic resonance images," *Lancet*, vol. 358, no. 9277, pp. 201-205, July 2001.

[95] K. van der Hiele, A. A. Vein, A. van der Welle, J. van der Grond, R. G. J. Westendorp, E. L. E. M. Bollen, M. A. van Buchem, J. G. van Dijk, and H. A. M. Middelkoop, "EEG and MRI correlates of mild cognitive impairment and Alzheimer's disease," *Neurobiology of Aging*, vol. 28, no. 9, pp. 1322-1329, Sept. 2007.

[96] M. M. Ter-Pogossian, M. E. Phelps, E. J. Hoffman, and N. A. Mullani, "Positron-emission transaxial tomograph for nuclear imaging (PETT)," *Radiology*, vol. 114, no. 1, pp. 89-98, Jan. 1975.

[97] Mayo Clinic, "Radiopharmaceutical," [Online] 2009: <http://www.mayoclinic.com/health/drug-information/DR602307>,

[98] V. Lowe, B. Kemp, C. Jack, M. Senjem, S. Weigand, M. Shiung, G. Smith, D. Knopman, B. Boeve, B. Mullan, and R. Petersen, "Comparison of 18F-FDG and PiB PET in Cognitive Impairment," *Journal of Nuclear Medicine*, vol. 50, no. 6, pp. 878-886, May 2009.

[99] M. T. Studenski, J. G. Parker, D. R. Gilland, S. Majewski, and B. Hammond, "Performance evaluation of a small field-of-view, mobile PET/SPECT system," in *IEEE Nuclear Science Symposium Conference,* 2007, pp. 3770-3773.

[100] University of Virginia Health System, "Positron Emission Tomography (PET)," [Online] 2007: <http://www.healthsystem.virginia.edu/uvahealth/adult_radiology/pet.cfm>, Last accessed: 9-30-2009

[101] A. B. Newberg and A. Alvi, "The role of PET imaging in the management of patients with central nervous system disorders," *Radiologic Clinics of North America*, vol. 43, no. 1, pp. 49-65, Jan. 2005.

[102] H. Gray, *Anatomy of the Human Body*, 20 ed. Philadelphia: Lea & Febiger, 1918.

[103] N. Foster, "PET Image Analysis for the University of Pennsylvania Tobacco Settlement Project," [Graphic] 2007.

[104] S. Minoshima and D. Cross, "NEUROSTAT," [Online] 2008: <http://128.95.65.28/~Download/>,

[105] K. Ishii, F. Willoch, S. Minoshima, A. Drzezga, E. P. Ficaro, D. J. Cross, D. E. Kuhl, and M. Schwaiger, "Statistical Brain Mapping of 18F-FDG PET in Alzheimer's Disease: Validation of Anatomic Standardization for Atrophied Brains," *Nuclear Medicine*, vol. 42, no. 4, pp. 548-557, Apr. 2001.

[106] N. L. Foster, J. L. Heidebrink, C. M. Clark, W. J. Jagust, S. E. Arnold, R. Barbas, S. DeCarli, Turner, A. Koepee, M. Higdon, and S. Minoshima, "FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease," *Brain*, vol. 130, no. 10, pp. 2616-2635, Aug. 2007.

[107] G. Jacques, "Daubechies and quadratic B-spline wavelets for automated early diagnosis of Alzheimer's disease." Engineering: Electrical, Master of Science, Rowan University, 2004.

[108] A. Topalis, "Multiresolution wavelet analysis of event-related EEG potentials using ensemble of classifier data fusion techniques for the early detection of Alzheimer's disease." Engineering: Electrical, Master of Science, Rowan University, 2006.

[109] N. Stepenosky, "Data fusion of complementary information from parietal and occipital event related potentials for early diagnosis of Alzheimer's disease." Engineering: Electrical, Master of Science, Rowan University, 2006.

[110] H. Gandhi, "Stacked generalization for early diagnosis of Alzheimer's disease." Engineering: Electrical, Master of Science, Rowan University, 2007.

[111] B. Balut, "Data fusion based optimal EEG electrode selection for early diagnosis of Alzheimer's disease." Engineering: Electrical, Master of Science, Rowan University, 2008.

[112] T. Patel, "Ensemble of Classifiers Based Data Fusion of EEG and MRI for Diagnosis of Neurodegenerative Disorders." Engineering: Electrical Master of Science, Rowan University, 2008.

[113] O. Hansson, H. Zetterberg, P. Buchhave, E. Londos, K. Blennow, and L. Minthon, "Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow up study," *Lancet Neurology*, vol. 5, no. 3, pp. 228-234, Mar. 2006.

[114] CERND, "Biomarkers of Late Life Dementia," [Online] 2009: <http://www.cernd.org/index.php>, Last accessed: 9-1-2009

[115] B. Dubois, H. H. Feldman, C. Jacova, S. T. Dekosky, P. Barberger-Gateau, and J. Cummings, "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria," *Lancet Neurology*, vol. 6, no. 8, pp. 734-746, Aug. 2007.

[116] S. Mallat, *A Wavelet Tour of Signals Processing: The Sparse Way*, 3 ed. San Diego: Academic Press, 2008.

[117] Y. Y. Tan, *Wavelet Theory and Its Application to Pattern Recognition (Machine Perception & Aritificial Intelligence)* World Scientific Publishing Company, 2000.

[118] D. F. Walnut, *An Introduction to Wavelet Analysis*. New York: Birkhäuser Boston, 2004.

[119] P. Fleet, *Discrete Wavelet Transformations: An Elementary Approach with Applications*. Hoboken: John Wiley & Sons, 2008.

[120] J. S. Walker, *A Primer on Wavelets and Their Scientific Applications*. Boca Raton: CRC Press, 2008.

[121] W. Howitt, "Sub-Band Coding," [Online] 1995: <http://www.otolith.com/otolith/olt/sbc.html>, Last accessed: 10-27-2009

[122] R. Polikar, "Wavelet Decomposition Example," [Graphic] 2009: <http://users.rowan.edu/~polikar>.

[123] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2 ed Wiley-Interscience, 2000.

[124] C. Lai, M. Reinders, and L. Wessels, "Random Subspace Method for multivariate feature selection," *Pattern Recognition Letters*, vol. 27, no. 10, pp. 1067-1076, July 2006.

[125] T. Ho, "The Random Subspace Method for Constructing Decision Forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, Aug. 1998.

[126] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1421-1439, Feb. 2003.

[127] A. Goldszal, C. Davatzikos, D. Pham, M. Yan, N. Bryan, and S. Resnick, "An Image-Processing System for Qualitative and Quantitative Volumetric Analysis of Brain Images," *Journal of Computer Assisted Tomography*, vol. 22, no. 5, pp. 827-837, Oct. 1998.

[128] D. Shen and C. Davatzikos, "Very High-Resolution Morphometry Using Mass-Preserving Deformations and HAMMER Elastic Registration," *Neuroimage*, vol. 18, no. 1, pp. 28-41, Jan. 2003.

[129] C. Davatzikos, A. Genc, D. Xu, and S. Resnick, "Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy," *Neuroimage*, vol. 14, no. 6, pp. 1361-1369, Dec. 2001.

