Rowan University

Rowan Digital Works

Theses and Dissertations

4-25-2007

Stacked generalization for early diagnosis of Alzheimer's disease

Hardik P. Gandhi
Rowan University

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

Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Gandhi, Hardik P., "Stacked generalization for early diagnosis of Alzheimer's disease" (2007). *Theses and Dissertations*. 780.

<https://rdw.rowan.edu/etd/780>

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

STACKED GENERALIZATION FOR EARLY DIAGNOSIS OF ALZHEIMER'S
DISEASE

by
Hardik P. Gandhi

A Thesis

Submitted in partial fulfillment of the requirements of the
Master of Science Degree
of
The Graduate School
at
Rowan University
April 25, 2007

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

~~Approved~~

~~Members~~ of the committee

~~Advison~~

~~For the Major Department~~

~~For the College~~

© 2007 Hardik P. Gandhi

ABSTRACT

Hardik P. Gandhi

STACKED GENERALIZATION FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE

2006/07

Dr. Robi Polikar

Master of Science in Electrical Engineering

Alzheimer's disease (AD) is a neurological disorder characterized by nerve degeneration and neuronal death. The diagnosis of AD at an early stage is a major concern due to growing number of elderly people affected by the disease, as well as the lack of a standard diagnosis procedure available to community clinics. A biomarker that will allow early diagnosis of the disease would be beneficial. Such biomarker should be non-invasive, simple to obtain, safe, inexpensive, accurate, and most importantly, must be made available to local health clinics for maximum effectiveness.

Recent studies have used wavelets and other signal processing methods to analyze EEG signals in an attempt to find a noninvasive biomarker for AD. In this study, multiresolution wavelet analysis was performed on event related potentials (ERPs) of electroencephalogram (EEG) signals. Extracted feature sets were then used to train the ensemble system, based on stacked generalization algorithm, and individual decisions were calculated and compared. The ensemble decisions from individual data sources (specific electrode – stimuli – frequency band combinations) were combined using different data fusion techniques for further analyses. Particular emphasis was made for the value of procedure in the diagnosis of the disease at its earliest stage.

TABLE OF CONTENTS

ABSTRACT	ii
LIST OF FIGURES	v
LIST OF TABLES.....	vii
CHAPTER 1 INTRODUCTION.....	1
1.1 BIOLOGICAL ASPECTS OF ALZHEIMER’S DISEASE	1
1.2 DIAGNOSIS.....	4
1.3 SCOPE OF THE THESIS	7
1.4 ORGANIZATION OF THE THESIS	7
CHAPTER 2 BACKGROUND.....	9
2.1 CHEMICAL BIOMARKERS OF AD	9
2.2 IMAGING TECHNIQUES	10
2.3 ELECTROENCEPHALOGRAM	10
2.4 PREVIOUS WORK	17
CHAPTER 3 METHODS.....	33
3.1 STUDY COHORT	33
3.2 EXPERIMENTAL SETUP	35
3.3 ODDBALL PARADIGM WITH NOVEL SOUNDS.....	36
3.4 TIME-FREQUENCY ANALYSIS FOR FEATURE EXTRACTION.....	38
3.5 AUTOMATED CLASSIFICATION USING PATTERN RECOGNITION TECHNIQUES	47
3.5.1 MLP NEURAL NETWORKS.....	47
3.5.2 BACKPROPAGATION LEARNING RULE.....	49
3.5.3 K-FOLD CROSS VALIDATION	52
3.5.4 ENSEMBLE OF CLASSIFIERS	54
3.5.5 STACKED GENERALIZATION.....	57
3.5.6 DATA FUSION.....	65
3.6 DIAGNOSTIC PERFORMANCE	68
3.7 METHOD OVERVIEW.....	70
CHAPTER 4 IMPLEMENTATION AND RESULTS	71
4.1 IMPLEMENTATION DETAILS.....	71
4.1.1 FEATURE EXTRACTION.....	71
4.1.2 CLASSIFICATION.....	72
4.2 RESULTS.....	74
4.2.1 TWO-CLASS PROBLEM	74
4.2.1a INDIVIDUAL PERFORMANCES BY STRAIGHT META- CLASSIFIER TRAINING (SMT) APPROACH.....	75

4.2.1b INDIVIDUAL PERFORMANCES BY HIERARCHICAL META- CLASSIFIER TRAINING (HMT) APPROACH.....	77
4.2.1c PERFORMANCES BY DECISION LEVEL DATA FUSION	80
4.2.2 THREE-CLASS PROBLEM.....	92
4.2.2a INDIVIDUAL PERFORMANCES	92
4.2.2b PERFORMANCES BY DECISION LEVEL DATA FUSION	95
CHAPTER 5 CONCLUSIONS	103
5.1 SUMMARY OF ACCOMPLISHMENTS	103
5.2 SOURCES OF ERROR.....	106
5.3 FUTURE WORK.....	106
REFERENCES	107
APPENDIXES	115
APPENDIX A ADDITIONAL TWO-CLASS FUSION RESULTS	115
APPENDIX B ADDITIONAL THREE-CLASS FUSION RESULTS.....	119

LIST OF FIGURES

<i>Figure 1.1: Illustration of amyloid plaques and neurofibrillary tangles[§] [5].....</i>	<i>2</i>
<i>Figure 1.2: Cross section view of brain at different stages of AD[§] [6].....</i>	<i>4</i>
<i>Figure 2.1: The international 10-20 system: (a) left view and (b) top view; (c)location and nomenclature of the intermediate 10% electrodes[§] [27].....</i>	<i>12</i>
<i>Figure 2.2: (a) & (b) Expected P300 behavior from normal and AD patients; (c) & (d) not all individual cases follow this expected behavior [44].....</i>	<i>16</i>
<i>Figure 3.1: 7-level DWT decomposition [61]</i>	<i>44</i>
<i>Figure 3.2: Daubechies-4 wavelet scaling function and wavelet function.....</i>	<i>45</i>
<i>Figure 3.3: 7-level wavelet decomposition for a probable AD patient and a cognitively normal patient.....</i>	<i>46</i>
<i>Figure 3.4: MLP neural network architecture [73].....</i>	<i>48</i>
<i>Figure 3.5: K-fold cross validation [73].....</i>	<i>53</i>
<i>Figure 3.6: Stacked generalization algorithm [44].....</i>	<i>57</i>
<i>Figure 3.7: Initial training by straight meta-classifier training (SMT) approach.....</i>	<i>61</i>
<i>Figure 3.8: Initial training by hierarchical meta-classifier training (HMT) approach.....</i>	<i>63</i>
<i>Figure 3.9: Pseudo-code for stacked generalization algorithm.....</i>	<i>64</i>
<i>Figure 3.10: Decision profile matrix [85].....</i>	<i>66</i>
<i>Figure 3.11: Overview of the analysis.....</i>	<i>70</i>
<i>Figure 4.1: Comparison of OGPs of SMV rule combinations in Table 4.14 with those of individual data sources, for two-class problem.....</i>	<i>82</i>
<i>Figure 4.2: Comparison of OGPs of SMV rule combinations in Table 4.15 with those of individual data sources, for two-class problem.....</i>	<i>83</i>
<i>Figure 4.3: Comparison of OGPs of WMV rule combinations in Table 4.16 with those of individual data sources, for two-class problem.....</i>	<i>84</i>
<i>Figure 4.4: Comparison of OGPs of WMV rule combinations in Table 4.17 with those of individual data sources, for two-class problem.....</i>	<i>86</i>
<i>Figure 4.5: Comparison of OGPs of product rule combinations in Table 4.18 with those of individual data sources, for two-class problem.</i>	<i>87</i>
<i>Figure 4.6: Comparison of OGPs of product rule combinations in Table 4.19 with those of individual data sources, for two-class problem.</i>	<i>88</i>
<i>Figure 4.7: Comparison of OGPs of sum rule combinations in Table 4.20 with those of individual data sources, for two-class problem.....</i>	<i>89</i>
<i>Figure 4.8: Comparison of OGPs of sum rule combinations in Table 4.21 with those of individual data sources, for two-class problem.....</i>	<i>91</i>
<i>Figure 4.9: Comparison of OGPs of WMV rule combinations in Table 4.29 with those of individual data sources, for three-class problem.</i>	<i>96</i>

<i>Figure 4.10: Comparison of OGPs of WMV rule combinations in Table 4.30 with those of individual data sources, for three-class problem.</i>	<i>97</i>
<i>Figure 4.11: Comparison of OGPs of product rule combinations in Table 4.31 with those of individual data sources, for three-class problem.</i>	<i>98</i>
<i>Figure 4.12: Comparison of OGPs of product rule combinations in Table 4.32 with those of individual data sources, for three-class problem.</i>	<i>99</i>
<i>Figure 4.13: Comparison of OGPs of sum rule combinations in Table 4.33 with those of individual data sources, for three-class problem.</i>	<i>100</i>
<i>Figure 4.14: Comparison of OGPs of sum rule combinations in Table 4.34 with those of individual data sources, for three-class problem.</i>	<i>101</i>
 <i>Figure 5.1: Relation of combinations of electrodes with physiological activities in brain</i>	<i>105</i>

LIST OF TABLES

<i>Table 2.1: Comparison of previous work.....</i>	<i>30</i>
<i>Table 3.1: Inclusion criteria for the research cohort.....</i>	<i>34</i>
<i>Table 3.2: Exclusion criteria for the research cohort.....</i>	<i>34</i>
<i>Table 3.3: Category for defining diagnostic performances</i>	<i>68</i>
<i>Table 4.1: Performance values for stacked generalization system trained using SMT approach on target-C_Z coefficients for four frequency bands, for two-class problem.....</i>	<i>75</i>
<i>Table 4.2: Performance values for stacked generalization system trained using SMT approach on target-F_Z coefficients for four frequency bands, for two-class problem.</i>	<i>75</i>
<i>Table 4.3: Performance values for stacked generalization system trained using SMT approach on target-P_Z coefficients for four frequency bands, for two-class problem.</i>	<i>75</i>
<i>Table 4.4: Performance values for stacked generalization system trained using SMT approach on novel-C_Z coefficients for four frequency bands, for two-class problem.....</i>	<i>76</i>
<i>Table 4.5: Performance values for stacked generalization system trained using SMT approach on novel-F_Z coefficients for four frequency bands, for two-class problem.</i>	<i>76</i>
<i>Table 4.6: Performance values for stacked generalization system trained using SMT approach on novel-P_Z coefficients for four frequency bands, for two-class problem.</i>	<i>76</i>
<i>Table 4.7: Performance values for stacked generalization system trained using HMT approach on target-C_Z coefficients for four frequency bands, for two-class problem.....</i>	<i>77</i>
<i>Table 4.8: Performance values for stacked generalization system trained using HMT approach on target-F_Z coefficients for four frequency bands, for two-class problem.</i>	<i>77</i>
<i>Table 4.9: Performance values for stacked generalization system trained using HMT approach on target-P_Z coefficients for four frequency bands, for two-class problem.</i>	<i>78</i>
<i>Table 4.10: Performance values for stacked generalization system trained using HMT approach on novel-C_Z coefficients for four frequency bands, for two-class problem.....</i>	<i>78</i>
<i>Table 4.11: Performance values for stacked generalization system trained using HMT approach on novel-F_Z coefficients for four frequency bands, for two-class problem.</i>	<i>78</i>
<i>Table 4.12: Performance values for stacked generalization system trained using HMT approach on novel-P_Z coefficients for four frequency bands, for two-class problem.</i>	<i>78</i>
<i>Table 4.13: Performance values for stacked generalization system trained using HMT approach on additional data sources, for two-class problem.</i>	<i>80</i>
<i>Table 4.14: Highest four performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using SMV rule, for two-class problem.</i>	<i>81</i>
<i>Table 4.15: Highest five performances from fusion combinations of outputs from all data sources using SMV rule, for two-class problem.</i>	<i>83</i>
<i>Table 4.16: Highest five performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using WMV rule, for two-class problem.</i>	<i>84</i>
<i>Table 4.17: Highest six performances from fusion combinations of outputs from all data sources using WMV rule, for two-class problem.....</i>	<i>85</i>

<i>Table 4.18: Highest five performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using product rule, for two class problem.</i>	<i>86</i>
<i>Table 4.19: Highest five performances from fusion combinations of outputs from all data sources using product rule, for two-class problem.</i>	<i>88</i>
<i>Table 4.20: Highest five performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using sum rule, for two-class problem.</i>	<i>89</i>
<i>Table 4.21: Highest five performances from fusion combinations of outputs from all data sources using sum rule, for two-class problem.</i>	<i>90</i>
<i>Table 4.22: Performance values for stacked generalization system, trained using HMT approach on target-C_Z coefficients for four frequency bands, for three-class problem.</i>	<i>92</i>
<i>Table 4.23: Performance values for stacked generalization system, trained using HMT approach on target-F_Z coefficients for four frequency bands, for three-class problem.</i>	<i>93</i>
<i>Table 4.24: Performance values for stacked generalization system, trained using HMT approach on target-P_Z coefficients for four frequency bands, for three-class problem.</i>	<i>93</i>
<i>Table 4.25: Performance values for stacked generalization system, trained using HMT approach on novel-C_Z coefficients for four frequency bands, for three-class problem.</i>	<i>93</i>
<i>Table 4.26: Performance values for stacked generalization system, trained using HMT approach on novel-F_Z coefficients for four frequency bands, for three-class problem.</i>	<i>93</i>
<i>Table 4.27: Performance values for stacked generalization system, trained using HMT approach on novel-P_Z coefficients for four frequency bands, for three-class problem.</i>	<i>93</i>
<i>Table 4.28: Performance values for stacked generalization system, trained using HMT approach on additional data sources, for three-class problem.</i>	<i>94</i>
<i>Table 4.29: Highest five performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using WMV rule, for three-class problem.</i>	<i>96</i>
<i>Table 4.30: Highest three performances from fusion combinations of outputs from all data sources using WMV rule, for three-class problem.</i>	<i>97</i>
<i>Table 4.31: Highest three performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using product rule, for three-class problem.</i>	<i>98</i>
<i>Table 4.32: Highest five performances from fusion combinations of outputs from all data sources using product rule, for three-class problem.</i>	<i>99</i>
<i>Table 4.33: Highest five performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using sum rule, for three-class problem.</i>	<i>100</i>
<i>Table 4.34: Highest five performances from fusion combinations of outputs from all data sources using sum rule, for three-class problem.</i>	<i>101</i>
<i>Table A.1: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using SMV rule, for two-class problem.</i>	<i>115</i>
<i>Table A.2: Additional performances from fusion combinations of outputs from all data sources using SMV rule, for two-class problem.</i>	<i>115</i>
<i>Table A.3: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using WMV rule, for two-class problem.</i>	<i>116</i>
<i>Table A.4: Additional performances from fusion combinations of outputs from all data sources using WMV rule, for two-class problem.</i>	<i>116</i>
<i>Table A.5: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using product rule, for two-class problem.</i>	<i>117</i>

<i>Table A.6: Additional performances from fusion combinations of outputs from all data sources using product rule, for two-class problem.</i>	<i>117</i>
<i>Table A.7: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using sum rule, for two-class problem.</i>	<i>118</i>
<i>Table A.8: Additional performances from fusion combinations of outputs from all data sources using sum rule, for two-class problem.</i>	<i>118</i>
 <i>Table B.1: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using WMV rule, for three-class problem.</i>	 <i>119</i>
<i>Table B.2: Additional performances from fusion combinations of outputs from all data sources using WMV rule, for three-class problem.</i>	<i>119</i>
<i>Table B.3: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using product rule, for three-class problem.</i>	<i>120</i>
<i>Table B.4: Additional performances from fusion combinations of outputs from all data sources using product rule, for three-class problem.</i>	<i>120</i>
<i>Table B.5: Additional performances from fusion combinations of outputs from data sources of C_Z, F_Z and P_Z electrodes using sum rule, for three-class problem.</i>	<i>121</i>
<i>Table B.6: Additional performances from fusion combinations of outputs from all data sources using sum rule, for three-class problem.</i>	<i>121</i>

CHAPTER 1

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease characterized by progressive cognitive deterioration along with neuropsychiatric symptoms. The affected individual shows symptoms like declining ability to perform activities of daily living and behavioral changes. AD is the most common type of dementia. Currently, there are an estimated 24 million people worldwide with dementia, and is expected to increase to about 81 million people by 2040 [1]. An estimate of 4.5 million Americans suffer from AD, and this number is expected to reach 12–16 million by 2050 [2]. The AD usually begins after the age of 60, and the risk increases with age. While younger people may also get AD, it is much less common. About 5% of men and women ages 65 to 74 have AD, and nearly half of those age 85 and older may have the disease [3]. AD not only affects its victim but also their caregivers, both socially and financially. The average lifetime cost of care for an individual with AD is estimated at \$174,000, whereas national direct and indirect annual costs of caring for individuals with AD are estimated to be over \$100 billion [2].

1.1 BIOLOGICAL ASPECTS OF ALZHEIMER'S DISEASE

The physiology of AD relates to the disturbance of the intricate messaging network in brain by means of two abnormal structures: amyloid plaques and neurofibrillary tangles. Human body normally produces protein fragments which are named by a general term –

amyloid. Beta-amyloid is a toxic molecule, which originates from a normal protein called amyloid precursor protein (APP). In a healthy brain, these protein fragments would be broken down and eliminated. In AD, these fragments accumulate to form hard, insoluble plaques. The beta-amyloid plaques block neurons from sending messages, and hence disturbing the brain's communication network [4, 5].

Nerve cells are made of structures called microtubules, which helps in transport of nutrients and other important substances from one part of the nerve cell to another. A protein called tau plays vital role in internal support structure of microtubules. In AD patients there is an abnormality in the tau protein, which causes the microtubules to collapse and form insoluble twisted fibers called neurofibrillary tangles. Further growth of such tangles results in the decomposition of the neuron [5]. The amyloid plaques and neurofibrillary tangles are illustrated in Figure 1.1.

Figure 1.1: Illustration of amyloid plaques and neurofibrillary tangles[§] [5]

[§] Illustration reproduced by permission

There is an overall shrinkage of brain tissue as AD progresses. As a result the ventricles, the chambers within the brain that contain cerebrospinal fluid, are noticeably enlarged. Figure 1.2 shows the cross-section view of brain at different stages of disease. In the early stages of AD, short-term memory begins to decline when the cells in the hippocampus degenerate. The hippocampus is a part of the limbic system which is primarily responsible for our emotional life, and formation of memories. As hippocampus plays important role in converting short-term memory into long-term memory, the ability to perform routine tasks also decline gradually. As AD spreads through the cerebral cortex (the outer layer of the brain), judgment declines, emotional outbursts may occur, and language is impaired. Progression of the disease leads to the death of more nerve cells and subsequent behavior changes, such as wandering and agitation. In the final stages the ability to recognize faces and to communicate is completely lost. Patients lose bowel and bladder control, and eventually need constant care. This stage of complete dependency may last for years before the patient dies [6]. Average duration of the disease is approximately 7–10 years, although cases are known where reaching the final stage occurs within 4–5 years, or in some reported cases up to 22 years [7]. The condition of disintegration of brain at its earliest stage can only be identified by examining the brain tissue under a microscope, leaving autopsy as the only method for positive diagnosis. Certain brain scan techniques like magnetic resonance imaging (MRI) and computed tomography (CT), which measures brain volume have been implemented to diagnose the disease [8, 9]. However these techniques are not extremely successful, as they are not reliable in diagnosis of the disease in its early stage, compared to declining brain volume in normal aging.

Figure 1.2: Cross section view of brain at different stages of AD[§] [6].

1.2 DIAGNOSIS

The cause of AD is unknown and there are no medical tests available to conclusively diagnose AD pre-mortem. Expert clinicians who specialize in memory disorders can now diagnose AD with an accuracy of 80–90% [10]. AD diagnosis is made primarily on the basis of history, clinical observation, and tests of memory and intellectual functioning over a series of weeks or months. These tests include various physical tests like blood tests, neuroimaging, etc. being performed to rule out alternate disease diagnoses.

Interviews with family members, and/or caregivers are extremely important in the initial

[§] Illustration reproduced by permission

assessment. Such interviews also provide important information on the affected individual's functional abilities, which are key indicators of the significance of the symptoms and the stage of dementia.

Initial suspicion of dementia like AD may be strengthened by performing neuropsychological tests, such as Clinical Dementia Rating (CDR), Mini-Mental State Examination (MMSE), Severe Impairment Battery (SIB), etc. These tests are part of the National Institute of Neurological and Communicative Disorders and Stroke - Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable AD [11]. Results of such neuropsychological tests can be helpful in establishing the presence and severity of dementia. Even after these longitudinal – process of elimination – tests performed to diagnose AD, the diagnosis is still referred as “probable” or “possible” AD. The definitive diagnosis of AD must still await microscopic examination of brain at autopsy.

Diagnostic markers plays important role in physiological systems, which helps in determining whether a subject is likely to develop AD. The incubation period of the disease is still unknown, but the damage begins several years before the obvious symptoms are detected [12]. A study by Goldman *et al.* showed that non-demented individuals who are autopsied had sufficient plaques seen during their postmortem analysis, although there was no cognitive impairment or decline in mental and physical abilities before death [13]. Hence, damaging plaques and tangles may begin to form several years before the symptoms of the disease appear. Considering that AD is a slowly progressing late life disease with an average symptomatic onset age of 74, and an average duration of 7 years, a biomarker would not only be an efficient tool for early

diagnosis, but could also assist in identification of early stage AD patients. Identifying a population of early stage AD patients could provide a valuable pool of individuals that could assist in the development of new medical treatments for intervention in the progress of AD. Initiation of early effective treatment could delay the time when the major symptoms of disease appear. The slow progression of AD may add years to their life and cause their death by other old age disorders, or at least a non-dementia death [14, 15]. This could eliminate the years associated with the poorest quality of life, and reduce the full-time care expenses of the final years of the illness.

The standard AD diagnostic procedures, including the clinical evaluation and the current biomarker techniques, are conducted at major university hospitals and research clinics. However, as discussed above, it is not unusual for certain patients to not show any symptoms during clinical evaluations. Moreover, even when they do, these symptoms can only be validated after lengthy longitudinal studies and typically only at university/research hospitals. Most patients are evaluated at community health clinics where the expertise and accuracy of detecting AD is uncertain. A recent study by Health Maintenance Organization – based physicians, reported an average overall accuracy of about 75% [16]. Consequently, there is an urgent need for such biomarkers, which can detect the disease at early stage; and are clinically useful, accurate, non-invasive and cost-effective; so that these procedures can be made be available at community health clinics where most people get their initial – sometimes only – evaluation.

1.3 SCOPE OF THE THESIS

In this work, analysis of event related potentials (ERPs), obtained from the electroencephalogram (EEG) was done, to obtain a feasible biomarker for the early diagnosis of Alzheimer's disease. The ERP signals were obtained using an oddball paradigm, and recorded with various electrodes using a standard – International 10/20 electrode configuration system. The analysis of the ERP signals was accomplished through a wavelet decomposition which gave the coefficient sets in successive spectral bands. Coefficients of each spectral band were used to train and test a neural network ensemble of classifiers, based on stacked generalization algorithm. The ensemble decisions were then combined using various fusion techniques to increase the overall generalization performance of the system. In this study, a two-class diagnostic problem (probable AD patients and cognitive normal subjects) and a three-class diagnostic problem (mild-moderate AD patients, mild AD patients and cognitive normal subjects) were addressed by using individual data sources (specific electrode – stimuli – frequency band combinations), and their decision level data fusions.

1.4 ORGANIZATION OF THE THESIS

A detailed literature review of the background on various methods for early diagnosis of AD is described in Chapter 2. The methods used in this study are provided in detail in Chapter 3. These methods include the acquisition of ERP signals through oddball paradigm, the wavelet transform for the feature extraction, the ensemble system based on the stacked generalization algorithm, different fusion techniques, and evaluation of diagnostic performance measures. Chapter 4 provides implementation details and results

on various experiments based on the methods in Chapter 3. The ensemble decisions from individual data sources (specific electrode – stimuli – frequency band combinations) were combined using different data fusion techniques, and the performances are described in detail in Chapter 4. Chapter 5 consists of discussion of results along with conclusions about the study, sources of error, and suggestions for future work.

CHAPTER 2

BACKGROUND

Several studies have been conducted to obtain specific biomarkers for the diagnosis of AD [17, 18]. These studies can be divided into three categories: analysis of chemical biomarkers, imaging techniques and analysis of electroencephalogram.

