

Multimodal EEG, MRI and PET Data Fusion for Alzheimer's Disease Diagnosis

Robi Polikar, *Senior Member, IEEE*, Christopher Tilley, Brendan Hillis, Chris M. Clark

Abstract— Alarming increasing prevalence of Alzheimer's disease (AD) due to the aging population in developing countries, combined with lack of standardized and conclusive diagnostic procedures, make early diagnosis of Alzheimer's disease a major public health concern. While no current medical treatment exists to stop or reverse this disease, recent dementia specific pharmacological advances can slow its progression, making early diagnosis all the more important. Several non-invasive biomarkers have been proposed, including P300 based EEG analysis, MRI volumetric analysis, PET based metabolic activity analysis, as alternatives to neuropsychological evaluation, the current gold standard of diagnosis. Each of these approaches, have shown some promising outcomes, however, a comprehensive data fusion analysis has not yet been conducted to investigate whether these different modalities carry complementary information, and if so, whether they can be combined to provide a more accurate analysis. In this effort, we provide a first look at such an analysis in combining EEG, MRI and PET data using an ensemble of classifiers based decision fusion approach, to determine whether a strategic combination of these different modalities can improve the diagnostic accuracy over any of the individual data sources when used with an automated classifier. Results show an improvement of up to 10%-20% using this approach compared to the classification performance obtained when using each individual data source.

I. INTRODUCTION

Alzheimer's disease is a neurodegenerative disorder, causing neuronal death that leads to cognitive function decline. Two misfolded proteins, β -amyloid that causes plaques and hyperphosphorylated- τ that causes neurofibrillary tangles are often blamed, yet, the genesis of these proteins, and in fact the true cause of the disease, are still unknown.

Perhaps due to overall life expectancy increase, the number of AD cases has been growing over the last decades. Vast majority of AD patients are over the age of 65, with 19% between 75 and 84, and 42% ~50% above the age of 85 [1]. With over 5.2 million people suffering from AD in the U.S. alone, along with its enormous financial (over \$140 billion / year) and emotional cost of the disease on the patient, family, and society, Alzheimer's disease is now justly considered a major health concern. Yet, there is still no procedure for conclusively and definitively diagnosing AD, and misdiagnosis is quite common. The only definitive diagnosis remains the post-mortem analysis of the brain tissue under the microscope for the presence of plaques and tangles.

The most common method of pre-mortem diagnosis, then, is neuropsychological clinical evaluation, performed over the course of a year or longer, which includes behavioral tests (such as memory tests) as well as interviews of the patient and their caretakers. The goal of such an evaluation is to monitor the cognitive decline over time and eliminate other possible disorders. Therefore, clinical evaluations require the expertise of a qualified neuropsychologist, can be quite subjective (some hospitals use a committee of neurologists to reduce the subjectivity), and due to longitudinal nature of the process, can be quite expensive. However, such an evaluation can achieve over 90% accuracy, when conducted by expert neurologists at large hospitals or dementia clinics, however, geographical and financial restrictions generally limit most patients to smaller hospitals and clinics, where the diagnostic accuracies are estimated to be around 75%, even with the benefit of longitudinal evaluations [2].

Because of the cost, subjectivity, and the general difficulty of clinical evaluations, several biomarkers have been researched and proposed over the years for detecting AD, with cerebrospinal fluid (CSF) concentrations of tau and β -amyloid proteins being the most promising. However, measuring these concentrations requires highly invasive, expensive, potentially painful lumbar puncture (spinal tap), which also requires specialty clinics, research or university hospitals – severely restricting the utility of these biomarkers. In an attempt to find non-invasive biomarkers, researchers have investigated the feasibility of standard neuroimaging tools, such as the MRI and PET imaging [3], as well as neurophysiological measurements using EEG.