[130] R. Polikar and M. Ahiskali, "DWT of Sample ERP," [Graphic] 2009: <http://users.rowan.edu/~polikar>.

[131] R. Polikar, A. Topalis, D. Green, J. Kounios, and C. M. Clark, "Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 542-558, Apr. 2007.

[132] R. Polikar, "Neuron," [Graphic] 2006: <http://users.rowan.edu/~polikar>.

[133] R. Polikar, A. Topalis, D. Parikh, D. Green, J. Frymiare, J. Kounios, and C. M. Clark, "An ensemble based data fusion approach for early diagnosis of Alzheimer's disease," *Information Fusion*, vol. 9, no. 1, pp. 83-95, Jan. 2008.

[134] S. Haykin, *Neural Networks: A Comprehensive foundation*, 2 ed. Upper Saddle River: Prentice Hall, 1999.

[135] Mathworks, "Neural Network Toolbox," [Online] 2009: <http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/>,

[136] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken: Wiley-Interscience, 2004.

[137] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4 ed. Burlington: Academic Press - Elsevier, 2008.

[138] D. Wolpert, "Stacked Generalization," *Neural Networks*, vol. 5, no. 2, pp. 241-259, 1992.

[139] R. Polikar and M. Ahiskali, "Augmented Stacked Generalization," [Graphic] 2009: <http://users.rowan.edu/~polikar>.

[140] M. Ahiskali, R. Polikar, J. Kounios, D. Green, and C. M. Clark, "Combining Multichannel ERP Data for Early Diagnosis of Alzheimer's Disease," in *IEEE EMBS Conference on Neural Engineering,* 2009, pp. 522-525.

[141] M. Ahiskali, D. Green, J. Kounios, C. M. Clark, and R. Polikar, "ERP Based Decision Fusion for AD Diagnosis Across Cohorts," in *31st Annual International Conference of the IEEE EMBS,* 2009, pp. 2494-2497.

[142] Neuronetrix, "COGNISION System," [Online] 2007: <http://www.neuronetrix.com/>, Last accessed: 12-1-2009

APPENDICES

APPENDIX A: EEG GRAND AVERAGES

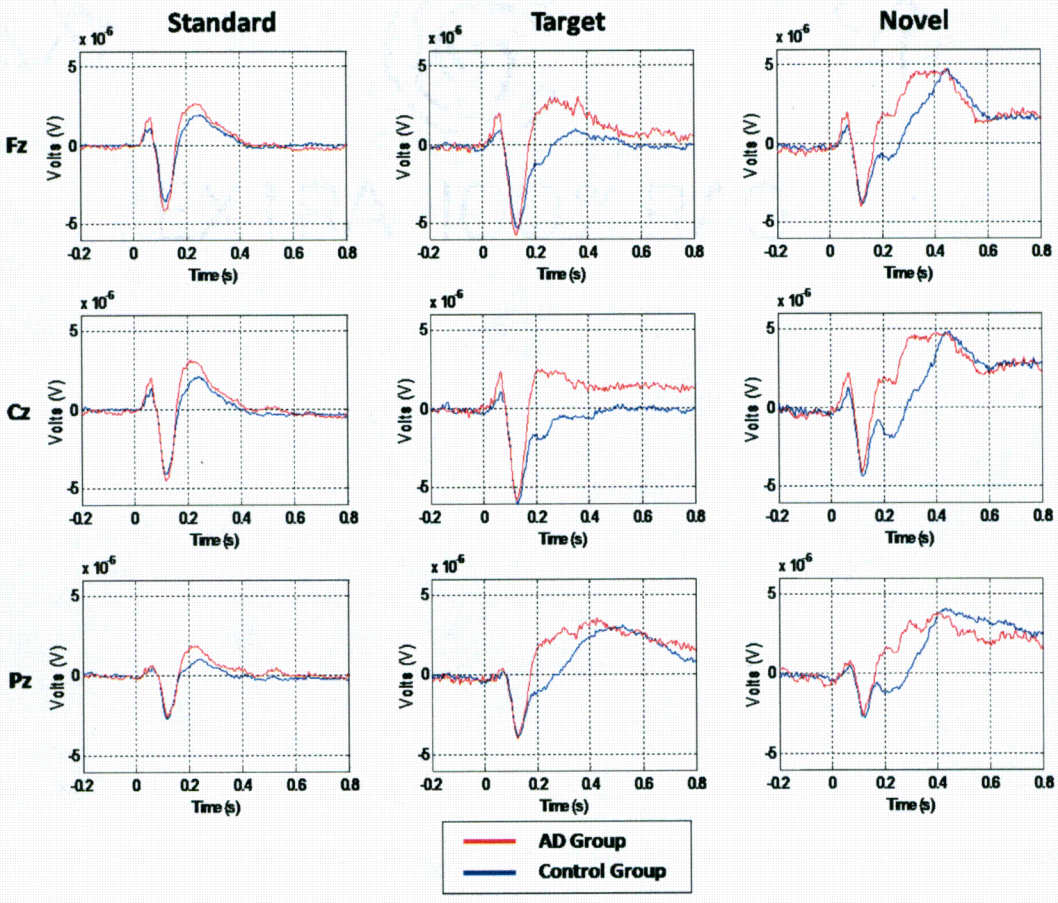

Figure 37 – Cohort A: Three-electrode ERP grand averages

All three stimuli grand averages across all available subjects in Cohort A (71 total subject averages, 34AD and 37CN) are shown for the FZ, CZ, and PZ electrodes. It can be seen that AD patients appear to have enlarged P200s, similar to Cohort B (following page) which have been shown to be indicative of additional effort and use of cognitive resources.