2.1 CHEMICAL BIOMARKERS OF AD

Some biomarkers are based on chemical analysis of fluids obtained from human body. Spinal fluids have been evaluated for AD diagnosis. One of the methods is to obtain cerebrospinal fluid tau protein (CSF-tau). The increase in amount of tau protein has been linked to AD. But this kind of abnormality can also be found in other dementias, like vascular dementia [19]. Beta-amyloid-derived diffusible ligands (ADDLs), made from beta-amyloid molecule, obtained from spinal fluid have also been linked to AD [20]. Cerebrospinal fluid samples for both of these methods are collected using a procedure known as spinal tap. This is done by lumbar puncture, which can be very painful, only available at research hospitals, and could be extremely costly. There is another easily obtained chemical biomarker for AD, called F2-isoprostane. F2-isoprostane is obtained from urine analysis [21]. However, the link of this method with AD is not as strong as CSF analysis, or analysis of brain tissue at postmortem.

2.2 IMAGING TECHNIQUES

Alzheimer's disease is a condition in which disintegration of brain tissue occurs, hence several imaging techniques which measures brain volume have been implemented to diagnose the disease. Methods such as magnetic resonance imaging (MRI) scans, which capture images of the lesions in the brain [8, 9], and detection of reduced N-acetyl aspartate by magnetic resonance spectroscopy [22], have been used for classification of Alzheimer's disease. Data from such studies show a relationship between decline in brain volume and cognitive performance. Moreover, the brain volume is decreased in both AD and healthy individuals with age, but the rate of change is more than 2.5 times faster in AD patients [8, 9]. Another study was done by Babiloni *et al.*, which used a low resolution brain electromagnetic tomography (LORETA) to compare mild AD with vascular dementia, and cognitively normal subjects. The low resolution brain electromagnetic tomography (LORETA-KEY) is a functional imaging technique that gives a 3-D distribution of electric neuronal activity, modeling the cortex as a collection of volume elements called voxels. Result showed the promising correlation between the diseases, and the stages of severity [23]. However, these imaging procedures are expensive, and their clinical usefulness and links with AD have not been studied in detail.

2.3 ELECTROENCEPHALOGRAM

Abnormalities in the brain tend to disrupt the brain's electrical signals and can be detected through electroencephalographic (EEG) signals. Electroencephalography is the neurophysiologic measurement of the electrical activity of the brain by recording from electrodes placed on the scalp or, in special cases, subdurally on the cerebral cortex. The

resulting tracings are known as an electroencephalogram (EEG) and represent an electrical signal from a large number of neurons. EEGs in various forms have been a useful tool for monitoring and diagnosing certain neurological disorders, such as epilepsy, syncope (fainting), sleep disorders, coma, and brain death.

Scalp recording technique is used for most EEG studies. The internationally standardized 10-20 system is usually employed for electrode placement for recording the EEG signals. In this system, 21 electrodes are placed on the surface of the scalp. Figure 2.1(a) shows the left view; and Figure 2.1(b) shows the top view of the electrode placement. The positions are determined by reference points: nasion – which is the delve at the top of the nose, level with the eyes; and inion – which is the bony lump at the base of the skull on the midline at the back of the head. From these points, the skull perimeters are measured in the transverse and median planes. Electrode locations are determined by dividing these perimeters into 10% and 20% intervals; hence it is named 10/20 electrode placement system. Three other electrodes are placed on each side equidistant from the neighboring points. Some of the nomenclatures for electrode placements are: A – ear lobe, C – central, Pg – nasopharyngeal, P – parietal, F – frontal, Fp – frontal polar and O – occipital [24, 25].

In addition to the 21 electrodes of the standard international 10-20 system, intermediate 10% electrode positions may be used. The locations and nomenclature of these electrodes are standardized by the American Electroencephalographic Society as in Figure 2.1(c). In this recommendation, four electrodes have different names compared to the 10-20 system; these are T_7 , T_8 , P_7 , and P_8 [26]. These electrodes are highlighted in the Figure 2.1(c).

Figure 2.1: The international 10-20 system: (a) left view and (b) top view; (c) location and nomenclature of the intermediate 10% electrodes[§] [27].

The EEG signal can be divided into several spectral bands which are associated with different brain functions [28, 29]. The five main frequency bands are as follows:

- The delta band (less than 4 Hz) can be seen in deep sleep stages, and is absent in awake healthy adults. It has been linked with certain encephalopathies and underlying lesions.
- The theta band (4–8 Hz) is associated with drowsiness, childhood, adolescence and young adulthood. It can be seen during hypnagogic states such as trances, hypnosis, deep day dreams, lucid dreaming and light sleep.

[§] Illustration reproduced by permission

- The alpha band (8–12 Hz) originates from the occipital lobe during period of relaxation, with eyes closed but still awake. Conversely alpha waves are attenuated with eyes open, as well as by drowsiness and sleep.
- The beta band (13–30 Hz) is associated with active, busy or anxious thinking and active concentration.
- The gamma band (greater than 30 Hz) is believed to be involved in high mental activities like perception, problem solving, fear and consciousness.

Different studies have shown several abnormalities in the EEG of AD patients. Jeong studied the EEG dynamics in patients with AD, showed that there was relative slowing and decreased complexity in EEG signal, and reduced degrees of functional connectivity in AD patients [30]. Mildly affected AD patients had increased theta activity compared to control groups, and more severely affected patients showed decreased alpha activity and increased delta activity [31]. AD patients have found to have increased delta and theta power but decreased alpha and beta mean power. The severity of dementia has also been linked with EEG rhythms [23, 32].

On other hand, EEG signals are said to be affected by coexisting medical illnesses, varying levels of anxiety and drowsiness during the recordings, hence amplifying the difficulty in identifying AD patients. Hence, EEGs were traditionally not included for the diagnosis of dementia, like AD [30].

In certain other studies, different protocols have been used for EEG acquisition, where the patients are exposed to a sensory stimulus in order to elicit a particular response, known as the event-related potentials (ERPs). There are three modalities typically used as stimuli: auditory, visual and somatosensory. For the auditory modality,

the stimuli are single tones of a preset frequency, or clicks with a broadband frequency distribution. For the visual modality, stimuli are produced by a single light or sometimes by the reversal of a pattern such as that of a checkerboard. For the somatosensory, peripheral nerves are stimulated using electrical stimuli. Bimodal stimulation, used in different experiments, is the combination of the visual and auditory modalities [33].

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

- No-task evoked potentials paradigm involves the subject in a relaxed state, instructed to perceive the stimuli without performing any task.
- The oddball paradigm involves two different stimuli presented in a pseudorandom order. The oddball tone (or target tone) is presented randomly in a series of frequently occurring (standard) tones. The standard tone is presented in 75–80% of the trials, and the oddball stimuli in the remaining 20–25% of the trials. The oddball stimulus is usually a different frequency from the standard, set far enough apart to be distinguishable from the frequent stimulus [33, 34].
- The “three-stimuli” paradigm involves a typical oddball paradigm with novel or “distracter” stimuli randomly added. These novel tones consist of a disruptive sound such as dog barking, color forms, etc. that disrupt the regular oddball paradigm routine [34].

Yamaguchi *et al*, 2000, developed a variation of this paradigm with the use of novel tones consisting of 60 unique environmental sounds, recorded from Disney movies

and edited to be 200ms in duration. In their protocol, frequent (standard) tones occur 65% of the time, oddball tones 20%, and the novel tones occur 15% of the time. Again, the subjects are asked to respond only to the oddball tones by performing a simple task defined at the onset of the experiment. This type of experiment is performed to differentiate between different types of dementia [35].

ERPs have the potential to provide a diagnostic biomarker for AD. Acquisition of ERP signal is a well-established, reliable, non invasive practice, and can be made available to local community clinics. Raw EEG signals are said to be affected by coexisting medical illnesses, varying levels of anxiety, drowsiness during the recordings, and other effects. On the other hand, components of ERP signals, which are based on cognitive reactions, are believed to be almost insensitive to these conditions [36–39].

ERPs are a series of positive and negative peaks that occur in response to a specific event to which the subject is asked to respond (detecting the target tone). P3 or P300 is a positive peak occurring around 300ms that is of particular interest. The P300 has been shown to occur only in response to oddball tones and has also been associated with mental activities. The latency and the amplitude of the P300 component have been shown to be related to age and the mental ability of the individual [38, 40]. The P300 is believed to directly reflect currents triggered by cortical postsynaptic potentials, and seems to be primarily generated in the temporo-parietal cortex, which is affected in dementias like AD [41]. Polich *et al.* have shown that increased latency and decreased amplitude of P300 is associated with AD [42, 43].

The P300 component, while useful to show statistical differences between AD and normal groups, is not discriminatory enough – on its own – to be able to identify

individual patients as having AD. ERPs from cognitively normal subjects may have a delayed or reduced P300, whereas that of AD may still have significant P300. This is explained in Figure 2.2, where Figure 2.2 (a) and (b) shows expected P300 behavior from normal and AD patients; whereas Figure 2.2 (c) and (d) shows that this is not true in all the cases. Hence primary thesis of this work is that instead of analyzing P300 specifically, frequency range in which P300 and other ERP components occur should be explored.

Figure 2.2: (a) & (b) Expected P300 behavior from normal and AD patients; (c) & (d) not all individual cases follow this expected behavior [44].

2.4 PREVIOUS WORK

Several studies have been carried out which analyses the components of EEG/ERP in relation with AD and other dementia. Most of these studies are based on statistical analyses of the EEG/ERP components. Lately very few studies have been carried out which works on the diagnosis problem, i.e. to classify between AD and/or other non-AD dementia with normal patients. Some studies which do the time domain analyses of the EEG/ERP signals are explained below:

Polich and Hoffman performed a test on a cohort of three groups: young adults, elderly normal subjects, and AD patients in the early stages of disease with 12 subjects in each group. ERPs were evoked by both auditory and visual oddball tasks of varied difficulties, and P300 components were analyzed. The results showed a great differentiation between AD and controls with relative easy visual tasks. P300 amplitudes were altered strongly for all three subject groups by difficult tasks. There was significant decline in P300 amplitudes and latency in visual task as compared to auditory task. Hence P300 analysis was proved to be a significant tool for assessment of AD [43].

Yamaguchi *et al.* performed a study with 16 AD patients, 16 vascular dementia (VaD) patients and 18 age-matched normal patients. The “three-stimuli” paradigm which involves a typical oddball paradigm with novel tones was utilized to evoke ERPs. In this study, the amplitude, latency and scalp topography of P300, for target and novel tones were compared among three groups of patients using ANOVA (a statistical analysis of variance test). The relationship between P300 measures and intelligence scores were evaluated by correlation analysis. The results indicated that the amplitude, latency and scalp topography of the target P300 were affected by both AD and VaD. However, the

amplitude of the novel P300 was significantly reduced in VaD, but not in AD, and the scalp topographies were different in the 3 groups. The amplitude was maximal at frontal sites in controls, at central sites in AD, and at parietal sites in VaD. The target P3 latency was prolonged in both AD and VaD, whereas the novelty P3 latency was only prolonged in VaD. Hence the study concluded that AD can be discriminated satisfactorily from VaD and normal [35].

Benvenuto *et al.* performed a study with 15 AD patients and 17 control subjects. ERPs evoked by brief flashes of light were analyzed using statistical k nearest neighbor technique. The results demonstrated that combinations of selected ERP time segments across different electrodes contained certain features which discriminated AD from control subjects with high sensitivity and specificity [45].

Phillips *et al.* investigated the sensitivity of P300 event-related potentials recorded during memory-demanding tasks. 14 AD patients, 16 mild cognitive impairment (MCI) patients and 15 age and education-matched individuals were tested on various memory measures. The ability of neuropsychological (delayed verbal memory), neuroanatomical (MRI-based hippocampal volume), and electrophysiological (memory search P300 amplitude) memory measures were explored to distinguish between the three subject groups using discriminant function analyses. Results showed that P300 amplitude was reduced in AD patients at all levels of memory load. Delayed verbal memory performance best separated the AD patients from MCI and control subjects, while delayed verbal memory and hippocampal volume best discriminated MCI subjects from control. Hence it supports the use of neuropsychological and neuroanatomical measures in diagnosing different dementia [46].

Spectral analysis has been widely used for the analysis of EEG and ERP signals to discriminate different types of dementia, and normal patients. Power spectral density and coherence analysis of EEG and ERP signals have been explored in following studies:

Signorino *et al.* did a statistical comparative study between two subgroups of patients with dementia. The study involved 22 AD patients and 36 vascular dementia (VaD) patients. EEG data were collected in resting eyes closed and again with eyes open condition. The analysis involved EEG spectral profile in low frequency powers and dominant activity in the 6.5–12 Hz band. The aim of the study was to discriminate the two groups by analysis of EEG spectral contents, demographic figures and mental deterioration scores. Results showed that spectral analysis can be a useful tool in differentiating two types of dementia, when a dynamic EEG (the power ratio of recordings during eyes closed and open) was processed. A more severe mental deterioration was found in the AD patients as compared to VaD [47].

Pucci *et al.* performed a study with 25 AD patients, 25 non-AD demented patients and 50 normal control subjects. EEG data were collected in resting – eyes closed/eyes opened – conditions. The data was subject to spectral analysis where the power was computed in five 5.5 Hz bands in the range of 1–28.5 Hz. Five unconventional frequency bands were explored: 1–6.5 Hz (B1), 6.5–12 Hz (B2), 12–17.5 Hz (B3), 17.5–23 Hz (B4), and 23–28.5 Hz (B5). The absolute and relative powers were calculated for all frequency bands, followed by derivations, and a logarithmic transform. They found a lack of dominant activity in the 6.5–12 Hz band in a significant number of the AD patients. In conclusion, the study provided further evidence that EEG spectral analysis

can allow the identification of a particular group of AD patients, from other AD patients, or other types of dementia [48].

Another study by Pucci *et al.* evaluated the relationship between various EEG spectral parameters and the age of onset of AD. The study involved EEG recordings during wakefulness with eyes opened and again with eyes closed. The test involved 150 AD subjects and 52 normal controls. The AD patient group was subdivided into three groups according to age of onset (OA): (1) $OA \leq 60$, (2) $61 < OA < 69$, and (3) $OA \geq 70$. The frequencies explored consisted of five 5.5 Hz bands in the range of 1–28.5 Hz as in the previous study. EEG spectral profiles were obtained from plots of the power spectral density values and were considered substitutes for the absolute powers. The results confirmed that this method can very well discriminate between AD patients and normal controls. In early AD, the EEG spectrum was characterized by lack of a dominant peak in the 6.5–12 Hz frequency range. The age of onset was found to correlate inversely with the 1–6.5 Hz relative powers and positively with the 6.5–12 Hz relative power. Evidence was shown for EEG changes due to early onset of AD when all three subgroups were compared [49].

Rodriguez *et al.* performed a study to determine whether a parameter such as the power spectral profile may be useful in the detection of AD. 48 patients with probable AD were recruited and divided into four groups based on global deterioration scores (GDR), which is a scale for assessment of dementia. 18 healthy control subjects were also recruited for the study. The EEG was recorded with eyes closed, in a resting state. The spectral profile for each patient was expressed by the relative power of seven different frequency bands (2–3.5 Hz, 4–5.5 Hz, 6–7.5 Hz, 8–9.5 Hz, 10–11.5 Hz, 12–13.5

Hz, and 14–22.5 Hz). Overall, the findings showed that 4–5.5 Hz and 10–11.5 Hz band powers displayed the highest significance statistically in differentiating between the groups. They concluded that spectral morphology significantly changes with the progression of AD, and since it was widely available, may be useful in a clinical setting for staging AD [50].

Hibino *et al.* performed analysis on EEG data collected (resting – eyes closed – condition) from 26 AD patients and 17 normal control subjects but with multiple EEG recordings for some of the AD patients, giving a total of 35 recordings for AD patients and the 17 from the control group. Power spectra was calculated using the Fast Fourier transform and separated into 9 frequency bands: delta (3–4 Hz), theta 1 (4–6 Hz), theta 2 (6–8 Hz), alpha 1 (8–9 Hz), alpha 2 (9–11 Hz), alpha 3 (11–13 Hz), beta 1 (13–15 Hz), beta 2 (15–20 Hz), and beta 3 (20–25 Hz). The relative power of each frequency band was calculated and the results for the alpha waves were averaged. The delta and beta waves were not used for the analysis because influences due to eye movement in the delta waves were not negligible and no effects on the beta waves due to cognitive deterioration have been reported. The theta 1, theta 2 and averages of the alpha waves were used as input into a fuzzy neural network (FNN), which used fuzzy inference to make a decision. Parameters were defined for two models, one for discrimination between AD and control subjects and the other to estimate MMS (Mini-mental state) scores into two groups, above 24 and less than or equal to 24. The authors concluded that this method provided promising results to stage the level of AD as well as its diagnosis [51].

A study by Babaloni *et al.*, conducted for the purpose of mapping, (i) the distributed EEG sources specific for mild AD compared to vascular dementia (VaD), or

to normal elderly subjects, and (ii) the distributed EEG sources sensitive to mild AD at different stages of severity. EEG was recorded using 10-20 electrode configuration from 19 electrodes in resting – eyes closed – condition. FFT-based power spectrum analysis was performed to compute the power density of the different frequency bands: delta (2–4 Hz), theta (4–8 Hz), alpha 1 (8–10.5 Hz), alpha 2 (10.5–13 Hz), beta 1 (13–20 Hz), and beta 2 (20–30 Hz). This study used LORETA-KEY to compute the power spectral profiles of cortical EEG sources. For study (i) mild AD showed higher decline in central, parietal, temporal and limbic alpha 1 sources to than in VaD and normal. Occipital alpha 1 source showed a significant decline in mild AD compared to VaD. The magnitude of the delta sources in VaD was greater than mild AD. For study (ii) showed a lower power of occipital alpha 1 sources in severe cases of AD patients. They concluded that the LORETA approach was a valid potential method for illustrating the power spectrum profiles of the level of cortical EEG sources. Further studies should be performed to evaluate the clinical usefulness of this method in early diagnosis, staging, and therapy monitoring of the disease [23].

Besthorn *et al.* recorded EEG signals from 50 AD patients and 42 normal controls with resting – eyes closed – condition from 17 electrodes. This study utilized coherence, which is the normalized cross-power spectrum per frequency of two signals that are recorded simultaneously from different electrodes on the scalp. Coherence was computed for 6 frequency bands: delta (1.5–3.5 Hz), theta (3.5–7.5 Hz), alpha 1 (7.5–9.5 Hz), alpha2 (9.5–12.5 Hz), beta 1 (12.5–17.5 Hz), and beta 2 (17.5–25.0 Hz), and the spatially averaged coherence was calculated. The results showed differences between the AD patients and controls in all frequencies except for the delta band. The effect was

most pronounced in the frontal and central derivations of the theta, alpha, and beta frequency bands. The authors indicated that the results showed the effects of neuronal loss and neocortical disconnection [52].

Locatelli *et al.* analyzed EEG signals from 10 probable AD patients and 10 normal controls. EEG was recorded in resting – eyes closed – condition from 16 electrodes with ears linked as reference. 50 epochs with duration of 1 second were used for analysis. Coherences for the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz) and beta (12–30 Hz) bands were calculated as the mean coherence values of the 50 epochs selected. Both “local” and “far” coherences were calculated to determine whether the changes due to AD were associated with the impairment of the short or long axonal fibers. Their findings showed a decreased coherence for EEG high frequencies in AD. For the regional coherence, alterations were more pronounced over the left temporo-parietal-occipital areas. For the long distance coherence, a decrease between pairs of “far” electrodes confirmed that long axons were more compromised than short ones in AD. Results also showed that, for the AD group there was a significant decrease of alpha inter-hemispheric coherence. The coherence decrease of alpha and beta bands began in the earliest stages of the disease, while the increased slow band coherence was evident in the later stages [53].

Hogan *et al.* examined memory-related EEG power and coherence over temporal and central recording sites in 10 mild AD patients and 10 normal controls. A memory test was performed during EEG data collection with 5 electrodes (Fz, Cz, Pz, T₃ and T₄, with auditory reference). The following spectrum bands were defined for further analysis: delta (0.5–3 Hz), theta (3–5 Hz), lower alpha 1 (5–7 Hz), lower alpha 2 (7–9

Hz), upper alpha (9–11 Hz), beta 1(15–20 Hz), and beta 2 (20–25 Hz). The frequency band at 12–14 Hz was not analyzed. Spectral power and coherence for the named frequency bands were analyzed. For both AD patients and controls, the power in the delta to upper alpha range was greater over the three central sites as opposed to the two temporal sites. For the beta frequency range, the opposite occurred, where the power was greater over the two temporal sites. This may have caused by the reference electrodes detecting activity from the underlying source, thus inflating the coherence. AD patients were found to have more power in lower alpha 1 than controls. This difference occurred primarily in central as opposed to temporal sites. For coherence, the upper alpha band produced significant group differences. The authors suggested that spectral power and coherence measures may offer insight into underlying cortical disruption [54].

Brunovsky *et al.* examined EEGs from 38 AD patients with varying degrees of dementia (mild, moderate, severe) using EEG spectra and coherences. Spectrum analysis was performed for 6 frequency bands: delta (0.5–3.5 Hz), theta (3.5–7.5 Hz), alpha 1 (7.5–9.5 Hz), alpha 2 (9.5–12.5 Hz), beta 1(12.5–17.5 Hz), and beta 2 (15.5–25 Hz). The mean amplitudes (square roots of the power values) were calculated for each band. For each of the 6 frequency bands, intrahemispheric (“local”) coherences were calculated for the anterior and posterior brain regions, and the interhemispheric (“far”) coherences consisted of pairs of distant electrodes on either side of the head. The results of the statistical evaluation indicated decreased “far” coherences in the alpha frequency bands to be highly correlated with the degree of dementia. This indicated that, the lower the coherence in this range, the higher the degree of dementia for the patient. The authors

concluded that a combination of EEG spectral values and coherence may be used to supplement clinical evaluations [55].

Since ERP is a time and frequency varying signal, a multi-resolution time – frequency based analysis, like wavelet transform have been widely used over the past years. Several studies focusing on the structural analysis of the components of ERPs, like P300, etc. have been done based on this approach. Some of them are briefly explained below:

In a study by Ademoglu *et al.*, quadratic B-spline wavelets were used in the analysis of pattern-reversal visual evoked potentials of normal and demented patients. This study was primarily performed to further explore the components of the ERPs, and not to distinguish between classes. The wavelet was shown to capture particular events occurring within the ERPs and was found suitable for further applications as a discrimination technique for cases of clinical problems associated with peak latency [56].

Demiralp *et al.* performed analysis of EEG signals from 10 healthy volunteers, aged 18-25 years, obtained through the use of the oddball paradigm. The purpose was to investigate the effectiveness of the wavelet transform in the analysis of ERPs. The transform was applied to both averaged and single-sweep responses. The results showed that the delta, theta, and alpha processing levels have some involvement in time during the oddball task. Analysis involving the topography pattern of the electrodes indicated that the wavelet coefficients reflect functional components at specific regions of the brain. The authors concluded that the wavelet transform was a powerful method for decomposing ERPs on the basis of frequency, temporal position and scalp topography. They suggested that signals in the same frequency ranges, but with specific temporal

position, and/or scalp topography, can be distinguished from each other. The processes overlapping in time can be separated by frequency, and/or scalp topography. The wavelet transform proved to be very useful as it was capable in separating various simultaneously occurring frequencies that could not be detected in the time domain signals [57].

In another study by Demiralp *et al.*, ERP study was done on 10 healthy volunteers between 18 and 55 years of age using an auditory oddball paradigm. The goal of this study was to classify single ERPs according to their characteristic response properties. Quadratic b-spline wavelets were used for their near optimal time-frequency localization properties. Wavelet decomposition yielded six sets of coefficients, each belonging to different frequency bands, 62–125 Hz (high gamma), 31–62 Hz (gamma), 16–31 Hz (beta), 8–16 Hz (alpha), 4–8 Hz (theta), and 0.5–4 Hz (delta). The grand average of the target and non-target response were calculated and compared. The delta response amplitude was found to be increased for the target response. The theta and alpha frequency ranges showed increased amplitude and prolonged target responses. The P300 response was detected in single trials using the wavelet coefficients for the delta frequency range. Instead of using single trials, the averaging technique should be used as it the most reliable technique for ERP estimation [58].

Quiroga *et al.* used quadratic B-spline wavelets and Fourier transform to analyze evoked potentials and compared results of the analysis. The evoked potentials were elicited through the use of two techniques: pattern visual (to review alpha band) and bimodal (to review gamma band). The results for analysis of the alpha band indicated that they were best localized in the occipital region. The gamma amplitude was the

largest in response to bimodal stimulation as compared to individual auditory or visual stimulation. In the comparison of the wavelet transform and the Fourier transform, the researchers found that wavelets had two advantages: (1) The wavelet transforms were capable of analyzing non-stationary signals, which were critical to avoid in the results of the Fourier transform, as the brain signals are highly non-stationary. (2) Better time – frequency resolution was achieved, so the ERP responses can be analyzed closely [59].

