Ensemble of classifiers based data fusion of EEG and MRI for diagnosis of neurodegenerative disorders

Tejash Patel
Rowan University
ENSEMBLE OF CLASSIFIERS BASED DATA FUSION OF EEG AND MRI
FOR DIAGNOSIS OF NEURODEGENERATIVE DISORDERS

by
Tejash Patel

A Thesis Submitted to the
Graduate Faculty in Partial Fulfillment
of the Requirements for the Degree of
MASTER OF SCIENCE

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

Approved: Members of the Committee

In Charge of Major Work

For the Major Department

For the College

Rowan University
Glassboro, New Jersey
2008
© Tejash Patel
The prevalence of Alzheimer’s disease (AD), Parkinson’s disease (PD), and mild cognitive impairment (MCI) are rising at an alarming rate as the average age of the population increases, especially in developing nations. The efficacy of the new medical treatments critically depends on the ability to diagnose these diseases at the earliest stages. To facilitate the availability of early diagnosis in community hospitals, an accurate, inexpensive, and noninvasive diagnostic tool must be made available. As biomarkers, the event related potentials (ERP) of the electroencephalogram (EEG) - which has previously shown promise in automated diagnosis - in addition to volumetric magnetic resonance imaging (MRI), are relatively low cost and readily available tools that can be used as an automated diagnosis tool.

16-electrode EEG data were collected from 175 subjects afflicted with Alzheimer's disease, Parkinson's disease, mild cognitive impairment, as well as non-disease (normal control) subjects. T2 weighted MRI volumetric data were also collected from 161 of these subjects. Feature extraction methods were used to separate diagnostic information from the raw data. The EEG signals were decomposed using the discrete wavelet transform in order to isolate informative frequency bands. The MRI images were processed through segmentation software to provide volumetric data of various brain regions in order to quantize potential brain tissue atrophy. Both of these data sources were utilized in a pattern recognition based classification algorithm to serve as a
diagnostic tool for Alzheimer’s and Parkinson’s disease. Support vector machine and multilayer perceptron classifiers were used to create a classification algorithm trained with the EEG and MRI data. Extracted features were used to train individual classifiers, each learning a particular subset of the training data, whose decisions were combined using decision level fusion. Additionally, a severity analysis was performed to diagnose between various stages of AD as well as a cognitively normal state.

The study found that EEG and MRI data hold complimentary information for the diagnosis of AD as well as PD. The use of both data types with a decision level fusion improves diagnostic accuracy over the diagnostic accuracy of each individual data source. In the case of AD only diagnosis, ERP data only provided a 78% diagnostic performance, MRI alone was 89% and ERP and MRI combined was 94%. For PD only diagnosis, ERP only performance was 67%, MRI only was 70%, and combined performance was 78%. MCI only diagnosis exhibited a similar effect with a 71% ERP performance, 82% MRI performance, and 85% combined performance. Diagnosis among three subject groups showed the same trend. For PD, AD, and normal diagnosis ERP only performance was 43%, MRI only was 66%, and combined performance was 71%. The severity analysis for mild AD, severe AD, and normal subjects showed the same combined effect.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 ALZHEIMER’S DISEASE</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 DIAGNOSIS</td>
<td>2</td>
</tr>
<tr>
<td>1.2 PARKINSON’S DISEASE</td>
<td>4</td>
</tr>
<tr>
<td>1.3 OBJECTIVES OF THIS STUDY</td>
<td>5</td>
</tr>
<tr>
<td>1.4 ORGANIZATION OF THIS THESIS</td>
<td>6</td>
</tr>
<tr>
<td>CHAPTER 2 BACKGROUND</td>
<td>7</td>
</tr>
<tr>
<td>2.1 CLINICAL DIAGNOSIS OF ALZHEIMER’S DISEASE</td>
<td>7</td>
</tr>
<tr>
<td>2.1.1 PATHOLOGY OF ALZHEIMER’S DISEASE</td>
<td>8</td>
</tr>
<tr>
<td>2.2 PATHOLOGY OF PARKINSON’S DISEASE</td>
<td>11</td>
</tr>
<tr>
<td>2.2.1 DIAGNOSIS OF PARKINSON’S DISEASE</td>
<td>12</td>
</tr>
<tr>
<td>2.3 BIOCHEMICAL BIOMARKERS</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1 AUTOMATED DIAGNOSIS WITH BIOCHEMICAL BIOMARKERS</td>
<td>14</td>
</tr>
<tr>
<td>2.4 MEDICAL IMAGING</td>
<td>16</td>
</tr>
<tr>
<td>2.4.1 MAGNETIC RESONANCE IMAGING</td>
<td>18</td>
</tr>
<tr>
<td>2.4.1a DIAGNOSIS WITH MRI</td>
<td>20</td>
</tr>
<tr>
<td>2.5 ELECTROENCEPHALOGRAM</td>
<td>22</td>
</tr>
<tr>
<td>2.5.1 SPECTRAL CONTENT OF THE EEG</td>
<td>24</td>
</tr>
<tr>
<td>2.6 ACQUISITION PROTOCOLS AND EVENT RELATED POTENTIALS</td>
<td>26</td>
</tr>
<tr>
<td>2.7 P300 HISTORY</td>
<td>30</td>
</tr>
<tr>
<td>2.8 P300 ANALYSIS USING WAVELETS</td>
<td>32</td>
</tr>
<tr>
<td>2.9 EVENT RELATED POTENTIALS IN ALZHEIMER’S DISEASE RESEARCH</td>
<td>33</td>
</tr>
<tr>
<td>2.10 ALZHEIMER’S DISEASE CLASSIFICATION</td>
<td>35</td>
</tr>
<tr>
<td>CHAPTER 3 METHODS</td>
<td>40</td>
</tr>
<tr>
<td>3.1 PREVIOUS WORK SPECIFIC TO THIS STUDY</td>
<td>40</td>
</tr>
<tr>
<td>3.2 CURRENT RESEARCH</td>
<td>42</td>
</tr>
<tr>
<td>3.3 STUDY COHORT</td>
<td>44</td>
</tr>
<tr>
<td>3.4 ERP ACQUISITION</td>
<td>47</td>
</tr>
<tr>
<td>3.5 FEATURE EXTRACTION</td>
<td>50</td>
</tr>
<tr>
<td>3.6 WAVELET TRANSFORMS</td>
<td>51</td>
</tr>
<tr>
<td>3.6.1 CONTINUOUS WAVELET TRANSFORM</td>
<td>52</td>
</tr>
<tr>
<td>3.6.2 DISCRETE WAVELET TRANSFORM</td>
<td>53</td>
</tr>
</tbody>
</table>
APPENDIX G: ERP GRAND AVERAGES FROM SEVERITY ANALYSIS DATA FUSION COHORT (MMSE CUTOFF = 23) ..... 205
APPENDIX H: ERP GRAND AVERAGES FROM SEVERITY ANALYSIS DATA FUSION COHORT (MMSE CUTOFF = 26) ..... 222
LIST OF FIGURES

Figure 2.1: Amyloid plaques and neurofibrillary tangles due to Alzheimer’s disease [12]. ....................................................... 9
Figure 2.2: Progressive shrinking of the brain shown in preclinical, mild, and severe Alzheimer’s Disease [10] .............................................. 11
Figure 2.3: T1 vs. T2 weighting. T1 & T2 weighted MRI of the brain [26] ......... 20
Figure 2.4: International 10/20 System of electrode placement and naming. . . . 23
Figure 2.5: ERP signal with prominent features labeled. .......................... 28
Figure 2.6: Expected P300 behavior, and P300 not following typical behavior . 30

Figure 3.1: The overall process of this study is shown followed by detailed depictions of each individual expert for each data type. ..................... 44
Figure 3.2: Artifact removal process. ............................................. 49
Figure 3.3: A diagram of the seven levels of discrete wavelet transform decomposition for an ERP signal with 128 Hz maximal frequency content. . . . 57
Figure 3.4: Wavelet and Scaling functions as well as decomposition and reconstruction filters for the Daubechies 4 wavelet. ......................... 59
Figure 3.5: Full wavelet decomposition of an ERP signal. .......................... 60
Figure 3.6: Model of a multilayer perceptron neural network ..................... 62
Figure 3.7: Maximal margin between two linearly separable classes. ............ 68
Figure 3.8: Decision boundaries in non-linearly separable data with slack variables. ................................................................................. 71
Figure 3.9: The sequence of processing of a raw MRI through a tissue segmentation step and through a region of interest analysis. ................. 75
Figure 3.10: K-Fold cross validation. Within each run the shaded data block is used for testing with the other data blocks used for training. ............... 77
Figure 3.11: Diagram of the ensemble based classification system employed. . 79
LIST OF TABLES

Table 3.1: Patient Cohort ................................... 47
Table 3.2: Explanation of diagnostic performance metrics .............. 81

Table 4.1: Results for the classification of the AD patient group from the Control patient group using weighted majority decision fusion .......... 83
Table 4.2: Results for the classification of the AD patient group from the Control patient group ........................................... 85
Table 4.3: Results for the classification of the MCI patient group from the Control patient group ........................................... 86
Table 4.4: Results for the classification of the PD patient group from the Control patient group ........................................... 87
Table 4.5: Results for the classification of the PD patient group and AD patient group from the Control patient group ........................................... 88
Table 4.6: Results for the classification of the mild and severe AD patient groups from the control patient group ........................................... 90
CHAPTER 1
INTRODUCTION

Modern medical technologies have enjoyed numerous advances over the past century. The widespread adoption of modern medical technologies has increased the average life expectancy from 47.3 to 77.8 years over the past century in the U.S. [1]. As the most populous generation in American history, consisting of 78.2 million people, enter their 60s, the prevalence of diseases that afflict the elderly is increasing dramatically. The number of patients afflicted with Alzheimer’s disease (AD) and Parkinson’s disease (PD) is not only increasing domestically but also worldwide. It is predicted that there are 26 million cases of Alzheimer’s disease and 4 million cases of Parkinson’s disease worldwide, which make these diseases a global health concern [2,3]. Currently, treatments can slow the progression of AD only if it is diagnosed at its earliest stages. However, current diagnostic techniques can be time consuming and expensive, requiring multiple visits to highly trained specialists who may not be able to identify the diseases in their earliest stages. This study explores whether a combination of EEG and MRI based features can improve the accuracy of an automated classification system for early diagnosis over either of the individual data sources.

1.1 ALZHEIMER’S DISEASE

There are an estimated 5.1 million cases of AD in the United States alone, with the number of global cases estimated to be drastically higher. Of the cases in the United States over 96% represent patients that are aged 65 and older. The prevalence of the disease increases rapidly with age, especially with patients who are over the age of 65.
The disease affects an average of 2% of those between the ages of 65 and 74, 19% of those between 75 and 84, and an alarming 42% of people over the age of 85 [4]. By 2050, the number of individuals in the United States age 65 and over with AD could rise to 11 to 16 million. By this date, more than 60 percent of the AD population will be aged 85 and above [4]. Alzheimer’s disease is already the fifth leading cause of death in the United States in those age 65 and older, after heart disease, cancers, cerebrovascular disease, and respiratory disease. With the prevalence expected to increase drastically, AD could become a much more severe cause of death in the coming decades [1].

AD not only imposes an emotional burden on patients and their caregivers, but it also inflicts major economic impact to all involved in caretaking. The global cost of dementia was estimated to be $315.4 billion annually in 2005. North American patients with dementia had annual costs averaging $24,000 [5].

1.1.1 DIAGNOSIS

Treatment of Alzheimer’s disease is currently hindered due to the lack of availability of a clinical test to diagnose the disease. At this time, the disease can only be diagnosed conclusively post-mortem. Since there is no disease modifying treatment, there exists a large desire for early diagnosis of AD so that new treatments can be administered and tested. Those being screened for AD can expect a diagnosis with an accuracy of 80-90% if examined by expert clinicians specializing in memory disorders [6]. These clinicians made their diagnosis primarily based on interviews with the patients and caregivers, medical history, clinical observation, and memory tests, all of which are administered over a period of weeks or months [7]. Other tests may be administered to rule out other diseases, but unfortunately, most patients do not reach this stage until their family and
caregivers start to notice external symptoms of the disease, which may be too late to apply any currently available treatments. Many patients cannot afford to or do not have access to expert clinicians. More importantly, the diagnostic accuracy of expert clinicians is not entirely replicable by general practitioners at local hospitals or community clinics. Diagnosis at these facilities is less exact, with an accuracy of 75%, sensitivity of 83%, and specificity of 55% [6].

Currently, the most accurate method is only viable post-mortem. In this post mortem test, brain tissue is extracted through biopsy and analyzed under a microscope to find protein markers of AD. However, this test is be extremely and unacceptably invasive and is unnecessarily risky in living patients. Therefore, biopsy is not in the normal testing regimen [8].

This study aims to examine three diagnostic biomarkers of Alzheimer’s disease: biochemical, anatomical, and physiological. These biomarkers are obtained through clinical procedures and may provide more nuanced information about the patient’s condition than a written evaluation. Biochemical markers, such as beta amyloid and hyperphosphorylated tau proteins, are found in cerebral spinal fluid, and are linked to AD pathology. The anatomical marker for AD is the atrophy of certain regions of interest in the brain, detected by magnetic resonance imaging (MRI). The physiological markers of the integrity of neuronal systems are provided by the detection and measurement of event related potentials (ERPs) of the electroencephalogram (EEG).
1.2 PARKINSON’S DISEASE

As the average age of the population increases, many elderly are susceptible to neurodegenerative diseases at a much higher rate. The number of patients afflicted with Parkinson's disease (PD) is not only increasing domestically but also worldwide. PD currently afflicts 1.5 million Americans with approximately 60,000 new cases diagnosed each year. The condition usually develops after the age of 65 but 15% of current cases are diagnosed before the age of 50 [9]. It is predicted that there are 4 million cases of Parkinson's disease globally, which makes PD a global health concern. Parkinson’s disease is a neurodegenerative disorder that occurs when neurons in the substantia nigra region of the brain begin to die. The neurons in this region are responsible for creating a chemical called dopamine, which, among other functions, is responsible for smooth, coordinated muscular movement. It is only when 80% of the dopamine producing cells become damaged that the symptoms of Parkinson’s begin to appear. The symptoms of the disease are mainly the loss of muscular ability and loss of coordination. The symptoms include shaking, stiffness, slow movement, and difficulty with balance.

Presently, treatments can slow the progression of PD but only after diagnosis. Current diagnosis can be time consuming and expensive since the main goal is to rule out other diseases, which usually is not done at the earliest stages. Most often, PD is not fully diagnosed until symptoms appear. This portion of the study explores whether a combination of EEG and MRI based features can improve the accuracy of an automated classification system for early diagnosis over either of the individual data sources.
1.3 OBJECTIVES OF THIS STUDY

The goal of this study has been to improve on previous work done to create a clinically available, automated, non-invasive, and readily available tool for the early diagnosis of Alzheimer's disease by using data fusion of heterogeneous data sources. In this study, event related potentials (ERPs) are acquired from the electroencephalogram (EEG) as well as volumetric magnetic resonance images from cohorts of normal, AD, PD, and MCI patients.

Two separate groups of data analysis were conducted. First, ERP only AD diagnosis was performed with a more expanded cohort than the previous study. The cohort in the ERP only group consisted of 49 normal and 49 AD patients. Second, data fusion analysis was done to diagnose various 2 and 3 group combinations of patient cohorts as follows: AD (49 patients) vs. Normal (34 patients); MCI (39 patients) vs. Normal (34 patients); PD (39 patients) vs. Normal (34 patients); and AD (49 patients) vs. PD (39 patients) vs. Normal (34 patients). The MRI volume data and the EEG spectral band wavelet coefficients were used to train and test an ensemble of neural network classifiers to form a data fusion based diagnosis of AD. The data fusion techniques developed to improve the AD diagnosis problem were applied to the diagnosis of other dementias such as Parkinson's Disease (PD) and mild cognitive impairment (MCI) in order determine the efficacy of this testing method for diagnosing various dementias. The data fusion techniques were applied to a severity analysis of AD to determine the possibility of distinguishing between mild and severe stage AD patients from normal.
1.4 ORGANIZATION OF THIS THESIS

A detailed literature review is presented in Chapter 2 reviewing biological markers for Alzheimer’s disease and other neurodegenerative diseases as well as current diagnosis techniques. Chapter 3 outlines the methodology used in this study, such as ERP signal acquisition, preprocessing, and feature extraction, MRI acquisition and feature extraction, as well as ensemble system and data fusion techniques. Chapter 4 provides implementation details as well as results. Chapter 5 provides a discussion of the presented results along with discussions of error sources, study conclusions, and suggestions for future work.
CHAPTER 2
BACKGROUND

Although the established method of diagnosis of Alzheimer’s disease centers on clinical interview of patient and caretakers, there is a drive to find reliable biomarkers to aid in diagnosis. New biomarkers that are under consideration typically fall into one of four categories: biochemical, anatomical, metabolic and physiological markers.