The MRI primarily provides a volumetric measurement, where the loss of brain tissue in predefined regions is measured over a longitudinal time frame. Since AD causes neuronal death, which in return causes loss of brain volume, MRI can easily identify such a change. PET imaging follows a different, but equally effective approach, by measuring the glucose metabolism of the brain. Hypometabolism, a drop in metabolic activity in the brain possibly can be easily measured by the PET scan. Finally, the event related potentials (ERPs) obtained from the electroencephalogram (EEG) provides another potential biomarker. Several studies using the well-known oddball paradigm have demonstrated that the decreased amplitudes and increased latencies of the so-called P₃₀₀ component of ERPs – a positive peak that occurs around 300 ms after a stimulus is observed by the subject – is linked to the cognitive decline [4-6]. Various signal processing approaches on the raw EEG or the P₃₀₀ has been conducted, verifying the presence of a statistical correlation, with limited success in patient specific diagnosis [7-9]. In fact, in our previous work, we have also shown that discrete wavelet coefficients of the ERPs, can be used as biomarkers in patient specific AD diagnosis, particularly when the ERPs in response to different types of stimuli are combined [10;11].

This work is supported by NIH grants AG022272 and AG10124, PA Dept. of Health, SAP4100027296, by NSF grant no ECCS 0239090 and NSF grant no ECCS 0926159.

C. Tilley, B. Hillis and R. Polikar are with Electrical and Computer Eng. Dept. at Rowan University, Glassboro, NJ 08028, USA. (e-mail: tilley89@students.rowan.edu, hillis78@students.rowan.edu).

C.M. Clark is with the Department of Neurology, University of Pennsylvania, Philadelphia, PA (e-mail: chris.clark@uphs.upenn.edu).

Contact author: Robi Polikar, phone: (856) 256 – 5372; fax: (856) 256-5241. E-mail: polikar@rowan.edu.

Previously, we have shown that an ensemble of classifiers based decision fusion approach, combining ERP data acquired from different electrodes is a feasible approach for automated early diagnosis. In this contribution, we investigate the obvious question of whether these three different modalities, MRI, PET and EEG, carry complementary information, and whether a strategic combination of these modalities can improve the diagnostic accuracy over any of the individual data sources when used with an automated classifier. While MRI, PET and CSF combination has been very recently explored [12], to the best of our knowledge, this is the first effort of its kind in investigating an MRI, PET and EEG data fusion for early diagnosis of AD. Initial results, reported in this paper, appear to be quite promising.

II. EXPERIMENTAL SETUP

A. Patient Cohort

The study cohort used in this analysis was chosen from a larger cohort of over 300 subjects who had undergone different biomarker tests. Not all subjects had all imaging tests done, hence a subset of 73 subjects, 37 with AD (15 male, 22 female, with mean age $\mu_{AGE(AD)}=74.5$) and 36 cognitively control (10 male and 26 female, $\mu_{AGE(CN)}=71.2$) were chosen based on the constrain that all selected subjects had at least one MRI scan, one PET scan and one EEG recording within relatively close period of each other.

The inclusion criteria for AD group was satisfying the NINCDS-ADRDA criteria [13] for probable AD, which includes a battery of memory tests, interviews with the subject and their caregivers, clinical dementia rating (CDR) score of 0.5 or higher for AD cohort ($\mu_{CDR(AD)}=1.12$) and 0 for the normal cohort. All subjects were over 60 years old. Exclusion criteria for both groups were evidence of any other central nervous system damage or use of sedatives, anxiolytic or antidepressants within 48 hours of ERP acquisition.

B. The Oddball Paradigm and ERP Acquisition

The ERPs were collected using an auditory oddball paradigm protocol. Electrodes were placed according to 10-20 standard (Fig 1), whose impedances were kept below 20 Ω . Each subject was tested for 30 minutes with approximately three minutes of rest for every five minutes of testing.