For AD specific diagnosis, there has been very few studies that use an appropriate time – frequency analysis, such as discrete wavelet transform, followed by an automated classifier like neural network. Some of them are briefly described below:

Petrosian *et al.* explored wavelet transform by using specifically designed and trained recurrent neural networks (RNNs) to discriminate between EEGs of 10 mild AD patients and 10 age-matched control subjects. The EEG recordings were taken during resting state without the use of a paradigm. The Daubechies 4 wavelet was chosen due to its good localizing properties in the time and frequency domain. The RNNs used in the study belonged to a type of discrete-time recurrent MLPs, which had better temporal capabilities than that of a regular feed-forward MLP, and was capable of representing and encoding strongly hidden states. Training on 3 AD subjects and 3 controls and testing on the remaining cohort yielded performance that was better than chance with 80% sensitivity and 100% specificity. Five out of seven of the AD subjects were correctly classified. The authors suggested that their approach may be extended to include more classes such as other types of dementia [60].

Polikar *et al.* applied the Daubechies 4 wavelet to EEG data collected from 14 normal subjects and 14 subjects diagnosed with probable AD. The ERP response in the

oddball paradigm was analyzed to determine if the use of the wavelet transform was feasible for the detection of AD with a multilayer perceptron (MLP) neural network. Half of the signals (7 AD and 7 normal) were used for training while the rest were used for testing the network. The generalization performance of the network was 93%. The results confirmed that the approach was feasible for classifying ERPs, but a more diverse database with a larger variety of signals, would help better in establishing statistical generalization for the problem [61].

Yagneswaran *et al.* performed a study to develop a feasible algorithm to differentiate between AD and control subjects. The study cohort consisted of 9 AD patients and 10 control subjects. The power frequency and wavelet analyses were done on EEG signals recorded from 9 channels using standard International 10-20 system. These characteristics were used to train Linear Vector Quantization (LVQ) based neural networks to classify the AD and control groups. The results demonstrated the feasibility of this approach as a potential and effective tool for early AD diagnosis [62].

Jacques *et al.* did analysis on a cohort of 10 probable AD patients and 18 cognitively normal controls, which were recruited from elderly patients over 60 years of age. Feature extraction of ERPs from P_Z node was done by wavelet transforms using two types of wavelets: Daubechies 4 (db4) and Quadratic b-spline wavelets. Five-fold cross validation was performed using the Learn++ algorithm yielding a performance of $85.3 \pm 5.2\%$ for the b-splines (using 15 MLP type classifiers, all with 30 hidden layer nodes and an error goal of 0.15), and $90 \pm 9.2\%$ for the db4 wavelet (5 classifiers, all with 30 hidden layer nodes and an error goal of 0.1). The sensitivities were 60% and 70%, specificities

were 100% and 100%, and the positive predictive values were 100% and 80%, for b-splines and db4 wavelets, respectively [63].

Stepenosky *et al.* performed a study on a cohort of 25 probable AD and 27 cognitive normal subjects to develop a feasible diagnostic algorithm. In this setup, target and novel responses from P_Z, C_Z and F_Z electrodes were analyzed. Feature extraction was done by wavelet transforms using db4 wavelet. Three independent multilayer perceptron (MLP) classifiers were developed and trained using different feature sets. Their results were fused using different combination rules like, weighted majority voting (WMV), sum rule, product rule, decision templates, etc. The best classification performance was obtained by product rule at 90%. The results demonstrated the feasibility of ensemble system based on data fusion, in the early diagnosis of AD [64].

In our previous study, an automated ensemble classification system, based on stacked generalization algorithm was introduced. This algorithm was used to discriminate between probable AD and cognitively normal subjects. This study included 22 probable AD and 22 cognitively normal subjects. The diagnostic performances were obtained on the wavelet features using db4 wavelet from target P_Z responses. Results demonstrated the feasibility of this approach in the diagnosis of AD [44].

A recent study by Polikar *et al.* analyzed a cohort of 28 probable AD patients and 24 cognitively normal individuals, to develop an ensemble of classifiers system based on data fusion approach which combined information from two or more sources. The emphasis was made to develop an ensemble system based on most discriminating information from each data source. The ERPs recorded from P_Z, C_Z and F_Z electrodes, for both target and novel stimuli were decomposed into different frequency bands using

wavelet analysis. Data fusion included developing multiple classifiers trained with strategically selected subsets of the training data from each source, which were then combined using modified weighted majority voting. The fusion of 1–2 Hz band of novel P_Z and 4–8 Hz band of target F_Z responses gave the best classification performance of 79.2%. The results obtained by this approach are promising and better than those achieved at community health clinics/ hospitals [65].

The studies described above differ from the current study in this thesis, due to several reasons: small cohort size, analysis based on statistical approach rather than diagnosis, EEG signals were used rather than ERP signals, and the most important one – the classification not targeted for early diagnosis of the disease.

The overall comparison of previous work is done in Table 2.1.

Table 2.1: Comparison of previous work

Authors	Year	Main Approach	Cohort	Main Results	Ref.
Bestorn <i>et al.</i>	1994	-Compare coherence from resting EEG between groups	50 AD, 42 N	-Statistical analysis -Differences between AD and N in all frequencies except delta band	[52]
Signorino <i>et al.</i>	1996	-Discriminate AD & VaD by analysis of EEG spectral contents, demographic figures, and mental deterioration scores	22-AD, 36 VaD	-Statistical analysis -EEG spectral analysis allows positive differentiation -Severe mental deterioration in AD compared to VaD	[47]
Polich <i>et al.</i>	1998	-Time domain analysis of P300 - ERPs by auditory & visual oddball paradigm with different level task are compared	12 young N, 12 eld. N, 12 AD	-Statistical analysis -Different task altered P300 -Overall Amplitude ↓ & latency ↑ -Visual task showed significant decline in P300 of AD	[43]
Pucci <i>et al.</i>	1998	-Analyze power spectra of resting EEG to discriminate between classes	25 AD, 25 non-AD, 50 N	-Statistical analysis -EEG spectral analysis allows the identification of a group of AD patients from other AD patients, or other types of dementia	[48]
Locatelli <i>et al.</i>	1998	-Compare coherence from resting EEG between groups	10 AD, 10 N	-Statistical analysis - The ‘local’ and ‘far’ coherences were compared -Significant differences in different groups and stages of AD	[53]

Authors	Year	Main Approach	Cohort	Main Results	Ref.
Ademoglu <i>et al.</i>	1998	-Used WT to explore components of pattern-reversal visual ERPs	Dem. & N	-Statistical analysis -QBS wavelet able to relate ERPs components with dementia	[56]
Pucci <i>et al.</i>	1999	-Evaluate relationship between EEG spectral parameters and age of onset of AD	150AD, 52 N	-Statistical analysis -Age of onset was found to correlate inversely with the 1-6.5 Hz relative powers, and positively with the 6.5-12 Hz relative power.	[49]
Rodriguez <i>et al.</i>	1999	-Determine whether a parameter such as power spectral profile may be used in detection of AD using resting state EEG	48-AD, 18 N	-Statistical analysis -4-5.5 Hz and 10-11.5 Hz band powers displayed the highest significance statistically in differentiating between groups	[50]
Demiralp <i>et al.</i>	1999	-Investigated effectiveness of WT in analysis of ERPs	10 N (18-25 yrs)	-Statistical analysis -Delta, theta and alpha waves involved in time during oddball tasks -WT useful for separating simultaneous frequencies	[57]
Demiralp <i>et al.</i>	1999	-Classify single ERPs based on their characteristic responses using WT	10 N (18-25 yrs)	-Statistical analysis -Delta amplitude ↑ for target responses -Theta and alpha amplitude ↑ and latency ↑ for target responses	[58]
Yamaguchi <i>et al.</i>	2000	-Three-stimuli oddball paradigm with novel tones -Amplitude, latency and scalp-topography of P300 compared using ANOVA	16 AD, 16 VaD, 18 N	-Statistical analysis -Amplitude, latency and scalp topography affected by both AD & VaD with significant differences	[35]
Hibino <i>et al.</i>	2001	-Discriminate between AD and controls and estimate MMSE using spectral analysis -theta 1, theta 2 and alpha bands used as input to fuzzy NN	26 AD, 17 N	-Diagnostic analysis -Promising results were obtained for diagnosis and staging of AD using MMSE scores	[51]
Quiroga <i>et al.</i>	2001	-Used WT & FT to review alpha and gamma bands of ERPs, -Compared WT vs. FT analysis		-Statistical analysis -Different bands shows different region localization -WT better than FT for analyzing ERPs	[59]
Petrosian <i>et al.</i>	2001	-WT of EEG to train RNNs and discriminate groups	10 mild AD, 10 N	-Diagnostic analysis -Performance better than chance -80% sensitivity & 100% specificity	[60]
Polikar <i>et al.</i>	2001	-WT of ERPs to train MLP NN and discriminate groups	14 AD, 14 N	-Diagnostic analysis -Generalization performance 93% -Feasible approach	[61]
Benvenuto <i>et al.</i>	2002	-k-nearest neighborhood used to analyze ERPs evoked by brief flashes of light	15 AD, 17 N	-Diagnostic analysis -Certain features discriminated AD from normal with high sensitivity & specificity	[45]
Yagneswaran <i>et al.</i>	2002	-FT and WT analysis of EEG to train LVQ NN and classify diff. groups	9 AD, 10 N	-Diagnostic analysis -Feasible approach for diagnosis of disease	[62]

Authors	Year	Main Approach	Cohort	Main Results	Ref.
Hogan <i>et al.</i>	2003	-Memory related power and coherence over temporal and central electrodes	10 AD, 10 N	-Statistical analysis -AD patients have more power in lower alpha1 compared to N -Coherence of upper alpha bands produced significant diff. in groups	[54]
Brunovsky <i>et al.</i>	2003	-Compare EEG using spectral and coherence analysis	38 AD (mild, mod., sev.)	-Statistical analysis -The 'local' and 'far' coherences compared -Decreased 'far' coherences related to degree of dementia: ↓ coherence indicates ↑the degree of dementia	[55]
Phillips <i>et al.</i>	2004	-Time domain analysis -ERPs evoked by memory demanding task of diff. levels -P300 sensitivity compared	14 AD, 16 MCI, 15 N	-Statistical analysis -P300 amplitude ↓ in AD for all memory task levels -AD, MCI and N were separated by using different memory tasks	[46]
Babiloni <i>et al.</i>	2004	-LORETA spectral profiles of various spectral bands -To define EEG sources specific for (i) mild AD compared to VaD or N, (ii) stages of AD	48 mild-AD, 20 VaD, 38 N	-Statistical analysis -Spectral profiles related differently to different groups	[23]
Jacques <i>et al.</i>	2005	-WT analysis of ERPs from target P _Z -Learn++ based classifier to discriminate groups	14 AD, 18 N	-Diagnostic analysis -High performances -Feasible approach for early diagnosis	[63]
Stepenosky <i>et al.</i>	2006	-WT analysis of ERPs from target and novel P _Z , C _Z and F _Z electrodes -Decision level fusion by combination rules	25 AD, 27 N	-Diagnostic analysis -Feasible approach for early AD diagnosis -Best performance by product rule at 90%	[64]
Gandhi <i>et al.</i>	2006	-WT analysis of ERPs from target P _Z electrode -Stacked generalization based ensemble system	22 AD, 22 N	-Diagnostic analysis -Feasible automated classification approach -Higher classification performances	[44]
Polikar <i>et al.</i>	2007	-WT analysis of ERPs from target and novel P _Z , C _Z and F _Z electrodes -Ensemble system based on decision level fusion by modified WMV rule	28 AD, 24 N	-Diagnostic analysis -Best performance of 79.2% by fusion of 1–2 Hz from novel P _Z and 4–8 Hz from target F _Z -Performs better than diagnosis at community clinics/ hospitals	[65]

CHAPTER 3

METHODS

The methods involved in this study are discussed in detail in this chapter. These methods include the selection process for the study cohort, acquisition of ERP signals through oddball paradigm followed by preprocessing, wavelet transform for the feature extraction, classification algorithm by ensemble system, different fusion techniques, and evaluation of diagnostic performance measures.

3.1 STUDY COHORT

The data set used for this project consisted of EEG/ERP signals acquired at Drexel University, from the patients recruited by the University of Pennsylvania for the development of an algorithm for the early diagnosis of AD. The neuropsychological AD diagnosis was done through a clinical evaluation, using a series of neuropsychological tests, which included interviews with the patient and their caregivers. Ninety patients were recruited by the Memory Disorders Clinic and Alzheimer's disease Research Center of University of Pennsylvania, according to the following inclusion and exclusion criteria:

Table 3.2: Inclusion criteria for the research cohort

INCLUSION CRITERIA		
	Normal Cohort	AD Cohort
Age	60 years or older	60 years or older
Clinical Dementia Rating (CDR) score	0	Greater than or equal to 0.5
Mini Mental State Exam (MMSE) score	Greater than 26	Less than or equal to 26
Cognitive and functional decline	None during the past 2 years	Present during the past 12 months.
NINCDS-ADRDA criteria[11]	Does not satisfy	Satisfies

Table 3.2: Exclusion criteria for the research cohort

EXCLUSION CRITERIA FOR BOTH GROUPS
Evidence of any central nervous system neurological disease (e.g. stroke, multiple sclerosis, Parkinson disease, or other form of disorder).
Use of sedatives or antidepressant medications within 48 hours of ERP acquisition.

All subjects received a thorough neurological exam and completed medical histories. Major demographic and medical information, including current medications (prescription, over – the – counter and alternative medications) were taken into consideration. The evaluation included standardized assessment for overall impairment, functional impairment, extra pyramidal signs (physical symptoms like tremors, rigidity, slowness of movement, etc.), behavior changes, and depression.

The recruited AD patients, included in this study were at the earliest stages of the disease. One of the inclusion criteria measure was the Clinical Dementia Rating (CDR) score used to determine AD progression. The CDR test rating is based on six cognitive-functional categories: memory, orientation, judgment, community affairs, home and hobbies, and personal care. A five point scale is used to rate the function in each category: normal (0), questionable impairment (0.5), mild impairment (1), moderate impairment (2), and severe impairment (3). An interview with the subject and their

caregivers was used to rate each category and special scoring rules were used to determine the overall CDR score [66].

Another test used for inclusion in the current study was the Mini Mental State Exam (MMSE), a standardized test widely used to assess cognitive mental status. The test assesses orientation, attention, immediate and short-term recall, language, and the ability to follow simple verbal and written commands. In addition, MMSE provides a total score of cognitive function on a scale of 0 to 30. A score less than 19 is considered to indicate cognitive impairment. Cognitive performance as measured by the MMSE shows an inverse relationship between MMSE scores and the age. The MMSE medians range from 29 for those 18 to 24 years of age, to 25 for individuals 80 years of age and older. Education levels are directly related to the MMSE scores. The median score is 29 for individuals with at least 9 years of schooling, 26 for those with 5 to 8 years of schooling, and 22 for those with 0 to 4 years of schooling [67]. Both, CDR and MMSE tests are not used alone for diagnosis of AD. They constitute as a part of the tests, designed by NINCDS-ADRDA for diagnosis of probable AD, and play important role in assessing the severity of the disease [11].

3.2 EXPERIMENTAL SETUP

The current project consisted of two related classification problems:

1 Two-class problem:

This consisted of a problem to distinguish patients with probable AD from cognitively normal subjects. The cohort class labels for this test were obtained from the University of Pennsylvania clinical evaluations.

2 Three-class problem:

Using MMSE scores, further class evaluation of probable AD patients into two more classes: mild AD patients and mild-moderate AD patients was done. Cut-off MMSE score used in this evaluation was 24. Hence a three-class problem was obtained which distinguish mild-moderate AD patients, mild AD patients, and cognitively normal subjects.

3.3 ODDBALL PARADIGM WITH NOVEL SOUNDS

The ERPs were obtained using an auditory oddball paradigm while the subjects were comfortably seated in a quiet room. The protocol described by Yamaguchi *et al.* with slight modifications, was used [35]. Binaural audiometric thresholds were first determined for each subject using a 1000 Hz tone. The evoked response stimulus was presented to both ears at 60 dB above each subject's auditory threshold. This procedure removed any bias to variations in hearing among patients. Each stimulus consisted of tone bursts 100 ms in duration, including 5 ms onset and offset envelopes. Tones at the frequency of 1000 Hz were presented 65% of the trials, and tones of 2000 Hz were presented 20% of the trials in a random sequence. The remaining 15% of the stimuli consisted of unexpected novel sounds that were presented randomly. These novel sounds included 60 unique environmental sounds that were recorded digitally and edited to 200 ms duration. A total of 1000 stimuli, including frequent 1000 Hz (n=650), infrequent 2000 Hz tones (n=200) and novel sounds (n=150) were delivered to each subject with an interstimulus interval of 1.0–1.3 seconds. The subjects were instructed to press a button each time they heard the 2000 Hz oddball (target) tone. Data collection typically took

about 30 minutes per subject, which included frequent breaks, approximately three minutes of rest every five minutes. The experimental session was preceded by a 1 minute practice session where the subjects sampled the difference between target and standard tones without the novel sounds.

ERPs were recorded from 19 tin electrodes embedded in an elastic cap with 2 reference electrodes on ears. The EEG recording was initiated 200 ms before the stimulus (to provide a baseline signal) and continued for 800 ms after the stimulus. The electrode impedances were kept below 20 k Ω to yield a good signal. Recordings which contained obvious artifacts were identified and rejected by the EEG technician. The remaining ERP potentials were then amplified, digitized at 256 Hz/channel (19 channels), and low-pass filtered. The ERPs for each channel were then averaged, notched filtered at 59–61Hz, and base-lined with respect to the prestimulus interval. The final signals consisted of 257 samples of one-second duration. The signals, for each tone, of multiple trials, were synchronized and averaged to remove variations in cortical activity. The averaging protocol involved averaging 90–250 recordings per patient (per tone). Furthermore, all responses with artifacts, responses to missed targets, etc. were removed from the average by the EEG technician, to obtain consistent ERP responses.

In this study, initially ERPs obtained from C_Z, F_Z and P_Z electrodes, and later from O₂, T₈, P₃, T₇, F_{P2} and F₈ additional electrodes were analyzed. The locations of these electrodes can be seen in Figure 2.1. The P_Z electrode was analyzed since ERPs are known to be most prominent in the central parietal regions of the cortex [68]. Since P300 is traditionally associated with the oddball tone; only responses to this tone are typically analyzed. In previous studies [44, 61, 63], analysis of oddball responses from P_Z

electrode were only conducted. But in this study, we have analyzed diagnostic information that may be found in the data from other electrodes in response to the target tones, as well the novel tones. Justification for analysis of C_Z and F_Z electrodes is their relative and symmetric proximity to P_Z electrode. Other electrodes in this study were chosen based on their individual performance in providing diagnostic information, as obtained from one of the colleague working on a different aspect of this project [69]. Justification for analyzing the responses to the novel tones is the potential information that may be present in other components of ERPs, which may be prominent in responses to the novel tones.

3.4 TIME-FREQUENCY ANALYSIS FOR FEATURE EXTRACTION

Frequency analysis is an alternate and informative technique for describing time domain signal. An advantage of the frequency domain representation over the time representation is its ability to give a clear visualization of the specific frequencies and their contributions within the signal. Frequency domain analysis is traditionally done by Fourier transform (FT), invented by Joseph Fourier. The continuous FT can be described as:

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

where, $X(f)$ represents $x(t)$ in the frequency domain. The function $x(t)$ is continuous and sufficiently smooth. The inverse transform can be calculated by:

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

for every real number t . The inverse FT is the recombination of all frequency components of $x(t)$. Fourier transform involves a correlation between the signal to be analyzed and complex exponentials of different frequencies. The appropriate interpretation of the FT requires the signal to be stationary as no time information is provided by the transformation.

The short-time Fourier transform (STFT), also known as Gabor transform (designed by Dennis Gabor), was developed to provide time-frequency representation of the signal. This transform uses the FT to determine frequency and phase content of localized sections of the signal. The signal to be transformed is multiplied by a window function. “Windowing” the complex sinusoids of the Fourier transform gives a time evolution of the frequencies by sliding the windows throughout the signal. The STFT can be calculated as:

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

where $x(t)$ is the signal to be analyzed, $w(t)$ is the windowing function, and τ is the translation of the window throughout the signal. This procedure gives an optimal time-frequency representation, but is limited by the Heisenberg Uncertainty Principle, which gives a lower bound of:

$$\Delta t \Delta f \geq \frac{1}{4\pi} \quad (3.4)$$

where, Δt is the time resolution and Δf is the frequency resolution. Δt indicates how well two spikes in time can be separated from one another in the transfer domain, and Δf indicates how well two spectral components can be separated from each other in time

domain. If the window is too narrow, there will be poor frequency localization, and if the window is too wide, there will be poor time localization [70].

Grossman and Morlet developed the wavelet transform to overcome the fixed resolution problems of the STFT. The main advantage of the wavelets is their varying window size, wide for low frequencies and narrow for the high frequencies, leading to an optimal time-frequency resolution in all frequency ranges. Wavelet transforms can be classified into two types: continuous wavelet transforms (CWT) and discrete wavelet transforms (DWT). CWT operates over every possible scale and translation, while the DWT uses a specific dyadic set of scale and translation values [70].

The CWT of a signal is defined as:

$$\gamma(s, \tau) = \int_{-\infty}^{\infty} x(t) \psi_{s, \tau}^* dt \quad (3.5)$$

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

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

The factor $1/\sqrt{s}$ is the energy normalization across different scales. The wavelet transform only deals with general properties of its calculation. Unlike the FT which uses a fixed basis function of complex exponentials; the wavelet basis functions are not fixed and several wavelets are available as they can be custom designed for specific application. Any function that satisfies the following two conditions can be a wavelet function:

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

and

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

The CWT is computed by continuously shifting a continuously scalable function over the entire signal, while calculating the correlation between the signal and the wavelet function. Hence CWT is unpractical for many real life applications. The DWT, on the other hand, overcomes this problem by scaling and translating in discrete steps.

The DWT utilizes two sets of functions, a scaling function $\phi(t)$, and a wavelet function $\psi(t)$, each associated with low-pass and high-pass filters, respectively. These functions can be obtained as a weighted sum of the scaled and shifted versions of the scaling function itself by:

$$\phi(t) = \sum_n h[n] \phi(2t - n) \quad (3.9)$$

$$\psi(t) = \sum_n g[n] \phi(2t - n) \quad (3.10)$$

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

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k) \quad (3.11)$$

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (3.12)$$

Different scale and translations of these functions allow differentiation of frequency and time localizations of the signal. The coefficients (weights), $h[n]$ and $g[n]$, that satisfy

Equations (3.9) and (3.10) constitute the impulse responses of the low-pass and high-pass filters used in the wavelet analysis, and define the type of the wavelet used in the analysis. Decomposition of the signal into different frequency bands is therefore accomplished by successive high-pass and low-pass filtering of the time domain signal.

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

$$d_1[k] = y_{high}[k] = \sum_n x[n] \cdot g[2k - n] \quad (3.13)$$

$$a_1[k] = y_{low}[k] = \sum_n x[n] \cdot h[2k - n] \quad (3.14)$$

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

passing $y_{low}[n]$ through another set of high-pass and low-pass filters, to obtain level 2 detail coefficients, d_2 and level 2 approximation coefficients, a_2 .

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

The overall wavelet decomposition algorithm is illustrated in Figure 3.1, where the frequency range analyzed with each set of coefficients is marked with 'F'. The length of each set of coefficients is also provided, which depends on the specific wavelet used in the analysis. The decomposition in Figure 3.1 is obtained using the Daubechies wavelets with 4 vanishing moments, whose corresponding filters $h[n]$ and $g[n]$ are of length 8.

The number of coefficients in each decomposition level is determined by the filter convolved with the original signal, then down sampled. In the first level of decomposition, the signal length is 257, and the filter length is 8, resulting in a filtered signal with 264 points. After down sampling, it is reduced to 132 points and the process repeats. An approximation signal $A_j(t)$ and a detail signal $D_j(t)$ can be reconstructed from level j approximation and detail coefficients as:

$$A_j(t) = \sum_k a_j[k] \cdot \phi_{j,k}(t) \quad (3.15)$$

$$D_j(t) = \sum_k d_j[k] \cdot \psi_{j,k}(t) \quad (3.16)$$

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

$$\begin{aligned}
x(t) &= A_j(t) + \sum_{j=-\infty}^j D_j(t) \\
&= \sum_k a_j[k] \cdot \phi_{j,k}(t) + \sum_{j=-\infty}^j \sum_k d_j[k] \cdot \psi_{j,k}(t)
\end{aligned} \tag{3.17}$$

A detailed review of wavelet transforms can be seen in “The Wavelet Tutorial,” developed by Robi Polikar [71].