2.1 CLINICAL DIAGNOSIS OF ALZHEIMER’S DISEASE

If there is reason to suspect that a patient may have dementia, the first step is usually to administer a battery of standardized neuropsychological tests. The most frequently used test is the Mini Mental State Exam (MMSE), which consists of a small number of simple questions about the current time, date, location, etc. The patient is also asked to perform simple tasks to test their cognitive state and short-term memory, such as counting or spelling backwards, copying a diagram, or writing a sentence. The tests are scored zero to thirty where zero is a vegetative state and 30 is normal cognitive ability. This test, just like any singular test from the diagnostic process, is not used independently to make a diagnosis. Similar tests include obtaining a patient’s Clinical Dementia Rating (CDR), Severe Impairment Battery (SIB), or other tests that target specific skills such as memory, cognitive ability, or psychomotor skills [8].
2.1.1 PATHOLOGY OF ALZHEIMER'S DISEASE

The human brain functions as a massively interconnected network of neurons. Each neuron can be a node in multiple pathways that work together to perform normal cognitive and motor tasks. The physiological occurrences that are hallmarks of Alzheimer's disease detriment this messaging network, which causes eventual loss of cognitive ability. Two misfolded proteins commonly referred to as plaques and tangles cause the neuronal blockages that lead to cognitive decline.

One of the misfolded proteins that lead to Alzheimer's disease is called beta-amyloid protein that consists of segmented fragments from a larger protein called amyloid precursor protein (APP). Protein fragments such as beta-amyloids are normally metabolized and removed in healthy brains. Although the APP protein that appears to be important in helping neurons grow and survive, its decomposed form, beta-amyloid, eventually develops into harmful plaques. Enzymes decompose the APP into beta-amyloids, which then begin to clump with other molecules to form insoluble amyloid plaques outside of neurons and stick to their membranes. These plaques tend to develop within the hippocampus, a memory center deep within the brain, as well as parts of the cerebral cortex used in thinking and making decisions. It is not known whether beta-amyloid plaques themselves cause AD or if they are a byproduct of some other neurological process of AD [10]. Alterations in the structure of APP is known to cause a rare, inherited form of AD, but it is also possible that large accumulations of these insoluble plaques on the membranes of neurons can cause blockages in the synapses' communication which eventually leads to cell death [11].
The other protein linked to AD is found within the structure of a neuron. Healthy neurons utilize microtubules as part of an organized internal support structure. The microtubules not only provide structural integrity, but they also help guide nutrients and molecules down the cell to far away axons. The hyperphosphorylated tau protein helps stabilize these microtubules. In AD, the tau proteins begin to pair with other threads of tau proteins and become tangled creating neurofibrillary tangles. This tangling destroys the structural integrity of the microtubules preventing vital molecules from being transported leading to communication malfunctions and eventually cell death [10]. The detrimental effects of amyloid plaques and neurofibrillary tangles are illustrated in Figure 2.1.

Figure 2.1: Amyloid plaques and neurofibrillary tangles due to Alzheimer's disease [12]

The neuronal death that occurs due to the plaques and/or tangles leads to the overall atrophy of the brain. The brain is affected progressively by the four stages of AD: preclinical, mild, moderate, and severe. The course of AD is not the same in each patient but symptoms seem to develop over the same general stages. In preclinical
stages, AD begins to affect the areas within the hippocampus that are instrumental in the formation of memories. These areas begin to atrophy. The onset of AD could go unnoticed for up to twenty years before the first symptoms of memory loss occur [10]. If a diagnosis is made at this stage, it is usually a diagnosis of mild cognitive impairment (MCI), which is hypothesized to be the transitional phase between a normal brain and one with AD. Not all MCI patients develop AD however, that rate stands at about 30% [13]. Once the first symptoms are apparent and an AD diagnosis is made, the average life expectancy of the patient is 8-10 years. Life expectancy varies amongst patients depending on the current stage of AD during diagnosis as well statistical risk factors such as age [13].

As the disease progresses and begins to affect the cerebral cortex, memory loss continues and begins to affect cognitive abilities. Poor judgment, loss of initiative, increased anxiety, and changes in mood and personality occur at this stage. Clinical diagnosis of AD is usually made during this stage as patients and caretakers begin to accept that memory and cognitive changes that are occurring are more than what would occur during normal aging. By the moderate stage of AD, areas of the cerebral cortex, frontal, and parietal regions that control language, reasoning, and consciousness are affected. At this stage patients have difficulty organizing thoughts logically, lose impulse control, and have problems recognizing friends and family. Severe AD is the final stage where plaques and tangles are widespread throughout the brain leading to massive brain atrophy. The patients cannot recognize people and cannot communicate cohesively. Physically, patients lose weight and lose the most basic motor skills such as swallowing as well as bladder and bowel control. The progressive shrinking of the brain, illustrated
in Figure 2.2, is then at its most extreme, and most cognitive and motor ability has been lost.

![Brain Shrinkage Diagrams](image)

**Figure 2.2** Progressive shrinking of the brain shown in preclinical, mild, and severe Alzheimer's Disease [10].

### 2.2 PATHOLOGY OF PARKINSON'S DISEASE

Dopamine is a neurotransmitter needed to facilitate inter-neuronal communication. The dopamine from the substantia nigra is used to communicate to the corpus striatum to produce smooth and controlled movement. In a Parkinson’s disease brain, the abnormal communication between the substantia nigra and the corpus striatum due to the lack of dopamine production results in abnormal nerve firing patterns within the brain that cause impaired movement. By the time symptoms of PD have appeared, it is possible that the
patient has lost 60-80% of their dopamine producing cells in the substantia nigra. It has been hypothesized that Lewy bodies are the missing link in fully understanding the brain degeneration in Parkinson’s disease. One of the main characteristics in a degenerating dopaminergic cell is the presence of intercellular sphere-shaped structures in the brain cells known as Lewy bodies. These structures may prevent the cell from functioning normally or may be just a byproduct of the degenerative process, but they are present in all PD patients at autopsy. Amyloid beta, the protein that plays a major role in Alzheimer’s disease also may cause atrophy in the substantia nigra region in PD patients, which may help explain the co-existence of AD and PD in many patients. Genetic markers have also been discovered which may help link familial cases of PD. However, even in familial cases, exposure to toxins or environmental factors may affect the progression of the disease, such as accelerated development of symptoms. Compounds derived from synthetic heroin as well as viruses such as influenza and herpes are known to cause Parkinson’s like symptoms due to the inflammation of the substantia nigra. [14].

2.2.1 DIAGNOSIS OF PARKINSON'S DISEASE

Current diagnostic methods for PD are only viable when symptoms become apparent. The physician will diagnose the disease after a thorough mental and physical examination. Blood tests as well as MRI can be used to rule out other neurodegenerative disorders but cannot directly diagnose PD [9]. Positron emission tomography (PET) techniques are being investigated for the diagnosis of PD. PET scans are practical since symptoms of PD result from a deterioration of the metabolic pathways that result in decreased dopamine production. PET sans can be used in conjunction with imaging drug
markers that act like dopamine to observe changes in dopaminergic systems. The uptake of the drug marker begins to occur very early in the course of PD, well before symptoms become apparent. This technique can be used to help eliminate other causes of physical symptoms and in the future may be able to conclusively diagnose PD [15].

2.3 BIOCHEMICAL BIOMARKERS

The biochemical markers for the early detection of Alzheimer's disease that are currently being studied are found in the cerebral spinal fluid (CSF). The CSF is the fluid surrounding the brain within the brain's ventricles, and also within the spinal column. The CSF can be obtained through a procedure called a lumbar puncture, which extracts CSF from the spinal column from the lumbar region of the spine. Although this procedure is highly invasive, painful to the patient, and expensive, CSF has proven to be one of the most reliable predictors of AD pathology.

Proteins such as β-amyloid and tau are found in the CSF and have been linked to AD pathology. Accumulation of these proteins in an AD affected brain interferes with signaling at the synapses, and eventually causes neuronal death. The levels of tau and β-amyloid can be detected in elevated levels in the CSF of patients with Alzheimer's disease obtained through a lumbar puncture, and hence CSF concentrations of these proteins have proven to be reliable in diagnosing AD in its earliest stages [16]. CSF biomarkers are not currently used in standard testing regimens due to their relative unavailability to rural patients who are not in proximity to a physician skilled in the procedure, its high cost, and to the variation in test results observed between various testing sites which prevents baseline standardization of the test [17].
2.3.1 AUTOMATED DIAGNOSIS WITH BIOCHEMICAL MARKERS

It is known that beta amyloid and tau proteins are present during AD onset. A 1999 study investigated the performance of the tau protein as a potential biomarker for AD. AD patients were from a community-based sample, including 407 AD patients (274 with probable AD and 133 with possible AD), 28 patients with depression, and 65 healthy elderly control subjects. Lumbar punctures were performed to obtain CSF. Hyperphosphorylated tau levels were elevated in probable and possible AD patients. Based on this analysis, a cutoff level of 302 pg/mL resulted in diagnostic performances of 93% sensitivity and 86% specificity. It was concluded that CSF tau levels remain stable over time with little inter-individual variation, and that this type of test could be beneficial to identify early onset AD [18].

Another 1999 study explored whether specific levels of platelet amyloid beta precursor protein (APP) are related to AD, and whether it shows a correlation with the progression of clinical symptoms. The subject cohort had 32 AD probable, 25 age matched control, and 16 non-AD dementia subjects. CSF levels of APP were evaluated by using Western blotting analysis and immuno-staining of whole platelets. The study concluded that occurrence of the APP biomarker is specifically altered in patients with AD. These alterations have a positive correlation with the progression of clinical symptoms, which supports the notion of the APP count being a diagnostic marker of the progression of AD [19].

The precursor to the beta amyloid protein, APP, has been shown to be a marker of the progression of AD in other studies as well. Di Luca, et al, 2005, evaluated the accuracy of artificial neural networks in automated AD diagnosis.
control subjects were enrolled and APP measures were taken. Fifteen different ANN models were trained and tested such as feed-forward and complex recurrent ANNs based on different learning laws (back propagation, sine-net, bi-modal) were compared to linear discriminant analysis (LDA). The best result was obtained with the Self Recurrent Network Dynamic Sine Net model, which reached a diagnostic performance of 93.08%. The corresponding result using LDA was 81.6%. This study concluded through extensive modeling and simulation that an automated procedure can be created to diagnose AD with beta amyloid precursor biochemical markers [20].

Maddelena, et al., 2003, also explored whether two protein biomarkers of AD would aid in diagnosis. Their study measured the levels of these two proteins in 100 patients with dementia and 31 control subjects. The study concluded that the ratio of hyperphosphorylated tau to beta amyloid was significantly increased in patients with AD. The measure provided a diagnostic accuracy in distinguishing AD patients of 86% sensitivity and 97% specificity. This test also showed subjects with non-AD dementias had diagnostic performances of 80% sensitivity and 73% specificity, and subjects with other neurological disorders had diagnostic performances of 80% sensitivity and 89% specificity [21].
2.4 MEDICAL IMAGING

Various forms of medical imaging can be utilized to quantify other markers of Alzheimer’s disease, and in fact have been used to monitor the progression and diagnosis of AD. The most common imaging modalities in use are magnetic resonance imaging (MRI), positron emission tomography (PET), and computed tomography (CT).

An anatomical marker for AD is the atrophy of certain regions of interest within the brain, which can be detected by MRI. MRI accentuates contrast between tissues based on the ratio of bound to unbound water molecules. Brain matter has more bound water molecules compared to the surrounding CSF, therefore using a T2 weighted MRI which shows water and fluid containing tissues as bright regions, the CSF is accentuated [22]. Volumetric software using region of interest analysis can isolate the low contrast brain matter from the high contrast CSF. This information can be used to compile slices of brain images to determine the overall brain volume, or of various regions, using segmentation software. Despite the possibility of a brain region volume based diagnosis with a single scan, time-course studies tend to provide a better visualization of brain atrophy [23,24].

Contrary to MRI, PET scan provides a metabolic marker, measuring the glucose metabolism in all regions of the brain. Since both the gray matter and white matter atrophy, and loss of metabolism are directly linked to neuronal death, monitoring these two metrics in combination can provide a better understanding of current brain state [25]. Furthermore, since the symptoms of Parkinson’s disease result from a deterioration of the metabolic pathways within the brain, PET scan data may also prove to be more beneficial for PD diagnosis [26,14].
The potential for these technologies to provide a more analytic based diagnosis over current methods is apparent; however, the high cost of these tests that require specialized facilities and physicians who are often not located in community clinics become limiting factors. Additionally, the time requirement to conduct a time course patient evaluation prevents these tests from being adopted in mainstream diagnostic regimens.

A recently developed imaging modality that shows some potential in low cost AD diagnosis is called near infrared spectroscopy (NIRS). This modality utilizes infrared light based imaging to measure changes in cortical hemoglobin oxygenation during various cognitive tasks. The imaging has the ability to noninvasively penetrate skin and bone to observe hemodynamic changes. Although common signal processing techniques are used in analysis, more research is necessary to develop a dependable diagnosis tool [27,28].

Fluorescence Anisotropy is another technique being developed for the early diagnosis of Alzheimer’s disease. This technique studies the aggregation of the beta amyloid protein. A small volume of fluorescently labeled beta amyloid protein is combined with unlabeled beta amyloid as well as other necessary chemicals and aggregation is observed. This method is able to detect beta amyloid aggregation well before other known methods. This technique could allow for high-throughput early diagnosis once developed, but may be hindered due to invasive nature of current beta amyloid collection techniques [29].

DNA microarray has recently proved to be an effective technique to analyze gene expression data in a high throughput fashion. RNA is extracted from tissue samples and
a two-step reverse transcription and fluorescent labeling is done. The resulting genetic information is hybridized to a microarray. Each marker emits at different wavelengths allowing detection of individual signals with a laser scanner. For the early diagnosis of Alzheimer's disease, hippocampus neurons are extracted via stereotactic needle biopsy and the few known genes related to beta amyloid and tau protein formation are hybridized. The gene expression is measured. Although the genetic characteristics of AD are not fully understood, microarray expression data have been able to differentiate cognitive impairment in animal models. There is potential that large-scale animal studies may provide useful information of AD diagnosis in humans, despite the fact that sample collection remains highly invasive [30].

2.4.1 MAGNETIC RESONANCE IMAGING

Perhaps the most conventional imaging modality that has been used for the early diagnosis of AD is magnetic resonance imaging (MRI). MRI utilizes magnetic fields, radio frequencies, and computers to produce cross-sectional images [31]. MRI is based on manipulating the intrinsic spin of hydrogen nuclei by placing the hydrogen nuclei in a large magnetic field and exposing them to radio frequency (RF) pulses. Although used as a topographic modality since its conception, MRI has become a very important source of volumetric data. Using computer software, a set of consecutive image slices can be used to create 3D models. Contrast between tissues is based primarily on the ratio of free to bound water in a given tissue. Since grey and white brain matter have different ratios of free to bound water than the surrounding CSF and bone, contrast between the
brain and its surroundings varies, allowing for the quantification of brain size within an MRI image.

Two modalities of MRI can be obtained, which are referred to as T1 and T2 imaging. T1 Relaxation refers to the return of the water molecule’s magnetization vector to its initial state after being rotated by the RF pulse. T1 Relaxation is the time in which the longitudinal component has grown to 63% of its initial value [31]. Tissues and substances in the brain that have a large ratio of bound to unbound water have short T1 times. Conversely, tissues with the opposite ratio have longer T1 times.

T2 Relaxation refers to the rotation of the water molecule’s magnetization vector away from the transverse axis after the RF pulse. T2 Relaxation is the time in which the transverse component has decayed to 37% after its peak following the RF pulse [31]. Tissues and substances in the brain that have a large ratio of unbound to bound water have short T2 times. Conversely, tissues with the opposite ratio have longer T2 times.

Figure 2.3 displays T1 and T2 weighted images, which result in considerably different contrast depending on tissue type [32]. It can be seen that the T1 weighted image displays good tissue contrast and that cerebrospinal fluid (CSF) appears dark. Conversely, the T2-weighted image displays less tissue spatial resolution and CSF appears bright.
2.4.1a DIAGNOSIS WITH MRI

MRI has been studied as a viable tool for the diagnosis of Alzheimer’s disease. Fritzsche, et al., 2006, conducted a study with 68 subjects consisting of 27 cognitively normal control subjects, 16 subjects with mild cognitive impairment, and 25 AD subjects. An automated volumetric analysis of the CSF distribution was conducted. MRI images were spatially oriented and intensities were normalized. A segmentation technique was used to classify areas of CSF, grey matter, and white matter. The distributions of abundance of various tissue types along with the overall brain volume atrophy were used to create six data features. Classification was performed using Fisher Linear Discriminant (FLD) as well as neural networks. AD subjects were classified correctly in 80% of cases and control subjects in 85% of cases. Classification between normal and MCI patients led to a sensitivity of 81% and a specificity of 80%. Classification between MCI and AD patients showed a performance of 59%. This study proved that automated classification of AD subjects is possible when considering volume metrics of various brain regions extracted from MRI data [33].
Texture data derived from MRI has also been used to discriminate between AD and control groups. Torabi, et al., 2006, used a cohort of 50 normal subjects and 25 AD subjects. MRI images were collected from each subject and the cohort was separated to a 60/40 training and test set. Images were preprocessed to fit the same overall brain image map. Texture feature are extracted by considering 16 ‘landmark’ pixels and analyzing the similarity of these pixels to neighboring pixels. This processing created a 336-feature data set from texture analysis. Feature size is reduced using principal component analysis, and the data are used to train a radial basis function neural network. Texture analysis was able to differentiate the two subject categories with a 95% performance [34].