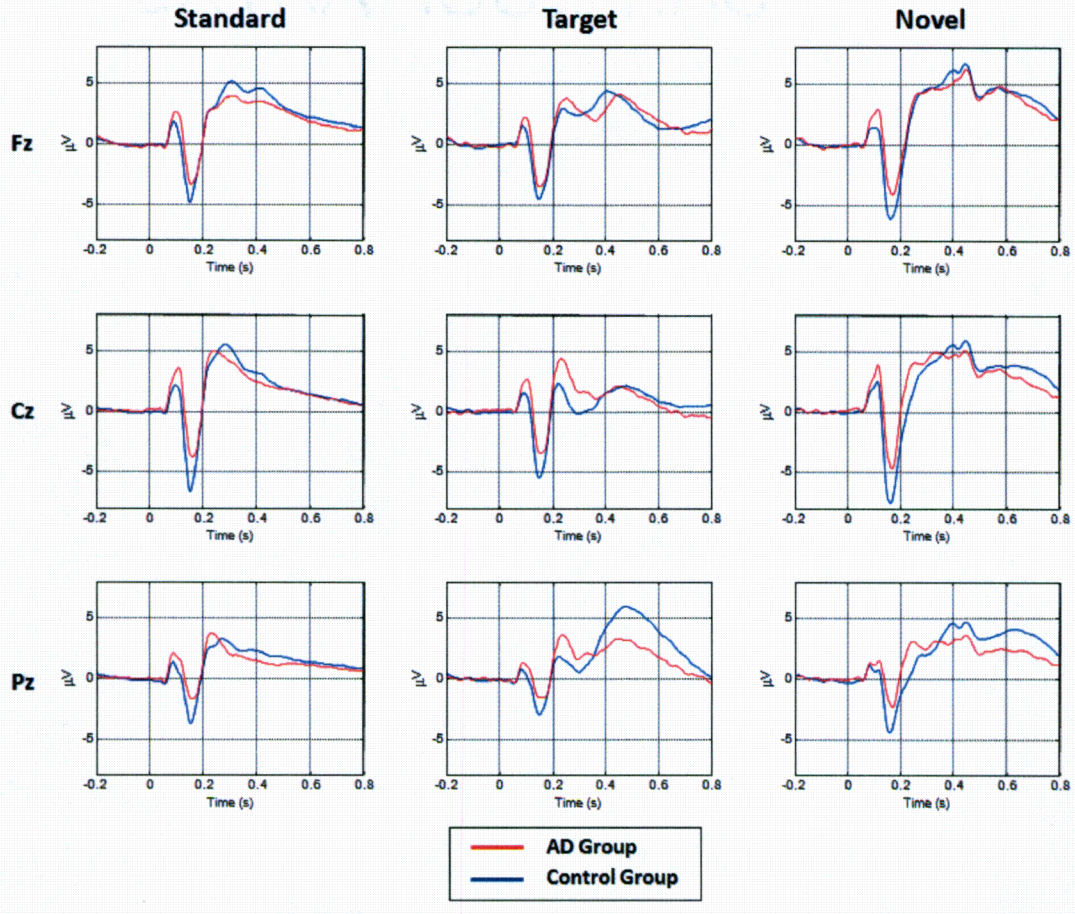

Figure 38 – CERND (Cohort B): Three-electrode ERP grand averages

All three stimuli grand averages across available subjects the CERND study (Cohort B, 197 total subject averages, 103AD and 94CN) are shown for the FZ, CZ, and PZ electrodes. It can be seen that AD patients appear to have enlarged P200s, similar to Cohort A which have been shown to be indicative of additional effort and use of cognitive resources.

APPENDIX B: RAW BUTTON PRESS ACCURACIES

Subject accuracies for all classes in pressing a hardware button during EEG recordings when presented with a target tone is shown in this section. Hardware failure of the button press mechanism in the actual setup occurred for a small percentage of subject trials; therefore, these trials in question were ignored for this analysis. Further explanation of omitted trials is provided in section B.6.

B.1 AD PROBABLE SUBJECTS

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
1638	ad	1	100.00	99.99	91.38	97.12	330	1
1796.01	ad	1	99.99	100.00	97.13	99.04	356	1
189.33	ad	1	99.98	99.95	87.36	95.76	359	1
2356	ad	1	99.91	99.91	12.93	70.92	147	0
2375	ad	1	99.80	99.87	33.99	77.89	593	0
2380	ad	1	99.96	99.98	5.91	68.62	109	0
2397	ad	1	99.98	99.93	48.77	82.89	265	0
3465	ad	1	99.99	99.99	99.43	99.80	356	1
3568	ad	1	100.00	100.00	93.68	97.89	336	1
3756	ad	1	99.97	99.98	2.30	67.42	60	0
3853	ad	1	100.00	99.99	99.43	99.81	361	1
3862	ad	1	99.97	99.96	4.83	68.25	65	0
3867	ad	1	99.99	99.97	91.38	97.11	347	1
3875	ad	1	99.99	99.99	95.98	98.65	355	1
3930	ad	1	100.00	99.99	98.85	99.61	354	1
3981	ad	1	100.00	99.99	91.38	97.12	348	1
4015	ad	1	99.99	100.00	97.70	99.23	361	1
4055	ad	1	100.00	100.00	0.00	66.67	0	0
4182	ad	1	99.99	99.98	90.23	96.73	342	1
4222	ad	1	99.99	99.99	97.13	99.03	362	1
4282	ad	1	99.99	100.00	83.33	94.44	308	1
4392	ad	1	99.99	99.97	80.46	93.47	311	1
4474	ad	1	100.00	100.00	97.13	99.04	351	1
4504	ad	1	100.00	100.00	98.85	99.62	355	1
4528	ad	1	99.99	99.99	96.55	98.84	356	1
4591	ad	1	99.99	100.00	98.03	99.34	421	1
4593	ad	1	100.00	100.00	98.03	99.34	407	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
4686	ad	1	100.00	99.99	95.40	98.46	344	1
4689	ad	1	99.99	99.98	72.99	90.99	284	0
4726	ad	1	100.00	100.00	0.00	66.67	0	0
4747	ad	1	99.93	99.88	71.26	90.36	375	0
4748	ad	1	99.81	99.85	11.49	70.39	520	0
4765	ad	1	99.97	99.96	63.79	87.91	301	0
4777	ad	1	99.89	99.81	24.14	74.61	360	0
4800	ad	1	99.98	99.97	81.03	93.66	340	1
4811	ad	1	99.97	99.98	73.56	91.17	315	0
4820	ad_NED							0
4857	ad	1	100.00	100.00	98.85	99.62	357	1
4903	ad	1	100.00	99.99	98.85	99.61	357	1
4908	ad	1	99.72	99.76	67.24	88.91	486	0
4925	ad	1	99.99	99.97	93.10	97.69	353	1
4940	ad_NED							0
4971	ad	1	99.90	99.92	60.34	86.72	413	0
4975	ad	1	99.98	99.95	32.18	77.37	144	0
4980	ad	1	99.90	99.94	22.41	74.08	243	0
5006	ad	1	99.98	99.98	84.48	94.82	327	1
5007	ad	1	99.99	99.42	59.77	86.39	409	0
5013	ad	1	100.00	100.00	94.58	98.19	396	1
5057	ad	1	99.99	99.97	28.72	76.23	282	0
5070	ad	1	99.95	99.72	8.62	69.43	221	0
5089	ad	1	99.99	99.88	88.51	96.12	367	1
5105	ad	1	100.00	100.00	0.00	66.67	0	0
5123	ad	1	99.95	99.95	94.83	98.24	359	1
5158	ad	1	99.88	99.75	36.21	78.61	398	0
5239	ad	1	99.98	100.00	95.98	98.65	366	1
5247	ad	1	99.92	99.91	74.14	91.32	213	0
5299	ad	1	99.99	99.95	93.10	97.68	370	1
5313	ad	1	99.96	99.92	69.54	89.81	336	0
5335	ad	1	100.00	100.00	97.70	99.23	353	1
5410	ad	1	99.97	99.97	80.46	93.47	339	1
5458	ad	1	100.00	100.00	96.55	98.85	352	1
5472	ad	1	99.97	99.94	59.20	86.37	271	0
5501	ad	1	100.00	100.00	95.40	98.47	340	1
5511	ad	1	100.00	100.00	98.85	99.62	360	1
5547	ad	1	99.99	99.99	92.12	97.36	396	1
5571	ad	1	99.89	99.87	23.56	74.44	301	0
5585	ad	1	99.98	99.99	94.25	98.07	353	1
5630	ad	1	99.95	99.85	45.98	81.93	272	0