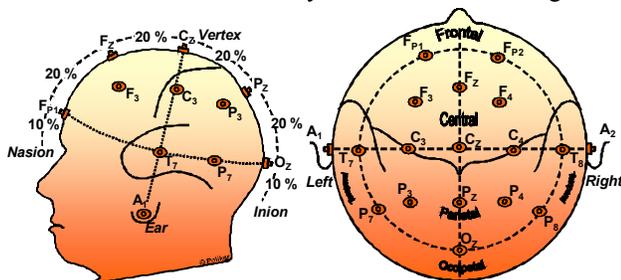


Fig. 1. Electrode placement

1,000 random stimuli were presented, 65% consisting of standard tones of 1 kHz, 20% as target tones of 2 kHz, and 15% as novel sounds. A random inter-stimulus interval of 1.0 to 1.3 seconds was inserted. Standard and target stimuli were presented in 100ms bursts with a 5ms on/offset envelope. The novel stimuli were environmental sounds about 200ms long, each unique and never repeated. The subjects were instructed

to press a button each time they heard the target tone only. The data was sampled at 256 samples /second.

The preprocessing of the EEG signals included notch filtering at 59-61 Hz, followed by removal of major artifacts using a 20th order derivative based thresholding. EEG recordings were epoched with respect stimulus type, averaged and time-locked with 200ms pre-stimulus and 800ms post-stimulus intervals. The pre-stimulus baseline was removed from the entire ERP, resulting in one second length of 256 samples per stimulus type, per electrode channel, per patient.

We have previously reported that discrete wavelet transform (DWT) coefficients corresponding to certain frequency bands are particularly informative [10;11;14]. Using the DWT with Daubechies 4 wavelet, signals were broken down into the following frequency bands for analysis:

- | | |
|-------------------------------|-----------------------------|
| d_1 : 64~128 Hz ($N=132$) | d_5 : 4 ~ 8 Hz ($N=14$) |
| d_2 : 32 ~ 64 Hz ($N=69$) | d_6 : 2 ~ 4 Hz ($N=10$) |
| d_3 : 16 ~ 32 Hz ($N=38$) | d_7 : 1 ~ 2 Hz ($N=8$) |
| d_4 : 8 ~ 16 Hz ($N=22$) | a_7 : 0 ~ 1 Hz ($N=8$). |

In this study, the decomposition levels 6,7 and the approximation level 7 were of primary focus. These bands were chosen as ERP signals generally reside within 0 – 4 Hz. 30 feature sets were generated, 15 from responses to novel tones and 15 from responses to target tones using 5 different electrode locations in the parietal region: P₃, P₄, P₇, P₈, and P_Z, with three frequency bands each. These locations were chosen due parietal region being the first affected region in AD, an outcome that was also verified in previous analyses.

C. MRI Acquisition & Processing

Different stages of AD can be characterized by the amount of atrophy in the brain. The change of volume in particular regions of the brain, as well as the overall brain volume, can be detected by volumetric MRI imaging techniques. MRI enhances the differences in tissue matter based on the ratio of bound to unbound water molecules. Since different types of brain tissue have different such ratios, a quantification of regional volumes can be obtained. In this effort, the so-called T₁ weighted MRI was used, which refers to the duration of the net magnetization vector to return to its initial state after being rotated by an RF pulse. Tissues that have a large ratio of bound to unbound water have short T₁ durations. Brain tissue has a high amount of bound water compared to the surrounding CSF and therefore appears accentuated in a T₁ image. Original raw MRI data consists of images of the brain taken in consecutive slices perpendicular to the coronal and parallel to the transverse planes. An elastic warping algorithm was used that standardizes images to a reference topology, while preserving the morphological characteristics of the individual brain [15]. The raw image is segmented into white and gray matter, cerebrospinal fluid, and ventricles. Automated region of interest (ROI) analysis then determines the brain regions visible in the image, allowing the areas corresponding to the skull to be removed and the volumes of specific areas to be computed. Fourteen such areas were identified on each side of the brain, giving 28 features extracted per image. The volumes corresponding to these 28 areas were normalized with respect to total intracranial volume before being used to train a classifier.