Time course MRI can also provide important diagnostic information for AD since the progression of AD is well established. Fox, et al., 2001, used a cohort of 20 AD diagnosed patients, 20 control patients, and 4 symptom-free individuals from families with rare autosomal dominant mutations known to cause early onset Alzheimer’s disease. All patients underwent serial MRI scans over a course of five to eight years. The first was used as a baseline and a non-linear fluid matching algorithm was used to compare repeat scans with baseline. Progressive atrophy was revealed in the presymptomatic individuals in the parietal and medial temporal lobes. In the AD diagnosed patients significant atrophy was observed in all clinically predictable regions. Based on these results, serial MRI can be used as an effective early diagnosis tool if the baseline scan is performed early enough in AD progression [35].

Although MRI seems to provide clinically significant diagnostic information for AD, there may be other readily available clinical data that could complement MRI data.
van der Hiele, *et al*, 2006, investigated whether cognitive function in normal and AD subjects is better reflected in MRI, EEG, or both. They established a cohort of 33 subjects, 10 with probable Alzheimer's disease, 11 with mild cognitive impairment, and 12 control subjects. EEG data were collected while subjects took part in memory tasks. MRI data were collected and analyzed to quantize the frequency of grey matter, white matter, and cerebral spinal fluid within the brain. Multiple linear regression analyses were performed between these two data sets. The study concluded that changes in brain function measured with EEG, and brain atrophy measured with MRI, are each associated with different aspects of cognitive decline in AD, and these data complement each other for the task of differentiating these two groups [36].

2.5 **ELECTROENCEPHALOGRAM**

The physiological markers of the integrity of neuronal systems can be detected and measured using the event related potentials (ERPs) of the electroencephalogram (EEG). Electrodes placed on the scalp measure summated electrical brain activity originating from a large number of individual neurons in proximity to the recording electrodes. Unlike biochemical, anatomical, and metabolic biomarkers, EEG signals provide measurement of the electrical activity within millisecond time resolution. EEG has become the biomarker of choice for analyzing brain activity in many disorders such as epilepsy and sleep disorders [37].

EEG is most often collected through electrodes placed directly on the scalp with reference electrodes placed on the ears or near the eyes. Conductive gel or paste between the electrode and the head ensures proper signal recording by reducing impedance.
caused by hair and dead skin cells. Electrodes are placed on the head using the International 10/20 specification for electrode placement. This ensures that a common system of electrode placement on the scalp to standardize signal collection. The electrode naming convention relies on relating electrode location to reference points on the head. A measurement is taken from the top of the nose (nasion) to the bony cleft at the back of the head (inion). The distances along the head between these points along the transverse and sagittal anatomical planes are divided into 10% and 20% intervals, respectively and serve as placement points for electrodes [38]. An illustration of the described electrode placement can be seen in Figure 2.4. The abbreviations in the electrode names correspond to the lobe being measured, as follows: F – frontal, C – central, P – parietal, T – temporal, O – occipital, and A – auditory reference. The numbers in the electrode name correspond to the location on the head with even numbered electrodes on the right side of the head, odd on the left size, and z electrodes along the centerline of the head along the transverse axis [39].

Figure 2.4: International 10/20 System of electrode placement and naming.
There are, however, some aspects of the EEG that need to be considered during analysis. The excellent time resolution of the EEG is apparent, however the spatial resolution is lost due to the inherently poor localization properties of the EEG. The signals measured at the surface of the skull originate from post-synaptic currents from every neuron firing during a neurological event. The surface electrodes can detect these signals in near real time, however, there may be repetitive information in neighboring electrodes and it is impossible to determine at exactly which synaptic point the signal originated. Typically, signals can be localized to a specific region. To increase spatial resolution, subdural electrodes have been utilized in some studies. Although signal localization is improved, this procedure is highly invasive and may result in infection or brain hemorrhaging [40].

Even in the most stable test environment, artifacts can easily alter the EEG. Artifacts can enter the signal due to poor contact with the scalp (due to hair and dead skin cells, improper gel application, or sweating), physical noise due to eye blinks or other muscular activity, and electrical noise. These artifacts must be removed from the EEG signals and by an EEG technician or through automated artifact rejection tools before data analysis.

2.5.1 SPECTRAL CONTENT OF THE EEG

Certain brain activities are associated with specific spectral bands of the EEG. Most brain activity occurs below the 60 Hz range and is typically separated into five specific frequency bands [37].
Delta rhythm ( < 4 Hz ) – The lowest constituent frequency band of the EEG also tends to be the highest in amplitude. The delta rhythm is the dominant rhythm in infants and occurs in deep sleep states in adults. It is also resultant of unconscious processes within the mind. In a cognitively normal individual, the magnitude of the delta band inversely relates to the amount of attention and focus the individual is employing. Presence in an adult is usually sign of cerebral damage or encephalopathy [37,41].

Theta rhythm ( 4 – 8 Hz ) – Is usually present during daydreaming and fantasizing periods as well drowsy periods before sleep. Activity in this band is abnormal in adults but is normal in children. Presence of this band in normal waking adults is an indication of stress or disease [37,41].

Alpha rhythm ( 8 – 12 Hz ) - Alpha rhythms are the most common type of activity in a healthy adult brain. This band originates in the occipital and frontal cortex regions and promotes mental resourcefulness, aids in the ability to mentally coordinate, and enhances over sense of relaxation. This is the major rhythm seen in normal relax state adults throughout life [37,41].

Mu rhythm ( 7 – 11 Hz ) – Mu rhythms, although in the same frequency range as alpha rhythms, are independent phenomena because of differences in source generation. Mu rhythms are generated due to output from the motor cortex, which controls movements of the extremities. The integration of environmental inputs into a motor process
generates the information in this band. The mu rhythm attenuates with contralateral extremity movement and it does not react to eye blinks [42].

Beta rhythm (12–36 Hz) – The beta rhythm represents fast activity within the brain. It is usually seen symmetrically in the brain and is most evident frontally. It has high activity during nervous or anxious states. This is the band observed in adults that are thinking, listening, or problem solving [37,41].

Gamma rhythm (>36 Hz) – The gamma band represents brain activity with the highest frequency. These rhythms usually correlate to simultaneous processing and communication between multiple regions of the brain. Well-regulated and efficient gamma activity tends to correlate to good memory whereas a weak gamma band creates learning disabilities [37,41].

2.6 ACQUISITION PROTOCOLS AND EVENT RELATED POTENTIALS
EEG recording can be obtained while employing specific protocols that provide certain stimuli to elicit particular responses. The stimuli can be auditory, visual, somatosensory, or olfactory. The auditory stimuli are typically a series of tones or sounds. Visual stimuli are typically patterns, colors, and words. Somatosensory stimuli typically present a targeted electrical pulse to a body part of a patient. Olfactory stimuli are odors presented to the patient [43]. The responses to these stimuli within the EEG are called event related potentials (ERP). Protocols can be constructed with a series of various stimuli or can be combined with a cognitive task to elicit ERPs.
In traditional auditory oddball paradigm there are two stimuli. A standard tone, typically of 1 kHz, is presented to the patient for most of the recording. Dispersed infrequently between the standard stimuli are the oddball stimuli, typically of 2 kHz. The two types of stimuli are presented in a random order with a random amount of inter-stimulus time. When the subject is presented with an oddball tone, they are instructed to either push a button or to tally the number of oddball tones [44].

Yamaguchi, et al., 2000 presented the modified auditory oddball paradigm that has been used in this study [45]. In addition to the standard and target (oddball) tones present in the standard protocol, the modified protocol presents a third type of stimulus consisting of novel sounds, such as a dog barking or a bell ringing. The patients were asked to respond, by pressing a button, every time an infrequent target stimuli (2 kHz tone) was delivered within a series of frequent standard stimuli (1 kHz tone) and infrequent novel sounds. The series of tones presented were 65% standard, 20% target, and 15% novel with a 1 – 1.3 second inter-stimulus interval (ISI). The subjects were not warned of the novel sounds nor were they instructed to respond to them. The resultant evoked potentials represent activity from different areas of the brain and may provide more discriminatory information between various dementias.

The EEG recordings collected from the same subject and same stimuli were averaged together to obtain the event related potentials. The minimum number of recordings (epochs) required for this type of averaging has been found to be twenty [46]. In practice, a higher number of ERPs are averaged in order to reduce the noise effects that are present in the averaged signal. The amplitudes that are observed in the ERP response are so minute that they may not be visible over the noise present in the EEG.
By averaging, the features of the ERP become more visible while noise is minimized. A typical ERP collected using the oddball paradigm can be seen in Figure 2.5.

![ERP signal with prominent features labeled.](image)

The various peaks and valleys in a typical ERP are measured based on amplitude and latency after the stimulus. The ERP is composed of the following prominent features: P100, P200, P300, N100, and N200. The numbers in the name correspond to the approximate latency, with respect to the stimulus, in the typical observed peak. Although the peaks latency and amplitude may differ between individual ERP, the order remains consistent. The P300 is the most commonly analyzed peak, relating to cognitive ability.
The P300 typically occurs 250-500 milliseconds after a stimulus is given, and has the largest amplitude of all the peaks in the ERP. The presence and amplitude of the P300 has been shown to relate to cognitive ability. Therefore, ERP protocols such as two and three stimuli oddball paradigms tend to elicit the P300. The P300 is slightly different for responses for novel and target stimuli and have been named accordingly. P3a for novel stimuli tends to have a lower latency than the P3b for target stimuli [47].

ERPs have not been directly considered as a diagnostic tool due to the difficulty evaluating a single subject’s ERP. When comparing ERP groups coming from cognitively normal and AD groups, there tend to be distinguishing features. Despite the fact that a high amplitude P300 may suggest a strong cognitive response, it is not uncommon to see a strong P300 in an AD patient, and conversely see a weak P300 response in a normal patient. Figure 2.6 displays some visual anomalies that can occur with the P300. Due to these inconsistent characteristics, a more detailed evaluation is necessary beyond visual analysis.
2.7 P300 HISTORY

First observed in 1965, the P300’s relation to cognition, its origin within the brain, and its research applications have been studied extensively over the past 40 years. The P300 was originally thought to originate deep within the brain in the hippocampus region. This idea was expanded by Polich, et al., 1995, who determined that there is a possibility of multiple sources for the P300 within the brain, and most likely generated from sources in the temporal-parietal junction [47]. In 1997, a single stimulus protocol was compared to a two-stimuli protocol. In this direct comparison it was determined that although a P300 created using the two stimuli oddball paradigm had a slightly stronger P300,
differences between the signals from either protocol were not statistically significant. Regardless of the protocol used, the response must originate from the same brain region since the patient, in both cases, is actively responding to a stimulus [44].

In 1999 a study was conducted to determine whether auditory or visual stimuli would elicit the strongest P300. It was determined that both auditory and visual stimuli presented in an active task condition (subject presses button upon stimuli) elicited a much stronger P300 than the passive task (no physical task upon stimuli) condition. The auditory stimuli elicited much stronger P300 signals in both the passive and active task conditions compared to the visual stimuli [48]. The oddball paradigm was compared again to the single stimulus protocol in 2002 with a focus on how altering the interstimulus interval (ISI) would alter the P300. The results showed that oddball paradigm produced a stronger P300 with short ISI of 2.5 seconds, whereas the single stimulus protocol showed minimal P300 amplitude. When the ISI was increased to 30 seconds the two protocols produced similar P300s with the oddball response producing a P300 with longer duration [49].

It is now believed that the P300 is formed from multiple simultaneous sources within the brain. The P300 is now considered to have two subcomponents, the P3a and P3b. They are elicited depending on the difficulty of a task and attention required to standard, target, and novel stimuli. The P3a seems to originate from frontal regions during attentive periods and task processing, whereas the P3b originates from the temporal and parietal regions associated with memory processing. Due to this variation the P3a and P3b are considered to carry different information from each other, although the P3a is only found in 10-15% of cognitively normal young adults [50].
Since different bands of the EEG represent different underlying processes, analysis of specific brain activity requires access to specific frequency information within the signal. However, when considering an ERP, which is a 1 second duration signal consisting of averaged EEG segments, the latency of the P300 peak is as important as isolating frequency bands. In order to extract specific regions of spectral information while retaining temporal resolution, the wavelet transform is used for analysis.

Basar-Eroglu, et al., 2001, compared three different protocols, the oddball paradigm, the single stimulus with varying ISI, and single stimulus with constant ISI with every fourth stimulus removed. The resultant ERPs were decomposed using the wavelet transform and compared across frequency bands. It was observed that the P300 component was visible in all electrode locations when the response was elicited with the oddball paradigm and had delta peaks around 2 Hz [51].

Demiralp, et al., 1999, observed that ERP analysis in the time domain alone was susceptible to the heterogeneous nature of P300 and other ERP components. It was hypothesized that this was due to the simultaneous brain processes that created the signal. By decomposing the signal into the time and frequency components, they were able to capture and precisely measure components from various brain events. The frequency components in the delta, theta, and alpha ranges reflect specific aspects of cognition. By decomposing ERPs into constituent time dependant frequency bands, they were able to observe components of simultaneous cognitive processing not able to be seen in the time domain alone [52].
Basar, et al., 2001, compared the utility of the wavelet transform in ERP data analysis to other signal processing techniques. They aimed to separate the superimposed frequency band signals present in the original signal, namely the alpha, theta, delta, and gamma bands. The wavelet technique extracted more information in the delta band and alpha band, which were not visible in the Fourier transform based spectral analysis. Advantages of using the wavelet transform over existing signal processing techniques were the time localization of frequency components, no need for a fixed time window, and significant data size reduction [53].

A five-octave wavelet transform was used by Demiralp, et al., 2001 to analyze single train as well as averaged ERPs. The study emphasized the ability of utilizing this transform to compare ERP components among subjects of various age and cognitive behavior. The study again confirmed the fact that the most distinguishing features of the ERP after wavelet decomposition lie in the delta band response [54].

2.9 EVENT RELATED POTENTIALS IN ALZHEIMER’S DISEASE RESEARCH
The feasibility of event related potentials for AD diagnosis has been researched over the last decade. The abundance of research in this field has lead to frequent attempts to consolidate the existing knowledge base on the topic [55,56]. The aspects from the existing body of work that are relevant to this thesis will be outlined in the following sections.

The established methodology used to analyze ERPs has been to visually inspect the amplitude and latency of the P300 peak. It has been shown that the latency of the P300 peak progressively increases in relation to age in cognitively normal patients.
In patients with Alzheimer’s disease it has been shown that this latency is larger compared to the P300 latency of cognitively normal subjects. However, this increased latency is not exclusive to AD and may be observed in various other dementias [59,60]. This relationship seemed also to be evident in the amplitude of the P300. It was shown the amplitude is higher in normal patients compared to AD patients although this finding was later refuted [61,62]. Although trends can be found in the P300 among large cohorts, it is difficult to determine statistically significant trends when only considering individual patients.

In order to better differentiate the P300 between cognitively normal and AD patients with a quantitative approach, coherence analysis was done in a study consisting of 14 AD patients and 10 normal patients. Changes of relative frequency bands where analyzed and coherence analysis was done between frequency bands. The theta power was significantly higher in AD patients compared to controls. Additionally, the alpha 2 band (9.5-11.0 Hz) showed an increase in the AD group post task compared to the slight decrease in the control group. Most significantly, the alpha 1 band (8.0-9.0 Hz) showed a discernible increase in control subjects compared to AD subjects [63].

Genetic factors that may affect AD have also been researched. Only a small number of cases of AD result from known genetic causes, many of which are associated with early onset diagnosis. 19 healthy subjects and 33 subjects with a known family history of AD were examined. The auditory oddball paradigm was used in this study. The amplitudes of the P50, P300, and N100 responses were higher in subjects with familial AD compared to control groups. This study demonstrated a link in between the
P300 and early cognitive decline with patients with a first-degree familial history of AD [64].

A similar study was conducted in 2005 with the biological children of people with AD to examine P300 as a preclinical marker. The oddball paradigm was used in the AD group, the children group, as well as two age and gender matched control groups. As expected the amplitudes of the P300 were significantly lower in the AD group compared to the control group. Interestingly, both the amplitude and latency of the offspring group were lower compared to the control group. This study extended the diagnostic ability of the P300 by identifying changes in the P300 at the preclinical electrophysiological level in a high-risk patient group [65].

2.10 ALZHEIMER’S DISEASE CLASSIFICATION

It has been established that particular characteristics within the ERP make it possible to discern Alzheimer’s disease patients from cognitively normal patients. Classification methods have been developed that can utilize this information to create an automated diagnostic method.

Polikar, et al., 1997, proposed a method of automated classification using artificial neural networks. EEG data was collected from a 28 patient cohort consisting of 14 AD patients and 14 normal patients using the oddball paradigm. Each tone was 200 ms in length and had an ISI of 1.5 seconds. 86% of the tones were non-target 1 kHz tones and 14% were target 2 kHz tones. Classification was initially performed on the raw time domain signals with a maximum classification of 64%. Classification was then performed using wavelet coefficients as the features leading to a classification accuracy
of 79% to 93%. The results from this study were promising but the procedure needed to be validated with a larger patient cohort [66].