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
5707	ad	1	99.99	99.98	91.95	97.31	340	1
5797	ad	1	99.99	100.00	98.28	99.42	356	1
5895	ad	1	100.00	100.00	96.55	98.85	346	1
5913	ad	1	99.98	99.98	94.83	98.26	371	1
5937	ad	1	100.00	100.00	0.00	66.67	0	0
5972	ad	1	99.99	99.99	90.80	96.93	337	1
6072	ad	1	100.00	100.00	0.00	66.67	0	0
6225	ad	1	99.98	99.98	5.75	68.57	66	0
6240	ad	1	99.99	99.96	96.55	98.83	431	1
189.33	ad	2	100.00	100.00	0.00	66.67	0	0
2356	ad	2	99.97	99.96	10.34	70.09	101	0
2375	ad	2	99.85	99.82	51.15	83.61	471	0
2380	ad	2	99.92	99.91	20.69	73.51	220	0
3465	ad	2	100.00	99.99	99.43	99.80	354	1
3853	ad	2	100.00	100.00	97.70	99.23	354	1
3862	ad	2	99.99	99.99	1.15	67.04	20	0
3875	ad	2	100.00	100.00	0.00	66.67	0	0
3930	ad	2	100.00	100.00	0.00	66.67	0	0
3973	ad	2	99.99	99.99	90.80	96.93	337	1
4392	ad	2	100.00	100.00	0.00	66.67	0	0
4474	ad	2	100.00	100.00	95.98	98.66	343	1
4504	ad	2	99.99	99.99	96.55	98.85	354	1
4528	ad	2	99.98	99.98	97.13	99.03	369	1
4591	ad	2	99.99	99.97	95.98	98.65	356	1
4686	ad	2	100.00	99.99	94.25	98.08	340	1
4712	ad	2	99.98	99.99	85.63	95.20	339	1
4747	ad	2	99.96	99.97	8.62	69.52	92	0
4765	ad	2	99.91	99.80	50.57	83.43	396	0
4820	ad	2	99.99	99.97	0.00	66.65	0	0
4832	ad	2	99.99	99.97	98.85	99.61	364	1
4857	ad	2	100.00	100.00	0.00	66.67	0	0
4903	ad	2	100.00	99.99	98.28	99.42	348	1
4925	ad	2	100.00	100.00	0.00	66.67	0	0
4955	ad	2	100.00	99.99	91.95	97.31	325	1
4971	ad	2	99.97	99.97	82.76	94.24	335	1
5046	ad	2	99.97	99.96	75.86	91.93	323	0
5070	ad	2	99.89	99.66	20.69	73.42	370	0
5089	ad	2	99.99	99.97	95.98	98.64	362	1
5123	ad	2	99.90	99.70	42.76	80.79	331	0
5247	ad	2	99.88	99.88	37.93	79.23	323	0
5410	ad	2	99.97	99.98	84.48	94.81	358	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
5458	ad	2	100.00	99.99	95.40	98.46	340	1
5547	ad	2	99.99	99.97	94.25	98.07	340	1
3853	ad	3	99.99	99.98	97.04	99.01	406	1
4528	ad	3	99.99	99.99	95.98	98.65	360	1
4821	ad	3	99.99	100.00	96.55	98.85	342	1
						Number of Sessions with TARGET perf > 80%:		**64**
						Total Number of AD Sessions:		**112**

B.2 NORMAL CONTROL SUBJECTS

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
10.22	cn	1	100.00	99.99	94.09	98.03	400	1
111.22	cn	1	99.99	99.99	96.55	98.84	344	1
1406	cn	1	100.00	99.97	96.55	98.84	349	1
182.33	cn	1	100.00	100.00	97.70	99.23	349	1
2316	cn	1	99.99	99.99	92.53	97.50	343	1
2330	cn	1	100.00	99.99	95.98	98.65	346	1
2344	cn	1	99.99	99.85	92.53	97.46	387	1
3092	cn	1	100.00	100.00	99.43	99.81	355	1
3178.02	cn	1	99.99	99.99	96.55	98.85	348	1
3197.01	cn	1	100.00	99.99	97.13	99.04	347	1
32.33	cn	1	100.00	100.00	95.98	98.66	347	1
3355	cn	1	100.00	100.00	98.85	99.61	351	1
354.01	cn	1	100.00	100.00	97.70	99.23	351	1
3569	cn	1	100.00	100.00	98.85	99.62	349	1
3661	cn	1	100.00	99.98	93.10	97.69	349	1
3678	cn	1	100.00	99.99	92.53	97.50	343	1
3691	cn	1	100.00	100.00	98.28	99.42	356	1
3720	cn	1	100.00	99.99	95.40	98.47	340	1
3723	cn	1	100.00	100.00	97.13	99.04	344	1
3725	cn	1	99.99	100.00	93.68	97.89	336	1
3726	cn	1	100.00	100.00	98.28	99.42	348	1
3741.01	cn	1	100.00	99.99	97.70	99.23	354	1
3798.01	cn_FAIL							0
3798.02	cn	1	100.00	100.00	98.28	99.42	352	1