D. PET Acquisition & Processing

Unlike MRI, Positron Emission Tomography (PET) is a nuclear imaging technique, based on the detection of gamma rays emitted by a radioactive tracer introduced into the body using a biologically active molecule. The most common molecule used for this purpose is fluorodeoxyglucose (FDG), and hence this imaging technique is also called FDG-PET. The images used in this study were FDG-PET images, where scanning began 30 minutes after the FDG injection and lasted also about 30 minutes. The imaging used the ADNI (Alzheimer's Disease Neuroimaging Initiative) protocol [16]. Stereotactic surface projection analysis was applied to these images, an approach that was shown to be very useful for AD diagnosis [17;18]. This analysis, made by an automated software library called Neurostat SSP (freely available for download at [19]) provide relative glucose metabolic rates in 43 predetermined regions of interest, which were used as the features to train the classifiers.

III. DATA ANALYSIS

A. Ensemble Based Classification Systems

An ensemble of classifiers based decision fusion approach was implemented for automated classification. An ensemble-based system is obtained by combining a set of diverse classifiers, where diversity is typically achieved by using different training parameters for each classifier. If proper diversity is achieved, different errors are made by the individual classifiers, strategic combination of which can reduce the total error. The diversity can be obtained in several ways, including using different subsets of the training data obtained by resampling, different parameters of a classifier model, different classifiers all together, or different subsets of the features of the given dataset. The latter one is commonly referred to as random subspace analysis.

In a sense, an ensemble based system implements a decision – fusion approach, where the decisions made by component classifiers are combined to arrive at a final decision. The goal is to improve the generalization performance over a single classifier based system. However, the structure of this approach also lends itself naturally to data fusion applications, where information from different sources can be combined. The goal in data fusion applications is to improve the accuracy of the final decision compared to a decision made based on a single source of data. Using the ensemble approach for a data fusion application is quite straightforward: a separate classifier can be trained on each dataset that come from a separate source, and the classifiers can then be combined using an appropriate combination rule [20].

In this analysis we combine the two uses of the ensemble systems, that is, we use an ensemble of classifiers both for improving the accuracy over a single classifier, as well as for achieving data fusion. In essence, we first train an ensemble of classifiers for each of the EEG, MRI and PET based data to obtain three “experts.” We then combine these ensembles of classifiers to achieve decision fusion based data fusion of EEG, MRI and PET data. Our goal is to investigate whether these different modalities provide complementary information for the diagnosis of Alzheimer’s disease. Figure 2 illustrates this general approach.

Once the individual classifiers are generated, they can be combined by any one of several combination rules. Among these, perhaps the most commonly used ones are the sum rule and the majority voting rule. Let us define the decision of the i^{th} classifier as $d_{i,j} \in \{0,1\}$ $i=1,\dots,L$ and $j=1,\dots,c$, where L is the number of classifiers and c is the number of classes. If i^{th} classifier chooses class j , then $d_{i,j} = 1$, and zero, otherwise. In majority voting, the ensemble decision is chosen as the class that receives the highest number of votes, that is, we first compute the total support S_j for class J as

$$S_j(\mathbf{x}) = \sum_{i=1}^L d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in \{0,1\} \quad (1)$$

and choose the class with this highest support as the winning class. Alternatively, if classifiers can provide continuous outputs for each class, representing the support provided for that class, then the sum rule can also be used to combine the classifiers, where the total support for class J is obtained as the sum of all classifiers’ class J outputs. The final decision is then the class that receives highest sum support

$$S_j(\mathbf{x}) = \sum_{i=1}^L d_{i,j}(\mathbf{x}), \quad d_{i,j}(\mathbf{x}) \in [0,1] \quad (2)$$

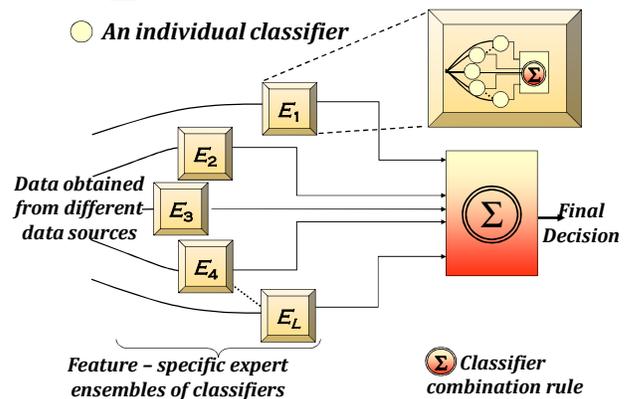


Fig. 2. Ensemble based decision fusion (In this application, $L=3$).