Petrosian, et al., 2001, examined the ability of a recurrent neural network (RNN) to discriminate between the EEG of patients with AD and an age matched control group. Two minute resting state EEG recordings from the parieto-occipital channels were collected from 10 early AD patients and 10 control patients, and preprocessed using the wavelet transform with the Daubechies 4 wavelet (db4). The RNN was trained using 3 AD and 3 control subjects. Of the 14 patients used for testing, 5 of 7 AD patients and 7 of 7 control patients were classified correctly creating a 80% sensitivity and 100% specificity [67].

In 2003, Cho, et al., proposed a technique to discern an AD subject based on single channel EEG recording using the genetic algorithm and an artificial neural network. Five minute resting EEG signals as well as event related potentials were recorded from the P4 electrode from 16 early AD subjects and 16 age-matched controls. The resultant EEG and ERPs were analyzed to produce 88 spectral features, 28 statistical features, 2 chaotic features, and 10 ERP features. This feature set was fed into a genetic algorithm to determine the minimal set of features that are most efficient to automatically classify the two groups. The most effective 35 features were used as training and testing inputs into the multilayer perceptron based ANN resulting in a classification accuracy of 81.9% [68].

Abasolo, et al., 2003 collected single channel EEG data from the P3 electrode from a cohort of 7 normal and 7 AD subjects. The resulting signals were split at 5 second intervals and approximate entropy (ApEn) analysis was performed, which
quantifies regularity in time series data, or producing a measure of the complexity of the data. It was shown using ANOVA testing that the ApEn, or complexity scores, were statistically significantly higher in normal patients than in AD patients. EEG combined with such complexity analysis could prove to be a useful tool in the diagnosis of AD [69].

Yagneswaran, et al., 2002, used a cohort of 9 probable AD subjects and 10 control subjects. Nine channel EEG data were recorded from each patient and the signal power frequency and wavelet coefficients were considered to make the diagnosis. The overall power spectra showed no discernable diagnostic information so a band pass FIR filter using a Hamming window was used to separate delta, theta, alpha, and beta bands. These bands as well as relative power and slower wave ratio were used to train and test a neural network. An additional neural network was trained and tested using the Daubechies 5 wavelet coefficients of the EEG data. The neural network trained with the power spectrum features obtained a 94.7% accuracy while the neural network trained on wavelet coefficients obtained an 89.4% accuracy [70].

Tao and Tian, 2005, recorded 21 site EEG signals from a cohort of 12 AD patients and 18 MCI patients. Five minutes of resting state eyes open EEG was recorded followed by 5 minutes of eyes closed. Next, 21 random numbers were shown to the subject in 30 seconds and the subject was asked to note the number of odd digits as their cognitive task. The Mexican hat wavelet was used to extract the gamma band from the collected data. Coherence analysis enables the quantification of linear dependencies between multichannel time series data. In other words it can be used to identify variations in signals with similar spectral properties. Coherence analysis was performed
for pairs of electrode channels. In the rest state with eyes open, coherence measures were similar in all three-subject groups. In the eyes closed resting state, the AD patients had significantly lower coherence compared to the MCI and control groups. In the cognitive task load state, AD subjects as well as MCI subjects showed lower coherence than the control group. Coherence analysis between pairs of electrodes during cognitive task load could prove to be a useful diagnostic measure of AD [71].

Chapman, et al., 2007, collected EEG data from a cohort of 12 AD subjects and 12 normal subjects using the number-letter paradigm where a visual recognition was used to elicit brain activity. The resultant ERPs were measured using principal component analysis (PCA) to extract eight features. The team tested whether a discriminant function based classification technique would provide adequate diagnostic accuracy. When trained and tested using 50% of data for training and 50% of data for testing performance was 92% with 100% sensitivity and 83% specificity. When leave one out cross validation was used, which uses one subject at a time for testing and the rest of the subjects for training, classification performance was 79%, sensitivity was 83%, and specificity was 75%. Although the split validation provided promising results, the leave one out cross validation is more representative of the performance potential with this technique [72].

Henderson, et al., 2006, collected EEG data from 30 probable AD patients, 6 vascular dementia patients, 3 mixed dementia patients, and 42 cognitively normal patients. A fractal dimension based approach and an approach based on a probability density function with zero-crossing intervals were both used for classification using raw EEG data. The fractal based approach showed sensitivity of 67% and specificity of
99.9% while the probability density based approach showed 78% sensitivity. These analysis techniques seem promising but may need to be tested thoroughly in two class classification problems before being expanded into a 3 class diagnostic tool [73].
CHAPTER 3
METHODS

3.1 PREVIOUS WORK SPECIFIC TO THIS STUDY

Earlier phases of this study involved a single cohort of 71 AD and cognitively normal patients to evaluate the diagnostic value of ERP based analysis. The overall approach in these studies involved signal decomposition using the discrete wavelet transforms of individual ERPs, and then using these coefficients for training a neural network type classification algorithm.

During the first phase, the cohort contained 32 subjects. A wavelet feature extraction technique was used using Daubechies 4 and quadratic B-spline wavelets to extract coefficients from the signals obtained from the PZ electrode. This data was used to train the Learn++ classification algorithm with MLP base classifiers to distinguish between AD and normal subjects. The overall performance achieved with this technique was in the low 80% range [74]. The Learn++ algorithm allows any classification algorithm learn incrementally from new data in the absence of old data. This allows networks to be trained on new data without having to re-learn old data.

The project was expanded as the cohort size reached its final size of 71 patients. The Learn++ algorithm was used for classification using ERP data decomposed using only the Daubechies 4 wavelet. The PZ electrode was again used with the addition of the CZ and FZ electrodes. Diagnostic performance for a single trial of the algorithm was 83.1% while a five trial average provided a 79.2% accuracy [75].
Study of the 71 patient cohort was later expanded to the investigation of additional electrode locations with an emphasis on the parietal region. The Daubechies 4 wavelet was used for decomposing the ERPs into spectral bands with the three lowest and most informative frequency bands being used as features. The ensemble of classifier based Learn++ algorithm was suitably modified to allow combining classifiers on different feature sets. A separate classifier was trained on each feature set, consisting of wavelet coefficients of different frequency bands, obtained from different electrodes and in response to different stimulus types. Decision level fusion was employed using sum rule, product rule, weighted majority voting, and decision templates to combine the individual classifiers. Diagnostic performance has then reached 83.1% for the full 71 patient cohort [76].

The stacked generalization algorithm was also used for AD classification with the full 71 patient cohort. The performance in distinguishing normal from AD patients was then 85.65%. Initial work was also performed on a severity analysis. The AD group was divided into mild and moderate subjects, and including the normal group, a three-class classification was performed resulting in a 71.34% performance [77].

The second phase of the study introduced a new cohort of 62 subjects. An ensemble based classification algorithm was used in conjunction with wavelet based feature extraction from all electrodes using the three lowest frequency bands. These were all analyzed individually to obtain the most complimentary feature sets for classification. The best combinations of classifiers were combined using decision level fusion using sum rule, product rule, weighted majority voting, and Dempster Shafer rule. The optimal
combinations in the algorithm were tested for both cohorts and yielded a mid 80% to low 90% performance [78].

3.2 CURRENT RESEARCH

Present work continues Phase II of this study. The new cohort has first been extended to 98 patients (49 normal and 49 AD) whose data were used to reassess the ERP only diagnosis work previously done. This cohort has finally been expanded to 161 patients (49 AD, 34 normal, 39 PD, 39 MCI). Each patient in this cohort also has several biomarker data including EEG, volumetric MRI data, and biochemical marker levels. An ensemble based classification algorithm has been implemented with multiple ‘experts,’ a group of classifiers trained on data from different biomarkers. Decision level fusion has been performed using simple majority voting, weighted majority voting, and sum rule in a cross validation setting. The classification performance of combining MRI and ERP based experts (classifiers) was compared with the classification performance of the biomarker data, currently the single most accurate diagnostic test.

This approach has been evaluated on a variety of cohort subsets. Two class problems compared subjects with Alzheimer’s disease, Parkinson’s disease, and mild cognitive impairment to cognitively normal controls. Three class problems were designed to distinguish between AD, PD, and normal; as well as AD, MCI, and normal. Finally, a severity analysis was performed with 3 subject groups: mild AD, moderate AD and cognitively normal. For purposes of this study, AD patients were separated into mild and severe groups based on an MMSE cutoff at scores of 20, 23, and 26. AD patients with the cutoff scores and above were considered mild AD and patients with scores
lower than the cutoff were considered severe AD. Also, since biochemical marker data is available for normal, AD, and MCI subjects, classifiers trained on such data were used as a baseline for classification performances for these groups. Figure 3.1 shows the overall approach followed.

**Overall Process**

<table>
<thead>
<tr>
<th>ERP Expert</th>
<th>MRI Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Fusion</td>
<td>Final Decision</td>
</tr>
<tr>
<td>Combination</td>
<td>Data Fusion</td>
</tr>
<tr>
<td>ERP Expert</td>
<td>MRI Expert</td>
</tr>
</tbody>
</table>

**ERP Expert**

- Data Acquisition (Oddball Paradigm)
- Pre-processing
  - Noise and artifact removal
- Feature Extraction
  - Discrete Wavelet Transform
- Ensemble Generation
- Decision Fusion

**MRI Expert**

- Data Acquisition (T1 MRI Scan)
- Pre-processing
  - Volumetric Analysis
- Feature Selection
  - Random subspace sampling
- Ensemble Generation
  - Decision Fusion
  - MRI Decision
3.3 STUDY COHORT

The data used in this study were collected in collaboration with the University of Pennsylvania. Subjects were given a neurological examination and diagnosis was made at the University of Pennsylvania. The University of Pennsylvania used proprietary segmentation software to determine the volumes of various brain regions from processed MRI images. Only the processed volume measures were provided to us at Rowan University. In an effort to maximize automation of data processing after collection, the EEG signals were provided in their raw form, and artifact removal and signal averaging was performed at Rowan. Biochemical marker levels were provided as levels of hyperphosphorylated tau and beta amyloid proteins in the CSF.

The first step in confirming the cognitive state of the subjects was to issue a battery of standardized neuropsychological tests. The Mini-mental state exam (MMSE) was administered as part of neurological testing. The MMSE is the simplest of the battery of tests used to diagnose dementia. MMSE examines memory and language skills. The test is scored from 0 to 30 with any score below of 19 indicating obvious impairment. The Clinical Dementia Rating (CDR) was also administered to determine
dementia progression. The CDR is based on six cognitive functional categories: memory, orientation, judgment, community affairs, home and hobbies, and personal care. Clinical rating in each category is compiled to create the CDR that is scored incrementally as 0, 0.5, 1, 2, or 3, with 0 being cognitively normal and 3 being severely demented. Patients were also given the Dementia Severity Rating Score (DSRS). The DSRS is a functional measurement, administered as a multiple choice test, incorporating both cognitive impairment and activities of daily living in which the higher the score the worse the cognitive impairment.

The following inclusion and exclusion criteria were used for the AD/MCI, PD and normal subject groups. Table 3.1 summarizes the demographics from each of the cohorts.

**Cognitively Normal Inclusion criteria**

- Age > 55
- Clinical Dementia Rating: Score = 0
- No indication of functional or cognitive decline during the two years prior to enrollment

**Cognitively Normal Exclusion criteria**

- Evidence of any central nervous system neurological disease (e.g. stroke, multiple sclerosis, Parkinson’s disease, etc) by history or exam
- Use of sedative, anxiolytic or anti-depressant medications 48 hours prior
Alzheimer’s Disease and Mild Cognitive Impairment Inclusion criteria

- Age >55
- Clinical Dementia Rating: Score = 0.5
- Mini Mental State Exam Score \( \leq 26 \)
- Presence of functional and cognitive decline over 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

Alzheimer’s Disease and Mild Cognitive Impairment Exclusion criteria

- Evidence of any central nervous system neurological disease (e.g. stroke, multiple sclerosis, Parkinson’s disease, etc) by history or exam
- Use of sedative, anxiolytic or anti-depressant medications 48 hours prior

Parkinson’s Disease Inclusion criteria

- Age >55
- Deterioration in cognitive function endorsed by a knowledgeable informant.
- Failure to meet National Institute of Neurological and Communicative Disorders and Stroke - Alzheimer’s Disease and Related Disorders Association and DMS-IV revised criteria for AD
- No clinical criteria suggesting the presence of PD, major depression, psychosis, or a non-AD dementia
Parkinson’s Disease Exclusion criteria

- Presence of co-morbidity or any medical condition that would preclude a full evaluation

Table 3.1: Statistics on the 161 patient cohort used for this study, including cohort size per diagnosis, average ages with standard deviations, and average MMSE scores with standard deviations

<table>
<thead>
<tr>
<th>161 Patient Cohort</th>
<th>Number of Subjects</th>
<th>Average Age</th>
<th>Standard Deviation</th>
<th>Average MMSE Score</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>49</td>
<td>75.35</td>
<td>9.30</td>
<td>18.25</td>
<td>6.34</td>
</tr>
<tr>
<td>PD</td>
<td>39</td>
<td>70.53</td>
<td>6.34</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MCI</td>
<td>39</td>
<td>70.24</td>
<td>8.08</td>
<td>26.16</td>
<td>2.87</td>
</tr>
<tr>
<td>Normal</td>
<td>49</td>
<td>70.87</td>
<td>10.11</td>
<td>28.93</td>
<td>1.09</td>
</tr>
</tbody>
</table>

3.4 ERP ACQUISITION

The ERPs were obtained using an auditory oddball paradigm. Subjects were comfortably seated in a quiet room wearing headphones. Before the commencement of data collection, a 1 kHz tone was presented to establish the patient’s unique hearing threshold and subsequently adjusting volume accordingly.

Each stimulus was presented at equal volume to both ears at 60 dB above each subject’s auditory threshold to remove bias between patients. Each stimulus was 100 ms in duration; consisting of a standard 1 kHz tone, target (oddball) 2 kHz tone, or novel environmental sounds (>200 ms duration). A total of 1000 stimuli were presented, 65% of which were standard tones, with the remaining split as 20% target tones and 15% novel sounds. Each stimulus and the associated EEG data is called an epoch. The inter-stimulus interval varied between 1.0-1.3 seconds. The subjects were instructed to press a
button each time they heard the target tone and not to respond for the standard and novel tones. The data collection was typically 30 minutes per subject, which included three minutes of rest every five minutes.

The EEG signals were recorded from 19 tin electrodes applied directly to the subjects’ scalp with two reference ear electrodes. EEG recording was continuous. The electrode impedances were kept below 20 kΩ. Signals were amplified and digitized at 256Hz/channel. The signals were notch filtered between 59-61 Hz to remove electrical noise. Baseline normalization with respect to the pre-stimulus interval was performed.

An automated, derivative based artifact removal technique was developed to remove artifactual recordings from the EEG before averaging the epochs. For each patient, the raw signals were separated based on stimulus. EEG analysis for each stimulus was initiated 200 ms prior to the stimulus and continued for 800 ms after the stimulus. Averaging all epochs for a particular stimulus/electrode combination helps to reduce the low signal to noise ratios sometimes seen in EEG and gives a better representation of the ERP response. A 20th order derivative was then used for artifact rejection: any one second segment, or epoch, whose derivative was above a certain set threshold was determined to be an artifactual recording, and was subsequently removed from the averaging. Equation 3.1 shows the artifact rejection approach and an example of the artifact removal is shown in Figure 3.2. The artifact free signals were averaged by stimulus type yielding three ERPs for each patient:

\[
f(x) = \begin{cases} 
  f(x), & \text{where } f(x + 20) - f(x) < \text{threshold} \\
  \text{remove, where } f(x + 20) - f(x) \geq \text{threshold}
\end{cases}
\]  

(3.1)
Figure 3.2: Artifact removal process. Vertical lines separate one-second epochs for each stimulus. Epochs exceeding a threshold are removed by further processing. The upper plot is raw EEG signal with artifacts. The lower plot shows the location of artifacts, identified by the derivative process. Corresponding epochs were removed.

Grand averages for each diagnosis group are presented in Appendix A-H in the following order: A) AD vs. Normal (ERP Only), B) AD vs. Normal, C) MCI vs. Normal, D) PD vs. Normal, E) AD vs. PD vs. CN, and F-H) Mild AD vs. Severe AD, vs. Normal with MMSE cutoff of 20, 23, 26. Grand averages are averages of each electrode from each patient in a diagnosis group. They illustrate overall trends in patient groups.
3.5 FEATURE EXTRACTION

The brain activity that is measured with the electroencephalogram contains a vast amount of information about cognitive ability. However, information obtained at a scalp-based electrode is an amalgam of information originating at various sources within the brain. This information may not be apparent in the raw signal, but isolating various frequency bands within the signal can separate these information sources. The frequency domain representation allows for a visualization of the frequency content of the signal, which is of the utmost relevance to EEG signals.