3798.03	cn	1	100.00	100.00	99.43	99.81	354	1
3868	cn	1	99.99	99.99	93.68	97.89	349	1
3879.01	cn	1	100.00	99.98	99.43	99.80	359	1
3950	cn	1	100.00	100.00	97.70	99.23	350	1
3980	cn	1	99.99	99.99	95.98	98.65	355	1
4031	cn	1	100.00	99.99	98.28	99.42	352	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
4110	cn	1	100.00	100.00	98.85	99.62	351	1
4128	cn	1	100.00	100.00	98.85	99.61	358	1
4138	cn	1	100.00	99.99	99.43	99.81	356	1
4219	cn	1	99.99	99.97	96.55	98.84	359	1
4380	cn	1	100.00	99.99	95.98	98.65	346	1
4635	cn	1	100.00	100.00	0.00	66.67	0	0
4636	cn	1	100.00	99.98	97.70	99.23	356	1
4662	cn	1	99.99	100.00	97.13	99.04	364	1
4717	cn	1	99.98	99.99	83.33	94.43	341	1
4850	cn	1	99.99	99.96	78.82	92.92	366	0
4907	cn	1	100.00	99.99	98.85	99.61	355	1
4968	cn	1	100.00	100.00	96.55	98.85	346	1
5106	cn	1	100.00	100.00	98.85	99.62	346	1
5108	cn	1	100.00	100.00	93.10	97.70	337	1
5195	cn	1	99.99	100.00	74.38	91.46	321	0
5241	cn	1	99.99	99.99	91.95	97.31	340	1
5322	cn	1	100.00	100.00	96.55	98.85	340	1
55.33	cn	1	99.99	99.93	90.23	96.72	359	1
5851	cn	1	100.00	99.98	97.70	99.23	359	1
6117	cn	1	100.00	100.00	0.00	66.67	0	0
6385	cn	1	100.00	100.00	0.00	66.67	0	0
6554	cn	1	100.00	99.99	97.13	99.04	348	1
67.33	cn	1	100.00	100.00	97.13	99.04	352	1
78.33	cn	1	100.00	100.00	98.85	99.62	353	1
10.22	cn	2	100.00	100.00	0.00	66.67	0	0
111.22	cn	2	100.00	100.00	0.00	66.67	0	0
1406	cn	2	100.00	99.99	98.85	99.61	352	1
2316	cn	2	99.99	99.78	94.25	98.01	417	1
3092	cn	2	100.00	100.00	97.70	99.23	349	1
32.33	cn	2	99.99	99.98	94.83	98.27	350	1
3355	cn	2	99.99	100.00	95.40	98.47	350	1
3569	cn	2	100.00	100.00	99.43	99.81	349	1
3661	cn	2	99.99	99.99	95.98	98.65	351	1
3678	cn	2	99.98	100.00	85.06	95.01	324	1
3691	cn	2	100.00	100.00	96.06	98.69	401	1
3726	cn	2	100.00	100.00	90.23	96.74	321	1
3741.01	cn	2	100.00	100.00	97.70	99.23	343	1
3798.01	cn	2	100.00	100.00	99.43	99.81	348	1
3798.02	cn	2	100.00	100.00	99.43	99.81	352	1
3798.03	cn	2	100.00	100.00	98.28	99.42	349	1
3868	cn	2	100.00	100.00	97.13	99.04	350	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
3950	cn	2	100.00	100.00	0.00	66.67	0	0
3980	cn	2	100.00	100.00	96.55	98.85	342	1
4138	cn	2	100.00	100.00	98.85	99.62	354	1
4219	cn	2	100.00	100.00	0.00	66.67	0	0
4380	cn	2	100.00	100.00	0.00	66.67	0	0
4662	cn	2	100.00	100.00	0.00	66.67	0	0
4684	cn	2	100.00	100.00	95.40	98.47	340	1
4717	cn	2	99.99	99.99	95.98	98.65	360	1
4968	cn	2	100.00	100.00	99.43	99.81	349	1
5106	cn	2	100.00	100.00	98.85	99.62	347	1
5108	cn	2	100.00	99.99	93.10	97.70	337	1
5195	cn	2	100.00	100.00	0.00	66.67	0	0
5322	cn	2	100.00	100.00	0.00	66.67	0	0
55.33	cn	2	99.99	99.99	94.83	98.27	350	1
67.33	cn	2	100.00	100.00	0.00	66.67	0	0
78.33	cn	2	100.00	99.99	98.85	99.61	354	1
1406	cn	3	100.00	100.00	98.85	99.62	348	1
3092	cn	3	100.00	100.00	98.85	99.62	348	1
3661	cn	3	99.99	100.00	90.23	96.74	326	1
3691	cn	3	100.00	100.00	98.85	99.62	353	1
3741.01	cn	3	100.00	100.00	97.70	99.23	347	1
3868	cn	3	100.00	100.00	95.40	98.47	334	1
4380	cn	3	100.00	100.00	98.28	99.43	346	1
4907	cn	3	100.00	99.99	97.13	99.04	342	1
							Number of Sessions with TARGET perf > 80%:	**80**
							Total Number of CN Sessions:	**94**