Figure 3.1: 7-level DWT decomposition [61]

In this study, the Daubechies wavelet with 4-vanishing moments (db4) was used for feature extraction of ERPs, obtained from each electrode, into successive frequency bands. This analysis was done based on previous studies in [61, 63]. Figure 3.2 shows the scaling function $\phi(t)$ as well as the wavelet function $\psi(t)$ for the db4 wavelet. The decomposition low and high-pass filters as well as the reconstruction low and high-pass filters are also shown in Figure 3.2.

Figure 3.2: Daubechies-4 wavelet scaling function and wavelet function

Figure 3.3 illustrates the reconstructed signals obtained at each level, by the 7-level wavelet decomposition of ERP signals, obtained from probable AD and cognitive normal subjects. The coefficients at each levels of decomposition constitute the frequency bands, which were individually analyzed in this study. Amplitudes of the signals indicate that that majority of the signal's energy lie in $d_6 \sim d_7$ and a_7 . Even though the signals from both AD and normal patients show significant P300 component, there are differences in their spectral bands of the wavelet decomposition. These differences lead to the correct classification of signal.

Figure 3.3: 7-level wavelet decomposition for a probable AD patient and a cognitively normal patient

3.5 AUTOMATED CLASSIFICATION USING PATTERN RECOGNITION TECHNIQUES

In automated classification applications, the distinctive features of signals to be identified (wavelet coefficients in this study) are first obtained, and are presented to the classification algorithm as training data. The trained system is then evaluated on data not seen during the training.

Artificial neural network (ANN) is a neural network model in artificial intelligence, which is modeled after cortical structures of the brain. It consists of interconnected processing elements called nodes or neurons that works together to produce an output. The output of the neural network relies on the co-operation of individual neurons within the network. An ANN is configured for pattern recognition (data classification) through a learning process. Learning involves adjustments to the synaptic connections that exist between the neurons [72]. In this project multilayer perceptron (MLP) neural networks were used as base classifiers for stacked generalization algorithm.

3.5.1 MLP NEURAL NETWORKS

The MLP neural network is a well-known and well-established classification technique used in classification of signals, whose classes are separated by a complex decision boundary. Neural networks are typically trained with a subset of the data for which the correct class information is available, and tested with another subset of data where the classes determined by the network are compared to true labels. The training allows the

network to identify key features in the signals, which are characteristics of particular classes.

An MLP is a feed-forward network, which has one or more hidden layers between the input and output layers. The MLP is trained using the supervised backpropagation learning rule to distinguish between the different classes [72]. The dimensionality of the signals to be classified determines the number of input nodes (wavelet coefficients, in this study), and the number of output nodes is determined by the number of classes. The architecture of a typical MLP neural network is shown in Figure 3.4. The input, $\mathbf{x}_i$, $i = 1, 2, \dots, d$, depends on the wavelet type or other features. The number of hidden layers and hidden layer nodes — HN_j , $j = 1, 2, \dots, H$ at each layer, depends on the application. The $\mathbf{z}_k$, $k = 1, 2, \dots, c$ are the values of the output nodes.

Figure 3.4: MLP neural network architecture [73]

All nodes are fully interconnected to nodes of the adjacent layers by a set of weights. The weights connecting the input nodes to the hidden layer nodes are denoted by w_{ji} , and the weights connecting the hidden layer to the output nodes are denoted by w_{kj} .

The outputs y_j and z_k are obtained as:

$$y_j = f(\text{net}_j) = f\left(\sum_{i=1}^d w_{ji}x_i\right) \quad (3.18)$$

$$z_k = f(\text{net}_k) = f\left(\sum_{j=1}^H w_{kj}y_j\right) \quad (3.19)$$

where $f(\cdot)$ is the activation function, net_j is the weighted sum of inputs at hidden layer node j , and net_k is the weighted sum of inputs at output node k . The weights are determined through the backpropagation training algorithm described in the next section.

3.5.2 BACKPROPAGATION LEARNING RULE

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

During the learning phase, the untrained network is presented with patterns, and the outputs of the network are compared with the known class information for those patterns. The difference between the outputs and the target values is the error or criterion

function which is a scalar function of the weights. This function is minimized when the output of the network equals the desired output, and thus the weights are adjusted to reduce the error. The training error is the sum over the output units of the squared difference between the desired output, t_k and the actual output, z_k and is given as:

$$J(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^c (t_k - z_k)^2 = \frac{1}{2} \|\mathbf{t} - \mathbf{z}\|^2 \quad (3.20)$$

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

$$\Delta \mathbf{w} = -\eta \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} \quad (3.21)$$

where, η is the learning rate, indicating the relative step size of the change in weights.

The algorithm takes a step in weight space that lowers the corresponding value of the criterion function. The algorithm takes a weight vector at iteration m and updates itself according to:

$$\mathbf{w}(m+1) = \mathbf{w}(m) + \Delta \mathbf{w}(m+1) \quad (3.22)$$

where, m is the index of the pattern presentation.

For a three layer network which has single hidden layer, considering the hidden to output weights to be w_{kj} , the weight update (learning) rule for the hidden to output unit k is obtained as:

$$\frac{\partial J(\mathbf{w})}{\partial w_{kj}} = \frac{\partial J(\mathbf{w})}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}} = -\delta_k \frac{\partial net_k}{\partial w_{kj}} \quad (3.23)$$

where, δ_k is the sensitivity of the output unit k , and is defined as:

$$\delta_k = -\frac{\partial J(\mathbf{w})}{\partial net_k} \quad (3.24)$$

Assuming that the activation function $f(\cdot)$ is non-linear, and hence differentiable, Equation (3.20) can be differentiated, so that δ_k becomes:

$$\delta_k = -\frac{\partial J(\mathbf{w})}{\partial net_k} = \frac{\partial J(\mathbf{w})}{\partial z_k} \frac{\partial z_k}{\partial net_k} = (t_k - z_k) f'(net_k) \quad (3.25)$$

The last derivative in Equation (3.23) can be obtained by deriving Equation (3.19):

$$\frac{\partial net_k}{\partial w_{kj}} = y_j \quad (3.26)$$

Substituting Equations (3.24), (3.25) and (3.26) in Equation (3.23) gives the weight update rule for the hidden to output weights by:

$$\Delta w_{kj} = \eta \delta_k y_j = \eta (t_k - z_k) f'(net_k) y_j \quad (3.27)$$

The learning rule for input to hidden unit can be obtained as

$$\frac{\partial J(\mathbf{w})}{\partial w_{ji}} = \frac{\partial J(\mathbf{w})}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad (3.28)$$

where,

$$\begin{aligned} \frac{\partial J(\mathbf{w})}{\partial y_j} &= \frac{\partial}{\partial y_j} \left[\frac{1}{2} \sum_{k=1}^c (t_k - z_k)^2 \right] \\ &= - \sum_{k=1}^c (t_k - z_k) \frac{\partial z_k}{\partial y_j} \\ &= - \sum_{k=1}^c (t_k - z_k) \frac{\partial z_k}{\partial net_k} \frac{\partial net_k}{\partial y_j} \\ &= - \sum_{k=1}^c (t_k - z_k) f'(net_k) w_{kj} \end{aligned} \quad (3.29)$$

The sensitivity for a hidden unit j is δ_j , and is defined as:

$$\delta_j \equiv f'(net_j) \sum_{k=1}^c w_{kj} \delta_k \quad (3.30)$$

Thus the learning rule for the input to hidden weights is given as:

$$\Delta w_{ji} = \eta \delta_j x_i = \eta \left[\sum_{k=1}^c w_{kj} \delta_k \right] f'(net_j) x_i \quad (3.31)$$

Equations (3.27) and (3.31), along with the training protocols, constitutes the backpropagation algorithm. It is also called as the “backpropagation of errors” algorithm because the error propagates back from the output layer to the hidden layer in order to update the input to hidden weights [72, 73].

3.5.3 K-FOLD CROSS VALIDATION

A cross validation process is implemented to provide a reasonably good estimate of the true generalization performance. In k -fold cross validation, the entire available training dataset is split into $k > 2$ partitions, creating k blocks of data. Of these k blocks, $k-1$ are used for training and the remaining k^{th} block is used for testing. This procedure is then repeated k times, using a different block for testing in each case. This is illustrated in Figure 3.5.

The k test performances so obtained are averaged, and the average test performance is declared as the estimate of the true generalization performance of the algorithm. This can be obtained as:

$$\bar{P} = \frac{1}{k} \sum_{i=1}^k P_i \quad (3.32)$$

Figure 3.5: *K-fold cross validation* [73]

The choice of k is data dependent: selecting k too large divides the data into too many partitions, allowing a larger amount of data for training and also large variation in individual performance P_i . On the other hand, choosing k too small causes not enough partitions to be made and reduces the amount of data available for training. The extreme case, where $k = N$ (total data instances), is known as leave-one-out validation where the entire data but one instance is used for each training session. The testing is done solely on the remaining instance. The entire process is repeated N times [73].

3.5.4 ENSEMBLE OF CLASSIFIERS

Ensemble of classifiers or combining classifiers is an established and rapidly developing technique in the fields of pattern recognition and machine learning. It is variously known as committees of learners, mixture of experts, classifier ensembles, multiple classifier system, consensus theory, etc. This approach has been developed because a highly accurate and reliable classification is required for practical applications.

If we have many different classifiers, it is appropriate to use them in some form of combination in the hope of increasing both reliability and the overall accuracy. Hence classifiers should be considered as a team of experts similarly to the way a person may consult a panel of experts before making a decision. A good example of this technique: a patient suffering from certain disease, not happy with the decision of one physician, may consult several doctors for their opinions before considering the diagnosis. This is the basic intuition behind multiple classifier system. The classifiers in an ensemble system must not only be accurate to contribute information, but also different enough from each other, or diverse to ensure that the provided information is beneficial. Hence, to design any ensemble system, two important points should be addressed: (1) How would individual classifier be designed, and (2) How will they differ from each other? Moreover, the increase in the overall accuracy by combination of classifiers is achieved at the expense of increased complexity [74].

Several ensemble system algorithms have been designed to solve pattern recognition problems. One of such algorithms is “bagging” by Breiman, which is one of the most intuitive, simplest to implement, and has very good performance. Diversity in bagging is obtained by using bootstrapped replicas of the training data, i.e. different

training data subsets randomly drawn – with replacement – from the entire training data. Each training data subset is used to train a different classifier of the same type. Individual classifiers are then combined by a majority vote of their decisions. For any given instance, the class chosen by most classifiers is the ensemble decision [75].

Another algorithm by Schapire, called “boosting” in which a “weak learner” – an algorithm that generates classifiers that can do just better than random guessing, can be turned into a “strong learner” by changing the distribution of training data. Similar to bagging, boosting creates an ensemble of classifiers by resampling the data, which are then combined by majority voting. In boosting, resampling is strategically done so that each classifier gets the most informative training data. This is done by creating three weak classifiers: the first classifier C_1 is trained with a random subset of the available training data. The second classifier C_2 is trained on a training data only half of which is correctly classified by C_1 , and the other half is misclassified. The third classifier C_3 is trained with features misclassified by both, C_1 and C_2 . The three classifiers are combined through a three-way majority voting. The error of this ensemble system is less than the error of the best classifier in it [76].

Later, Freund and Schapire introduced “AdaBoost” algorithm [77], which is a variant of original boosting algorithm. In this algorithm, distribution of weights is done over the training data, which can be modified as required, at each iteration. Initially the distribution of the training data is uniform. The weak classifier sub-samples the training data based on the distribution and generate hypothesis for trained model. If the current classifier find a certain feature difficult to classify, than that feature will have higher weight for the next iteration, conversely if the current classifier finds a certain feature

easy to classify, than that object will have less weight in the next iteration. The final performance is obtained by the weighted votes of individual hypotheses according to their performances.

Other variations of ensemble of classifier system constitute of stacked generalization algorithm [78], developed by Wolpert, and mixture of experts [79] by Jacobs *et al.* These algorithms have multiple-level classification models. The stacked generalization algorithm constitute of multiple level (generally 2 levels) of classifiers. The first level constitute of set of classifiers, $C_1, \dots, C_T$, and the second level consists of a meta-classifier C_{T+1} . The second level meta-classifier is trained on the outputs of the first level classifiers. The second level classifier learns the correct and incorrect predictions of the first level classifiers. This algorithm is explained in more details in next section.

The mixture of experts works on conceptually similar technique, which has set of classifiers, $C_1, \dots, C_T$ in the first level. The second level classifier C_{T+1} serve as a gating network, which assigns weights to the outputs of the first level of classifiers. The gating network is trained by expectation maximization (EM) algorithm [80]. The inputs of the gating network are actually training data instances, and not the outputs of the first level of classifiers as in stacked generalization algorithm. The outputs of the first level of classifiers are combined using simple combination rules like weighted majority voting, sum rule, etc. The output of the algorithm depends on this combination rule which has instance specific weights provided by the gating network.

In previous studies, ensemble systems based on variations of boosting algorithm have been explored [63, 65]. In this study we have explored a new ensemble system based on stacked generalization algorithm, which has been predicted to work better than

the single level models. The stacked generalization algorithm and its applications are explained in detail in next section.

3.5.5 STACKED GENERALIZATION

Stacked generalization is an ensemble of classifiers approach for minimizing the generalization error rate of one or more classifiers. Stacked generalization works by reducing the biases of the classifiers with respect to the training dataset. Hence, when used with multiple classifiers, it is a more sophisticated version of cross-validation [78].

In stacked generalization algorithm, classifiers $C_1, \dots, C_T$ are trained using training parameters $\theta_1, \dots, \theta_T$ to obtain output hypotheses $h_1, \dots, h_T$. The outputs of these classifiers and the corresponding true classes are then used to train the second level meta-classifier, C_{T+1} . This is illustrated in Figure 3.6.

Figure 3.6: Stacked generalization algorithm [44]

The training parameter θ_{T+1} for the meta-classifier C_{T+1} , predicts the output decision y . It can be defined as:

$$y = h_{T+1}((h_1, h_2, \dots, h_T), \theta_{T+1}) \quad (3.33)$$

There are two major issues in designing a stacked system, initially described as “black art” by Wolpert [78]. The first is the choice of classifier at both first-level and meta-level that will lead to the best empirical results. The second issue concerns the combination of the predictions of the first-level classifiers and their mapping to generate the training features for the meta-classifier. Typically the class predictions of the first level are used as training features for meta-classifier. Various researches have been done to resolve these issues:

Chan experimented with various representations including the “class-attribute-combiners” scheme, where the class predictions of the first-level classifiers were appended with the input signal, and these features together with the correct class labels contributed the training data for the second level meta-classifier. He also experimented with an “arbiter” scheme, where the meta-classifier was trained only on the subset of the first-level classifiers, which disagreed on their predictions. In another scheme called “hybrid” scheme, the meta-classifier was trained only on the subset of the first-level classifiers, where the first-level classifiers disagree, by the “class-attribute-combiners” scheme. The “class-attribute-combiner” scheme gave the best results in his experiments [81].

Another study by Ting and Witten suggested the use of output class probabilities, where each level-one classifier predicted a probability distribution vectors over all

classes, instead of just using class labels. The individual vectors from T classifiers in first-level were concatenated, thus resulted in $T * Q$ attributes at the meta-level, where Q was the number of relevant classes. They also suggested using “multi-response linear regression (MLR)” for meta-classifier learning. MLR is an adaptation of linear regression designed by Breiman [82], which transforms the classification problem into Q different binary prediction problems. For each class, a linear equation is constructed to predict ‘1’ if class value equals the class under consideration and ‘0’ otherwise. Their result showed that using these approaches, the implementation of stacked generalization algorithm gave better predictive accuracy than that obtained by single level learning model (single strong classifier), and other generalization techniques like cross-validation and majority voting [83].

Ghorbani and Owrangh implemented stacked generalization using neural networks as the base classifiers for a statistically neutral problem (mostly two class problem where probability of mapping an input onto an output is always the chance value of 0.5). Results showed that stacked generalization scheme provided better classification performance and generalization accuracy over single level cross-validation model [84].

The outcome of all these studies showed that stacked generalization typically outperformed other techniques. However, stacking involves training of multiple level-1 classifiers and a meta-classifier, and hence it could be computationally expensive. Considering all the effects from previous studies, a stacked algorithm was designed, which would effectively work in this application of diagnostic classification.

The stacked generalization algorithm designed for the study in this thesis, worked in three steps: initial training, intermediate training and testing. Initial training of stacked

generalization involved a k -fold cross validation process of the entire training dataset to obtain the training data for the meta-classifier C_{T+1} . The entire training data of length N was first divided into K nearly equal blocks, each of length N/K (extreme case being $K = N$ – leave-one out validation). Each classifier in the ensemble C_1 through C_T was trained K times, using $K-1$ blocks of the training data. For each trial in k -fold ($k = 1, \dots, K$), a block of data was not seen by any of the classifiers in first level. The outputs for that block from first level classifiers along with their correct class labels constituted as the training data for the second level meta-classifier C_{T+1} . Each fold provided a set of training data for the second level classifier. The outputs from the first level classifiers were in continuous form.

In this study, the second level meta-classifier was trained according to two different approaches of initial training:

1 Straight meta-classifier training (SMT):

In this approach, the continuous outputs of each classifier in ensemble C_1 through C_T (first level of classifiers) were concatenated and used to train the second level meta-classifier [81]. This process of training the meta-classifier is illustrated in Figure 3.7. The entire training data was divided into K blocks, each of size $d \times [N/K]$ (where d was the number of data points in each training instance, i.e. 8, 10 or 14 in this thesis), to contribute in k -fold cross validation. For each fold, $K-1$ blocks were used to train all the classifiers in level 1, and the remaining block was used to test those classifiers. The outputs for such block for each first level classifier are represented by $\mathbf{y}_{tk}$ in Figure 3.7, where $t = 1, \dots, T$ represents a particular classifier in 1st level ($C_1, \dots, C_T$), and k represents the test block in k -fold validation ($k = 1, \dots, K$). Each $\mathbf{y}_{tk}$ was of size $\mathcal{C} \times [N/K]$, where $\mathcal{C}$

was the total number of classes (2 or 3 in this thesis). This process was repeated K times each time using a different block for testing. For each fold, outputs from all classifiers in first level were concatenated. Hence by k -fold validation ($k = 1, \dots, K$), K blocks of outputs, each of length N/K , gave N output instances. These N instances (each of size $[e \cdot T] \times 1$) along with the corresponding class labels were used to train the second level meta-classifier C_{T+1} .

Figure 3.7: Initial training by straight meta-classifier training (SMT) approach

2 Hierarchical meta-classifier training (HMT):

In this approach, the continuous outputs of each classifier in ensemble C_1 through C_T (first level of classifiers) along with the original data, which gave those outputs, were concatenated and used to train the second level meta-classifier [81, 84]. This process of training the meta-classifier is illustrated in Figure 3.8. Similarly as in SMT approach, the entire training data was divided into K -blocks, each of size $d \times [N/K]$ (where d was the number of data points in each training instance, i.e. 8, 10 or 14 in this thesis), to contribute in k -fold cross validation. For each fold, $K-1$ blocks were used to train all the classifiers in level 1, and the remaining block was used to test those classifiers. The outputs for such block for each classifier in first level are represented by $\mathbf{y}_{tk}$ in Figure 3.8, where $t = 1, \dots, T$ represents a particular classifier in 1st level ($C_1, \dots, C_T$), and k represents the test block in k -fold validation ($k = 1, \dots, K$). Each $\mathbf{y}_{tk}$ was of size $\mathcal{C} \times [N/K]$, where $\mathcal{C}$ was the total number of classes (2 or 3 in this thesis). This process was repeated K times each time using a different block for testing. For each fold, outputs from all classifiers in first level were concatenated along with the input data from that block. Hence by k -fold validation, K blocks of outputs, each of length N/K , gave N output instances. These N instances (each of size $[\mathcal{C} \cdot T + d] \times 1$) along with the corresponding class labels were used to train the second level meta-classifier C_{T+1} .

Figure 3.8: Initial training by hierarchical meta-classifier training (HMT) approach

Once C_{T+1} was trained, all data were pooled. In intermediate training, individual classifiers C_1 through C_T were trained again on the entire training database, using a suitable resampling method. In testing, class predictions for the test dataset was done by the trained algorithm.

The pseudo-code for the stacked generalization algorithm using above training approaches is described in Figure 3.9.

Stacked Generalization

Input:

- Training data set: N examples: $L = [(\vec{x}_n, y_n)]$, $n = 1, \dots, N$, each $\vec{x}_n$ having d data points, and the class labels $y_n \in \Omega$, where $\Omega = \{\omega_1, \dots, \omega_{\mathcal{C}}\}$
- A (supervised) base classifier; and the number of level-1 classifiers, T , to be generated.

Training:

Initial Training

Training of first level classifiers:

- Divide entire training data L , into K blocks of size N/K , i.e. $L_1, L_2, \dots, L_K$ (extreme case $K = N$ – Leave-one-out validation)
- **Do for** $k = 1, 2, \dots, K$ (k fold cross validation)
 1. Define L_k and $L^{(-k)} = L - L_k$;
 2. Train first-level classifiers $C_1, \dots, C_T$ using $L^{(-k)}$;
 3. Test classifiers $C_1, \dots, C_T$ on L_k ;
 4. Let $\mathbf{y}_{ik} \in \mathbb{R}^{e \times [N/K]}$ represent the continuous outputs from each level-1 classifiers, C_b , $t = 1, \dots, T$, and $\mathcal{C}$ is number of classes.

End

- At the end of entire cross-validation, K blocks of outputs (each of size $\mathcal{C} \times [N/K]$) from T level-1 classifiers gives a total of N output instances, $\vec{y}_n$, where $t = 1, \dots, T$ and $n = 1, \dots, N$ (each of size $\mathcal{C} \times 1$).

Training of meta-classifier:

Straight meta-classifier training (SMT):

- Concatenate the outputs $\vec{y}_{1n}, \dots, \vec{y}_{Tn}$, $n = 1, \dots, N$ for all the classifiers in level-1 to form N training instance for meta-classifier C_{T+1} :

$$\vec{Y}_n = [\vec{y}_{1n}, \dots, \vec{y}_{Tn}]^T, n = 1, \dots, N \text{ (each of size } [\mathcal{C} \cdot T] \times 1 \text{)}$$
- Train meta-classifier C_{T+1} using these N instances along with their original class labels.

OR

Hierarchical meta-classifier training (HMT):

- Concatenate the outputs $\vec{y}_{1n}, \dots, \vec{y}_{Tn}$, $n = 1, \dots, N$, along with the corresponding input data $\vec{x}_n$ to form N training instance for meta-classifier C_{T+1} :

$$\vec{Y}_n = [\vec{y}_{1n}, \dots, \vec{y}_{Tn}, \vec{x}_n]^T, n = 1, \dots, N \text{ (each of size } [\mathcal{C} \cdot T + d] \times 1 \text{)}$$
- Train meta-classifier C_{T+1} using these N instances along with their original class labels.

Intermediate training

- Pool all data;
- Retrain $C_1, \dots, C_T$ on entire training data L .

Testing:

- Given an unlabeled instance $\vec{x}$;
- Classify $\vec{x}$ using classifiers $(C_1, \dots, C_T) \rightarrow C_{T+1} \rightarrow y$;
- The output y from C_{T+1} predicts the correct class for $\vec{x}$.

Figure 3.9: Pseudo-code for stacked generalization algorithm

3.5.6 DATA FUSION

One of the simplest and most commonly used approaches in ensemble based system is to perform decision level data fusion. The outputs of the ensemble system for individual data sources (specific electrode – stimuli – frequency band combinations) were fused using different combination rules. In this study, we have used simple majority voting (SMV), weighted majority voting (WMV), product, and sum combination rules for decision level data fusion.

Let us assume that only class labels are available from classifier outputs. The decision of the t^{th} classifier defined as $d_{t,j} \in \{0,1\}$, $t=1,\dots,T$ and $j=1,\dots,C$, where T is the number of classifiers and C is the number of classes. If t^{th} classifier chooses class ω_j , then $d_{t,j} = 1$, and 0, otherwise.

In simple majority voting (SMV), the ensemble decision is chosen as the class that receives the highest number of votes, that is, we choose class ω_k , if:

$$\sum_{t=1}^T d_{t,k} = \max_{j=1}^C \sum_{t=1}^T d_{t,j} \quad (3.34)$$

If some of the classifiers are more qualified than others, then weighting their decisions more heavily may further improve the overall performance. In weighed majority voting (WMV), weight w_t is assigned to classifier h_t in proportion to its estimated performance. The ensemble decision will choose class ω_k , if:

$$\sum_{t=1}^T w_t d_{t,k} = \max_{j=1}^C \sum_{t=1}^T w_t d_{t,j} \quad (3.35)$$

In most applications, the weight w_t assigned to classifier h_t is typically proportional to its individual accuracy p_t :

$$w_t \propto \log \frac{p_t}{1-p_t} \quad (3.36)$$

For the following discussion, let us assume that outputs of the classifiers are interpreted as in continuous form. The continuous output provided by a classifier for a given class is interpreted as the degree of support given to that class. Kuncheva *et al.* defined a *decision profile matrix* in [85]. The decision profile matrix $DP(\mathbf{x})$, for an instance $\mathbf{x}$ consists of elements $d_{t,j} \in [0,1]$, which represent the support given by the t^{th} classifier to class ω_j . The rows of $DP(\mathbf{x})$ represents the support given by individual classifiers to each of the classes, where the columns represent the support for particular class from all the classifiers. The decision profile matrix is illustrated in Figure 3.10.