An advantage of frequency domain representation is its ability to distinctly represent the specific frequencies and their contributions within the signal. The most commonly used method to achieve frequency domain representation is the Fourier transform (FT). The continuous FT is described in Equation 3.2 and the inverse FT is described in Equation 3.3, where $X(f)$ is the frequency representation of the original time domain signal $x(t)$. In Fourier transform, all time dependant information is lost in the commonly utilized magnitude spectrum. The time information remains in the phase response of the FT but is not practically useful in analysis. Therefore, FT is an appropriate approach primarily for signals whose frequency content does not change in time (stationary signal). Since the spectral content of ERPs vary in time, FT is not an optimal analysis approach as the frequency information cannot be localized in time.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} \, dt$$ \hspace{1cm} (3.2)

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} \, df$$ \hspace{1cm} (3.3)
The short-time Fourier transform (STFT) overcomes the temporal limitation of the FT to provide time-frequency representation of the signal by using a windowing approach. The FT is used to determine frequency and phase content of consecutive segments of the original time based signal, where each segment is obtained through an appropriate windowing function. Equation 3.4 describes the STFT, where \( x(t) \) is the original time based signal, \( w(t) \) is the windowing function, and \( \tau \) is the translation of the window. The width of the windowing function, which is fixed in STFT, determines the time and frequency resolutions that can be obtained by the STFT.

\[
STFT_{\tau}^w(\tau,f) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j2\pi ft}dt
\]  
(3.4)

The Heisenberg Uncertainty Principle, which states that the exact time and the exact frequency cannot be known concurrently, limits the trade-off between the time & frequency resolutions. A narrow window allows for excellent time resolution but poor frequency resolution; conversely, a wide window allows for excellent frequency resolution but poor time resolution. This tradeoff becomes an issue when dealing with non-stationary signals such as the EEG; the window size will remain constant for a given STFT [79]. The wavelet transform is a more effective time-frequency representation tool for this application.

3.6 WAVELET TRANSFORMS

The wavelet transform (WT) addresses the fixed time-resolution problem associated with STFT by varying the window size depending on the frequency band being processed. Several versions of the wavelet transform are available for different types of signals. The continuous wavelet transform (CWT), a discrete version of the CWT called
the wavelet series, and the discrete wavelet transform (DWT) which is used for feature extraction in this study [80].

3.6.1 CONTINUOUS WAVELET TRANSFORM

The CWT calculates the time-frequency representation of the signal using basis functions whose support is varied based on scale. Specifically, the CWT does not use the FT within each window, but rather a correlation metric with the wavelet function whose width varies based on the scale (or frequency) being utilized. The CWT is defined as

\[ \Psi_x^\psi (\tau, s) = \int x(t) \psi^*(t) dt \]  

(3.5)

where \( x(t) \) represents the original signal to be analyzed, \( \psi^* \) denotes complex conjugation, \( \tau \) and \( s \) represent the translation and scale variables for \( \psi \), the mother wavelet. The mother wavelet is scaled and translated from the prototype wavelet function given by:

\[ \psi_{\tau, s}(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t - \tau}{s} \right) \]

(3.6)

Unlike the Fourier transform that uses fixed basis functions, the wavelet transform can use any wavelet function that adheres to two admissibility conditions shown in Equations 3.7 and 3.8. The wavelet function must be a wave whose area is zero; and it must be of finite duration and finite energy. [79].

\[ \int_{-\infty}^{\infty} \psi(t) dt = 0 \]  

(3.7)

\[ \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \]  

(3.8)

To calculate CWT, the mother wavelet is held at constant scale and translated until the wavelet transverses the entire signal. The scale is then incremented and the
process of compression and dilation along with transitional changes is repeated for all values of the scale and translation. However, the CWT utilizes a continuously scalable wavelet function, which makes it impractical for many real life applications that rely on discrete processing. The discrete wavelet transform overcomes this issue by scaling and translating in discrete intervals.

3.6.2 DISCRETE WAVELET TRANSFORM

CWT is not a true discrete transform that can be executed on a computer; also the CWT leaves redundant information in the signal after being transformed. The solution for redundancy removal is the discrete wavelet transform (DWT), which provides non-redundant information for decomposition and reconstruction of the original signal. It also is easier to implement and less computationally expensive when compared to CWT. Discrete wavelet transformation is possible with multiresolution analysis and subband coding.

3.6.3 MULTiresOLUTION ANALYSIS (MRA)

Multiresolution analysis processes a function at various levels of approximation or resolution. This analysis allows a complex function to be divided into several simpler functions that can then be analyzed individually. Time localization of spectral components is available with this technique. During the discretization process, there is also the ability to change sampling rates allowing for significant reduction in the signal length for each level of representation [81]. The signal is sampled repetitively with each successive sampling at half the prior frequency. This procedure is completed for a set
number of iterations resulting in the signal approximation, which is the low frequency representation of the original signal. The information removed between sample points represents the higher frequency detail coefficients at each level of the signal.

During each such sampling, the approximation signal is allocated to the approximation subspace $A_s$, and the detail information removed is allocated to the wavelet subspace $W_s$. Any function $x_s(t)$ can be reconstructed as a linear combination of these two subspaces. Equations 3.9 and 3.10 show how these two subspaces are generated. Any function can be represented as linear combinations of $\phi_{k,s}$ and $\psi_{k,s}$. $A_s$ is generated by the base:

$$\phi_{k,s} : 2^{s/2} \phi(2^s t - k) \quad k \in \mathbb{Z}$$  \hspace{1cm} (3.9)

whereas $W_s$ is generated by the base:

$$\psi_{k,s} : 2^{s/2} \psi(2^s t - k) \quad k \in \mathbb{Z}$$  \hspace{1cm} (3.10)

Equation 3.11 shows the finite energy scaling function used. The scaling function is associated with the extraction of the low frequency portion of the signal. This concept denotes that energy in the signal at a particular level is nested in the approximation at the higher level as seen in Equation 3.12.

$$\phi(t) \in L^2(\mathbb{R})$$  \hspace{1cm} (3.11)

$$\{0\} \leftarrow \ldots \subset A_{-1} \subset A_0 \subset A_1 \subset \ldots \rightarrow L^2$$  \hspace{1cm} (3.12)

The scaling function shown in Equation 3.11 must satisfy the following dilation equation:

$$\phi(t) = \sum_k g[k] \phi(at - k)$$  \hspace{1cm} (3.13)
\( a \) is greater than zero with 2 being the usually selected value. The scaling function is a scaled and translated version of itself, therefore it can be seen that:

\[
x(t) \in A_s \iff x(2t) \in A_{s+1}
\]

\[
x(t) \in A_s \iff x(t + 2^{-s}) \in A_s
\]

\( A_s \) is a subspace of \( A_{s+1} \) for each value of \( s \). \( W_s \) is complimentary with \( A_s \); when the \( W_s \) detail information is combined with \( A_s \) approximation information, it forms \( A_{s+1} \), the higher level approximation as seen in equation 3.16 where \( \oplus \) represents direct summation.

\[
A_s \oplus W_s = A_{s+1}
\]

These subspaces are directly related to the approximation and detail of the signal. The subspaces can be directly related to \( x(t) \) and \( y(t) \) such that \( x_s(t) \in A_s \) and \( y_s(t) \in W_s \). Equations 3.17 and 3.18 show how they are calculated [82].

\[
x_s(t) = \sum_k a_{k,s}(2^s t - k)
\]

\[
y_s(t) = \sum_k w_{k,s}(2^s t - k)
\]

3.6.4 SUBBAND CODING

DWT implementation of multiresolution analysis can be obtained through subband coding using digital filters. The filters are quadrature mirror filters, which are half band lowpass and half-band highpass filters that are odd indexed, alternated, and reversed versions of each other. The wavelet being used will stipulate what the low pass filter, with an impulse response \( h[n] \), will be for decomposition. To create the high pass
version of the filter, the following conversion is performed on the impulse response of the low pass filter, where \( L \) is the length of the filter.

\[
g[L - 1 - n] = (-1)^n h[n]
\] (3.19)

Signal decomposition occurs when the original signal is filtered by the low pass filter \( h[n] \) or the high pass filter \( g[n] \). Filtering is conducted via convolution in the discrete time space as shown for the low pass filtering case in Equation 3.20.

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

After applying the high pass and low pass filters to the signal, the two filtered signals hold the approximation and detail coefficients of the original signal. Assuming that the highest frequency in the original signal is \( \pi \), the signal is sampled at \( 2\pi \) in accordance to the Nyquist criterion. After applying half band filters, the highest frequency retained in the resulting signals is then \( \pi/2 \), therefore the signal need only be sampled at \( \pi \). When sampled at this lower rate, the redundant information for the higher sampling rate is removed. This, effectively, is down sampling the signal by 2.

The signal resulting from the low pass filtering is the approximation of the original signal and has frequency band \([0-\pi/2]\). The resultant signal from high pass filtering of the original signal has frequency band \([\pi/2-\pi]\). Both signals are down sampled by 2. Equations 3.21 and 3.22 show the low pass and high pass filtering followed by downsampling, respectively.

\[
y_{low} = \sum_{n} x[n] \cdot h[2k - n]
\] (3.21)

\[
y_{high} = \sum_{n} x[n] \cdot g[2k - n]
\] (3.22)
Upon the first iteration of this filtering, the resultant signals represent the level 1 approximation and detail coefficients. The filtering is repeated using the level 1 approximation coefficients resulting in the level 2 approximation and detail coefficients. This process is continued until the final frequency band is reached. The single sample coefficient will not allow further filtering at this point. The group of detail coefficients at all levels represents the DWT of the original signal. In each successive level, subsampling causes time resolution to be reduced by half and the halved bandwidth causes frequency resolution to double. Figure 3.3 shows a block diagram of the DWT process where \( g[n] \) is a highpass filter and \( h[n] \) is a lowpass filter defined by the wavelet being used.

![Figure 3.3: A diagram of the seven levels of discrete wavelet transform decomposition for an ERP signal with 128 Hz maximal frequency content.](image)

3.6.5 SIGNAL RECONSTRUCTION

Wavelet decomposition is a fully reversible procedure. If all levels of approximation and detail coefficients are available, a perfect reconstruction of the original signal can be calculated. An approximation of the signal at any level can be obtained by summing the detail and approximation coefficients for that level as seen in Equation 3.23.

\[
x(t) + y(t) = \sum_k a_{k} \phi_{k}(t) + \sum_k w_{k} \psi_{k}(t) = x_{r-1}(t)
\]  

(3.23)
where \( s \) represents the level of the reconstruction, \( a \) and \( w \) are approximation and detail coefficients, respectively, \( \phi \) is the scale function and, \( \psi \) is the wavelet function.

For a discrete signal, the high and low frequency reconstruction filters are used. These are easily obtained from the original decomposition filters. A similar level summation is shown in Equation 3.24 for the discrete case. In order to obtain the original signal, this operation is performed at each level and then summed.

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

(3.24)

3.6.6 DAUBECHIES WAVELET

The Daubechies wavelet allows for perfect reconstruction of the original signal. To decompose a signal accurately the wavelet must match the frequency characteristics of the signal. In this study, the Daubechies wavelet with 4 vanishing moments (db4) was used due to its optimal frequency similarities with the ERP such as its smooth wavelet function. The length of the db4 lends itself well to ERP analysis since this 8 point filter is much smaller than the 256 point ERP signal. The wavelet and scaling functions as well as the decomposition and reconstruction filters are shown in Figure 3.4 and a full ERP wavelet decomposition is shown in Figure 3.5.
Figure 3.4: Wavelet and Scaling functions as well as decomposition and reconstruction filters for the Daubechies 4 wavelet.
3.7 CLASSIFICATION

An automated classification algorithm is a set of instructions that can learn the hidden patterns in data to identify data coming from different underlying distributions, or different categories (classes). Feature extraction is usually the first step to determine the indicators in the data that will distinguish among different classes. The classifier is then "trained" by showing previously labeled examples of different classes. The training is
essentially funding a function that maps a given set of features to their corresponding true labels. As such, a classifier is therefore a “function approximator.”

Artificial neural networks are a commonly used group of classification algorithms that can adjust their internal parameters based on training data, and then identify the class/category of instances of the previously unseen field data. The multilayer perceptron and support vector machine are examples of automated classification algorithms.

3.7.1 MULTILAYER PERCEPTRON

The multilayer perceptron (MLP) is a feed-forward network, consisting of one or more hidden layers, with constituent nodes, between the input and output layers. The common structure of an MLP is shown in Figure 3.6. The first layer of nodes is the input layer. This layer accepts input data and sends the signal forward through each node of the hidden layers of the network and reaches the output layer. The number of input nodes is the number of features in the data, and the number of output nodes is the number of classes present. The output is encoded in a binary numbering scheme. The input layer nodes serve only to accept data into the network, while nodes in the hidden layers and output layer conduct computations on the data based on synaptic weights they create. These weights are created and optimized using the error back propagation algorithm.
3.7.1a BACKPROPAGATION ALGORITHM

The back propagation algorithm consists of four stages: initialization, presentation of training data, forward computation, and backward computation [83]. Each pair of nodes are linked together with a synaptic weight, where the network’s knowledge resides. During initialization, the synaptic weights are randomly initialized from a zero-mean uniform distribution. The variance of the initial synaptic weights is chosen so that the standard deviation of the induced local field of a node lies in the transition area between the linear and saturated regions of the sigmoid activation function. The logarithmic sigmoid is used as the activation function in this study. The activation function determines the output of a node given a certain set of inputs.

Figure 3.6: Model of a multilayer perceptron neural network.
Once the network is created and synaptic weights are initialized, the network is presented with training data as input, along with the correct labels for the training data. An iterative process of forward and backward computation is performed for each instance of training data until the maximum number of iterations or the error goal is reached. In forward computation a single training example is denoted as \((x(n), d(n))\), with the input data \(x(n)\) being presented to the input layer nodes, and the \(d(n)\) (desired and correct class information encoded as a binary vector) presented to the output layer nodes.

Proceeding forward through the network, the induced local fields and function signals of the network are computed one layer at a time. Equation 3.25 shows the induced local field \(v_j^{(l)}(n)\) for node \(j\) in layer \(l\).

\[
v_j^{(l)}(n) = \sum_{i=0}^{m_l} w_{ij}^{(l)}(n)y_i^{(l-1)}(n)
\]

where \(ij\) represents the output of node \(i\) being passed to node \(j\), \(m_l\) represents the number of nodes in layer \(l\), and \(y_i^{(l-1)}(n)\) is the output of node \(i\) in the previous layer \(l-1\) at iteration \(n\), and \(w_{ij}^{(l)}(n)\) is the synaptic weight of node \(j\) at layer \(l\) that is fed from node \(i\) in layer \(l-1\). For \(i=0, y_i^{(l-1)}(n)=+1\) and \(w_{j0}^{(l)}(n)=b_j^{(l)}(n)\) is the bias applied to node \(j\) in layer \(l\). The output signal of node \(j\) in layer \(l\) is:

\[
y_j^{(l)} = \varphi_j(v_j(n))
\]

The activation function is usually a logarithmic sigmoid represented by:

\[
\varphi(n) = \frac{1}{1 + e^{-n}}
\]

If \(j\) is in the first layer where \(l=1\):

\[
y_j^{(0)}(n) = x_j(n)
\]

\(x_j(n)\) represents the \(j\)th element of the input vector \(x(n)\).
If node $j$ is in the output layer, where $l=L$, is the maximum number of layers in the network:

$$y^{(L)}_j(n) = o_j(n) \quad (3.29)$$

The error signal can then be computed as:

$$e_j(n) = d_j(n) - o_j(n) \quad (3.30)$$

The error signal contains important information about the network’s current degree of ‘learning’ of the training data. Instantaneous error energy can be computed from the error signal when $C$ is the set of all possible output nodes:

$$e(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (3.31)$$

Backward computation aims to minimize this error by adjusting the synaptic weights throughout the network using the following equation:

$$v_j(n) = \sum_{i=1}^{m} w_{ji}(n)y_j(n) \quad (3.32)$$

The output signal at node $j$ of iteration $n$ is shown in Equation 3.33 with its activation function shown in Equation 3.34:

$$y_j(n) = \varphi(v_j(n)) \quad (3.33)$$

$$\varphi(v_j) = \frac{1}{1 + e^{-av_j}} \quad (3.34)$$

Error minimization is conducted via the gradient decent algorithm. The gradient of the error must be calculated to determine the direction of the steepest decent as shown in equations 3.35 and 3.36. A step size $\eta$ is taken in the direction opposite to the highest gradient.
\[
\frac{\partial e(n)}{\partial w_{ji}(n)} = \frac{\partial e(n)}{\partial j} \frac{\partial j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}
\] (3.35)

\[
\frac{\partial e(n)}{\partial w_{ji}(n)} = -e_j(n) \varphi_j'(v_j(n)) y_j(n)
\] (3.36)

The synaptic weight updating is now done as shown in Equations 3.37 and 3.38.

Where output node \( j \) with local gradient \( \delta(n) \) is the derivative of Equation 3.35.

\[
\Delta w_{ji}(n) = -\eta \frac{\partial e(n)}{\partial w_{ji}(n)}
\] (3.37)

\[
\delta_j(n) = a \left[ d_j(n) - y_j(n) \right] y_j(n) \left[ 1 - y_j(n) \right]
\] (3.38)

For a hidden layer node, the gradient function, sometimes called sensitivity, is computed as follows:

\[
\delta_j(n) = \frac{\partial e(n)}{\partial y_j(n)} = -\frac{\partial e(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}
\] (3.39)

in which

\[
\frac{\partial e(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial y_k(n)} \frac{\partial y_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)}
\] (3.40)

and

\[
\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n)
\] (3.41)

The local gradient is calculated based on the gradient of the following layer:

\[
\delta_j(n) = \varphi_j'(v_j(n)) \sum_k \delta_k(n) w_{kj}(n)
\] (3.42)

The actual adjustment of weights is done through the generalized delta rule (Equation 3.43) which contains a learning rate parameter, \( \eta \), and a defined momentum constant, \( \alpha \).

\[
w_{ji}^{(i)}(n + 1) = w_{ji}^{(i)}(n) + \alpha \left[ w_{ji}^{(i)}(n - 1) \right] + \eta \delta_j^{(i)}(n) y_j^{(i-1)}(n)
\] (3.43)
The forward and backward computation are then iterated by presenting new sets of training examples to the network until either the maximum number of iterations or the error goal is reached.