B.3 MILD COGNITIVE IMPAIRMENT SUBJECTS

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
11.33	mci	1	100.00	99.99	94.25	98.08	345	1
135.33	mci	1	100.00	100.00	96.55	98.85	351	1
149.22	mci	1	99.99	99.97	90.80	96.92	351	1
160.22	mci	1	99.98	99.98	81.03	93.66	323	1
160.33	mci	1	99.99	100.00	95.98	98.66	354	1
2104	mci	1	100.00	99.99	92.12	97.37	381	1
2385	mci	1	99.98	99.96	66.09	88.68	308	0
2387	mci	1	100.00	100.00	96.55	98.85	347	1
3325	mci	1	100.00	100.00	0.00	66.67	0	0
3505	mci	1	99.96	99.95	70.69	90.20	310	0
3536	mci	1	100.00	100.00	0.00	66.67	0	0

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
36.22	mci	1	99.98	99.55	4.60	68.04	217	0
3886	mci	1	100.00	100.00	0.00	66.67	0	0
3973	mci	1	99.99	100.00	94.83	98.27	347	1
4019	mci	1	99.99	100.00	97.70	99.23	354	1
4053	mci	1	99.99	99.97	91.38	97.11	346	1
4186	mci	1	99.99	100.00	93.10	97.70	335	1
4308	mci	1	100.00	99.97	95.40	98.46	346	1
4357	mci	1	100.00	100.00	95.98	98.66	344	1
4360	mci	1	99.95	99.79	36.78	78.84	261	0
4376	mci	1	99.98	99.97	52.30	84.08	254	0
4501	mci	1	100.00	99.98	98.28	99.42	357	1
4548	mci	1	99.99	99.64	87.93	95.85	458	1
4596	mci	1	99.99	100.00	94.83	98.27	350	1
4684	mci	1	99.99	100.00	94.83	98.27	347	1
4688	mci	1	99.99	100.00	97.13	99.04	354	1
4712	mci	1	99.99	100.00	89.66	96.55	338	1
4729	mci_FAIL							0
4761	mci	1	100.00	99.99	97.70	99.23	351	1
4795	mci	1	100.00	100.00	0.00	66.67	0	0
4810	mci	1	99.97	99.89	92.53	97.46	413	1
4821	mci	1	100.00	100.00	99.43	99.81	356	1
4823	mci	1	100.00	99.98	99.43	99.80	362	1
4839	mci	1	100.00	99.99	98.28	99.42	357	1
4842	mci	1	100.00	99.99	98.28	99.42	361	1
4843	mci	1	100.00	100.00	0.00	66.67	0	0
4853	mci	1	99.92	99.74	10.34	70.00	321	0
4901	mci	1	99.99	99.98	93.68	97.88	350	1
4910	mci	1	100.00	100.00	96.55	98.85	345	1
4955	mci	1	100.00	100.00	93.68	97.89	336	1
4961	mci	1	99.99	100.00	99.43	99.81	357	1
5003	mci	1	100.00	100.00	0.00	66.67	0	0
5021	mci	1	99.99	100.00	95.98	98.66	348	1
5044	mci	1	100.00	100.00	97.70	99.23	354	1
5046	mci	1	100.00	100.00	94.25	98.08	332	1
5067	mci	1	99.98	99.97	93.10	97.69	425	1
5147	mci	1	100.00	99.99	98.85	99.61	352	1
5219	mci	1	100.00	99.99	98.85	99.61	353	1
5231	mci	1	99.99	99.97	92.53	97.50	348	1
5240	mci	1	100.00	100.00	98.85	99.62	352	1
5250	mci	1	99.99	100.00	95.98	98.66	349	1
5296	mci	1	99.99	100.00	97.13	99.04	353	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
5319	mci	1	99.98	99.98	66.67	88.88	239	0
5439	mci	1	99.99	99.99	86.21	95.40	321	1
5446	mci	1	99.99	99.93	90.15	96.69	417	1
5474	mci	1	99.98	99.98	94.25	98.07	357	1
5477	mci	1	99.96	99.95	57.14	85.69	324	0
5569	mci	1	100.00	100.00	98.28	99.42	351	1
5647	mci	1	99.98	99.99	89.66	96.54	415	1
5671	mci	1	100.00	100.00	97.13	99.04	340	1
5718	mci	1	100.00	100.00	0.00	66.67	0	0
5722	mci	1	99.99	99.99	97.70	99.23	354	1
5822	mci	1	99.98	99.97	90.23	96.73	357	1
5853	mci	1	100.00	100.00	0.00	66.67	0	0
5855	mci	1	99.99	99.99	94.25	98.08	338	1
5864	mci	1	99.99	99.98	99.43	99.80	365	1
5907	mci	1	99.98	100.00	96.55	98.84	427	1
5946	mci	1	100.00	100.00	0.00	66.67	0	0
6069	mci	1	100.00	100.00	0.00	66.67	0	0
6077	mci	1	99.99	99.98	65.52	88.50	252	0
6123	mci	1	100.00	100.00	0.00	66.67	0	0
6139	mci	1	100.00	100.00	0.00	66.67	0	0
6170	mci	1	100.00	100.00	95.98	98.66	338	1
6428	mci	1	100.00	100.00	99.43	99.81	348	1
6511	mci	1	100.00	100.00	98.85	99.62	345	1
6692	mci	1	100.00	100.00	95.98	98.66	340	1
6720	mci	1	99.99	100.00	94.83	98.27	341	1
11.33	mci	2	99.99	99.96	95.40	98.45	356	1
149.22	mci	2	99.99	99.99	94.25	98.08	356	1
2344	mci	2	100.00	100.00	0.00	66.67	0	0
3505	mci	2	99.97	99.97	58.62	86.19	265	0
36.22	mci	2	99.92	99.55	16.38	71.95	204	0
3981	mci	2	100.00	99.99	93.10	97.70	283	1
4015	mci	2	99.99	99.97	0.00	66.66	0	0
4308	mci	2	100.00	100.00	0.00	66.67	0	0
4360	mci	2	99.98	99.98	14.37	71.44	85	0
4376	mci	2	99.98	99.99	54.02	84.67	215	0
4548	mci	2	100.00	99.99	98.85	99.61	354	1
4688	mci	2	100.00	100.00	0.00	66.67	0	0
4729	mci	2	100.00	100.00	97.13	99.04	345	1
4761	mci	2	100.00	99.99	97.70	99.23	340	1
4821	mci	2	100.00	99.99	95.98	98.65	346	1
4839	mci	2	99.99	99.99	98.85	99.61	364	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
4842	mci	2	100.00	99.99	99.43	99.81	355	1
4907	mci	2	100.00	100.00	98.85	99.61	352	1
4910	mci	2	100.00	100.00	97.13	99.04	347	1
4961	mci	2	99.99	99.99	94.25	98.08	345	1
4975	mci	2	99.98	99.56	32.18	77.24	317	0
5219	mci	2	100.00	100.00	0.00	66.67	0	0
5231	mci	2	99.98	99.96	90.80	96.91	362	1
5241	mci	2	100.00	100.00	0.00	66.67	0	0
5439	mci	2	100.00	100.00	90.80	96.93	320	1
5477	mci	2	99.99	99.98	90.80	96.93	334	1
5855	mci	2	100.00	100.00	97.13	99.04	342	1
3505	mci	3	100.00	100.00	0.00	66.67	0	0
4729	mci	3	100.00	99.99	98.85	99.61	356	1
							Number of Sessions with TARGET perf > 80%:	**72**
							Total Number of MCI Sessions:	**105**