B. Classification Protocols

We used multilayer perceptron (MLP) as the component classifier, although any supervised learning algorithm can be used. All MLPs were trained with gradient descend based backpropagation with an error goal of 0.01 and a momentum term of 0.95. As mentioned above, the individual features were i) the (8~10) DWT coefficients of the ERPs for specific 0-1, 1-2 and 2-4 Hz frequency bands for five parietal electrode locations; ii) the volumetric readings of 28 regions of interest for MRI; and iii) glucose metabolisms from 43 regions of interest for the PET. A total of 30 feature sets generated for the EEG data allowed us to interpret each feature set as a different source of data and use an internal EEG specific data fusion to determine whether there is complementary information among the EEG channels. Each MLP for the EEG data had 10 hidden layer nodes, and 30 such MLPs were generated, one for each feature set. Using different feature sets provided a natural diversity for the EEG based classifiers. The MRI and PET data did not have individually separate datasets. However, in order to give equal total voting weight to EEG, MRI and PET classifiers, we decided to train 30 MRI classifiers and 30 PET classifiers. In order to achieve the diversity for MRI and PET classifiers, we used the random subspace method, training each MRI and PET

classifier with 2/3 of the total available features (18 of 28 for MRI, and 28 of 43 for PET). Sum and simple majority voting were used as combination rules. All experiments were repeated using a five-fold cross validation, whose results are provided below.

We should mention that in this preliminary study for combining EEG, MRI and PET data, our goal was not so much to obtain the best classification performance, but rather to determine whether these modalities provide complementary information, and whether the ensemble based approach can extract such information. Therefore, individual classifiers' free parameters (e.g. number of hidden layer nodes) were not optimized, but used with common sense default values.

IV. RESULTS

The average diagnostic performances of the EEG, MRI and PET classifiers were 56.7%, 67.22% and 70.3%, respectively, in identifying the test dataset subjects as AD vs. control. We have then used the sum and simple majority voting (SMV) rules to obtain EEG+MRI, EEG+PET, MRI+PET and EEG+MRI+PET data fusion diagnostic accuracies. Two sets of each experiment were done, once using all 30 classifiers for each data source, and once using the top 15 classifiers for each data source. The top 15 EEG classifiers would then correspond to those channel and frequency bands that performed better than the others, and the MRI and PET classifiers would represent those feature subsets whose combination provided better performance. The results are tabulated in Table 1.

TABLE 1. DATA FUSION PERFORMANCES (% ACCURACY)

	All Classifiers		Top 15 in each category	
	SUM	SMV	SUM	SMV
EEG+MRI	76.05	73.19	80.08	78.66
EEG+PET	74.37	77.23	81.26	81.26
MRI+PET	76.05	77.48	80.08	78.91
EEG+MRI+PET	80.08	76.05	85.55	82.94

We note from Table 1 that each combination provided better diagnostic performance than the corresponding individual data sources, and the combination of all three sources performed better than the other combinations. This outcome indicates that each of these modalities do carry complementary information and the simple combination of classifiers trained on these different modalities can improve the diagnostic performance. Finally, the sum rule appears to provide better classification performance than majority voting rule.

V. CONCLUSIONS

We described a simple data fusion approach for combining data from different sources with a specific application on diagnosis of Alzheimer's disease from EEG, MRI and PET data. The approach generates an ensemble of classifiers for data obtained from each modality. For EEG, where different channels and frequency bands provided a natural individual data subsources, a single classifier was trained with each such source. For MRI and PET, random subspace analysis was used to generate the ensembles. In all cases, the combinations showed better performances than the classifiers trained on individual data sources. Considering that each

modality actually measures a very different (electrophysiological vs. anatomical vs. metabolic) quantity, this outcome is in fact reassuring, whereas the ability of the approach to extract such complementary information is very promising. While the goal of this study was not to obtain the best possible classification (and hence the individual classifiers were not optimized), the results are nevertheless shows significant potential.