3.7.2 SUPPORT VECTOR MACHINES

Despite the apparent effectiveness of the MLP, the MLP has two major drawbacks. First, the process of training an MLP network through the iterative back propagation algorithm is very computationally intensive. If the number of epochs specified is too numerous, the training error goal is set too low, or the network size is large, training of an MLP can take many hours. The second major drawback of an MLP is the random initialization of hidden layer node weights. Since the back propagation algorithm merely alters the existing weight in a node, the random initialization of weights results in different decision boundaries in every MLP trained on the same data set. Therefore, the MLP has to be trained and tested multiple times on each dataset and averaged to fully extract performance measures.

The underlying principles of the support vector machine (SVM) allow it to avoid these issues. It is computationally efficient and performs calculations higher dimensional computation in lower dimensional spaces. In these spaces, hyperplanes are created and optimized to maximize margins between classes. Since this process is optimal with respect to an error function, and there is no randomization in the SVM design, the same decision boundaries will be created each time an SVM is trained on a particular training set. Additionally, the SVM does not get stuck in local minima during the optimization process, which can happen in the MLP.
An SVM classifier separates classes by maximizing the margins between classes. Equation 3.44 shows the hyper plane that separates two classes. \( \frac{b}{||w||} \) is the perpendicular distance from the hyperplane to the origin, and \( ||w|| \) is the Euclidean of \( w \), where \( w \) is the normal to the hyperplane.

\[
wx + b = 0 \quad (3.44)
\]

SVMs accept training data labeled with class information as \( y_i \in [-1,1] \). The distances from the hyperplane to the closest data point on the positive and negative sides of the hyperplane are called \( d_+ \) and \( d_- \), respectively. Therefore, the margin of a separating hyperplane between the two classes is \( d_+ + d_- \). All data points must satisfy the following constraints:

\[
x_i \cdot w + b \geq 1 \quad \text{for} \quad y_i = 1
\]
\[
x_i \cdot w + b \leq -1 \quad \text{for} \quad y_i = -1 \quad (3.45)
\]

These two constraints can be combined into one inequality:

\[
y_i (x_i \cdot w + b) - 1 \geq 0 \quad \forall i \quad (3.46)
\]

The margin between the two linearly separable classes is defined in Equation 3.47 and shown in Figure 3.7.

\[
m = \frac{2}{||w||} \quad (3.47)
\]
The two hyperplanes are maximally separating the two classes. The points that lie directly on - and actually create - the hyperplanes are called the support vectors.

Changing the support vectors will completely change the decision boundaries. Lagrange multipliers are used to optimize the inequality constraints that help define the boundaries in Equation 3.45:

$$L_p = \frac{1}{2}||w||^2 - \sum_{i=1}^{j} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{j} \alpha_i$$  \hspace{1cm} (3.48)

$L_p$ is minimized with respect to $w, b$, with a concurrent requirement that the derivatives of $L_p$ with respect to all $\alpha$ vanish. Since the gradient will have a very small value, the
minimum of $L_p$ will be attained. The solution to this minimization is seen in Equations 3.49 and 3.50.

$$w = \sum_i \alpha_i y_i x_i$$ \hspace{1cm} (3.49)

$$\sum_i \alpha_i y_i = 0$$ \hspace{1cm} (3.50)

Equations 3.49 and 3.50 can be substituted into Equation 3.54 to obtain the Dual Lagrangian:

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j$$ \hspace{1cm} (3.51)

Support vector training constitutes maximizing $L_D$ with respect to $\alpha_i$ subject to constraints in Equation 3.50. The Lagrange multiplier $\alpha_i$ exists for each training point. When the Lagrange multiplier of a training point has a nonzero value, it is considered a support vector that lies on one of two hyperplanes. If the Lagrange multiplier of a training point is equal to zero, it lies on the margin.

The Karsh-Kuhn-Tucker (KKT) conditions must be met for any optimization problem that has constraints. These conditions are a system of equations and inequalities that the solution of a nonlinear programming problem must satisfy when the objective function and the constraint functions are differentiable. The KKT conditions that must be met for the Lagrangian are as follows:

$$\frac{\partial}{\partial w_v} L_p = w_v - \sum_i \alpha_i y_i x_i v = 0 \quad v = 1, \ldots, d$$ \hspace{1cm} (3.52)

$$\frac{\partial}{\partial \alpha_i} L_p = - \sum_i \alpha_i y_i = 0$$ \hspace{1cm} (3.53)

$$y_i (x_i \cdot w + b) - 1 \geq 0 \quad i = 1, \ldots, l$$ \hspace{1cm} (3.54)
\[ \alpha_i \geq 0 \quad \forall i \]  \hspace{1cm} (3.55)

\[ \alpha_i (y_i (x_i \cdot w + b) - 1) = 0 \quad \forall i \]  \hspace{1cm} (3.56)

When the above conditions, which are designed for separable data, are applied to a non-separable dataset, no feasible solution will be found. To compensate for non-separable data, slack variables are introduced which alter the constraints:

\[ x_i \cdot w + b \geq 1 - \xi \quad \text{for} \quad y_i = +1 \]

\[ x_i \cdot w + b \leq -1 + \xi \quad \text{for} \quad y_i = -1 \]  \hspace{1cm} (3.57)

\[ \xi_i \geq 0 \quad \forall i \]

where \( \xi \) is a slack variable.
Figure 3.8 shows the decision boundaries and hyperplanes in a non-linearly separable case with slack variables.

The SVM still aims to maximize the margin between the hyperplanes, while minimizing the number of data points between the hyperplanes. The Lagrangian constraints are now:

\[ 0 \leq \alpha_i \leq C \]
\[ \sum_i \alpha_i y_i = 0 \]  (3.58)
The variable $C$ assigns a higher penalty to errors, which essentially controls the width of the margin. The only difference from the optimal hyperplane is the addition of the upper bound of $C$ on $\alpha_i$. The primal Lagrangian is:

$$L_p = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \sum_i \alpha_i \{y_i (x_i \cdot w + b) - 1 + \xi_i\} - \sum_i \mu_i \xi_i \quad (3.59)$$

The Lagrange multipliers $\mu_i$ ensure that the slack variables $\xi_i$ remain positive.

The following modified KKT conditions must be satisfied since the constraint function now contains the slack variable:

$${\frac{\partial}{\partial w_v} L_p} = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad v = 1, \ldots, d \quad i = 1, \ldots, l \quad (3.60)$$

$${\frac{\partial}{\partial \alpha_i} L_p} = -\sum_i \alpha_i y_i = 0 \quad (3.61)$$

$${\frac{\partial}{\partial \xi_i} L_p} = C - \alpha_i - \mu_i \quad (3.62)$$

$$y_i (x_i \cdot w + b) - 1 + \xi_i \geq 0 \quad (3.63)$$

$$\xi_i \geq 0 \quad (3.64)$$

$$\alpha_i \geq 0 \quad (3.65)$$

$$\mu_i \geq 0 \quad (3.66)$$

$$\alpha_i (y_i (x_i \cdot w + b) - 1 + \xi_i) = 0 \quad (3.67)$$

$$\mu_i \xi_i = 0 \quad (3.68)$$

The KKT conditions can be interpreted in terms of a decision boundary: A point $x_i$ satisfying Equation 3.60 with equality is on the margin, which satisfies Equation 3.65 with inequality. A point $x_i$ satisfying Equation 3.60 with inequality must have Equation
3.65 at equality. Therefore, all points that are on the boundary (support vectors) have $\alpha_i > 0$, while all other points have $\alpha_i = 0$.

The KKT conditions are solved using quadratic programming (QP). It involves maximizing or minimizing the quadratic function of several variables subject to linear constraints. This optimization can be solved with iterative numerical techniques. The KKT conditions only allow for an otherwise linear decision boundary handle non-separable data. The SVM does not create nonlinear decision boundaries. The SVM can only create linear decision boundaries. When a classification is not linearly separable, the SVM is able to still create a linear decision boundary, but in a higher dimensional space. Transforming $x_i$ through a (kernel) transformation function allows the construction of a linear boundary. The kernel allows the SVM to perform a nonlinear mapping from the input space to the higher dimensional space that is hidden from both the input and output. This creates an optimal (linear) hyperplane in the high dimensional space. Conveniently, no computation need be done in the higher dimensional space due to the kernel trick. The following kernel is used to replace $x_i \cdot x_j$ in the Lagrangian:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$$

(3.69)

Equation 3.69 represents the general form of a kernel. Kernels can be expressed more explicitly. The Gaussian kernel with width $\sigma$ is used in this study:

$$K(x, y) = e^{-\frac{|x-y|^2}{2\sigma^2}}$$

(3.70)

All kernels chosen must satisfy Mercer’s condition to ensure that function carried out in a higher dimensional space gives the same result in the lower dimensional space. The following expansion:
\[ K(x,y) = \sum_i \Phi(x)_i \Phi(y)_i \] (3.71)

for any arbitrary function \( g(x) \), for which:

\[ \int g(x)^2 \, dx < \infty \] (3.72)

then allows the following expansion to be valid:

\[ \int K(x,y)g(x)g(y) \, dx \, dy \geq 0 \] (3.73)

The decision boundary is:

\[ g(x) = \sum_{j=1}^n \alpha_j y_j K(x_j,x) + b = \sum_{x_i \in S} \alpha_i y_i K(x_i,x) + b = 0 \] (3.74)

The final decision boundary is a weighted sum of the kernel function evaluated on dot products of the test data with the support vectors [84]. The original Lagrangian function (Equation 3.51) becomes:

\[ L_D = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i,x_j) \] (3.75)

3.7.3 CLASSIFIER TRAINING

The input data into the data fusion classification algorithm are features extracted from ERP and MRI data. The MRI data in its raw form are topographic images of the brain taken in consecutive slices perpendicular to the coronal and parallel to the transverse planes. To ensure a standardized analysis across patients, an elastic warping algorithm places images into a standardized space while preserving the morphological characteristics of the individual brain. In order to extract the quantized volumetric data of various brain regions, the raw image is then segmented into white matter, gray matter, cerebrospinal fluid, and ventricles, and a density map of the image is calculated. This
analysis allows the determination of which tissues are experiencing neuronal atrophy. Once the density map is composed, an automated region of interest (ROI) analysis then determines the brain regions visible in the image. This sequence removes skull regions as well. This is repeated for each image slice and then consecutive slices are compiled. Voxel analysis is used to quantize the volume of the various brain images. Figure 3.9 displays the image processing sequence.

![Image](image-url)

*Figure 3.9: The sequence of processing of a raw MRI through a tissue segmentation step and through a region of interest analysis*

Thus volumes were measured for 14 unique anatomically defined regions in each hemisphere of the brain, enabling us to examine volume differences in regions of interest among various diagnostic groups. The features used for classification are the volumes of the following brain regions: Left/right lateral ventricle, left/right hippocampal, left/right parietal lobe white matter, left/right parietal lobe grey matter, left/right temporal lobe white matter, left/right temporal lobe grey matter, left/right frontal lobe white matter, left/right frontal lobe grey matter, left/right anterior cingulated, left/right posterior cingulated, left/right medial temporal lobe, left/right insula [85-88].

Of the 24 volumetric MRI features, a random subspace sampling method was utilized in order to create an ensemble of diverse classifiers. Random subsets of 18
features were taken during training in order to match the number of ERP classifiers (30 or 42 classifiers) to ensure similar voting weights, as will be explained further in this section.

The ERP data used was extracted using the discrete wavelet transform with the Daubechies 4 wavelet. The DWT of a 1 second length ERP sampled at 256 Hz results in 7 levels of detail coefficients. Each epoch begins 200 ms prior to stimuli and ends 800 ms after stimuli, which are then segmented and averaged with respect to stimulus type. Since a majority of the distinguishing information within the ERP is present only from 0-600 ms post stimulus, the information from the first and last 200 ms of the signal are removed. After processing, 6 detail coefficients from level five, and 4 coefficients from levels six and seven are produced. These are the coefficients used as input into the classification algorithm. The algorithm accepts data from target and novel stimuli from either 5 or 7 electrodes at a time and 3 frequency bands resulting in 30 or 42 classifiers trained on distinct electrode/stimulus/frequency band combinations.

The CSF biomarker data used as input into a separate benchmark classification algorithm consists of 6 quantized levels of hyperphosphorylated tau and beta amyloid proteins. These data were obtained from cerebral spinal fluid collected during a lumbar puncture of each patient. These data are not used in any type of data fusion or ensemble based classification algorithm. However, they are used in a leave one out classification algorithm as a reliability benchmark for the other algorithms since the CSF biomarkers are the currently the best method of diagnosis available.

For all fusion types, support vector machines were used except for sum rule, for which the multilayer perceptron was used since it provides a continuous range of support
weights for classification rather than discrete labels like the SVM. For all SVM implementation the same parameters were used. A C value of 1000 with a Gaussian kernel with spread of .5 was used. This spread was used because all data was normalized to between 0 and 1 and therefore had a mean close to .5. For the sum rule implementation the MLP was used with 15 hidden layer nodes in a single layer and an error goal of 0.15.

3.7.4 K-FOLD CROSS-VALIDATION

Cross validation is used for training and testing in order to provide a reasonably good estimate of the true generalization performance. The entire available dataset is divided into K blocks. Of these blocks, K-1 blocks are used for training and the Kth block is used for testing. This process is repeated K times using a different block for testing in each iteration. The performances of the trained algorithm on all K testing blocks are averaged for the best estimation of the dataset’s performance [89]. Figure 3.10 shows the K-fold cross validation process.

Figure 3.10: K-Fold cross validation. Within each run the shaded data block is used for testing with the other data blocks used for training.
Choosing a small $K$ value reduces the amount of data for training, which can have a detrimental impact on small datasets. Choosing a large $K$ value results in a better-trained classifier, but at the expense of high computation time and greater performance variability. When $K$ is equal to the number of data instances, a leave-one-out (LOO) cross validation is obtained. In the case of this study, one patient is left out for testing while the remaining patients in the dataset were used for classifier training. Each subject is used exactly once for testing. This is the best cross validation technique for small datasets. This study uses and compares LOO, dual LOO, and 70% / 15% / 15% cross validation technique. Dual LOO uses 1 subject to obtain classifier parameters through validation approaches, 1 subject for testing, and the remainder for training. The 70% / 15% / 15% cross validation technique uses 15% of the data to obtain classifier parameters, 15% for testing, and 70% for training. Classifier parameters only need to be obtained during weighted majority voting.

3.8 ENSEMBLE OF CLASSIFIER BASED DATA FUSION

Data fusion entails combining data sources from multiple information sources in order to achieve a more informed classification decision. Data fusion in this study was done at the decision level due to the heterogeneous data types. The decision level data fusion was done by creating an ensemble of classifiers in a mixture of experts form. Multiple classifiers each providing a decision on a single subject’s data can be considered akin to a person consulting a panel of experts before making a decision, ensuring increased reliability and overall accuracy of the final decision. Each classifier generates distinct decision boundaries on the training data, each with its own error. Combining the
decisions from an ensemble of classifiers helps reduce error. An ensemble of classifiers must contain sufficiently diverse classifiers trained on complementary training data.

Within this study, the main classification algorithm consists of two ‘experts’, one trained on MRI data and one trained on ERP data. Each ‘expert’ consists of 30 or 42 classifiers (depending on a 5 or 7 electrode combination), each trained on a particular subset of the data: for ERP, a particular stimulus/electrode/frequency band combination, and for MRI a particular random 18-feature subset of the 24-featureset. Figure 3.11 shows a model of this ensemble. Classifiers within each expert contribute their decision, which are then combined to create a final decision. Regardless of the type of combination rule used, the ERP and MRI experts always have an equal overall weight in the final decision.