B.4 PARKINSON'S DISEASE SUBJECTS

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
4565	pd	1	99.99	99.99	97.13	99.04	356	1
4569	pd	1	100.00	100.00	98.85	99.62	354	1
4594	pd	1	100.00	100.00	97.70	99.23	352	1
4598	pd	1	99.99	100.00	97.13	99.04	362	1
4605	pd	1	99.99	99.99	97.70	99.23	359	1
4663	pd	1	100.00	100.00	95.40	98.47	349	1
4727	pd	1	99.99	99.89	91.63	97.17	451	1
4754	pd	1	99.99	99.98	95.98	98.65	361	1
4760	pd	1	100.00	99.97	85.06	95.01	319	1
4808	pd	1	100.00	100.00	85.06	95.02	309	1
4854	pd	1	99.99	99.92	92.12	97.34	423	1
4877	pd	1	99.98	99.96	75.29	91.74	313	0
4878	pd	1	99.99	99.99	91.95	97.31	352	1
4886	pd	1	100.00	100.00	98.85	99.62	349	1
4902	pd	1	100.00	99.99	99.43	99.81	359	1
4956	pd	1	99.99	99.99	94.25	98.08	342	1
4963	pd	1	99.99	100.00	94.83	98.27	340	1
4967	pd	1	99.99	100.00	98.28	99.42	352	1
4979	pd	1	100.00	100.00	99.43	99.81	350	1
4985	pd	1	100.00	100.00	98.85	99.61	349	1
500.99	pd	1	100.00	100.00	97.13	99.04	352	1
5045	pd	1	100.00	99.99	98.28	99.42	352	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
5049	pd	1	99.98	99.99	90.23	96.73	347	1
5054	pd	1	100.00	100.00	99.43	99.81	352	1
5056	pd	1	100.00	100.00	97.13	99.04	348	1
5069	pd	1	99.99	99.98	92.53	97.50	347	1
5079	pd	1	100.00	100.00	97.70	99.23	348	1
5104	pd	1	100.00	100.00	98.85	99.62	351	1
5160	pd	1	99.99	99.97	87.93	95.96	340	1
5183	pd	1	100.00	100.00	97.70	99.23	345	1
5198	pd	1	99.99	99.99	92.53	97.50	339	1
5207	pd	1	100.00	100.00	98.85	99.62	352	1
5224	pd	1	100.00	100.00	99.43	99.81	351	1
5226	pd	1	99.99	99.96	97.13	99.03	365	1
5234	pd	1	99.98	99.99	51.72	83.90	202	0
5235	pd	1	100.00	100.00	91.95	97.32	326	1
5256	pd	1	99.99	100.00	95.98	98.66	345	1
5262	pd	1	100.00	100.00	98.28	99.42	356	1
5263	pd	1	100.00	100.00	94.83	98.28	343	1
5265	pd	1	100.00	99.99	95.98	98.66	348	1
5271	pd	1	99.99	99.98	92.53	97.50	343	1
5272	pd	1	100.00	99.99	97.13	99.04	349	1
5277	pd	1	99.99	99.98	90.23	96.74	328	1
5278	pd	1	100.00	100.00	98.85	99.62	353	1
5298	pd	1	99.99	99.99	95.98	98.65	355	1
5324	pd	1	100.00	100.00	97.13	99.04	351	1
5331	pd	1	100.00	99.99	98.85	99.61	355	1
5341	pd	1	100.00	100.00	99.43	99.81	355	1
5346	pd	1	100.00	99.99	97.13	99.04	358	1
5366	pd	1	100.00	100.00	99.43	99.81	350	1
5421	pd	1	100.00	100.00	98.85	99.62	348	1
5426	pd	1	99.99	100.00	93.10	97.70	341	1
5430	pd	1	100.00	100.00	99.43	99.81	354	1
5438	pd	1	100.00	100.00	99.01	99.67	412	1
5457	pd	1	99.99	99.98	95.40	98.46	353	1
5487	pd	1	100.00	100.00	99.43	99.81	350	1
5494	pd	1	99.99	100.00	90.80	96.93	333	1
5500	pd	1	99.98	99.97	94.25	98.07	364	1
5503	pd	1	100.00	100.00	93.68	97.89	335	1
5507	pd	1	99.99	100.00	96.55	98.85	355	1
5525	pd	1	99.99	99.97	95.98	98.65	363	1
5553	pd	1	99.99	100.00	98.85	99.62	357	1
5554	pd	1	100.00	100.00	97.70	99.23	351	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
5584	pd	1	100.00	99.99	97.70	99.23	353	1
5594	pd	1	99.98	99.97	67.82	89.26	259	0
5609	pd	1	100.00	99.99	95.40	98.46	343	1
5627	pd	1	99.99	99.98	82.18	94.05	306	1
5635	pd	1	100.00	99.99	97.70	99.23	353	1
5672	pd	1	100.00	100.00	98.28	99.43	350	1
5696	pd	1	100.00	100.00	98.85	99.62	351	1
5723	pd	1	100.00	99.99	98.85	99.61	359	1
5725	pd	1	100.00	99.99	96.55	98.85	345	1
5744	pd	1	100.00	100.00	98.85	99.62	348	1
5754	pd	1	100.00	99.99	93.68	97.89	334	1
5756	pd	1	99.99	100.00	97.70	99.23	350	1
5767	pd	1	99.99	100.00	97.70	99.23	357	1
5770	pd	1	100.00	100.00	99.43	99.81	352	1
5774	pd	1	100.00	99.99	99.43	99.81	353	1
5794	pd	1	100.00	100.00	98.28	99.43	350	1
5795	pd	1	100.00	100.00	98.85	99.62	353	1
5804	pd	1	100.00	100.00	96.55	98.85	341	1
5835	pd	1	99.98	99.98	73.56	91.17	277	0
5866	pd	1	100.00	100.00	98.62	99.54	294	1
5880	pd	1	98.64	98.87	26.60	74.70	6134	0
5891	pd	1	100.00	100.00	98.85	99.62	354	1
5903	pd	1	99.95	99.92	83.33	94.40	373	1
5959	pd	1	99.98	99.97	89.08	96.34	340	1
5961	pd	1	100.00	100.00	98.28	99.42	351	1
5997	pd	1	100.00	99.99	99.43	99.80	356	1
6064	pd	1	100.00	100.00	94.25	98.08	335	1
6087	pd	1	100.00	100.00	95.98	98.66	346	1
6093	pd	1	100.00	100.00	99.43	99.81	355	1
6148	pd	1	100.00	100.00	0.00	66.67	0	0
6151	pd	1	100.00	100.00	0.00	66.67	0	0
6152	pd	1	100.00	100.00	95.40	98.47	352	1
6160	pd	1	100.00	99.99	95.98	98.65	346	1
6188	pd	1	100.00	100.00	97.13	99.04	345	1
6226	pd	1	100.00	100.00	99.43	99.81	356	1
6258	pd	1	100.00	100.00	0.00	66.67	0	0
6332	pd	1	100.00	100.00	0.00	66.67	0	0
6392	pd	1	100.00	100.00	0.00	66.67	0	0
6401	pd	1	100.00	100.00	0.00	66.67	0	0
6425	pd	1	100.00	100.00	0.00	66.67	0	0
6460	pd	1	100.00	99.99	89.08	96.36	315	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
6483	pd	1	99.99	99.99	93.68	97.89	335	1
6509	pd	1	100.00	100.00	0.00	66.67	0	0
6538	pd	1	100.00	99.99	98.85	99.61	350	1
6553	pd	1	100.00	100.00	98.85	99.62	349	1
6567	pd	1	99.99	100.00	95.98	98.66	342	1
3725	pd	2	99.99	99.99	85.06	95.02	311	1
4565	pd	2	100.00	100.00	0.00	66.67	0	0
4594	pd	2	100.00	100.00	0.00	66.67	0	0
4605	pd	2	100.00	100.00	81.61	93.87	290	1
4663	pd	2	100.00	100.00	99.43	99.81	352	1
4727	pd	2	99.99	99.99	95.40	98.46	346	1
4760	pd	2	99.99	99.96	89.66	96.54	346	1
4808	pd	2	100.00	100.00	0.00	66.67	0	0
4878	pd	2	99.46	99.76	41.38	80.20	3969	0
4886	pd	2	100.00	100.00	0.00	66.67	0	0
4902	pd	2	100.00	100.00	97.13	99.04	345	1
4963	pd	2	100.00	100.00	0.00	66.67	0	0
4967	pd	2	99.92	99.74	31.03	76.90	318	0
4979	pd	2	100.00	100.00	99.43	99.81	349	0
500.99	pd	2	100.00	100.00	0.00	66.67	0	1
5045	pd	2	100.00	99.98	95.98	98.65	343	0
5054	pd	2	100.00	100.00	99.43	99.81	347	1
5056	pd	2	100.00	99.99	97.13	99.04	344	1
5079	pd	2	100.00	100.00	0.00	66.67	0	1
5104	pd	2	100.00	100.00	0.00	66.67	0	0
5183	pd	2	100.00	100.00	85.63	95.21	296	0
5198	pd	2	99.99	99.98	87.36	95.77	328	0
5234	pd	2	99.98	99.98	35.06	78.34	154	1
5235	pd	2	100.00	100.00	99.43	99.81	351	1
5256	pd	2	99.99	99.99	89.08	96.35	330	0
5262	pd	2	100.00	100.00	0.00	66.67	0	1
5265	pd	2	99.99	99.99	91.38	97.12	336	1
5272	pd	2	100.00	100.00	94.83	98.27	336	0
5278	pd	2	100.00	100.00	97.13	99.04	343	1
5298	pd	2	100.00	100.00	94.25	98.08	336	1
5324	pd	2	100.00	100.00	97.70	99.23	343	1
5331	pd	2	100.00	99.99	97.70	99.23	345	1
5346	pd	2	100.00	99.98	96.55	98.84	349	1
5421	pd	2	100.00	100.00	99.43	99.81	348	1
5430	pd	2	100.00	100.00	99.43	99.81	349	1