REFERENCES

- [1] Alzheimer's Assoc., "2009 Alzheimer's disease facts and figures," *Alzheimer's and Dementia*, vol. 5, no. 3, pp. 234-270, 2009.
- [2] A. Lim, et al., "Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series," *J Am Geriatr Soc*, vol. 47, no. 5, pp. 564-569, May1999.
- [3] C. M. Clark, et al., "Biomarkers for Early Detection of Alzheimer Pathology," *Neurosignals*, vol. 16, no. 1, pp. 11-18, 2008.
- [4] J. Polich, C. Ladish, and F. E. Bloom, "P300 assessment of early Alzheimer's disease," *Electroencephalogr. Clin. Neurophysiol.*, vol. 77, no. 3, pp. 179-189, May1990.
- [5] S. Yamaguchi, et al., "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, vol. 111, no. 2, pp. 195-203, Feb.2000.
- [6] D. J. Linden, "The 300: where in the brain is it produced and what does it tell us?," *Neuroscientist*, vol. 11, no. 6, pp. 563-576, 2005.
- [7] M. Lindau, et al., "Quantitative EEG abnormalities and cognitive dysfunctions in frontotemporal dementia and Alzheimer's disease," *Dement. Geriatr. Cogn Disord.*, vol. 15, no. 2, pp. 106-114, 2003.
- [8] A. A. Petrosian, D. V. Prokhorov, W. Lajara-Nanson, and R. B. Schiffer, "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG," *Clin. Neurophysiol.*, vol. 112, no. 8, pp. 1378-1387, Aug.2001.
- [9] M. Karrasch, et al., "Brain oscillatory responses to an auditory-verbal working memory task in mild cognitive impairment and Alzheimer's disease," *International Journal of Psychophysiology*, vol. 59, no. 2, pp. 168-178, Feb.2006.
- [10] R. Polikar, A. Topalis, D. Parikh, D. Green, J. Frymiare, J. Kounios, and C. M. Clark, "An ensemble based data fusion approach for early diagnosis of Alzheimer's disease," *Information Fusion*, vol. 9, no. 1, pp. 83-95, Jan.2008.
- [11] R. Polikar, A. Topalis, D. Green, J. Kounios, and C. M. Clark, "Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 542-558, Apr.2007.
- [12] K. B. Walhovd, et al. "Combining MR Imaging, Positron-Emission Tomography, and CSF Biomarkers in the Diagnosis and Prognosis of Alzheimer Disease," *AJNR Am J Neuroradiol*, vol. 31, no. 2, pp. 347-354, Feb.2010.
- [13] G. McKhann, et al., "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*, vol. 34, no. 7, pp. 939-944, 1984.
- [14] G. Jacques, J. L. Frymiare, J. Kounios, C. Clark, and R. Polikar, "Multiresolution analysis for early diagnosis of Alzheimer's disease," *IEEE Engineering in Medicine and Biology Conference (EMBC 2004)*, vol. 1, pp. 251-254, 2004.
- [15] T. Liu, D. Shen, and C. Davatzikos, "Deformable registration of cortical structures via hybrid volumetric and surface warping," *NeuroImage*, vol. 22, no. 4, pp. 1790-1801, Aug.2004.
- [16] R. Petersen and M. W. Weiner, "ADNI Protocol," 2005.
- [17] S. Minoshima, et al., "A Diagnostic Approach in Alzheimer's Disease Using Three-Dimensional Stereotactic Surface Projections of Fluorine-18-FDG PET," *J Nucl Med*, vol. 36, no. 7, pp. 1238-1248, 1995.
- [18] N. L. Foster, et al., "FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease," *Brain*, vol. 130, no. 10, pp. 2616-2635, Oct.2007.
- [19] S. Minoshima and D. Cross, "Neurostat/3D - SSP," Available at: <http://128.95.65.28/~Download/>. Last access:06/24/2010, 2008.
- [20] D. Parikh and R. Polikar, "An Ensemble Based Incremental Learning Approach to Data Fusion," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 37, no. 2, pp. 437-450, 2007.