Figure 3.10: Decision profile matrix [85]

Algebraic combiners are simply, non-trainable combiners of continuous outputs.

The total support received by class ω_j (the j^{th} column of decision profile matrix $DP(\mathbf{x})$) is given by:

$$\mu_j(\mathbf{x}) = \mathfrak{I}[d_{1,j}(\mathbf{x}), \dots, d_{T,j}(\mathbf{x})] \quad (3.37)$$

where $\mathfrak{I}(\cdot)$ is the combination function.

For sum combination rule, the support for ω_j is obtained as the sum of all classifiers' j^{th} outputs:

$$\mu_j(\mathbf{x}) = \sum_{t=1}^T d_{t,j}(\mathbf{x}) \quad (3.38)$$

The ensemble decision is taken as the class ω_j , for which the total support $\mu_j(\mathbf{x})$ is maximum:

$$E(\mathbf{x}) = \omega_m \mid m = \arg \max_j (\mu_j(\mathbf{x})) \quad (3.39)$$

In product rule, the supports provided by the classifiers are simply multiplied.

The support for class ω_j is:

$$\mu_j(\mathbf{x}) = \frac{1}{T} \prod_{t=1}^T d_{t,j}(\mathbf{x}) \quad (3.40)$$

Similarly as in sum rule, the ensemble decision is taken as the class whose total support is greatest.

Other combination rules can be obtained by using different algebraic combination function $\mathfrak{I}(\cdot)$.

A detailed review of different combiners can be done in [86].

3.6 DIAGNOSTIC PERFORMANCE

The overall performance measures achieved from the learning algorithms lacks the class specific performances. In medical tests and procedures, class descriptive measures are necessary for patient evaluation. Sensitivity, specificity, positive predictive value and negative predictive value are four commonly used quantities in medical diagnostics.

Table 3.3 gives the class information on which medical diagnostic performances can be defined for this study.

Table 3.3: Category for defining diagnostic performances

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

In Table 3.3, A is the number of patients classified as AD, and diagnosed as probable AD by the clinical evaluation. On the other hand, B is the number of patients who are also classified as AD, but clinically diagnosed as cognitively normal. C is the number of patients who are classified as normal, but clinically diagnosed as probable AD. Finally, D is the number of patients who are classified as cognitively normal, and clinically diagnosed as cognitively normal. $A+B+C+D$ is the total number of patients.

The overall generalization performance (OGP) represents the average probability of a correct decision. OGP describes the ratio of patients the classification system has correctly identified. It is calculated as:

$$OGP = \frac{A + D}{A + B + C + D} \quad (3.41)$$

Sensitivity represents the probability of a positive diagnosis, when the condition is present in the patient. In this study, it is the ratio of the number of AD patients correctly diagnosed by classification system. It is calculated as:

$$Sensitivity = \frac{A}{A + C} \quad (3.42)$$

Specificity represents the probability of a negative diagnosis, when the condition is not present in the patient. In this study, it is the ratio of cognitively normal patients correctly diagnosed by the classification system. It is calculated as:

$$Specificity = \frac{D}{B + D} \quad (3.43)$$

Positive Predictive Value (PPV) represents the probability that the patient has the disease of those patients who are diagnosed positive for the disease. In this study, PPV is the proportion of patients who actually have AD, to those who are diagnosed as AD patients by the classification system. It is calculated as:

$$PPV = \frac{A}{A + B} \quad (3.44)$$

Negative Predictive Value (NPV) represents the probability that the patient does not have the disease of those patients who are diagnosed as not to have the disease. In this study, NPV is the proportion of patients who are actually cognitively normal, to those who are diagnosed as cognitively normal by the classification system. It is calculated as:

$$NPV = \frac{D}{C + D} \quad (3.45)$$

3.7 METHOD OVERVIEW

The experimental setup included the criteria to recruit cohort for the study. The cohort was subject to the oddball paradigm procedure, which invoked event related potentials. The ERPs were subject to preprocessing by artifact removal, filtering, averaging and base-line correction. The ERP signals were then decomposed, using the wavelet transform, into successive spectral bands. The coefficients from each spectral band were used as inputs to the ensemble system based on stacked generalization algorithm. The output performances along with diagnostic performance measures were calculated and compared. The results obtained using individual data sources were combined using various combination rules and their performances were compared. The entire process is illustrated in Figure 3.11.

Figure 3.11: Overview of the analysis

CHAPTER 4

IMPLEMENTATION AND RESULTS

This chapter describes the implementation detail and results on various experiments included in this work.

4.1 IMPLEMENTATION DETAILS

Out of 90 patients recruited for the project, 71 were used in this analysis. Reasons for excluding 19 patients varied for several reasons: the patient not being able to complete the test, patient did not satisfy the early stage requirement, etc. The final cohort for two-class problem consisted of 34 probable AD patients ($\mu_{AGE}=74$, $\mu_{MMSE}=24.7$), and 37 cognitively normal subjects ($\mu_{AGE}=76$, $\mu_{MMSE}=29.2$). The three-class problem cohort consisted of 18 mild-moderate AD patients ($\mu_{AGE}=76$, $\mu_{MMSE}=22.5$), 16 mild AD patients ($\mu_{AGE}=73$, $\mu_{MMSE}=27.1$), and the same 37 cognitively normal subjects.

4.1.1 FEATURE EXTRACTION

For each patient, ERPs were extracted and averaged from all the electrodes used in this study (initially from C_Z, F_Z and P_Z, electrodes, and later from O₂, T₈, P₃, T₇, F_{P2} and F₈ electrodes), for responses to target and novel tones. All averaged ERPs were decomposed into seven levels using Daubechies wavelet with four vanishing moments (db4) [61, 63]. Of the eight frequency bands created by the decomposition, the following

bands were used individually for further analyses: approximation at 0–1 Hz, and details at 1–2, 2–4 and 4–8Hz. Detail coefficients at 16–32, 32–64 and 64–128 Hz were not considered for analyses, as ERPs are known not to include any relevant frequency components in those intervals. In fact, the P300, which has been earlier linked with AD, is known to reside in 0–4 Hz interval, primarily around 3 Hz.

4.1.2 CLASSIFICATION

Data sources (specific electrode – stimuli – frequency band combination) were individually used to train the stacked generalization based ensemble system. All classifiers in ensemble system were multilayer perceptron type classifiers. Several experiments were performed to determine the architectural parameters of the stacked generalization ensemble system such as: the number of classifiers in first level in the 5–9 range, and for each classifier, the number of hidden layer nodes in the 5–50 range, and the error goal in the 0.005–0.1 range. After several trials, the following parameters were decided as the common architecture for all experiments: the first level consisted of an ensemble of 9 classifiers with each having one hidden layer and 10 hidden layer nodes. The second level meta-classifier had one hidden layer with 30 hidden layer nodes. The error goal was set to be 0.01 for all the classifiers.

For each data source, 5 fold cross validation technique was utilized for estimation of true generalization performance. Hence by 5 fold validation, 71 patients were divided into 5 blocks (4 blocks of 14 patients and 1 block of 15 patients). The stacked generalization algorithm worked in three steps: Initial training, intermediate training, and testing. In initial training of stacked generalization, for each fold, the training data set (consisted of 56-57 patients) was used in leave-one-out format to train the first level of

classifiers. Of 56 (or 57) patients' data, 55 (or 56) were used to train all classifiers in the first level, and the remaining one was used for testing. This was repeated 56 (or 57) times, each time using a different instance to test. The second level meta-classifier was trained according to two different training approaches:

1 Straight meta-classifier training (SMT):

In SMT approach, the outputs of each classifier in the first level of classifiers were concatenated, and along with the corresponding class labels are used to train the second level classifier. For each fold validation, there were 56 (or 57) training instances for the second level meta-classifier.

2 Hierarchical meta-classifier training (HMT):

In this approach, the second level classifier was trained by concatenation of outputs of each classifier in first level of classifiers, along with the corresponding original signal. Similarly as in SMT approach, for each fold validation, there were 56 (or 57) training instances for the second level meta-classifier.

This concluded the initial training of the stacked generalization algorithm.

Once the meta-classifier was trained, all data were pooled. In intermediate training, the first level classifiers were retrained again using all 56 (or 57) data instances. After the training was complete, the test fold was used to validate the trained ensemble. For each fold validation, a test performance was obtained. The 5 fold cross validation process gave 5 test performances, which were then averaged, and the average test performance was declared as the true generalization performance of the algorithm.

This experiment was repeated 10 times and the overall generalization performance (OGP) was obtained as the average of 10 such runs.

4.2 RESULTS

Performances of ensemble systems, based on stacked generalization algorithm, for two-class and three-class diagnostic problems are summarized in this section. The outputs from different data sources (specific electrode – stimuli – frequency band combinations) were fused using different combination rules, and their performances were calculated and compared. For two-class problem, along with generalization performances, commonly used medical diagnostic quantities like sensitivity, specificity, PPV and NPV were also obtained.

The terminologies used in result tables are as follows:

- *OGP* is the overall generalization performance averaged over 10 trials along with its 95% confidence interval.
- *Best 5 trial GP* is the generalization performance obtained from the best 5 trials of those 10 trials, along with its 95% confidence interval over those 5 trials.
- *Best GP* is the best generalization performance obtained from the 10 trials.
- The sensitivity, specificity, PPV and NPV are also listed along with their 95% confidence intervals over 10 trials.

4.2.1 TWO-CLASS PROBLEM

In two-class problem, classification of probable AD patients and cognitively normal subjects was done. Along with performance figures, diagnostic performances like sensitivity, specificity, PPV and NPV were also calculated.

4.2.1a INDIVIDUAL PERFORMANCES BY STRAIGHT META-CLASSIFIER

TRAINING (SMT) APPROACH

This section consists of the results from the stacked generalization ensemble system trained using SMT approach for two-class diagnosis problem. Tables 4.1–4.3 describe the performance figures obtained from target responses from C_Z, F_Z and P_Z electrodes, for 0–1, 1–2, 2–4 and 4–8 Hz spectral bands. The best performing frequency band for each electrode is highlighted in tables.

Table 4.1: Performance values for stacked generalization system trained using SMT approach on target-C_Z coefficients for four frequency bands, for two-class problem.

Target C _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	56.06±3.83	60.19±3.57	64.95	54.41±6.21	57.30±3.38	53.74±4.00	58.00±3.79
1–2 Hz	59.78±3.98	64.19±2.10	66.57	53.24±5.56	65.41±5.05	58.67±4.68	60.43±3.50
2–4 Hz	59.82±5.05	64.82±4.92	71.81	61.47±6.53	58.38±5.84	57.64±4.80	62.39±5.52
4–8 Hz	62.02±3.45	65.89±3.07	68.00	46.47±5.13	75.95±6.00	64.52±5.27	60.68±2.58

Table 4.2: Performance values for stacked generalization system trained using SMT approach on target-F_Z coefficients for four frequency bands, for two-class problem.

Target F _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	60.50±3.09	63.77±4.63	70.19	64.41±4.02	56.76±5.83	58.05±3.30	63.39±3.03
1–2 Hz	58.52±3.49	61.85±5.14	69.14	50.88±5.06	65.14±4.68	57.37±4.74	59.12±2.89
2–4 Hz	58.67±4.63	63.62±6.80	71.90	54.71±7.75	61.89±3.46	56.50±4.70	60.18±4.76
4–8 Hz	57.35±2.98	60.80±2.77	63.52	58.24±4.52	56.22±4.53	55.07±2.83	59.47±3.16

Table 4.3: Performance values for stacked generalization system trained using SMT approach on target-P_Z coefficients for four frequency bands, for two-class problem.

Target P _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	65.13±2.69	67.96±3.50	72.10	61.76±3.57	68.11±6.92	64.75±4.40	65.92±1.76
1–2 Hz	55.53±3.71	59.37±3.72	63.33	55.59±5.56	55.14±7.41	53.54±3.50	57.28±3.87
2–4 Hz	65.93±4.23	70.93±4.82	76.48	66.47±7.36	65.14±6.21	63.82±4.30	68.26±4.46
4–8 Hz	53.56±4.71	57.96±3.75	62.29	49.41±4.73	57.03±6.84	51.70±4.90	54.90±4.44

For target responses, the best performance was obtained from target P_Z electrode for 2–4 Hz spectral band with the OGP of 65.93% $\pm$ 4.23%, and the Best GP of 76.48%. Tables 4.4–4.6 describe the performance figures obtained from novel C_Z, F_Z and P_Z electrodes, for 0–1, 1–2, 2–4 and 4–8 Hz spectral bands. The best novel performance was obtained from P_Z electrode for 2–4 Hz spectral band with the OGP of 71.71% $\pm$ 2.53%.

Table 4.4: Performance values for stacked generalization system trained using SMT approach on novel-C_Z coefficients for four frequency bands, for two-class problem.

Novel C _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	63.28 $\pm$ 3.85	67.31 $\pm$ 4.16	73.24	64.41 $\pm$ 7.04	62.16 $\pm$ 4.46	60.95 $\pm$ 3.53	65.97 $\pm$ 4.86
1–2 Hz	59.02 $\pm$ 7.02	66.11 $\pm$ 11.28	80.38	47.65 $\pm$ 11.0	69.19 $\pm$ 7.46	58.25 $\pm$ 8.40	59.50 $\pm$ 6.39
2–4 Hz	61.99 $\pm$ 4.21	66.59 $\pm$ 4.15	71.81	58.24 $\pm$ 6.25	65.14 $\pm$ 4.77	60.51 $\pm$ 4.39	63.08 $\pm$ 4.16
4–8 Hz	54.00 $\pm$ 4.10	58.48 $\pm$ 1.92	60.57	52.65 $\pm$ 7.38	54.86 $\pm$ 3.16	51.37 $\pm$ 4.36	56.15 $\pm$ 4.27

Table 4.5: Performance values for stacked generalization system trained using SMT approach on novel-F_Z coefficients for four frequency bands, for two-class problem.

Novel F _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	59.21 $\pm$ 2.97	61.81 $\pm$ 2.22	64.86	52.35 $\pm$ 4.73	65.41 $\pm$ 5.05	58.39 $\pm$ 3.38	59.94 $\pm$ 2.77
1–2 Hz	58.78 $\pm$ 5.07	62.70 $\pm$ 10.58	77.71	60.00 $\pm$ 5.26	57.57 $\pm$ 6.94	56.81 $\pm$ 5.61	60.91 $\pm$ 4.42
2–4 Hz	60.92 $\pm$ 2.82	63.54 $\pm$ 4.71	69.14	59.71 $\pm$ 6.72	61.89 $\pm$ 7.02	59.49 $\pm$ 3.74	62.80 $\pm$ 2.95
4–8 Hz	56.20 $\pm$ 3.55	59.85 $\pm$ 5.42	64.95	45.88 $\pm$ 6.44	65.41 $\pm$ 4.15	54.71 $\pm$ 4.32	56.97 $\pm$ 3.10

Table 4.6: Performance values for stacked generalization system trained using SMT approach on novel-P_Z coefficients for four frequency bands, for two-class problem.

Novel P _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	68.35 $\pm$ 4.53	73.33 $\pm$ 5.27	80.19	67.94 $\pm$ 6.14	68.65 $\pm$ 7.24	67.04 $\pm$ 5.17	70.13 $\pm$ 4.45
1–2 Hz	65.02 $\pm$ 3.16	68.32 $\pm$ 4.71	72.10	61.76 $\pm$ 7.80	67.84 $\pm$ 5.11	63.82 $\pm$ 2.76	66.41 $\pm$ 3.86
2–4 Hz	71.71 $\pm$ 2.53	74.59 $\pm$ 2.32	77.71	72.94 $\pm$ 3.40	70.27 $\pm$ 5.46	69.65 $\pm$ 3.48	73.93 $\pm$ 2.07
4–8 Hz	62.71 $\pm$ 3.33	66.57 $\pm$ 2.55	68.95	60.29 $\pm$ 7.09	64.86 $\pm$ 6.93	61.69 $\pm$ 4.23	64.31 $\pm$ 3.61

Comparing the classification performances from target and novel responses, it was observed that performances obtained from novel responses were significantly better. The medical diagnostic performances were also higher for novel responses.

4.2.1b INDIVIDUAL PERFORMANCES BY HIERARCHICAL META-CLASSIFIER TRAINING (HMT) APPROACH

The results from stacked generalization ensemble system trained using HMT approach for individual data sources for two-class diagnosis problem are described in this section. Tables 4.7–4.12 describe the performance figures obtained from target and novel responses from C_Z, F_Z and P_Z electrodes, for 0–1, 1–2, 2–4 and 4–8 Hz spectral bands. The best performing frequency band for each electrode is highlighted in Tables 4.7–4.12.

Table 4.7: Performance values for stacked generalization system trained using HMT approach on target-C_Z coefficients for four frequency bands, for two-class problem.

Target C _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	61.89±3.03	65.39±2.99	68.86	66.47±2.84	57.57±4.82	59.19±3.18	65.04±2.89
1–2 Hz	57.44±2.79	60.00±1.59	62.00	47.94±4.55	65.95±6.26	56.92±4.13	57.91±2.20
2–4 Hz	62.48±3.57	66.00±2.47	68.95	63.24±3.99	61.62±4.53	60.34±3.86	64.58±3.41
4–8 Hz	57.52±4.02	62.19±2.26	64.86	46.18±5.78	67.57±4.46	56.59±5.16	57.82±3.39

Table 4.8: Performance values for stacked generalization system trained using HMT approach on target-F_Z coefficients for four frequency bands, for two-class problem.

Target F _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	59.70±3.97	64.42±4.28	69.24	55.59±4.90	63.24±5.32	58.38±4.83	60.81±3.56
1–2 Hz	61.63±5.07	67.12±6.21	75.90	61.47±7.04	61.62±6.36	59.64±5.05	63.66±5.26
2–4 Hz	60.35±3.92	64.57±5.51	69.14	60.29±5.45	60.27±3.29	58.12±3.76	62.45±4.14
4–8 Hz	60.04±3.85	63.92±5.10	70.76	54.41±5.88	64.86±4.07	58.68±4.00	60.93±3.81

Table 4.9: Performance values for stacked generalization system trained using HMT approach on target- P_z coefficients for four frequency bands, for two-class problem.

Target P_z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	71.00±3.97	75.54±3.23	78.95	70.29±6.61	71.62±5.01	69.59±4.30	72.81±4.48
1–2 Hz	62.72±2.72	65.49±2.81	67.71	59.41±4.18	65.41±5.67	61.56±3.49	63.70±2.51
2–4 Hz	67.55±4.01	72.21±3.98	77.62	63.82±6.03	70.81±4.62	66.85±4.48	68.26±3.97
4–8 Hz	47.50±5.17	52.90±4.93	57.81	48.24±7.02	46.49±4.62	45.11±5.08	49.65±5.34

Table 4.10: Performance values for stacked generalization system trained using HMT approach on novel- C_z coefficients for four frequency bands, for two-class problem.

Novel C_z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	68.71±2.93	72.29±1.52	73.43	65.00±3.89	71.89±5.48	68.39±3.89	69.13±2.47
1–2 Hz	74.20±3.13	77.47±3.90	80.86	66.18±5.36	82.16±2.76	77.26±3.42	72.74±3.51
2–4 Hz	73.21±2.40	75.81±1.42	77.52	74.41±4.21	72.16±5.07	71.42±3.25	75.60±2.63
4–8 Hz	59.79±3.78	63.77±5.15	70.67	57.94±5.34	61.35±6.18	58.21±4.37	61.37±3.33

Table 4.11: Performance values for stacked generalization system trained using HMT approach on novel- F_z coefficients for four frequency bands, for two-class problem.

Novel F_z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	65.51±4.19	69.30±4.00	74.67	61.47±5.73	69.19±3.43	64.57±4.47	66.32±4.13
1–2 Hz	65.09±2.69	68.44±1.87	70.67	57.94±3.29	71.08±6.04	65.44±4.42	64.71±1.87
2–4 Hz	66.50±3.13	69.96±3.21	73.24	61.47±3.63	71.08±4.28	66.32±3.89	66.75±2.61
4–8 Hz	57.25±3.26	61.14±2.97	63.43	47.06±6.72	66.49±4.92	56.36±3.87	57.94±3.18

Table 4.12: Performance values for stacked generalization system trained using HMT approach on novel- P_z coefficients for four frequency bands, for two-class problem.

Novel P_z	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0–1 Hz	68.24±3.86	72.34±4.37	77.52	66.76±6.58	69.19±2.44	66.34±3.41	69.74±4.53
1–2 Hz	72.17±2.25	74.25±1.70	76.38	66.76±2.98	76.76±4.39	72.85±3.29	71.55±1.88
2–4 Hz	72.30±2.03	74.57±2.00	76.38	72.94±4.63	71.35±5.17	70.47±3.37	74.42±2.67
4–8 Hz	62.51±3.47	66.36±4.26	70.86	56.47±5.93	68.11±4.97	62.09±4.38	63.19±3.57

Using this approach, the best performance was obtained from novel C_z electrode for 1–2 Hz spectral band with the OGP value of 74.20% ± 3.13%, and the Best GP of 80.86%. HMT approach also gave better classification performances for novel responses,

particularly for novel C_Z and novel P_Z responses. It can be seen that other diagnostic quantities like sensitivity, specificity, etc. were also higher for novel responses, providing 74.41% sensitivity, 82.16% specificity, 77.26% PPV, and 75.60% NPV. This is one of the most surprising outcomes of this study, as the intention of using novel tones was to improve robustness of P300 component in target responses, rather than the diagnosis of the disease.

Comparing the results obtained from SMT (Tables 4.1–4.6) and HMT (Tables 4.7–4.12) approaches of training, it can be seen that stacked generalization ensemble system trained on HMT approach gave better overall performances, which matches with the previous experiments in [81]. The diagnostic performances were also improved by using HMT approach. Hence for further analyses in this thesis, stacked generalization training was done only by using HMT approach.

Additional data sources, which gave better individual performances in the research of a peer working on a different aspect of this problem [69], were also taken into consideration. Table 4.13 lists the performance figures for those additional sources (electrode –stimuli – frequency band combinations) from the stacked generalization ensemble system trained using HMT approach. The better performances are highlighted in Table 4.13. The 1–2 Hz spectral band from novel T₈ electrode gave the best performance in those additional data sources with the OGP of $72.22\% \pm 2.57\%$, and the Best GP of 76.19%.

Table 4.13: Performance values for stacked generalization system trained using HMT approach on additional data sources, for two-class problem.

Data sources	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Target P ₃ : 1–2Hz	64.63±1.97	66.90±1.61	68.00	65.59±4.86	63.51±5.56	62.66±2.86	66.93±2.23
Target P ₃ : 2–4 Hz	67.21±2.82	70.04±2.99	73.52	67.65±4.75	66.49±4.48	65.13±3.22	69.28±3.10
Target T ₇ : 2–4 Hz	60.09±2.70	62.95±3.37	66.38	58.24±5.13	61.62±4.62	58.30±2.84	61.73±2.70
Target T ₇ : 4–8 Hz	57.57±4.29	61.96±6.85	69.43	56.76±7.48	57.84±3.66	55.02±4.22	59.70±4.69
Target F _{P2} : 1–2 Hz	70.03±2.86	73.39±2.15	74.95	70.59±3.84	69.46±4.47	68.18±3.19	72.07±2.76
Target F ₈ : 1–2 Hz	69.55±3.55	73.12±5.69	80.38	67.94±5.82	70.81±3.95	68.20±3.47	70.91±4.35
Novel O ₂ : 1–2 Hz	62.05±2.97	64.44±1.56	66.38	61.18±4.30	62.43±5.19	60.17±3.28	63.67±3.00
Novel T ₈ : 1–2 Hz	72.22±2.57	74.80±1.64	76.19	74.41±5.15	70.00±2.13	69.44±2.07	75.17±3.66

In addition to classification performances for individual data sources, decision level data fusions were done by fusing the outputs of ensemble system for individual data sources. Data fusions were done to check whether there were any improvements in the overall classification performances by combining the diverse outputs of different data sources.

4.2.1c PERFORMANCES BY DECISION LEVEL DATA FUSION

The outputs of the stacked generalization ensemble system, trained using HMT approach, for individual data sources, were combined using various fusion techniques. For two-class problem, simple majority voting (SMV), weighed majority voting (WMV), product and sum combination rules were used for decision level data fusion.