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
5457	pd	2	100.00	99.99	98.28	99.42	348	1
5487	pd	2	100.00	100.00	99.43	99.81	349	1
5554	pd	2	100.00	100.00	95.40	98.47	340	1
5594	pd	2	99.94	99.98	46.55	82.16	264	1
5672	pd	2	100.00	100.00	96.55	98.85	340	1
5696	pd	2	100.00	100.00	98.85	99.62	348	0
5795	pd	2	100.00	100.00	98.28	99.42	348	1
5835	pd	2	99.97	99.97	70.11	90.02	282	1
5891	pd	2	99.99	100.00	97.13	99.04	349	1
5959	pd	2	99.97	99.97	85.63	95.19	334	0
5961	pd	2	100.00	100.00	99.43	99.81	348	1
						Number of Sessions with TARGET perf > 80%:		**126**
						Total Number of PD Sessions:		**155**

B.5 PARKINSON'S DISEASE WITH DEMENTIA SUBJECTS

Patient	Class	Visit	STD	NOV	TAR	AVG	Events	Thresh
238.99	pdd	1	100.00	100.00	0.00	66.67	0	0
4728	pdd	1	99.89	99.78	32.18	77.28	364	0
4818	pdd	1	99.98	99.84	52.30	84.04	295	0
4972	pdd	1	99.77	99.58	57.47	85.61	705	0
5043	pdd	1	99.99	100.00	67.82	89.27	275	0
5264	pdd	1	99.98	99.97	85.71	95.22	397	1
5312	pdd	1	100.00	100.00	0.00	66.67	0	0
5476	pdd	1	99.99	99.99	94.25	98.08	345	1
5552	pdd	1	100.00	100.00	0.00	66.67	0	0
5755	pdd	1	99.96	99.97	81.61	93.85	340	1
5935	pdd	1	100.00	100.00	0.00	66.67	0	0
6030	pdd	1	99.98	99.98	78.16	92.71	307	0
6195	pdd	1	100.00	100.00	0.00	66.67	0	0
6237	pdd	1	100.00	100.00	0.00	66.67	0	0
6289	pdd	1	99.99	99.99	97.13	99.04	358	1
4598	pdd	2	99.99	99.99	95.40	98.46	349	1
4754	pdd	2	100.00	100.00	0.00	66.67	0	0
5043	pdd	2	99.99	99.99	37.36	79.11	150	0
5069	pdd	2	100.00	99.99	94.83	98.27	346	1
5494	pdd	2	99.98	99.99	70.11	90.03	292	0
						Number of Sessions with TARGET perf > 80%:		**6**
						Total Number of PDD Sessions:		**20**

B.6 BUTTON PRESS ACCURACY STATISTICS (COHORT B)

Sessions with Target Perf > 80%: **348** **71.60%** of total

Session Counts:		AD	CN	MCI	PD	PDD
1st:	328	75	53	76	109	15
2nd:	145	34	33	27	5	46
3rd:	13	3	8	2	0	0
Total Checked:	**486**					

Four sessions not checked (total sessions is 490). This is because of either not enough trials for the subject were present, or due to corrupt files unable to be loaded in MATLAB for some reason.

NED: not enough data (too few trials, needs at least 4 trials)

FAIL: could not load data file in MATLAB (corrupt)

Total Sessions Available for Each Class:

AD	CN	MCI	PD	PDD	Overall
112	94	105	155	20	**486**

Average Target Button Press Performance:

AD	CN	MCI	PD	PDD	Overall
78.05	95.90	75.58	93.01	64.78	**81.47**

Note: Any trials with 'events' less than 200 were omitted from these average calculations--this was done due to the hardware failure that occurred with the button mechanism at one point. Any event-count less than 200 for any given trial generally indicates button failure (not patient mistakes).

APPENDIX C: ERP SUBJECT ANALYSIS

Analysis was performed for all subjects used in this study to determine the clarity and accuracy of their respective ERPs. A complete collection of all 202 AD/CN and 272 MCI/PD/PDD figures is provided in both an online (compressed ZIP archive) and hard copy (CD media) format, upon request (please contact via email mahiskali@ieee.org or polikar@rowan.edu for a full electronic copy of all patient figures). A sample of such a figure is shown in Figure 39.

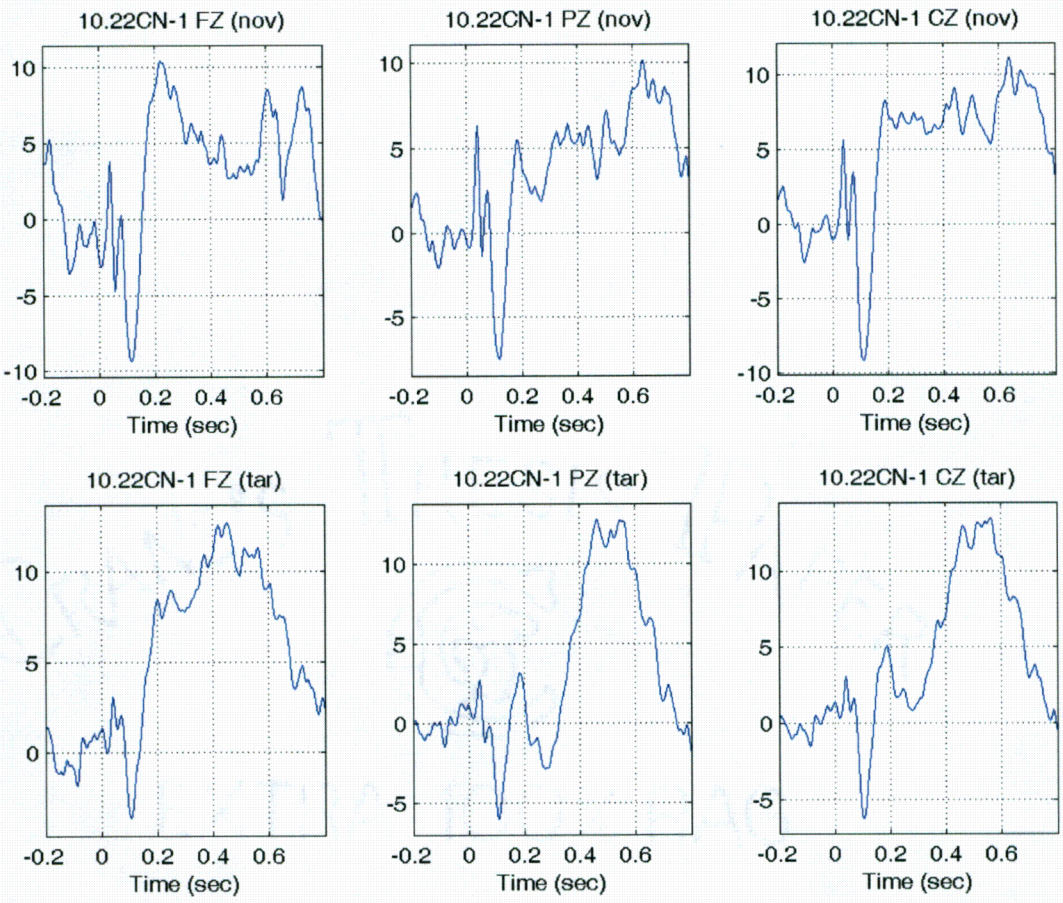

Figure 39 – Sample ERP for subject analysis