![Diagram of the ensemble based classification system employed.](image)

*Figure 3.11: Diagram of the ensemble based classification system employed.*
3.8.1 COMBINATION RULES

The combination methods used in this study include simple majority voting, weighted majority voting, and the sum rule. Sum rule is the only combination technique that requires use of the MLP. An MLP gives a continuous output, \( d_{i,j} \in [0,1] \), support given by classifier \( C_i \) to class \( j \), where \( i=1,\ldots,N \) and \( j=1,\ldots,c \). \( N \) is the total number of classifiers and \( c \) is the total number of classes. The supports of each class from a given ensemble are summed, and the class with the highest sum becomes the decision as described in Equation 3.76.

\[
\mu_j(x) = \frac{1}{N} \sum_{j=1}^{N} d_{i,j}(x) \quad d_{i,j}(x) \in [0,1]
\] (3.76)

To calculate simple majority voting, the final class decision is based on which class the majority of the classifiers selected. SVMs give class decisions in binary form. The support for each class is then:

\[
\mu_j(x) = \sum_{j=1}^{N} d_{i,j}(x) \quad d_{i,j}(x) \in \{0,1\}
\] (3.77)

Weighted majority voting is similar to simple majority voting except a weighting factor is introduced based on validation performance as:

\[
\mu_j(x) = \sum_{j=1}^{N} W(T_i)d_{i,j}(x) \quad d_{i,j}(x) \in \{0,1\}
\] (3.78)

3.9 CLINICAL DIAGNOSTIC MEASURES

There are several measures of diagnostic accuracy. The percentage of correct classification can be a performance measure. For medical review, a number of more informative diagnostic performance measures must be reported. These measures are
positive predictive value, sensitivity, and specificity, calculated based on the number of true/false positive/negative diagnoses obtained as shown in Table 3.2.

<table>
<thead>
<tr>
<th>Test Outcome</th>
<th>Condition</th>
<th>Present</th>
<th>Not Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive</td>
<td>(TP)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative (FN)</td>
<td>True Negative (TN)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Explanation of diagnostic performance metrics.

Positive predictive value (PPV) is the probability that the subject has a disease given they tested positive. Sensitivity is the probability of a positive diagnosis given the condition is present in the subject. Specificity is the probability of a negative diagnosis when the subject does not have the condition. Each of these measures provides a unique insight into the efficacy of a diagnostic approach. Equations 3.79, 3.80, and 3.81 show formulations for these measures.

\[
PPV = \frac{TP}{TP + FP} \tag{3.79}
\]

\[
Sensitivity = \frac{TP}{TP + FN} \tag{3.80}
\]

\[
Specificity = \frac{TN}{TN + FP} \tag{3.81}
\]
CHAPTER 4

RESULTS

The results presented in this section are divided into several cohort comparisons. Within each group, ERP data, processed with the discrete wavelet transform, along with MRI volumetric brain data will be used to train an ensemble of MLP or SVM classifiers. Cross validation was used in addition to multiple decision level fusion techniques. Comparisons containing the Alzheimer’s Disease group have CSF biomarker based diagnostic performances, which is considered as the ‘gold standard’ diagnostic performance. All groups are examined with the ERP and MRI data fusion approach unless otherwise specified. The diagnostic cohort comparison groups are as follows:

- Alzheimer’s Disease vs. Control (ERP Only)
- Alzheimer’s Disease vs. Control
- Parkinson’s Disease vs. Control
- Mild Cognitive Impairment vs. Control
- Alzheimer’s Disease vs. Parkinson’s Disease vs. Control
- Mild Alzheimer’s Disease vs. Severe Alzheimer’s Disease vs. Control

4.1 ALZHEIMER’S DISEASE vs. CONTROL

The Alzheimer’s disease diagnosis based on the ERP only cohort has been performed with a weighted majority voting algorithm. In one iteration of the weighted majority voting (WMV) one patient was used for testing, one for validation to determine voting
weights, and the remainder for training. This was performed in a dual LOO fashion such that each patient was used for testing and validation once but never at the same time.

ERP features were used to train an ensemble of classifiers, which were combined using decision level fusion. This analysis was done to compare the diagnostic performances using various electrode combinations that were chosen based on their high performances on previous work with this cohort. The electrode combination that results in the highest performance based on the ERP data alone was the electrode combination used later in the ERP + MRI data fusion analysis. The three lowest frequency levels of wavelet coefficients for each electrode from novel and target stimuli were used as input features into the diagnosis algorithm. Table 4.1 shows the diagnostic performances and 95% confidence intervals for the AD/Normal classification. Combining the 3 lowest level frequency bands from the OZ, P7, P8, PZ, CZ, FZ, and FP2 electrodes results in the highest diagnostic accuracy in the AD vs. normal ERP only patient group. This electrode combination is therefore used in all further ERP analysis.

Table 4.1: Results for the classification of the AD patient group from the Control patient group using weighted majority decision fusion.

<table>
<thead>
<tr>
<th>Electrode Combination</th>
<th>Performance (%)</th>
<th>PPV (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P7, PZ, P8, T7, T8, CZ, FZ</td>
<td>77.55 ± 1.90</td>
<td>81.63</td>
<td>75.47</td>
<td>80.00</td>
</tr>
<tr>
<td>OZ, P3, PZ, P4, CZ, T7, T8</td>
<td>75.51 ±1.77</td>
<td>83.67</td>
<td>71.93</td>
<td>80.49</td>
</tr>
<tr>
<td>P7, PZ, P8, C3, C4, CZ, FZ</td>
<td>78.57 ±1.48</td>
<td>71.43</td>
<td>83.33</td>
<td>75.00</td>
</tr>
<tr>
<td>OZ, P7, P8, PZ, CZ, FZ, FP2</td>
<td>80.61 ±1.67</td>
<td>81.63</td>
<td>80.00</td>
<td>81.25</td>
</tr>
<tr>
<td>P3, P4, P7, PZ, P8</td>
<td>75.50 ± 1.61</td>
<td>77.78</td>
<td>71.43</td>
<td>79.59</td>
</tr>
</tbody>
</table>

The Alzheimer’s disease diagnosis based on ERP and MRI data fusion has been performed with multiple algorithm architectures to obtain diagnostic performances. The AD diagnosis cohort consists of 34 normal and 49 AD patients. The cohort size is smaller than the ERP only cohort due to the data constraints. All patients in this cohort
were required to have ERP, MRI, and CSF data. First, the biochemical marker data, which are a measure of hyperphosphorylated tau and beta amyloid proteins in the CSF, were used to train an SVM based leave-one-out classification algorithm to obtain a baseline performance. AD diagnosis via CSF biomarkers is currently the single most accurate method for diagnosis and will be considered as the benchmark for the data fusion classification. The data fusion trials utilized the ERP and MRI volumetric data. Multiple decision level fusion techniques were used. In one iteration of the weighted majority voting (WMV) diagnostic scheme, one patient was used for testing, one for validation to determine voting weights, and the remainder for training. This was performed in a dual LOO fashion so each patient was used for testing and validation once, but never at the same time. The weighted majority voting was also performed with a modified cross validation technique referred to as 70/15/15 in which 70% of patients are used for training, 15% for validation, and 15% for testing. This procedure was repeated 100 times for all trials, with random training, testing, and validation groups being drawn during each iteration. Both WMV schemes use SVM base classifiers. Simple majority voting was performed with a LOO cross validation using an SVM base classifier. The sum rule used a LOO cross validation technique with an MLP base classifier due to its inherent ability to provide supports to each class. Table 4.2 shows the diagnostic performances and 95% confidence intervals for AD diagnosis.
Table 4.2: Results for the classification of the AD patient group (49 subjects) from the Control patient group (34 subjects) using multiple decision fusion and cross validation techniques. Biochemical diagnostic performances shown as benchmark.

<table>
<thead>
<tr>
<th>Weighted Majority Voting (LOO Cross-Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>78.31±1.67</td>
<td>88.97±2.32</td>
<td>93.99±1.96</td>
<td>92.77±2.1</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>77.55</td>
<td>89.80</td>
<td>93.88</td>
<td>91.84</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>84.44</td>
<td>91.67</td>
<td>95.83</td>
<td>95.74</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>71.05</td>
<td>85.71</td>
<td>94.13</td>
<td>88.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weighted Majority Voting (70/15/15 Cross-Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>70.76±.83</td>
<td>78.91±.86</td>
<td>83.73±.70</td>
<td>92.77±2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simple Majority Voting (LOO Cross-Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>72.29±4.58</td>
<td>87.95±4.61</td>
<td>91.57±.0536</td>
<td>92.77±2.1</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>67.35</td>
<td>85.71</td>
<td>91.84</td>
<td>91.84</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>82.50</td>
<td>93.33</td>
<td>93.75</td>
<td>95.74</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>62.79</td>
<td>81.58</td>
<td>88.57</td>
<td>88.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum Rule (LOO Cross-Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>71.08±2.7</td>
<td>85.54±2.8</td>
<td>89.16±2.1</td>
<td>92.77±2.1</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>73.47</td>
<td>83.67</td>
<td>93.88</td>
<td>91.84</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>76.60</td>
<td>91.11</td>
<td>88.46</td>
<td>95.74</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>63.89</td>
<td>78.95</td>
<td>90.32</td>
<td>88.89</td>
</tr>
</tbody>
</table>
4.2 MILD COGNITIVE IMPAIRMENT vs. CONTROL

The Mild Cognitive Impairment diagnosis has been performed with the same algorithm architectures to obtain diagnostic performances. They follow the same setup as described above as the AD diagnosis. Table 4.3 shows the diagnostic performances and 95% confidence intervals of the MCI diagnoses.

Table 4.3: Results for the classification of the MCI patient group (39 subjects) from the Control patient group (34 subjects) using multiple decision fusion and cross validation techniques. Biochemical diagnostic performances shown as benchmark.

<table>
<thead>
<tr>
<th>Weighted Majority Voting (LOO Cross - Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>71.23 ± 2.66</td>
<td>82.19 ± 2.97</td>
<td>84.93 ± 2.78</td>
<td>89.04 ± 2.03</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>76.47</td>
<td>84.21</td>
<td>86.84</td>
<td>91.89</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>66.77</td>
<td>82.05</td>
<td>84.62</td>
<td>87.18</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>76.47</td>
<td>82.35</td>
<td>85.29</td>
<td>91.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weighted Majority Voting (70/15/15 Cross - Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>64.11 ± 1.72</td>
<td>71.40 ± 1.57</td>
<td>77.62 ± 1.04</td>
<td>89.04 ± 2.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simple Majority Voting (LOO Cross - Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>67.12 ± 3.29</td>
<td>73.96 ± 3.02</td>
<td>71.23 +/-3.01</td>
<td>89.04 ± 2.03</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>70.27</td>
<td>73.81</td>
<td>78.12</td>
<td>91.89</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>66.67</td>
<td>79.49</td>
<td>64.10</td>
<td>87.18</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>67.67</td>
<td>67.65</td>
<td>79.41</td>
<td>91.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum Rule (LOO Cross - Validation)</th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
<th>Biochemical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance (%)</td>
<td>65.75 ± 2.99</td>
<td>78.08 ± 2.84</td>
<td>82.19 ± 0.329</td>
<td>89.04 ± 2.03</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>71.88</td>
<td>81.08</td>
<td>88.24</td>
<td>91.89</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>58.97</td>
<td>76.82</td>
<td>76.92</td>
<td>87.18</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>73.53</td>
<td>79.41</td>
<td>88.24</td>
<td>91.18</td>
</tr>
</tbody>
</table>
4.3 PARKINSON’S DISEASE vs. CONTROL

The Parkinson’s disease diagnosis was performed with multiple algorithm architectures to obtain diagnostic performances. They follow the same setup as described above as the AD diagnosis. For the Parkinson’s diagnostics, however, there were no biochemical benchmark data. The hyperphosphorylated tau and beta amyloid proteins only occur in AD. There is no similar benchmark data available for PD in this study. Table 4.4 shows the diagnostic performances and 95% confidence intervals of the PD diagnoses.

*Table 4.4: Results for the classification of the PD patient group (39 subjects) from the Control patient group (34 subjects) using multiple decision fusion and cross validation techniques.*

<table>
<thead>
<tr>
<th></th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted Majority Voting (LOO Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>64.38 ± 3.74</td>
<td>71.23 ± 3.56</td>
<td>75.34 ± 3.15</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>70.97</td>
<td>75.00</td>
<td>78.38</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>56.41</td>
<td>69.23</td>
<td>74.36</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>73.53</td>
<td>73.53</td>
<td>76.47</td>
</tr>
<tr>
<td>Weighted Majority Voting (70/15/15 Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>58.61 ± 1.92</td>
<td>60.02 ± 1.13</td>
<td>62.85 ± 1.57</td>
</tr>
<tr>
<td>Simple Majority Voting (LOO Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>57.53 ± 4.53</td>
<td>69.85 ± 4.75</td>
<td>67.12 ± 4.04</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>59.09</td>
<td>72.97</td>
<td>74.19</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>66.67</td>
<td>69.23</td>
<td>58.97</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>47.06</td>
<td>70.59</td>
<td>76.47</td>
</tr>
<tr>
<td>Sum Rule (LOO Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>67.12 ± 3.40</td>
<td>69.86 ± 2.87</td>
<td>78.08 ± 2.54</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>71.43</td>
<td>72.97</td>
<td>82.86</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>64.10</td>
<td>69.23</td>
<td>74.36</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>70.59</td>
<td>70.59</td>
<td>82.35</td>
</tr>
</tbody>
</table>
4.4 THREE CLASS PROBLEMS

The efficacy of our approach to differentiate between two different dementia states and normal controls was also examined in the three-class problem setting.

4.4.1 PARKINSON’S DISEASE vs. ALZHEIMER’S DISEASE vs. CONTROL

The three class classification problem of differentiating between PD, AD, and control follows the same setup as described above as the AD diagnosis. Since PD patients do not have biochemical benchmark data, there will be no such benchmark. Additionally, clinical diagnostic measures will not be viable in this setting since those measures are inherently designed for a two class, disease and non-disease, diagnosis. Table 4.5 shows the diagnostic performances and 95% confidence intervals of this three-class problem.

Table 4.5: Results for the classification of the PD patient group (39 subjects) and AD patient group (49 subjects) from the Control patient group (34 subjects) using multiple decision fusion and cross validation techniques.

<table>
<thead>
<tr>
<th></th>
<th>ERP</th>
<th>MRI</th>
<th>ERP + MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted Majority Voting (LOO Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>43.06±2.94</td>
<td>66.39±3.21</td>
<td>71.31±2.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Majority Voting (70/15/15 Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>40.98 ± 1.56</td>
<td>58.20±1.68</td>
<td>62.78±2.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple Majority Voting (LOO Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>45.08 ± 3.64</td>
<td>69.67±3.41</td>
<td>64.75±3.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum Rule (LOO Cross - Validation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance (%)</td>
<td>42.62 ±2.22</td>
<td>67.21 +/- 1.90</td>
<td>68.85 +/- 1.83</td>
</tr>
</tbody>
</table>

88
The three class classification problem of differentiating between mild AD, severe AD, and control has been performed with the two highest performing algorithm architectures from the two class study. AD patients used for training are considered as mild AD subjects if they have an MMSE score above a certain cutoff, and those with MMSE scores lower than the cutoff are considered severe AD. The cutoff MMSE scores were 20, 23, and 26. The cohort with MMSE cutoff of 20 contains of 29 mild AD, 19 severe AD, and 34 normal subjects. The cohort with MMSE cutoff of 23 contains 19 mild AD, 29 severe AD, and 34 normal subject. The cohort with MMSE cutoff of 26 contains 11 mild AD, 38 severe AD, and 34 control subjects. This test was conducted using the best performing algorithms from the previous parts of the study, which were weighted majority voting with leave-one-out and 70/15/15 cross validations. Additionally, clinical diagnostic measures will are not viable in this setting since these measures are inherently designed for a two class, disease and non-disease state diagnosis. Table 4.6 shows the diagnostic performances and 95% confidence intervals of this three-class problem.
Table 4.6: Results for the classification of the mild and severe AD patient groups from the control patient group using decision fusion and two cross validation techniques with three MMSE cutoffs.

<table>
<thead>
<tr>
<th>MMSE CUTOFF = 20</th>
<th>Weighted Majority Voting (LOO Cross-Validation)</th>
<th>Weighted Majority Voting (70/15/15 Cross-Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERP</td>
<td>MRI</td>
<td>ERP + MRI</td>
</tr>
<tr>
<td>Performance (%)</td>
<td>46.34±2.57</td>
<td>40.24±2.74</td>
</tr>
<tr>
<td>ERP</td>
<td>MRI</td>
<td>ERP + MRI</td>
</tr>
<tr>
<td>Performance (%)</td>
<td>52.08±2.60</td>
<td>44.32±1.46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MMSE CUTOFF = 23</th>
<th>Weighted Majority Voting (LOO Cross-Validation)</th>
<th>Weighted Majority Voting (70/15/15 Cross-Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERP</td>
<td>MRI</td>
<td>ERP + MRI</td>
</tr>
<tr>
<td>Performance (%)</td>
<td>43.21±2.62</td>
<td>45.80±2.56</td>
</tr>
<tr>
<td>ERP</td>
<td>MRI</td>
<td>ERP + MRI</td>
</tr>
<tr>
<td>Performance (%)</td>
<td>49.92±2.82</td>
<td>45.74±2.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MMSE CUTOFF = 26</th>
<th>Weighted Majority Voting (LOO Cross-Validation)</th>
<th>Weighted Majority Voting (70/15/15 Cross-Validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERP</td>
<td>MRI</td>
<td>ERP + MRI</td>
</tr>
<tr>
<td>Performance (%)</td>
<td>41.70±2.74</td>
<td>52.10±2.43</td>
</tr>
<tr>
<td>ERP</td>
<td>MRI</td>
<td>ERP + MRI</td>
</tr>
<tr>
<td>Performance (%)</td>
<td>53.50±2.63</td>
<td>45.25±2.10</td>
</tr>
</tbody>
</table>

The interpretations of these results are given in the next chapter.
CHAPTER 5
CONCLUSIONS

The goal of this work was to expand the previous efforts to develop an automated method of diagnosing Alzheimer’s disease. Specifically, the approaches developed in AD diagnosis have now been applied to the diagnosis of Parkinson’s disease and mild cognitive impairment as well as AD severity analysis. Also new in this effort is the integration of ERP and MRI based features for heterogeneous data fusion. ERPs were processed with artifact rejection, averaging, and wavelet feature extraction from novel and target stimuli. Wavelet coefficients were extracted from the 1-2 Hz, 2-4 Hz, and 4-8 Hz frequency bands. MRI images were processed using segmentation software that was able to determine volumes of various brain regions. ERP and MRI features were also obtained for PD and MCI diagnosis, as well as an AD vs. PD vs. Normal three-class problem, and an AD severity analysis. Some patient groups also had biochemical data whose diagnosis performance was used as a benchmark performance. An automated decision level data fusion classification was developed using SVM and MLP classifiers.