For each combination rule, all possible combinations of individual data sources were performed. Only those combinations which gave better performances than the best individual performance are included in this report. The best performances for each fusion

rule are listed in this section of results. They are listed in an order with the best performing combination at the bottom of each table. Additional better combinations are listed in Appendix A.

To interpret with the tables for abbreviations of data sources in combinations, following technique is followed: *NCZ12* is interpreted as results from novel- C_Z electrode for 1 to 2 Hz frequency band, and *TPZ24* is interpreted as results from target- P_Z electrode for 2 to 4 Hz frequency band. For each fusion rule, the combinations specified in the first column are the fusion combination of outputs from those data sources.

For SMV rule data fusion, the highest four performances obtained by combining the results of individual data sources from C_Z , F_Z and P_Z electrodes are provided in Table 4.14. The best performance for this setup was obtained by the SMV rule combination of TPZ01, NCZ12, NCZ24, NPZ12, NPZ24, NCZ01, NPZ01, TPZ24 and NFZ24 data sources which gives the OGP of $80.29\% \pm 2.83\%$, and the Best GP of 84.67%.

Table 4.14: Highest four performances from fusion combinations of outputs from data sources of C_Z , F_Z and P_Z electrodes using SMV rule, for two-class problem.

SMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)NCZ12+NCZ24+NPZ12	76.95 $\pm$ 3.27	80.08 $\pm$ 1.39	81.81	71.18 $\pm$ 4.30	82.70 $\pm$ 4.57	79.42 $\pm$ 4.61	75.83 $\pm$ 3.13
(2)NCZ24+NPZ12+NPZ24	77.60 $\pm$ 1.61	79.30 $\pm$ 2.41	82.00	77.06 $\pm$ 2.58	77.84 $\pm$ 3.50	76.37 $\pm$ 2.68	78.77 $\pm$ 1.62
(3)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24	79.83 $\pm$ 3.36	83.56 $\pm$ 2.86	86.00	76.76 $\pm$ 4.48	82.70 $\pm$ 4.57	80.61 $\pm$ 4.54	79.63 $\pm$ 3.51
(4)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24+NCZ01+NPZ01+TPZ24+NFZ24	80.29 $\pm$ 2.83	83.16 $\pm$ 2.16	84.67	75.88 $\pm$ 4.30	84.59 $\pm$ 3.16	82.01 $\pm$ 3.36	79.38 $\pm$ 3.19

Figure 4.1 illustrates the comparison of OGPs of individual data sources with those of above combinations. From the figure, it can be seen that performance values

have improved substantially by performing decision level data fusion. Moreover, the diagnostic performances have also improved by performing data fusion.

Figure 4.1: Comparison of OGPs of SMV rule combinations in Table 4.14 with those of individual data sources, for two-class problem.

For SMV combination rule, using all data sources in this study, the highest five performances obtained are listed in Table 4.15. The best performance in this setup was obtained by the combination of outputs of NCZ12, NCZ24, NPZ24, NT812, NPZ12, TFP212, TPZ01, TF812 and TP324 data sources. For this combination, the OGP obtained was $82.92\% \pm 2.47\%$, and the Best GP was 87.52%.

Table 4.15: Highest five performances from fusion combinations of outputs from all data sources using SMV rule, for two-class problem.

SMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)NCZ24+NPZ24+NT812+NPZ12+TFP212+TF812+TP324+TP312+NPZ48	80.98±1.50	82.65±1.55	84.86	80.88±4.88	81.08±2.58	79.86±1.76	82.56±3.35
(2)NCZ12+NCZ24+NPZ24+NT812+NPZ12+TFP212+TPZ01	81.59±2.85	84.55±2.22	86.76	80.29±4.96	83.24±3.84	81.67±3.81	82.38±3.74
(3)NCZ12+NCZ24+NPZ24+NT812+NPZ12+TFP212+TPZ01+TF812+TP324+TP312+NPZ01	82.28±2.88	85.09±3.59	90.19	80.59±3.86	84.05±4.22	82.56±4.15	82.61±3.05
(4)NCZ12+NCZ24+NPZ24+NT812+NPZ12+TFP212+TPZ01+TF812+TP324+TP312+NPZ01+NPZ48 NO212	82.37±2.40	85.24±0.37	85.52	80.29±4.55	84.59±3.41	82.93±3.36	82.61±3.41
(5)NCZ12+NCZ24+NPZ24+NT812+NPZ12+TFP212+TPZ01+TF812+TP324	82.92±2.47	85.71±2.27	87.52	81.47±3.29	84.59±3.41	83.12±3.21	83.36±2.60

Figure 4.2 illustrates the comparison of OGP's of individual data sources with those of decision level fusion combinations described in Table 4.15. From the figure, it can be seen that performance values have also increased significantly by data fusion.

Figure 4.2: Comparison of OGP's of SMV rule combinations in Table 4.15 with those of individual data sources, for two-class problem.

For WMV rule data fusion, the highest five performances, obtained by combining the results of individual data sources from C_z , F_z and P_z electrodes, are provided in Table 4.16. The best performance in this case, was obtained by the WMV combination of the outputs of TPZ01, NCZ12 and NPZ24 data sources which gave the OGP of $80.20\% \pm 3.94\%$, and the Best GP of 87.43% . Figure 4.3 does the OGP comparison.

Table 4.16: Highest five performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using WMV rule, for two-class problem.

WMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)NCZ12+NCZ24+NPZ24	78.24 ± 2.56	80.69 ± 2.61	84.00	73.53 ± 3.71	82.70 ± 4.48	80.00 ± 4.26	77.38 ± 2.66
(2)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24	79.08 ± 2.23	81.45 ± 2.28	84.57	75.00 ± 4.23	82.70 ± 2.27	79.97 ± 2.15	78.43 ± 2.88
(3)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24+NCZ01+NPZ01+TPZ24+NPZ24	79.40 ± 3.48	83.01 ± 2.13	85.33	73.53 ± 5.61	85.14 ± 4.58	82.32 ± 5.04	78.03 ± 3.90
(4)TPZ01+NCZ12+NCZ24+NPZ24	79.84 ± 2.75	82.76 ± 2.89	85.90	77.06 ± 4.30	82.70 ± 3.55	80.55 ± 3.57	79.85 ± 3.29
(5)TPZ01+NCZ12+NPZ24	80.20 ± 3.94	84.51 ± 3.31	87.43	75.59 ± 4.44	84.59 ± 5.31	82.21 ± 5.67	79.08 ± 3.68

Figure 4.3: Comparison of OGPs of WMV rule combinations in Table 4.16 with those of individual data sources, for two-class problem.

For WMV rule, using all data sources in this study, the highest six performing combinations are listed in Table 4.17. The best performance in this setup was obtained by the combination of outputs of NPZ12, NPZ24, NCZ24, NT812, TFP212, TPZ01 and NCZ12 data sources with the OGP of $83.73\% \pm 3.49\%$. Figure 4.4 illustrates the OGP comparison.

Table 4.17: Highest six performances from fusion combinations of outputs from all data sources using WMV rule, for two-class problem.

WMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812 +TPZ01+NCZ12+NPZ01	81.12±3.17	84.65±4.13	89.43	78.24±4.67	84.05±3.21	81.87±3.57	80.93±3.71
(2)NPZ12+NPZ24+NCZ24 +NT812+TFP212+NCZ12	81.73±2.86	84.51±4.36	90.19	80.29±4.55	83.24±3.38	81.61±3.30	82.33±3.65
(3)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TPZ01	82.10±2.66	84.91±3.77	88.76	81.47±5.34	82.43±2.77	81.06±2.29	83.26±4.27
(4)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TPZ01 +NCZ12+NPZ01	82.20±3.01	85.71±2.43	88.29	79.71±4.90	84.86±3.05	82.89±3.41	82.18±3.85
(5)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812 +TPZ01+NCZ12	82.91±3.47	86.78±4.58	92.48	81.18±4.97	84.59±4.37	83.10±4.28	83.21±3.96
(6)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TPZ01 +NCZ12	83.73±3.49	87.87±2.60	89.71	82.35±5.24	85.41±3.99	83.94±4.37	84.27±4.47

Figure 4.4: Comparison of OGPs of WMV rule combinations in Table 4.17 with those of individual data sources, for two-class problem.

For product rule data fusion, the highest five performances, obtained by combining the results of individual data sources from C_z , F_z and P_z electrodes, are provided in Table 4.18. The best performance in this case, was obtained by the combination of TPZ01, NCZ12, NCZ24, NPZ12, NPZ24, NPZ01, TPZ24 and NFZ24 data sources, which gave the OGP of $80.31\% \pm 1.89\%$, and the Best GP of 84.19%.

Table 4.18: Highest five performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using product rule, for two class problem.

Product rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)TPZ01+NCZ12+NCZ24	78.91±3.35	82.32±3.28	86.29	76.18±4.90	81.89±4.82	79.80±5.00	79.11±3.59
(2)NCZ12+NCZ24+NPZ12+NPZ24	78.97±2.99	82.42±2.56	84.67	74.12±4.52	83.24±3.73	80.45±3.81	77.93±3.16
(3)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24	79.94±2.36	82.13±2.17	84.57	75.59±4.32	83.78±2.41	81.13±2.44	79.07±2.98
(4)TPZ01+NCZ12+NPZ24	79.99±3.99	84.36±4.00	86.95	76.47±4.95	83.78±4.73	81.41±5.20	79.56±3.91
(5)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24+NPZ01+TPZ24+NFZ24	80.31±1.89	82.42±1.53	84.19	74.12±4.30	86.22±3.21	83.43±3.14	78.57±2.52

Figure 4.5 illustrates the comparison of OGP values for individual data sources and product rule combinations in Table 4.18.

Figure 4.5: Comparison of OGPs of product rule combinations in Table 4.18 with those of individual data sources, for two-class problem.

For product rule, using all data sources in this study, the highest five performances obtained are listed in Table 4.19. Figure 4.6 illustrates the OGP comparison of these combinations with individual data sources. The best performance in this setup was obtained by product combination of outputs of NPZ12, NPZ24, NCZ24, NT812, TFP212, TF812, TPZ01 and NCZ12 data sources, giving the of OGP of 83.19% ± 3.55%, and the Best GP of 90.86%.

Table 4.19: Highest five performances from fusion combinations of outputs from all data sources using product rule, for two-class problem.

Product rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)NPZ12+NPZ24+NCZ24+NT812+TFP212+TF812	80.16±2.57	82.84±2.99	86.10	79.71±4.37	80.54±2.85	79.08±2.62	81.42±3.44
(2)NPZ12+NPZ24+NCZ24+NT812+TFP212+TF812+TPZ01	80.87±2.87	83.71±4.65	87.33	81.47±4.76	80.27±2.58	79.13±2.49	82.75±3.95
(3)NPZ12+NPZ24+NCZ24+NT812+TFP212+TF812+NCZ12	81.19±2.71	84.25±2.15	86.00	80.00±3.94	82.16±3.18	80.57±3.22	81.85±3.13
(4)NPZ12+NPZ24+NCZ24+NT812+TFP212+TF812+TPZ01+NCZ12+NPZ01	81.48±2.65	84.38±2.63	86.86	79.12±4.26	84.05±3.21	82.14±3.44	81.58±3.35
(5)NPZ12+NPZ24+NCZ24+NT812+TFP212+TF812+TPZ01+NCZ12	83.19±3.55	87.31±2.84	90.86	82.65±5.29	83.78±4.17	82.56±4.20	84.24±4.33

Figure 4.6: Comparison of OGPs of product rule combinations in Table 4.19 with those of individual data sources, for two-class problem.

For sum rule data fusion, the highest five performances, obtained by combining the results of individual data sources from C_z , F_z and P_z electrodes, are provided in Table 4.20. The best performance in this case, is obtained by the combination of outputs of TPZ01, NCZ12, NCZ24, NPZ12, NPZ24, NPZ01, TPZ24 and NFZ24 data sources which gave the OGP of $83.00\% \pm 2.94\%$, and the Best GP of 88.29% . Figure 4.7 illustrates the OGP comparison.

Table 4.20: Highest five performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using sum rule, for two-class problem.

Sum rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)TPZ01+NCZ12+NCZ24	77.52 ± 3.58	81.14 ± 5.28	85.90	75.29 ± 5.17	79.73 ± 3.89	77.45 ± 3.78	78.04 ± 4.00
(2)NCZ12+NCZ24	77.68 ± 2.97	81.05 ± 3.70	85.90	75.59 ± 4.86	79.73 ± 4.58	77.74 ± 3.86	78.26 ± 3.43
(3)NCZ12+NCZ24+NPZ12+NPZ24	78.70 ± 2.28	80.97 ± 3.20	85.43	74.12 ± 4.30	82.97 ± 3.41	80.20 ± 3.13	77.90 ± 2.83
(4)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24	81.16 ± 2.32	83.68 ± 1.97	85.90	76.18 ± 4.14	85.68 ± 2.74	83.12 ± 2.80	79.81 ± 2.87
(5)TPZ01+NCZ12+NCZ24+NPZ12+NPZ24+NPZ01+TPZ24+NFZ24	83.00 ± 2.94	86.27 ± 2.06	88.29	78.24 ± 4.45	87.57 ± 3.66	85.44 ± 4.00	81.55 ± 3.12

Figure 4.7: Comparison of OGPs of sum rule combinations in Table 4.20 with those of individual data sources, for two-class problem.

For sum rule, using all data sources in this study, the highest five performances obtained are listed in Table 4.21. Figure 4.8 illustrates the OGP comparison of these combinations with individual data sources. The best performance in this setup was obtained by combination of outputs of NPZ12, NPZ24, NCZ24, NT812, TFP212, TF812, TPZ01 and NCZ12 data sources, with the OGP of $85.65\% \pm 2.14\%$, and the Best GP of 91.62%.

Table 4.21: Highest five performances from fusion combinations of outputs from all data sources using sum rule, for two-class problem.

Sum rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
(1)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812	82.11 $\pm$ 2.50	84.86 $\pm$ 2.27	87.43	82.65 $\pm$ 4.80	81.62 $\pm$ 3.95	80.78 $\pm$ 3.34	84.01 $\pm$ 3.82
(2)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812 +NCZ12	82.57 $\pm$ 3.13	86.19 $\pm$ 2.21	88.76	81.18 $\pm$ 3.01	84.32 $\pm$ 4.71	82.92 $\pm$ 4.55	82.95 $\pm$ 2.69
(3)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812 +TPZ01+NCZ12+NPZ01	83.35 $\pm$ 3.44	86.61 $\pm$ 2.59	89.43	80.59 $\pm$ 4.87	86.49 $\pm$ 4.73	84.88 $\pm$ 5.22	83.07 $\pm$ 3.72
(4)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812 +TPZ01	84.72 $\pm$ 2.45	87.14 $\pm$ 4.17	91.14	85.59 $\pm$ 4.37	84.59 $\pm$ 4.65	83.99 $\pm$ 3.82	86.72 $\pm$ 3.42
(5)NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812 +TPZ01+NCZ12	85.65 $\pm$ 2.14	87.77 $\pm$ 2.93	91.62	84.12 $\pm$ 4.34	87.57 $\pm$ 4.39	86.63 $\pm$ 4.43	85.98 $\pm$ 3.06

Figure 4.8: Comparison of OGPs of sum rule combinations in Table 4.21 with those of individual data sources, for two-class problem.

Out of all these fusion combinations for classification of the two-class problem, the best results were obtained by sum rule data fusion. The sum rule data fusion of the outputs of NPZ12, NPZ24, NCZ24, NT812, TFP212, TF812, TPZ01 and NCZ12 data sources gave the best classification performances (the OGP value of $85.65\% \pm 2.14\%$, and the Best GP of 91.62%). Moreover in general case, fusion combinations of data sources gave better classification performances than those of individual data sources. The medical diagnostic performances were also significantly improved by data fusion. The idea behind the improved classification performances lies in the fact that, different data sources may have significantly diverse class related information, and their combinations by data fusion increased the overall classification performances. For two-class problem, most of the best performing combinations are obtained by fusion of data

sources from P_Z, C_Z, T₈, F_{P2} and F₈ electrodes. Hence, decision level data fusion is definitely a better approach to increase the accuracy of classification in the two-class diagnostic problem.

4.2.2 THREE-CLASS PROBLEM

In this problem, classification of mild-moderate AD patients, mild AD patients and cognitively normal subjects was done. This problem had mild AD patients, who were in the earliest stage of the disease. Their MMSE scores ($\mu_{\text{MMSE}}=27.1$) were in the range of that obtained from the normal cohort. Hence patients in mild AD group behaved more or less similarly as those in normal group. This similarity in the cohort made this diagnostic problem, a substantially difficult one.

4.2.2a INDIVIDUAL PERFORMANCES

The results from the stacked generalization ensemble system trained using HMT approach, on individual data sources, for three-class problem are described in this section. Tables 4.22–4.27 describe the performance figures obtained from target and novel responses from C_Z, F_Z and P_Z electrodes, for 0–1, 1–2, 2–4 and 4–8 Hz spectral bands, for three-class problem. The best performing frequency band for each electrode is highlighted in Tables 4.22–4.27.

Table 4.22: Performance values for stacked generalization system, trained using HMT approach on target-C_Z coefficients for four frequency bands, for three-class problem.

Target C _Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)
0–1 Hz	54.76±2.44	57.60±1.97	59.43
1–2 Hz	51.11±3.49	54.46±4.03	59.05
2–4 Hz	50.63±2.94	53.60±4.05	59.14
4–8 Hz	48.78±4.19	52.88±5.02	59.14

Table 4.23: Performance values for stacked generalization system, trained using HMT approach on target- F_Z coefficients for four frequency bands, for three-class problem.

Target F_Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)
0–1 Hz	53.88±3.08	57.20±2.71	59.43
1–2 Hz	51.54±3.62	55.60±3.13	57.90
2–4 Hz	48.29±2.71	51.16±3.45	54.19
4–8 Hz	52.27±3.10	55.50±3.39	59.43

Table 4.24: Performance values for stacked generalization system, trained using HMT approach on target- P_Z coefficients for four frequency bands, for three-class problem.

Target P_Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)
0–1 Hz	51.50±3.39	55.22±4.38	59.43
1–2 Hz	45.77±3.49	49.56±3.87	55.05
2–4 Hz	56.74±1.48	58.13±1.92	60.76
4–8 Hz	43.07±3.26	46.19±1.97	48.10

Table 4.25: Performance values for stacked generalization system, trained using HMT approach on novel- C_Z coefficients for four frequency bands, for three-class problem.

Novel C_Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)
0–1 Hz	56.43±3.81	60.10±2.11	62.19
1–2 Hz	58.87±5.80	64.69±8.01	74.29
2–4 Hz	56.85±4.10	61.16±3.15	63.43
4–8 Hz	44.61±2.80	47.54±4.03	52.29

Table 4.26: Performance values for stacked generalization system, trained using HMT approach on novel- F_Z coefficients for four frequency bands, for three-class problem.

Novel F_Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)
0–1 Hz	55.12±2.45	57.98±1.09	59.24
1–2 Hz	58.50±2.92	61.30±0.76	62.38
2–4 Hz	50.56±2.77	53.37±3.27	57.81
4–8 Hz	47.68±5.80	54.02±7.80	62.29

Table 4.27: Performance values for stacked generalization system, trained using HMT approach on novel- P_Z coefficients for four frequency bands, for three-class problem.

Novel P_Z	OGP (%)	Best 5 trial GP (%)	Best GP (%)
0–1 Hz	55.03±2.40	57.75±2.68	60.95
1–2 Hz	58.84±3.15	62.23±2.71	65.05
2–4 Hz	57.65±4.34	62.25±6.51	70.67
4–8 Hz	48.44±3.21	51.71±1.91	53.43

In this approach, the best performance was obtained from novel C_Z electrode for 1–2 Hz spectral band with the OGP of $58.87\% \pm 5.80\%$, and the Best GP of 74.29%. Similarly as in two-class problem, the three-class problem also gave better performances for novel responses, particularly for novel C_Z and novel P_Z responses.

Additional data sources, which gave better individual performances in the research of a peer working on a different aspect of this problem [69], were utilized in this problem as well. Table 4.28 lists the performance figures for those additional sources (specific electrode –stimuli – frequency band combinations) from the stacked generalization algorithm trained using HMT approach. The better performances are highlighted in Table 4.28. The 1–2 Hz spectral band from target F_{P2} electrode gave the best performance in those additional data sources with the OGP of $58.83\% \pm 2.62\%$, and the Best GP of 63.62%.

Table 4.28: Performance values for stacked generalization system, trained using HMT approach on additional data sources, for three-class problem.

Target	OGP (%)	Best 5 trial GP (%)	Best GP (%)
Target P ₃ : 1–2 Hz	49.81±4.17	54.72±3.87	57.81
Target P ₃ : 2–4 Hz	57.74±2.50	60.38±2.77	63.33
Target T ₇ : 2–4 Hz	53.84±3.28	57.30±3.71	60.76
Target T ₇ : 4–8 Hz	50.51±2.09	52.88±0.96	53.81
Target F _{P2} : 1–2 Hz	58.83±2.62	61.60±2.31	63.62
Target F ₈ : 1–Hz	47.28±1.87	49.41±2.20	52.29
Novel O ₂ : 1–2 Hz	50.35±4.22	55.22±5.21	62.48
Novel T ₈ : 1–2 Hz	51.04±2.92	54.30±2.63	57.90

In comparison with the two-class problem, the classification performances in the three-class problem were poor. This was due to the increased level of classification difficulty in the three-class problem, as discussed earlier.

Decision level data fusions were also done in the three-class problem to check whether the combinations of individual decisions caused any improvements in the classification performances.

4.2.2b PERFORMANCES BY DECISION LEVEL DATA FUSION

The outputs of the stacked generalization ensemble system, trained using HMT approach, for individual data sources, were combined using various fusion techniques. In three-class problem, weighed majority voting (WMV), product and sum combination rules were used for decision level data fusion.

Similarly as in two-class problem, for each combination rule, all possible combinations of individual data sources were performed. Only those combinations which gave better performances than the best individual performance are included in this report. The best performances for each fusion rule are listed in this section of results. They are listed in an order with the best performing combination at the bottom of each table. Additional better combinations are listed in Appendix B. The technique used in two-class data fusion was repeated in this section to interpret with the abbreviations of the data sources for combinations.

For WMV rule data fusion, the highest five performances obtained by combining the results of individual data sources from C_Z, F_Z and P_Z electrodes are provided in Table 4.29. The best performance in this setup was obtained by the combination of outputs of TPZ24, NCZ01, NCZ12, NCZ24, NFZ01, NFZ12, NPZ12 and NPZ24 data sources which gave the OGP of $64.81\% \pm 2.53\%$, and the Best GP of 72.00%.

Table 4.29: Highest five performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using WMV rule, for three-class problem.

WMV rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
(1)IPZ24+NCZ01+NCZ12+NCZ24+NFZ12+NPZ12+NPZ24	64.23±2.73	66.78±3.69	72.00
(2)IPZ24+NCZ12+ NCZ24+NFZ12+NPZ12+NPZ24	64.68±2.69	67.66±2.80	70.57
(3)IPZ24+NCZ01+NCZ12+NFZ12+NPZ12+NPZ24	64.78±2.17	67.03±2.74	69.43
(4)IPZ24+NCZ01+NCZ12+NCZ24+NFZ01+NFZ12+NPZ01+NPZ12+NPZ24	64.80±3.04	67.92±4.02	73.43
(5)IPZ24+NCZ01+NCZ12+NCZ24+NFZ01+NFZ12+NPZ12+NPZ24	64.81±2.53	67.35±3.93	72.00

The OGPs of WMV rule combinations and those of individual data sources are compared in Figure 4.9. From the figure, it can be seen that performance values have significantly increased by data fusion.

Figure 4.9: Comparison of OGPs of WMV rule combinations in Table 4.29 with those of individual data sources, for three-class problem.

For WMV rule, using all data sources in this study, the highest three performances obtained are listed in Table 4.30. The best performance in this setup was obtained from the combination of outputs of NPZ12, NPZ24, NCZ24, TT724, TFP212, TPZ24, NCZ01 and NCZ12 data sources with the OGP of $67.01\% \pm 1.44\%$. Figure 4.10 illustrates the OGP comparison.

Table 4.30: Highest three performances from fusion combinations of outputs from all data sources using WMV rule, for three-class problem.

WMV rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
(1)NPZ12+NPZ24+NCZ24+TT724+TFP212+TPZ24	65.36 ± 2.51	67.77 ± 3.65	71.90
(2)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+TPZ24+NCZ01+NCZ12	66.86 ± 1.58	68.42 ± 0.93	69.24
(3)NPZ12+NPZ24+NCZ24+TT724+TFP212+TPZ24+NCZ01+NCZ12	67.01 ± 1.44	68.42 ± 1.98	70.76

Figure 4.10: Comparison of OGPs of WMV rule combinations in Table 4.30 with those of individual data sources, for three-class problem.

For product rule data fusion, the highest three performances, obtained by combining the results of individual data sources from C_Z , F_Z and P_Z electrodes, are provided in Table 4.31. The best performance in this case was obtained by the combination of outputs of TPZ24, NCZ01, NCZ12, NCZ24, NPZ12 and NPZ24 data sources which gave the OGP of $69.85\% \pm 3.89\%$, and Best GP of 77.81%.

Table 4.31: Highest three performances from fusion combinations of outputs from data sources of C_Z , F_Z and P_Z electrodes using product rule, for three-class problem.

Product rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
(1)TPZ24+NCZ01+NCZ12+NFZ12+NPZ12+NPZ24	67.15 ± 2.63	70.13 ± 3.00	72.19
(2)TPZ24+NCZ01+NCZ12+NCZ24+NFZ12+NPZ12+NPZ24	68.27 ± 1.97	70.11 ± 3.13	73.52
(3)TPZ24+NCZ01+NCZ12+NCZ24+NPZ12+NPZ24	69.85 ± 3.89	74.08 ± 3.23	77.81

Figure 4.11 illustrates the comparison of OGP values for individual data sources and those of product rule combinations described in Table 4.31.

Figure 4.11: Comparison of OGPs of product rule combinations in Table 4.31 with those of individual data sources, for three-class problem.

For product rule, using all data sources in this study, the highest five performances obtained are listed in Table 4.32. The best performance in this setup was obtained from the combination of outputs of NPZ12, NPZ24, NCZ24, TP324, TT724, TFP212, NCZ01 and NCZ12 data sources with the OGP of $69.89\% \pm 2.45\%$. Figure 4.12 illustrates the OGP comparison.

Table 4.32: Highest five performances from fusion combinations of outputs from all data sources using product rule, for three-class problem.

Product rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
(1)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+TPZ24	67.84 ± 3.46	71.07 ± 4.88	76.29
(2)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+NCZ01	67.85 ± 2.94	71.41 ± 3.00	73.52
(3)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+NCZ12	68.61 ± 2.83	71.66 ± 2.19	73.62
(4)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+TPZ24+NCZ01+NCZ12	69.86 ± 3.64	73.89 ± 3.78	78.57
(5)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+NCZ01+NCZ12	69.89 ± 2.45	72.51 ± 3.45	76.10

Figure 4.12: Comparison of OGPs of product rule combinations in Table 4.32 with those of individual data sources, for three-class problem.

For sum rule data fusion, the highest five performances, obtained by combining the results of individual data sources from C_Z , F_Z and P_Z electrodes, are provided in Table 4.33. The best performance in this case, was obtained by the sum combination of outputs of TPZ24, NCZ01, NCZ12, NCZ24, NFZ12, NPZ12 and NPZ24 data sources which gave the OGP of $69.43\% \pm 2.52\%$. Figure 4.13 does the OGP comparison.

Table 4.33: Highest five performances from fusion combinations of outputs from data sources of C_Z , F_Z and P_Z electrodes using sum rule, for three-class problem.

Sum rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
(1)TPZ24+NCZ01+NCZ12+NCZ24+NFZ01+NFZ12+NPZ01+NPZ12+NPZ24	68.89 ± 1.97	71.03 ± 2.25	73.62
(2)TPZ24+NCZ01+NCZ12+NCZ24+NFZ01+NFZ12+NPZ12+NPZ24	69.16 ± 1.59	70.97 ± 1.45	72.19
(3)TPZ24+NCZ12+NCZ24+NPZ12+NPZ24	69.27 ± 2.89	71.81 ± 3.61	76.38
(4)TPZ24+NCZ01+NCZ12+NFZ12+NPZ12+NPZ24	69.28 ± 2.91	71.75 ± 2.30	74.86
(5)TPZ24+NCZ01+NCZ12+NCZ24+NFZ12+NPZ12+NPZ24	69.43 ± 2.52	71.54 ± 2.28	73.52

Figure 4.13: Comparison of OGPs of sum rule combinations in Table 4.33 with those of individual data sources, for three-class problem.

For sum rule, using all data sources in this study, the highest six performances obtained are listed in Table 4.34. The best performance in this setup was obtained from the combination of outputs of NPZ12, NPZ24, NCZ24, TT724, TFP212, TPZ24, NCZ01 and NCZ12 data sources with the OGP of $71.34\% \pm 2.94\%$. Figure 4.14 illustrates the OGP comparison.

Table 4.34: Highest five performances from fusion combinations of outputs from all data sources using sum rule, for three-class problem.

Sum rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
(1)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+TPZ24	69.75 ± 2.90	72.80 ± 2.69	74.86
(2)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+NCZ01	69.93 ± 2.39	72.63 ± 2.55	76.10
(3)NPZ12+NPZ24+NCZ24+TT724+TFP212+TPZ24	70.21 ± 2.77	73.18 ± 2.68	76.38
(4)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+NCZ12	70.73 ± 2.90	74.21 ± 2.80	77.43
(5)NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212+TPZ24+NCZ01+NCZ12	70.81 ± 2.08	73.12 ± 2.22	74.76
(6)NPZ12+NPZ24+NCZ24+TT724+TFP212+TPZ24+NCZ01+NCZ12	71.34 ± 2.94	74.23 ± 0.97	74.86

Figure 4.14: Comparison of OGPs of sum rule combinations in Table 4.34 with those of individual data sources, for three-class problem.

Out of all these fusion combinations for classification in the three-class problem, the best results were obtained by sum rule data fusion. The sum rule data fusion of outputs of NPZ12, NPZ24, NCZ24, TT724, TFP212, TPZ24, NCZ01 and NCZ12 data sources gave the best classification performance (OGP value of $71.34\% \pm 2.94\%$). In general, fusion combinations gave better classification performances than that of individual data sources. For three-class problem, most of the best performing combinations are obtained by fusion of data sources from P_Z, P₃, C_Z, T₇, and F_{P2} electrodes. Hence as in two-class problem, decision level data fusion is a better approach to improve the accuracy of classification in the three-class diagnostic problem as well.

CHAPTER 5

CONCLUSIONS

The application of this work consisted of an automated approach to obtain diagnostic identification of AD vs. normal patients based on their ERP recording. The problem of diagnosis of AD is widely recognized as a difficult problem, not only for machine learning algorithms, but for most neurophysiologists and neuropsychologists. The difficulty of the problem lies in the early diagnosis of the disease, during which the symptoms are often not much different than those that are associated with normal aging.

5.1 SUMMARY OF ACCOMPLISHMENTS

The proposed approach has a combination of some well-established techniques. This included oddball paradigm – used to invoke event related potentials (ERPs) in the EEG signals, ERP analysis using multiresolution wavelets, and the most recent developments in machine learning, the ensemble systems. ERP signals from task related target tones and unexpected novel tones were decomposed by wavelet decomposition in a set of coefficients for successive frequency bands. The classification was achieved by a sophisticated ensemble system, the stacked generalization algorithm, for all data sources (electrode – stimuli – frequency band combinations). The ensemble training was done using 5 fold cross validation technique to obtain a reasonably good estimate of the true generalization performance. This study addresses two diagnostic problems: a two-class problem (AD vs. normal subjects), and a three-class problem (mild-moderate AD vs.

mild AD vs. normal subjects). Implementation of three-class problem, further classified AD patients into mild AD patients and mild-moderate AD patients, which helped in the diagnosis of AD at the earliest possible stage.

For both two-class and three-class problems, the best individual band classification was obtained from 1–2 Hz spectral band of novel responses, acquired from C_Z electrode. One of the most surprising outcomes of this study was the ability of novel tones to discriminate the ERPs of different groups. The performances obtained with novel tones were significantly better than the performances obtained with target tones.

In addition to individual performances, decision level data fusions were also performed. The decision level data fusion by sum combination rule gave the best performance values for both, two-class and three-class problems. In general, performances by decision level data fusion were higher than those from individual data sources, and sometimes by wide margins. This demonstrates the feasibility of ensemble system in the diagnosis of early AD. The improved classification performances were achieved by the combination of information from different data sources, which have diverse class related information in their outputs. Most of these data sources came from C_Z, P_Z, P₃, T₇, T₈, F_{P2} and F₈ electrodes, whose locations are shown in Figure 5.1. In early stage AD patients, symptoms like memory loss, confusion about the location of familiar places, and poor judgment which are associated with functions of temporo-parietal region, and other symptoms like decline in cognitive abilities, mood and personality changes and increased anxiety, associated with functions of frontal region are detected. The selection of the electrodes described above makes sense, as they come

from temporal, central-parietal and frontal regions, which are generally affected in early stage AD.

The results obtained in this study are quite satisfactory by computational intelligence perspective. This is because the ensemble generalization performances are in 80% range, which exceeds the 75% diagnostic performances obtained by longitudinal clinical evaluations by trained physician at community clinics/ hospitals.

Figure 5.1: Relation of combinations of electrodes with physiological activities in brain

Overall, the results presented in this work are clinically significant because EEG-based automated classification system is non-invasive, accurate, cost-effective than longitudinal clinical evaluations, and can be easily implemented in community clinics, where most patients are initially evaluated.

5.2 SOURCES OF ERROR

The patients recruited for this study were diagnosed using clinical evaluation. This type of evaluation has an accuracy of around 80–90%. The classification algorithms used in this study were trained as if this was 100% (the “gold standard”). The original misdiagnosis of a test subject is a potential source of error in this study. The only way to obtain a diagnosis with 100% accuracy is through an autopsy. Inclusion of post mortem analyses of test subjects to obtain the true “gold standard” would minimize this error.

5.3 FUTURE WORK

Additional electrodes should be explored in details, for both two-class and three-class problems, to find more significant data sources which might help to improve the classification performances. Moreover different configuration of stacked generalization algorithm by trying different base classifier models, should also be explored. Future work should also assess the performance of this algorithm in distinguishing between AD and different types of non-AD patients, specifically patients with other forms of dementia such as vascular dementia (VaD), Parkinson’s disease, etc.

Since several fusion combinations gave better classification performances, it is recommended to get diagnostic results based on voting of those different results. The diversity in each case may improve the overall classification performance. Moreover, output from each combination can be weighted based on its individual performance to strategically increase the overall accuracy of the diagnosis.

REFERENCES

- [1] Alzheimer's disease International, "Statistics," [online], 2006:
<<http://www.alz.co.uk/research/statistics.html>>, Last accessed: Apr. 2007.
- [2] Alzheimer's Association, "Alzheimer's disease Statistics," [online], 2006:
<<http://www.alz.org/Resources/FactSheets/FSAzheimerstats.pdf>>, Last accessed:
Nov. 2006.
- [3] National Institute of Aging, "Alzheimer's Disease education and referral center,"
[online], 2006:
<<http://www.nia.nih.gov/Alzheimers/AlzheimersInformation/GeneralInfo/>>, Last
accessed: Apr. 2007.
- [4] M. Hashimoto, E. Rockenstein, L. Crews and E. Masliah, "Role of protein aggregation
in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's
diseases," *Neuromolecular Med.*, 2003; 4(1-2): 21-36.
- [5] Alzheimer's Disease Research, a program of the American Health Assistance
Foundation, "Alzheimer's Disease Research," [online], 2006:
<<http://www.ahaf.org/alzdis/about/AmyloidPlaques.htm>>, Last accessed: Apr. 2007.
- [6] Alzheimer's Disease Research, a program of the American Health Assistance
Foundation, "How the Brain and Nerve Cells Change during Alzheimer's Disease,"
[online], Nov 2006:
<http://www.ahaf.org/alzdis/about/Brain_Neurons_AD_Normal.htm>, Last accessed:
Apr. 2007.
- [7] Help-guide, "Aging Issues — Diagnosis and Risk," [online], 2006:
<http://www.helpguide.org/elder/alzheimers_diagnosis_risk_factors.htm>, Last
accessed: Apr. 2007.
- [8] C. Jack, D. Dickson, J. Parisi, Y. Xu, R. Cha, P. O'Brien, S. Edland, G. Smith, B.
Boeve, E. Tangalos, E. Kokmen and R. Petersen, "Antemortem MRI findings
correlate with hippocampal neuropathology in typical aging and dementia,"
Neurology, 2002; 58(5): 750-757.

-
- [9] M. Bobinski, M. de Leon, J. Wegiel, S. Desanti, A. Convit, L. Saint Louis, H. Rusinek and H. Wisniewski, "The histological validation of postmortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease," *Neuroscience*, 2000; 95(3): 721-725.
- [10] Help-guide, "Aging Issues- Expert, Non-commercial Information on Mental Health & Lifelong Wellness," [online], 2006:
<http://www.helpguide.org/elder/ahzheimers_dementia.htm>, Last accessed Jan. 2007.
- [11] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E.M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology*, 1984; 34: 939-944.
- [12] J. Morris and A. Price, "Pathologic correlates of nondemented aging, mild cognitive impairment, and early-stage Alzheimer's disease," *J. Mol. Neurosci.*, 2001; 2: 101-118.
- [13] W. Goldman, J. Price, M. Storandt, E. Grant, D. McKeel, E. Rubin and J. Morris, "Absence of cognitive impairment or decline in preclinical Alzheimer's disease," *Neurology*, 2001; 56(3): 361-367.
- [14] Z. Khachaturian, "The five-five, ten-ten plan for Alzheimer's disease [editorial]," *Neurobiology of Aging*, 1992; 13: 197-198.
- [15] T. Radebaugh, N. Buckholtz and Z. Khachaturian, "Fisher symposium: strategies for the prevention of Alzheimer's disease – overview of research planning meeting III," *Alzheimer Disease & Associated Disorders*, 1996; 10(1): 1-5.
- [16] A. Lim, W. Kukull, D. Nochlin, J. Leverenz, W. McCormick, J. Bowen, L. Teri, J. Thompson, E. Peskind, M. Raskind and E. Larson "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series," *Journal of the American Geriatrics Society*, 1999; 47(5): 564-569.
- [17] J. Growdon, "Biomarkers of Alzheimer's Disease," *Archives of Neurology*, 1999; 56: 281-283.
- [18] G. Reckess, "Antecedent Biomarkers in Alzheimer's Disease," [online], 2003:
<<http://www.alzforum.org/res/enab/workshops/biomarkers.asp>>, Last accessed: Apr. 2007.

-
- [19] N. Andreasen,, E. Vanmechelen, A. Van de Voorde, P. Davidsson, C. Hesse, S. Tarvonen, I. Raiha, L. Sourander, B. Winblad and K. Blennow, "Cerebrospinal fluid tau protein as a biochemical marker for Alzheimer's disease: a community based follow up study," *Journal of Neurology, Neurosurgery, and Psychiatry*, 1998; 64(3): 298-305.
- [20] L. Kalaugher, "Nanosensor could help Alzheimer's diagnosis," [online], 2004: <<http://nanotechweb.org/articles/news/3/5/6/1>>, Last accessed: Apr. 2007.
- [21] D. Praticò, V. Lee, J. Trojanowski, J. Rokach and G. Fitzgerald. "Increased F₂-isoprostanes in Alzheimer's disease: evidence for enhanced lipid peroxidation in-vivo," *FASEB Journal*, 1998; 12(15):1777-1783.
- [22] M. Valenzuela and P. Sachdev, "Magnetic resonance spectroscopy in AD," *Neurology*, 2001; 56:592-598.
- [23] C. Babiloni, *et al.*, "Mapping distributed sources of cortical rhythms in mild Alzheimer's disease, A multicentric EEG study," *Neuroimage*, 2004; 22:57-67.
- [24] H. Jasper, "Report of the Committee on Methods of Clinical Examination in Electroencephalography," *Electroenceph. Clin. Neurophysiol.*, 1958; 10: 370-371.
- [25] R. Cooper, J. Osselton and J. Shaw, "EEG Technology," *Butterworths, London*, 1969; 2nd Edition: 275.
- [26] F. Sharbrough, G. Chatrian, R. Lesser, H. Lüders, M. Nuwer and T. Picton, "American Electroencephalographic Society Guidelines for Standard Electrode Position," *J. Clin. Neurophysiol.*, 1991; 8: 200-202.
- [27] J. Malmivuo and R. Plonsey, "Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields," *Oxford University Press, New York*, 1995; 13. Available online at <<http://butler.cc.tut.fi/~malmivuo/bem/bembook/13/13.htm>>, Last accessed: Apr. 2007.
- [28] S. Louis, "EEG waves as defines by frequency," [online]: <http://www.brown.edu/Departments/Clinical_Neurosciences/louis/eegfreq.html>, Last accessed Apr. 2007.
- [29] C. Cripe, " Brainwaves and EEG – The language of the brain," [online]: <<http://www.crossroadsinstitute.org/eeg.html>>, Last accessed Apr. 2007.
- [30] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clinical Neurophysiology*, 2004; 115: 1490-505.

-
- [31] M. Penntila, J Partanen, H, Soininen and P Riekkinen, "Quantitative analysis of occipital EEG in different stages of Alzheimer's disease," *Electroencephalography and clinical Neurophysiology*, 1985; 60(1): 1-6.
- [32] U. Schreiter-Gasser, T. Gasser and P. Zeigler, "Quantitative EEG analysis in early onset Alzheimer's disease: correlations with severity, clinical characteristics, visual EEG and CCT," *Electroencephalography and clinical Neurophysiology*, 1994; 90: 267-272.
- [33] R. Quiroga, "Quantitative analysis of EEG signals: Time-frequency methods and Chaos theory," *Thesis: Institute of Physiology and Institute of Signal Processing - Medical University Lubeck*, 1998.
- [34] E. Niedermeyer and F. Lopez da Silva, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields," *Lippincott, Williams and Wilkins, Philadelphia*, 2005; 5th Edition.
- [35] S. Yamaguchi, H. Tsuchiya, S. Yamagata, G. Toyoda and S. Kobayashi, "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, 2000; 111: 195-203.
- [36] J. Polich, "P300 and Alzheimer's disease," *Biomed. Pharmacother*, 1989; 43: 493-499.
- [37] J. Ford, N. Askari, J. Gabrieli, D. Mathalon, J. Tinklenberg, V. Menon and J. Yesavage, "Event-Related Brain Potential Evidence of Spared Knowledge in Alzheimer's Disease," *Psychology and Aging*, 2001; 16: 161-176.
- [38] J. Olichney and D. Hillert, "Clinical applications of cognitive event-related potentials in Alzheimer's disease," *Phys. Med. Rehabil. Clin. N. Am.*, 2005; 15: 205-233.
- [39] D.J. Linden, "The p300: where in the brain is it produced and what does it tell us?" *Neuroscientist*, 2005; 11: 563-576.
- [40] J. Polich and A. Kok, "Cognitive and biological determinants of P300: an integrative review," *Biological Psychology*, 1995; 41: 103-146.
- [41] T. Frodl, H. Hampel, G. Juckel, K. Bürger, F. Padberg, R. Engel, H. Möller and U. Hegerl, "Value of event-related P300 subcomponents in the clinical diagnosis of mild cognitive impairment and Alzheimer's Disease," *Psychophysiology*, 2002; 39: 175-181.

-
- [42] J. Polich, C. Ladish and F. Bloom, "P300 assessment of early Alzheimer's disease," *Electroencephalogr. Clin. Neurophysiol.*, 1990; 77: 179-189.
- [43] J. Polich and L. Hoffman, "Alzheimer's Disease and P300: evaluation of modality and task difficulty," *Brain Topography Today*, 1998: 527-536.
- [44] H. Gandhi, D. Green, J. Kounios, C. Clark and R. Polikar, "Stacked Generalization for Early Diagnosis of Alzheimer's Disease," *28th Annual Int. Conf. of IEEE Engineering in Medicine and Biology Society (EMBS)*, 2006; 1: 5350-5353.
- [45] J. Benvenuto, Y. Jin, M. Casale, G. Lynch and R. Granger, "Identification of diagnostic evoked response potential segments in Alzheimer's disease," *Exp. Neurol.*, 2002; 176: 269-276.
- [46] N. Phillips, H. Chertkow, M. Leblanc, H. Pim and S. Murtha, "Functional and anatomical memory indices in patients with or at risk for Alzheimer's disease," *J. Int. Neuropsychol. Soc.*, 2004; 10: 200-210.
- [47] M. Signorino, E. Pucci, E. Brizioli, G. Cacchio, G. Nolfè and F. Angeleri, "EEG power spectrum typical of vascular dementia in subgroup of Alzheimer patients," *Archives of Gerontology and Geriatrics*, 1996; 23: 139-151.
- [48] E. Pucci, G. Cacchio, R. Angeloni, N. Berladinelli, G. Nolfè, M. Signorino and F. Angeleri, "EEG spectral analysis in Alzheimer's disease and different degenerative dementias," *Archives of Gerontology and Geriatrics*, 1998; 26: 283-291.
- [49] E. Pucci, N. Belardinelli, G. Cacchio, M. Signorino and F. Angeleri, "EEG power spectrum differences in early and late onset forms of Alzheimer's disease," *Clinical Neurophysiology*, 1999; 110: 621-631.
- [50] G. Rodriguez, F. Copello, P. Vitali, G. Perego and F. Nobili, "EEG spectral profile to stage Alzheimer's disease," *Clinical Neurophysiology*, 1999; 110: 1831-1837.
- [51] S. Hibino, T. Hanai, E. Nagata, M. Matsubara, K. Fukagawa, T. Shirataki, H. Honda and T. Kobayashi, "Fuzzy Neural Network Model for Assessment of Alzheimer-Type Dementia," *Journal of Chemical Engineering of Japan*, 2001; 34(7): 936-942.
- [52] C. Besthorn, H. Forstl, C. Geiger-Kabisch, H. Sattel, T. Gasser and U. Schreiter-Gasser, "EEG coherence in Alzheimer's disease", *Electoencepalogr. and Clinical Neurophysiology*, 1994; 90: 242-245.
- [53] T. Locatelli, M. Cursi, D. Liberati, M. Franceschi and G. Comi, "EEG coherence in Alzheimer's disease", *Electoencepalogr. and Clinical Neurophysiology*, 1998; 108: 229-237.

-
- [54] M. Hogan, G. Swanwick, J. Kaiser, M. Rowan, B. Lawlor, "Memory-related EEG power and coherence reductions in mild Alzheimer's disease," *International journal of Psychophysiology*, 2003; 112: 1378-1387.
- [55] M. Brunovsky, M. Matouseka, A. Edmanc, K. Cervenaa and V. Krajcab, "Objective Assessment of the Degree of Dementia by Means of EEG," *Neuropsychobiology*, 2003; 48: 19-26.
- [56] A. Ademoglu, T. Demiralp, J. Yordanova, V. Kolev and M. Devrim, "Decomposition of event-related brain potentials into multicomponents using wavelet transform," *Applied Signal Processing*, 1998; 5: 142-151.
- [57] T. Demiralp, J. Yordanova; V. Kolev; A. Ademoglu, M. Devrim and V. Samar, "Time-Frequency Analysis of Single-Sweep Event-Related Potentials by Means of Fast Wavelet Transform," *Brain and Language*, 1999; 66: 129-145.
- [58] T. Demiralp, A. Ademoglu, M. Schurmann, C. Baser-Eroglu and E. Baser, "Detection of P300 waves in single trials by the wavelet transform (WT)," *Brain and Language*, 1999; 66:108-128.
- [59] R. Quiroga, O. Sakowitz, E. Basar and M. Schurmann, "Wavelet Transform in the analysis of the frequency composition of evoked potentials," *Brain Research Protocols*, 2001; 8: 16-24.
- [60] A. Petrosian, D. Prokhorov, W. Lajara-Nanson and R. Schiffer, "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG," *Clinical Neurophysiology*, 2001; 112: 1378-1387.
- [61] R. Polikar, F. Keinert and M. Greer, "Wavelet analysis of event related potentials for early diagnosis of Alzheimer's disease," *A. Petrosian and F. Meyer (Ed.), Wavelets in Signal and Image Analysis, From Theory to Practice, Kluwer Academic Publishers, Boston*, 2001: 453-478.
- [62] S. Yagneswaran, M. Baker and A. Petrosian, "Power frequency and wavelet characteristics in differentiating between normal and Alzheimer EEG," *24th Annual Conference of IEEE Engineering in Medicine and Biology, and the Fall Meeting of the Biomedical Engineering Society (BMES / EMBS)*, 2002; 1: 46-47.
- [63] G. Jacques, J. Frymiare, J. Kounios, C. Clark and R. Polikar, "Multiresolution wavelet analysis and ensemble of classifiers for early diagnosis of Alzheimer's disease," *Proc. of 30th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2005; 5: 389-392.