5.1 SUMMARY OF ACCOMPLISHMENTS

Previous work has focused on developing the most optimal feature sets to diagnose Alzheimer’s disease using ERPs. The previous work centered on an exhaustive search to determine the best combination of electrodes, stimuli, and wavelet frequency bands that would result in the highest diagnostic accuracy. Additionally, the performances of various classification algorithms were investigated. This study aimed to expand on
previous work by introducing a heterogeneous data source in the form of volumetric MRI data, which was used along with ERP data to diagnose between various dementias such as AD, PD, and MCI, and well as perform an AD severity analysis.

Previous studies determined the optimal combination of individual electrodes, stimuli, and frequency bands for maximal diagnostic performance using ERPs. In this study we first sought to update the performance metrics on the previously used cohort since it has now grown in size to 49 AD and 49 cognitively normal subjects. Various electrode combinations were tested for the best performance; however, unlike previous studies, the frequency bands and stimuli used remained fixed throughout all tests. Both novel and target stimuli from the 1-2 Hz, 2-4 Hz, and 4-8 Hz frequency bands were used for each electrode. A weighted majority voting algorithm using leave-one-out validation was used to obtain diagnostic performance from the ERPs. The optimal performance came from an electrode combination that included the parietal as well as the right frontal region of the brain, with a performance of 80.61%. This ERP only AD diagnosis analysis may not have reached the performance of previous studies due the limitation of fixed input features, but was still well above the 75% diagnostic accuracy that is currently available in community health clinics.

The highest performing electrode combination was used in various other diagnosis tests along with MRI data. These two heterogeneous data sources were analyzed individually and together in various scenarios to determine if combining these two data types would provide a statistically significant increase in diagnostic performance compared to each alone. Weighted majority voting with LOO and 70/15/15 cross validation, simple majority voting, and sum rule were used for the decision level
data fusion. WMV and SMV were both performed using support vector machines. The LOO classification algorithm that used the sum rule for decision fusion was the only test that used the multilayer perceptron due to its readily available class support information. Some subject groups had access to biochemical data in the form of hyperphosphorylated tau and amyloid beta protein levels from the cerebral spinal fluid. This test is the single most accurate test available to diagnose AD, however it is an expensive and painful test so it will only be used as a benchmark performance in these tests. In groups with biochemical data, the patients were classified using the biochemical markers using a LOO validation, which served as a baseline for the maximal performance currently available from a single test. The AD data fusion cohort has 49 AD and 34 cognitively normal subjects that had ERP, MRI, and biochemical data. The biochemical analysis can diagnose AD with 92.77% accuracy. If the data fusion method can attain such a performance, then it would be an inexpensive and non-invasive diagnostic option. The highest ERP performance of 78.31% and the highest MRI performance of 89.16% were achieved using the weighted majority voting with LOO cross validation. Simple majority voting and sum rule had similar results, while WMV with 70/15/15 cross validation had lower performance figures, but also a significantly smaller confidence interval. When data fusion was performed on the LOO WMV the AD diagnostic accuracy improved dramatically to 93.99%, which is a statistically significant jump from the combined ensemble performances of the individual data types. Additionally, the average data fusion accuracy is higher than the average benchmark biochemical performance of 92.77%, although without statistical significance. Even with the strong MRI-only
diagnostic performance, the lower performing ERP data complemented the MRI data to perform a statistically significant data fusion.

The same group of tests was performed to diagnose MCI with a cohort of 39 MCI and 34 normal subjects. A biochemical benchmark was available for this cohort as well. Diagnostic performance obtained by biochemical data was 89%. The LOO WMV held the highest performances with the sum rule test performing only slightly lower. Again, we observed that the 70/15/15 WMV allows for significantly smaller confidence intervals due to a higher number of patients being utilized for determining algorithm parameters. ERP combined accuracy was 71.23%, MRI combined accuracy was 82.19%, and the combined data fusion diagnostic accuracy was 84.93%. This combined performance approaches the biochemical benchmark but remains statistically significantly lower. Again, it can be noticed that the MRI performs very strongly individually but is still able to complement the ERP data to boost combined performance. MRI seems to be a more exact diagnostic measure compared to ERP, which is susceptible to noise as well as other implementation anomalies that require significantly more preprocessing to be transitioned into a form suitable for dementia diagnosis.

Parkinson’s disease (PD) is a dementia that does not exhibit the same type of cognitive decline as AD or MCI. PD mainly degenerates motor skills in a patient. Therefore, the PD diagnosis test will determine if the ERP and MRI biomarkers, which showed strong performance in AD diagnosis, will be able to diagnose PD, despite the fact that it is pathologically dissimilar. The PD group does not have biochemical benchmark data since hyperphosphorylated tau and amyloid beta proteins do not occur in PD. The cohort used in this test contains 39 PD and 34 normal subjects. The overall
trend between the validation methods was similar to the previous tests. LOO WMV and sum rule performed similarly, while the 70/15/15 WMV and SMV performed lower. Again, the 70/15/15 test was able to reduce 95% confidence intervals by allowing more data for determining algorithm parameters. The highest data fusion diagnosis for PD occurred with the sum rule. ERP combined accuracy was 67.12%, MRI was 69.86%, with a combined data fusion diagnostic accuracy of 78.08%. Some anomalies begin to be seen; the highest individual MRI accuracy of 71.23% did not occur with the highest data fusion performance. Also, the ERP and MRI scores become quite close in some scenarios such as 70/15/15. Additionally, MRI actually exceeded the data fusion performance in the simple majority voting. Although the sum rule proved that data fusion could increase diagnosis performance, it is clear that either the data types or algorithms individually may not be optimal solution in diagnosing PD.

Next, three class analysis were performed to introduce more 'real-world' diagnostic scenarios. The first three-class problem involved a cohort of 49 AD, 39 PD, and 34 normal subjects. The highest ERP and the highest MRI combined performances came from simple majority voting of 45.08% and 69.67% respectively. Although other individual performances were very similar. The highest data fusion performance, however, came from the leave-one-out weighted majority voting of 71.31%. This proves that in a clinical diagnostic setting, there is a possibility of diagnosing Alzheimer’s disease from Parkinson’s disease from normal patients with a single all encompassing test.

Finally, a severity analysis was performed with the AD data fusion cohort. The severity analysis aims to be able to not only determine if a patient has AD or not, but to
also determine what stage of progression the patients AD is in. The cohort contains 34 normal and 49 AD subjects. The AD group was separated into two subgroups to create the three individual classes. This separation will be using the patients’ MMSE scores as a threshold. AD patients with an MMSE score below the threshold were considered severe AD while patients with an MMSE score above the threshold were considered mild AD. Three MMSE cutoffs were used; there were 20, 23, and 26. The combined performances for these three cutoffs were 60.92%, 68.17%, and 76.58%. In each case LOO and 70/15/15 weighted majority voting was employed. In each test, the 70/15/15 weighted majority voting consistently outperformed the LOO WMV in combined performance, although individual ERP and MRI performance were sometimes higher in the LOO test. It can be seen that as the size of the severe patient group increased with the increasing MMSE cutoff, the overall performance was improved. This could be for many reasons. The mild AD group data may have many similarities with cognitively normal data along with the severe AD group, and therefore more difficult to classify. This ambiguous class may have been lowering performance until its size was minimized. A more likely scenario is that the severe and mild AD groups are difficult to distinguish until the proper ‘fault-line’ was determined. In this case it can be seen that the higher MMSE score allows for the highest performance. However, there may be other metrics that can be used to more effectively separate the mild and severe AD group.
5.2 SOURCES OF ERROR

The sources of error in this study stem mainly from the data acquisition. The ERP data collection is prone to noise due to the very small magnitudes of the signals collected. Also, we know that some AD patients with very low cognitive ability were not able to properly perform the oddball paradigm test. In this case, there was no 'strong' and 'weak' P300 to investigate; there may not have been a P300 response at all. More importantly, the class information used for the ERP, MRI, and biochemical data came from clinical evaluation from expert neurologists with an expected accuracy of 90%. However, we accepted the given diagnosis as the correct class information for our tests. If some of the class labels were indeed incorrect, this could significantly affect the ability of the algorithms to learn the data. Autopsy is the only fully accurate way to diagnose dementias; so updating correct diagnosis information in the future could improve overall performance of these datasets and algorithms.

Also, the data itself was an amalgam from different test sites and processing centers. The MRI data was processed into its volumetric form using proprietary software that extracted only specific features at an off site facility. The ERP data was collected and preprocessed with an in-house automated technique. The use of professionally available software or having access to an expert electroencephalographer to analyze the raw data before testing may have helped reduce this error.
5.3 RECOMMENDATIONS FOR FUTURE WORK

The overall study from which the patients and data for this study were collected is a vast ongoing project. There is a continuing patient recruitment. All tests done in this study can be reanalyzed in the future for larger cohort sizes. Additionally, each two-class diagnosis problem should be optimized to the dementia being diagnosed. For example, the PD group should have the inclusion of PET scan or functional MRI (fMRI) data since they are more telling biomarkers and will become available as this study continues. Additionally, more work to be done to capitalize on the strengths of certain data types, such as MRI data in AD classification. Finally, more severity analysis problems should be addressed to emulate a more real world diagnostic environment. Distinguishing many dementias and dementia severity levels based on a data fusion of disease specific biomarkers is the future of this study.
REFERENCES


APPENDIX A

ERP GRAND AVERAGES FROM ERP ONLY COHORT

The graphs included in Appendix A are grand averages from the ERP only cohort for each electrode. This cohort contains 49 control and 49 AD patients. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control and AD patient groups are shown in each plot.

Figure A.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the ERP only cohort.
Figure A.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the ERP only cohort.
Figure A.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the ERP only cohort.
Figure A.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the ERP only cohort.
Figure A.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the ERP only cohort.
Figure A.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the ERP only cohort.
Figure A.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the ERP only cohort.
Figure A.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the ERP only cohort.
Figure A.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the ERP only cohort.
Figure A.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the ERP only cohort.
Figure A.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the ERP only cohort.
Figure A.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the ERP only cohort.
Figure A.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the ERP only cohort.
Figure A.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the ERP only cohort.
Figure A.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the ERP only cohort.
Figure A.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the ERP only cohort.
APPENDIX B

ERP GRAND AVERAGES FROM AD DATA FUSION COHORT

The graphs included in Appendix B are grand averages from the AD data fusion cohort for each electrode. This cohort contains 34 control and 49 AD patients. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control and AD patient groups are shown in each plot.

Figure B.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the AD data fusion cohort.
Figure B.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the AD data fusion cohort.
Figure B.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the AD data fusion cohort.
Figure B.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the AD data fusion cohort.
Figure B.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the AD data fusion cohort.
Figure B.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the AD data fusion cohort.
Figure B.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the AD data fusion cohort.
Figure B.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the AD data fusion cohort.
Figure B.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the AD data fusion cohort.
Figure B.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the AD data fusion cohort.
Figure B.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the AD data fusion cohort.
Figure B.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the AD data fusion cohort.
Figure B.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the AD data fusion cohort.
Figure B.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the AD data fusion cohort.
Figure B.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the AD data fusion cohort.
Figure B.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the AD data fusion cohort.
The graphs included in Appendix C are grand averages from the MCI data fusion cohort for each electrode. This cohort contains 34 control and 39 MCI patients. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control and MCI patient groups are shown in each plot.

Figure C.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the MCI data fusion cohort.
Figure C.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the MCI data fusion cohort.
Figure C.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the MCI data fusion cohort.
Figure C.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the MCI data fusion cohort.
Figure C.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the MCI data fusion cohort.
Figure C.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the MCI data fusion cohort.
Figure C.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the MCI data fusion cohort.
Figure C.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the MCI data fusion cohort.
Figure C.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the MCI data fusion cohort.
Figure C.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the MCI data fusion cohort.
Figure C.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the MCI data fusion cohort.
Figure C.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the MCI data fusion cohort.
Figure C.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the MCI data fusion cohort.
Figure C.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the MCI data fusion cohort.
Figure C.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the MCI data fusion cohort.
Figure C.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the MCI data fusion cohort.
APPENDIX D

ERP GRAND AVERAGES FROM PD DATA FUSION COHORT

The graphs included in Appendix D are grand averages from the PD data fusion cohort for each electrode. This cohort contains 34 control and 39 PD patients. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control and PD patient groups are shown in each plot.

Figure D.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the PD data fusion cohort.
Figure D.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the PD data fusion cohort.
Figure D.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the PD data fusion cohort.
Figure D.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the PD data fusion cohort.
Figure D.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the PD data fusion cohort.
Figure D.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the PD data fusion cohort.
Figure D.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the PD data fusion cohort.
Figure D.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the PD data fusion cohort.
Figure D.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the PD data fusion cohort.
Figure D.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the PD data fusion cohort.
Figure D.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the PD data fusion cohort.
Figure D.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the PD data fusion cohort.
Figure D.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the PD data fusion cohort.
Figure D.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the PD data fusion cohort.
Figure D.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the PD data fusion cohort.
Figure D.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the PD data fusion cohort.
APPENDIX E

ERP GRAND AVERAGES FROM AD + PD DATA FUSION COHORT

The graphs included in Appendix E are grand averages from the AD and PD data fusion cohort for each electrode. This cohort contains 34 control, 39 PD, and 49 AD patients. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control, PD, and AD patient groups are shown in each plot.
Figure E.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the AD and PD data fusion cohort.
Figure E.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the AD and PD data fusion cohort.
Figure E.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the AD and PD data fusion cohort.
Figure E.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the AD and PD data fusion cohort.
Figure E.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the AD and PD data fusion cohort.
Figure E.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the AD and PD data fusion cohort.
Figure E.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the AD and PD data fusion cohort.
Figure E.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the AD and PD data fusion cohort.
Figure E.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the AD and PD data fusion cohort.
Figure E.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the AD and PD data fusion cohort.
Figure E.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the AD and PD data fusion cohort.
Figure E.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the AD and PD data fusion cohort.
Figure E.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the AD and PD data fusion cohort.
Figure E.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the AD and PD data fusion cohort.
Figure E.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the AD and PD data fusion cohort.
Figure E.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the AD and PD data fusion cohort.
APPENDIX F

ERP GRAND AVERAGES FROM SEVERITY ANALYSIS

DATA FUSION COHORT (MMSE CUTOFF = 20)

The graphs included in Appendix F are grand averages from the AD severity data fusion cohort for each electrode. This cohort contains 34 control, 30 mild AD, and 19 severe AD patients. The mild and severe AD groups were separated at a MMSE score of 20. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control, mild AD, and severe AD patient groups are shown in each plot.
Figure F.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the 20 MMSE AD severity data fusion cohort.
Figure F.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the 20 MMSE AD severity data fusion cohort.
APPENDIX G

ERP GRAND AVERAGES FROM SEVERITY ANALYSIS

DATA FUSION COHORT (MMSE CUTOFF = 23)

The graphs included in Appendix G are grand averages from the AD severity data fusion cohort for each electrode. This cohort contains 34 control, 19 mild AD, and 30 severe AD patients. The mild and severe AD groups were separated at a MMSE score of 23. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control, mild AD, and severe AD patient groups are shown in each plot.
Figure G.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T7 electrode from the 23 MMSE AD severity data fusion cohort.
Figure G.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the 23 MMSE AD severity data fusion cohort.
APPENDIX H

ERP GRAND AVERAGES FROM SEVERITY ANALYSIS

DATA FUSION COHORT (MMSE CUTOFF = 26)

The graphs included in Appendix H are grand averages from the AD severity data fusion cohort for each electrode. This cohort contains 34 control, 11 mild AD, and 38 severe AD patients. The mild and severe AD groups were separated at a MMSE score of 26. The top plot shows the ERP response to the target stimuli and the bottom plot shows the ERP response to the novel stimuli. Grand averages from the control, mild AD, and severe AD patient groups are shown in each plot.
Figure H.1: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C3 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.2: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the C4 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.3: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the CZ electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.4: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP1 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.5: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FP2 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.6: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F3 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.7: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the F4 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.8: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the FZ electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.9: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the OZ electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.10: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P3 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.11: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P4 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.12: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P7 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.13: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the P8 electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.14: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the PZ electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.15: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the TZ electrode from the 26 MMSE AD severity data fusion cohort.
Figure H.16: Grand average ERP from responses to target (top) and novel (bottom) stimuli for the T8 electrode from the 26 MMSE AD severity data fusion cohort.