-
- [64] N. Stepenosky, J. Kounios, C. Clark and R. Polikar, "Ensemble Techniques with Weighted Combination Rules for Early Diagnosis of Alzheimer's Disease," *IEEE Int. Joint Conf. on Neural Networks (IJCNN)*, 2006: 1935-1942.
- [65] R. Polikar, A. Topalis, D. Green, J. Kounios and C. Clark, "Ensemble based data fusion for early diagnosis of Alzheimer's disease," *Information Fusion*, accepted / in print for 2006. Preprint available online at <http://users.rowan.edu/~polikar/RESEARCH/PUBLICATIONS/if07.pdf>.
- [66] Alzheimer's Disease Research Center, "Clinical Dementia Rating," [online], 2006: <http://alzheimer.wustl.edu/cdr/default.htm> Last accessed: Apr. 2007.
- [67] Tufts-New England Medical Center: Department of Psychiatry, "The Mini Mental State Examination," [online]: <http://www.tufts-nemc.org/psych/mmse.asp>, Last accessed: Apr. 2006.
- [68] B. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho and K. Sundaresan, "An exploratory study of factors affecting single trial P300 detection," *IEEE Trans. on Biomedical Eng.*, 2004; 51: 975-978.
- [69] B. Balut, "An ensemble-based data fusion approach that extracts complementary information from multichannel event related potentials for early diagnosis of neurodegenerative disorders," *Masters' Thesis, Rowan University*, 2007.
- [70] R. Polikar, "The story of wavelets," *N. Mastorakis (Ed.), Physics and Modern Topics in Mechanical and Electrical Engineering, World Scientific and Eng. Society Press*, 1999: 192-197.
- [71] R. Polikar, "The wavelet tutorial – The engineer's ultimate guide to wavelet analysis," [online], Jan. 2001: <http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>, Last accessed Apr. 2007.
- [72] R. Duda, P. Hart and D. Stork, "Pattern Classification," *John Wiley and Sons Inc., New York*, 2001, 2nd Edition.
- [73] R. Polikar, "Pattern Recognition," *M. Akay (Ed.), Wiley Encyclopedia of Biomedical Engineering, Wiley, New York*, 2006.
- [74] L. I. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms," *Wiley, New York*, 2004.
- [75] L. Breiman, "Bagging predictors," *Machine Learning*, 1996; 24(2): 123-140.

-
- [76] R. Schapire, "The strength of weak learnability," *Machine Learning*, 1990; 5: 197-227.
- [77] Y. Freund and R. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," *Comput. Syst. Sci.*, 1997; 57(1):119-139.
- [78] D.H. Wolpert, "Stacked generalization," *Neural Networks*, 1992; 5(2): 241-260.
- [79] R. Jacobs, M. Jordan, S. Nowlan, and G. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, 1991; 3:79-87.
- [80] M. Jordan and R. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, 1994; 6(2):181-214.
- [81] P. Chan, "An Extensive Meta-Learning Approach for Scalable and Accurate Inductive Learning," *PhD Thesis, Columbia University*, 1996.
- [82] L. Breiman, "Stacked Regressions," *Machine Learning*, 1996; 24: 41-48.
- [83] K. Ting and M. Witten, "Issues in stacked generalization," *Journal of Artificial Intelligence Research (JAIR)*, 1999; 10: 271-289.
- [84] A. Ghorbani and K. Owrangh, "Stacked Generalization in Neural Networks: Generalization on Statically Neutral Problems," *IEEE/INNS International Joint conference on Neural Networks (IJCNN)*, 2001; 1:1715-1720.
- [85] L. Kuncheva, J. Bezdek and R. Duin, "Decision templates for multiple classifier fusion: An experimental comparison," *Pattern Recognition*, 2001; 34(2): 299-314.
- [86] Polikar R., "Esemble based systems in decision making," *IEEE Circuits and Systems Mag.*, 2006; 6(3): 21-45.

APPENDIX A

ADDITIONAL TWO-CLASS FUSION RESULTS

Additional performances from fusion combinations for two-class problem in lieu with Section 4.2.1c are listed in Tables A.1–A.8.

Table A.1: Additional performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using SMV rule, for two-class problem.

SMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
TPZ01+NCZ12+NCZ24	76.41±3.45	80.11±3.24	83.05	72.94±4.52	80.00±4.19	77.19±3.98	76.40±3.46
TPZ01+NCZ24+NPZ12	76.59±3.00	79.90±3.71	83.05	75.88±3.94	77.57±4.82	75.92±3.90	77.86±3.01
TPZ01+NCZ24+NPZ24	76.88±2.91	79.70±4.13	85.52	77.35±4.65	76.49±4.99	75.48±3.76	78.82±3.30

Table A.2: Additional performances from fusion combinations of outputs from all data sources using SMV rule, for two-class problem.

SMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
NPZ24+NPZ48+NO212+NCZ24+TP312+TF812+TT748	71.55±2.33	73.96±1.68	76.38	71.47±4.86	71.35±3.99	69.82±2.87	73.39±2.95
NPZ48+NO212+NT812+TP312+TP324+TFP212+TT724	72.31±2.37	74.51±3.26	78.76	71.18±6.01	73.24±3.80	71.09±2.36	73.84±3.31
NPZ24+NPZ48+NO212+NT812+TP312+TT724+TT748	71.89±3.03	75.07±3.27	77.90	69.71±7.93	73.51±5.67	71.33±4.34	73.19±4.34
NPZ12+NPZ48+NO212+NCZ24+NT812+TP312+TP324	75.75±1.10	76.80±1.98	79.14	74.41±4.32	76.76±3.78	74.90±2.20	76.85±2.46
NPZ24+NPZ48+NO212+TP312+TP324+TFP212+TT724	71.49±2.39	73.90±1.91	75.90	71.76±5.62	71.08±4.08	69.72±2.68	73.61±3.41
NPZ24+NPZ48+NO212+NCZ24+TF812+TFP212+TT748	74.33±1.75	76.21±1.26	77.52	72.94±5.50	75.41±3.80	73.35±1.93	75.60±2.96
NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TT748	74.15±3.34	77.43±2.52	80.19	75.00±5.80	73.24±5.93	72.54±4.53	76.47±4.04
NPZ24+NPZ48+NO212+NT812+TP312+TP324+TT724	74.31±3.16	77.58±2.92	80.57	73.82±6.38	74.59±3.78	72.85±3.33	76.08±4.31
NPZ24+NPZ48+NO212+NCZ24+TF812+TT724+TT748	70.11±2.62	72.51±4.59	79.05	66.47±4.34	73.24±4.68	69.83±3.36	70.50±2.51
NPZ24+NPZ48+NO212+NCZ24+TP312+TP324+TFP212	75.29±3.77	79.24±3.52	83.43	76.47±5.43	74.05±7.12	73.69±4.70	77.64±3.90
NCZ24+NPZ24+NT812+NPZ12+TFP212+TF812+TP324+TP312+NPZ48+NO212+TT724	78.70±1.86	80.67±2.39	83.33	77.65±3.73	79.46±4.66	78.05±3.31	79.66±2.12
NCZ24+NPZ24+NT812+NPZ12+TFP212+TF812+TP324	79.56±1.81	81.58±1.94	83.43	81.18±3.86	77.84±2.19	77.14±1.60	82.03±2.88

NCZ24+NPZ24+NT812 +NPZ12+TFP212	79.35±2.58	81.85±1.22	83.24	81.47±3.58	77.30±3.18	76.81±2.79	82.09±3.17
NCZ12+NCZ24+NPZ24 +NT812+NPZ12	80.00±2.15	82.21±2.74	85.52	78.53±1.99	81.62±3.62	79.87±3.44	80.51±1.79
NCZ12+NCZ24+NPZ24+NT812 +NPZ12+TFP212+TPZ01+TF812 +TP324+TP312+NPZ01+NPZ48 +NO212+TT724+TT748	78.97±2.85	81.30±3.65	86.10	75.88±4.63	81.62±3.26	79.25±3.21	78.85±3.40

Table A.3: Additional performances from fusion combinations of outputs from data sources of C_Z , F_Z and P_Z electrodes using WMV rule, for two-class problem.

WMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
NCZ12+NCZ24	78.03±2.95	81.49±3.14	85.90	75.88±5.03	80.27±4.82	78.35±4.26	78.62±3.50
NCZ12+NPZ24	77.33±4.66	82.36±2.95	85.62	70.00±4.41	85.14±7.52	82.29±8.89	75.43±3.81
TPZ01+NCZ12+NCZ24	78.11±3.40	81.87±3.42	84.48	75.00±4.67	81.35±4.86	79.07±4.82	78.15±3.48
TPZ01+NCZ12 +NCZ24+NPZ12	78.18±3.05	81.33±2.74	84.19	73.53±3.43	82.70±4.48	79.94±4.67	77.31±2.62
TPZ01+NCZ24 +NPZ12NPZ24	78.25±2.73	80.88±4.03	84.67	74.12±5.85	81.89±1.83	78.94±1.86	77.85±3.91

Table A.4: Additional performances from fusion combinations of outputs from all data sources using WMV rule, for two-class problem.

WMV rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
NPZ24+NPZ48+NO212+NCZ24 +TP312+TF812+TT748	73.75±3.42	77.22±5.23	83.52	72.35±6.81	74.59±3.66	72.40±2.95	75.10±4.63
NPZ48+NO212+NT812+TP312 +TP324+TFP212+TT724	71.70±3.00	74.78±1.81	76.38	70.00±4.18	72.97±4.99	70.74±4.06	72.69±2.92
NPZ24+NPZ48+NO212+NT812 +TP312+TT724+TT748	72.31±2.90	75.41±3.79	79.43	71.47±4.09	72.70±5.86	71.1±4.00	73.60±2.57
NPZ12+NPZ48+NO212+NCZ24 +NT812+TP312+TP324	73.66±3.20	77.68±1.17	78.95	69.41±3.60	77.30±4.66	74.07±4.41	73.35±2.73
NPZ24+NPZ48+NO212+TP312 +TP324+TFP212+TT724	71.28±2.75	74.53±2.31	77.62	69.71±5.25	72.43±4.62	70.16±3.46	72.52±3.38
NPZ24+NPZ48+NO212+NCZ24 +TF812+TFP212+TT748	74.82±2.91	78.21±2.84	80.76	70.88±5.47	78.11±4.59	75.18±3.85	74.79±3.66
NPZ24+NPZ48+NO212+NCZ24 +NT812+TP312+TT748	73.30±1.91	75.43±2.08	77.90	72.65±4.44	73.51±3.26	71.71±2.09	74.75±2.55
NPZ24+NPZ48+NO212+NT812 +TP312+TP324+TT724	73.51±2.64	76.70±0.94	77.71	71.76±3.32	74.86±3.87	72.56±3.26	74.31±2.59
NPZ24+NPZ48+NO212+NCZ24 +TF812+TT724+TT748	73.27±2.39	75.71±3.51	78.76	68.24±5.32	77.57±4.47	74.03±3.65	72.93±3.04
NPZ24+NPZ48+NO212+NCZ24 +TP312+TP324+TFP212	75.25±3.18	78.78±2.32	81.90	73.82±6.38	76.22±4.44	74.24±3.53	76.54±4.60
NPZ12+NPZ24+NPZ48+NO212 +NCZ24+NT812+TP312+TP324 +TT724+TT748+TFP212+TF812	79.11±2.20	81.45±2.04	83.33	76.18±3.05	81.62±2.70	79.28±2.69	78.92±2.27
NPZ12+NPZ24+NCZ24 +NT812+TFP212+TF812	79.96±1.96	81.83±3.12	86.19	79.12±3.77	80.54±3.13	79.03±2.57	80.95±2.70
NPZ12+NPZ24+NCZ24+NT812 +TFP212+TF812+TPZ01	80.59±2.87	83.87±3.58	86.95	81.47±5.25	79.73±2.92	78.73±2.62	2.74±4.27
NPZ12+NPZ24+NCZ24+T812 +TFP212+TF812+NCZ12	80.75±1.87	82.78±2.42	86.00	79.12±3.63	82.16±2.91	80.45±2.48	81.23±2.61

Table A.5: Additional performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using product rule, for two-class problem.

Product rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
TPZ01+NCZ12	76.50±2.97	80.00±1.42	81.62	70.29±5.19	82.16±3.31	78.47±3.35	75.27±3.40
NCZ12+NCZ24	78.04±2.98	81.05±4.08	85.90	75.88±5.03	80.27±4.08	78.18±3.56	78.61±3.62
NPZ12+NPZ24	76.22±3.22	79.85±4.20	83.33	74.71±4.77	77.30±5.92	75.74±4.94	77.07±3.20
NCZ12+NCZ24+NPZ12+NPZ24	78.97±2.99	82.42±2.56	84.67	74.12±4.52	83.24±3.73	80.45±3.81	77.93±3.16

Table A.6: Additional performances from fusion combinations of outputs from all data sources using product rule, for two-class problem.

Product rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
NPZ24+NPZ48+NO212+NCZ24+TP312+TF812+TT748	73.82±3.16	77.41±2.97	80.67	72.94±6.17	74.32±4.48	72.55±3.50	75.37±4.31
NPZ48+NO212+NT812+TP312+TP324+TFP212+TT724	72.20±3.13	75.31±2.06	77.62	69.41±4.56	74.59±4.29	71.74±3.96	72.76±3.27
NPZ24+NPZ48+NO212+NT812+TP312+TT724+TT748	73.25±2.59	75.96±3.59	80.86	72.06±4.12	74.05±5.48	72.29±3.73	74.41±2.52
NPZ12+NPZ48+NO212+NCZ24+NT812+TP312+TP324	74.26±2.82	77.50±1.20	78.86	70.00±3.40	78.11±4.86	75.01±4.31	73.95±2.44
NPZ24+NPZ48+NO212+TP312+TP324+TFP212+TT724	71.28±2.29	73.94±1.95	76.10	68.82±5.62	73.24±3.80	70.42±2.55	72.23±3.25
NPZ24+NPZ48+NO212+NCZ24+TF812+TFP212+TT748	74.46±3.41	78.36±3.90	82.76	71.76±6.88	76.76±3.66	74.03±3.56	75.29±5.05
NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TT748	74.10±2.62	76.50±4.39	82.19	72.94±4.52	74.86±3.03	72.76±2.45	75.26±3.15
NPZ24+NPZ48+NO212+NT812+TP312+TP324+TT724	73.23±2.51	76.13±2.03	77.52	70.88±3.77	75.14±3.95	72.53±3.03	73.83±2.54
NPZ24+NPZ48+NO212+NCZ24+TF812+TT724+TT748	73.25±2.80	75.96±2.84	79.24	70.00±6.63	75.95±4.50	73.09±3.52	73.87±4.06
NPZ24+NPZ48+NO212+NCZ24+TP312+TP324+TFP212	77.45±3.54	80.99±4.38	84.67	75.88±4.30	78.65±4.68	76.79±4.31	78.12±3.55
NPZ12+NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TP324+TT724+TT748+TFP212+TF812	79.11±2.04	81.47±1.48	83.33	76.47±3.71	81.35±2.80	79.12±2.38	79.16±2.60

Table A.7: Additional performances from fusion combinations of outputs from data sources of C_Z , F_Z and P_Z electrodes using sum rule, for two-class problem.

Sum rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
TPZ01+NCZ12	75.65±4.05	80.29±1.75	81.81	70.00±5.76	80.81±5.34	77.40±5.09	74.74±4.02
NPZ12+NPZ24	76.05±3.08	79.54±3.30	82.00	72.94±4.06	78.65±5.50	76.40±4.81	76.05±2.76
TPZ01+NPZ01+NCZ01	74.75±2.87	77.85±2.33	80.48	75.00±5.27	74.32±3.05	72.88±2.83	76.69±3.77
TPZ01+NPZ12+NCZ12	76.70±2.69	79.68±3.11	83.24	69.71±5.78	83.24±2.54	79.32±2.43	75.28±3.78
TPZ01+NCZ12+TPZ24	77.10±3.94	81.58±3.93	84.48	72.35±6.88	81.62±2.37	78.14±3.23	76.66±4.80

Table A.8: Additional performances from fusion combinations of outputs from all data sources using sum rule, for two-class problem.

Sum rule Combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
NPZ24+NPZ48+NO212+NCZ24+TP312+TF812+TT748	74.50±3.40	77.92±5.11	84.57	73.53±7.28	75.14±3.50	73.10±2.70	76.15±4.89
NPZ48+NO212+NT812+TP312+TP324+TFP212+TT724	71.94±2.25	74.25±2.28	77.52	69.41±4.67	74.05±3.99	71.29±2.80	72.70±2.68
NPZ24+NPZ48+NO212+NT812+TP312+TT724+TT748	73.10±2.72	75.89±4.01	81.52	70.59±7.99	75.14±5.29	72.82±3.89	74.31±4.67
NPZ12+NPZ48+NO212+NCZ24+NT812+TP312+TP324	76.32±1.79	78.19±1.52	80.38	73.82±3.89	78.38±4.37	76.22±3.30	76.69±2.23
NPZ24+NPZ48+NO212+TP312+TP324+TFP212+TT724	71.07±2.88	73.64±2.25	76.29	68.53±6.43	73.24±3.91	70.29±3.00	72.08±3.63
NPZ24+NPZ48+NO212+NCZ24+TF812+TFP212+TT748	75.95±2.69	78.90±2.11	80.57	72.35±7.36	79.19±2.58	76.16±1.54	76.34±4.41
NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TT748	74.04±3.15	77.54±3.42	82.10	74.12±6.63	73.78±4.18	72.34±3.01	76.13±4.40
NPZ24+NPZ48+NO212+NT812+TP312+TP324+TT724	74.03±3.16	77.26±2.82	80.10	71.18±5.41	76.49±4.18	73.79±3.89	74.51±3.64
NPZ24+NPZ48+NO212+NCZ24+TF812+TT724+TT748	74.90±2.97	77.85±2.35	80.48	69.41±6.36	79.73±3.31	75.99±3.16	74.32±3.78
NPZ24+NPZ48+NO212+NCZ24+TP312+TP324+TFP212	75.36±3.97	79.39±2.52	81.71	76.18±5.19	74.59±6.39	73.85±5.04	77.50±3.99
NPZ12+NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TP324+TT724+TT748+TFP212+TF812	79.59±3.08	82.61±2.81	85.52	76.76±5.00	82.16±3.18	79.88±3.30	79.58±3.61

APPENDIX B

ADDITIONAL THREE-CLASS FUSION RESULTS

Additional performances for fusion combinations for three-class problem in lieu with Section 4.2.2b are listed in Tables B.1–B.6.

Table B.1: Additional performances from fusion combinations of outputs from data sources of C_Z , F_Z and P_Z electrodes using WMV rule, for three-class problem.

WMV rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
TPZ24+NCZ01+NCZ12	61.70±2.99	64.95±3.55	67.81
NPZ01+NPZ12+NPZ24	60.12±3.04	63.03±3.66	67.90
TPZ24+NCZ24+NFZ12+NPZ24	62.65±3.34	65.90±4.25	71.90

Table B.2: Additional performances from fusion combinations of outputs from all data sources using WMV rule, for three-class problem.

WMV rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
NPZ24+NPZ48+NO212+NCZ24+TP312+TF812+TT748	59.49±2.10	61.62±2.63	64.95
NPZ48+NO212+NT812+TP312+TP324+TFP212+TT724	60.30±1.41	61.60±2.51	64.76
NPZ24+NPZ48+NO212+NT812+TP312+TT724+TT748	59.01±2.99	62.08±2.77	65.90
NPZ12+NPZ48+NO212+NCZ24+NT812+TP312+TP324	61.01±1.56	62.70±1.56	63.62
NPZ24+NPZ48+NO212+TP312+TP324+TFP212+TT724	61.01±2.73	64.11±3.44	67.81
NPZ24+NPZ48+NO212+NCZ24+TF812+TFP212+TT748	60.75±1.94	62.99±1.91	64.86
NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TT748	60.35±2.30	63.05±0.93	63.71
NPZ24+NPZ48+NO212+NT812+TP312+TP324+TT724	62.53±3.58	66.10±2.85	69.24
NPZ24+NPZ48+NO212+NCZ24+TF812+TT724+TT748	60.97±3.80	65.12±4.49	70.48
NPZ24+NPZ48+NO212+NCZ24+TP312+TP324+TFP212	62.13±2.94	65.22±2.90	67.81
NPZ12+NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TP324+TT724+TT748+TFP212+TF812	63.87±1.14	65.03±1.69	66.48
NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212	64.82±2.23	67.24±2.03	69.33

Table B.3: Additional performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using product rule, for three-class problem.

Product rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
TPZ24+NCZ01+NCZ12	63.41±3.07	66.40±5.22	73.62
TPZ24+NCZ24+NFZ12+NPZ24	63.29±3.50	67.54±2.88	70.67
TPZ24+NCZ12+NCZ24+NPZ12+NPZ24	66.81±3.38	70.55±0.91	71.33
TPZ24+NCZ12+NCZ24+NFZ12+NPZ12+NPZ24	66.03±2.12	68.13±2.86	72.10
TPZ24+NCZ01+NCZ12+NCZ24+NFZ01+NFZ12+NPZ01+NPZ12+NPZ24	66.89±2.83	69.87±4.04	74.95

Table B.4: Additional performances from fusion combinations of outputs from all data sources using product rule, for three-class problem.

Product rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
NPZ24+NPZ48+NO212+NCZ24+TP312+TF812+TT748	55.41±2.48	57.94±1.78	59.52
NPZ48+NO212+NT812+TP312+TP324+TFP212+TT724	58.45±2.63	61.58±2.66	63.52
NPZ24+NPZ48+NO212+NT812+TP312+TT724+TT748	58.96±3.65	63.35±3.32	67.81
NPZ12+NPZ48+NO212+NCZ24+NT812+TP312+TP324	60.42±2.48	63.24±2.90	66.38
NPZ24+NPZ48+NO212+TP312+TP324+TFP212+TT724	61.27±2.00	63.05±0.98	63.81
NPZ24+NPZ48+NO212+NCZ24+TF812+TFP212+TT748	57.37±2.19	59.92±0.90	60.76
NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TT748	59.20±2.36	61.37±2.32	63.62
NPZ24+NPZ48+NO212+NT812+TP312+TP324+TT724	60.45±2.71	63.03±2.66	65.05
NPZ24+NPZ48+NO212+NCZ24+TF812+TT724+TT748	58.08±4.30	62.19±4.10	67.81
NPZ24+NPZ48+NO212+NCZ24+TP312+TP324+TFP212	60.68±2.22	63.22±1.78	64.76
NPZ12+NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TP324+TT724+TT748+TFP212+TF812	63.17±2.42	65.83±3.27	67.81
NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212	66.75±2.86	69.70±2.89	73.24

Table B.5: Additional performances from fusion combinations of outputs from data sources of C_z , F_z and P_z electrodes using sum rule, for three-class problem.

Sum rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
NCZ01+NCZ12+NCZ24	61.08±3.53	64.76±4.44	69.24
NPZ01+NPZ12+NPZ24	64.16±2.86	67.26±1.98	69.24
TPZ24+NCZ24+NFZ12+NPZ24	66.86±3.09	70.38±2.47	73.33
TPZ24+NCZ24+NPZ12+NPZ24	67.95±3.11	71.52±1.92	73.43

Table B.6: Additional performances from fusion combinations of outputs from all data sources using sum rule, for three-class problem.

Sum rule combinations	OGP (%)	Best 5 trial GP (%)	Best GP (%)
NPZ24+NPZ48+NO212+NCZ24+TP312+TF812+TT748	59.63±2.62	61.96±2.35	65.14
NPZ48+NO212+NT812+TP312+TP324+TFP212+TT724	60.30±2.03	62.44±2.81	64.95
NPZ24+NPZ48+NO212+NT812+TP312+TT724+TT748	60.50±3.29	63.90±4.42	69.33
NPZ12+NPZ48+NO212+NCZ24+NT812+TP312+TP324	62.45±1.33	63.89±1.43	65.14
NPZ24+NPZ48+NO212+TP312+TP324+TFP212+TT724	63.07±2.62	65.73±3.39	69.14
NPZ24+NPZ48+NO212+NCZ24+TF812+TFP212+TT748	62.02±2.52	64.70±2.21	66.48
NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TT748	61.62±2.41	63.89±3.26	67.71
NPZ24+NPZ48+NO212+NT812+TP312+TP324+TT724	62.12±2.27	64.44±2.68	67.90
NPZ24+NPZ48+NO212+NCZ24+TF812+TT724+TT748	62.30±3.67	66.38±2.62	69.05
NPZ24+NPZ48+NO212+NCZ24+TP312+TP324+TFP212	64.92±2.62	67.49±3.20	70.76
NPZ12+NPZ24+NPZ48+NO212+NCZ24+NT812+TP312+TP324+TT724+TT748+TFP212+TF812	65.70±2.39	67.79±4.14	73.52
NPZ12+NPZ24+NCZ24+TP324+TT724+TFP212	69.75±1.91	71.71±2.27	74.86