30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

EEG and MRI Data Fusion for Early Diagnosis of Alzheimer's Disease

Tejash Patel, Robi Polikar, *Senior Member, IEEE*, Christos Davatzikos, Christopher M. Clark

***Abstract*—The prevalence of Alzheimer's disease (AD) is rising alarmingly as the average age of our population increases. There is no treatment to halt or slow the pathology responsible for AD, however, new drugs are promising to reduce the rate of progression. On the other hand, the efficacy of these new medications critically depends on our ability to diagnose AD at the earliest stage. Currently AD is diagnosed through longitudinal clinical evaluations, which are available only at specialized dementia clinics, hence beyond financial and geographic reach of most patients. Automated diagnosis tools that can be made available to community hospitals would therefore be very beneficial. To that end, we have previously shown that the event related potentials obtained from different scalp locations can be effectively used for early diagnosis of AD using an ensemble of classifiers based decision fusion approach. In this study, we expand our data fusion approach to include MRI based measures of regional brain atrophy. Our initial results indicate that ERPs and MRI carry complementary information, and the combination of these heterogeneous data sources using a decision fusion approach can significantly improve diagnostic accuracy.**

I. INTRODUCTION

Alzheimer's disease is a neurodegenerative disorder characterized by progressive cognitive deterioration caused by neuronal death. AD causes gradual loss of memory, cognitive ability, and motor skills. For many years, AD was not recognized as a major public health problem primarily because a majority of the population never lived long enough to become susceptible. However, as overall life expectancy increases, particularly in developed countries, so does the prevalence of AD. The Alzheimer's Association estimates that there are over five million AD patients in the U.S. alone. The prevalence of the disease increases rapidly with age, especially with patients that are over the age of 65. The disease affects an average of 2% of those under the age of 74, 19% of those between 75 and 84, and an alarming 42 % of people over the age of 85. The devastating effect of the disease on its victims, coupled with the enormous grief caused to care givers and its steep financial toll on the society ($148 billion annually), makes AD a major public health concern [1].

The enormity of this concern is multiplied by the fact that there is neither a disease modifying treatment, nor even a reliable laboratory test for diagnosis that can be used premortem. The disease is characterized by accumulation of two misfolded proteins, β-amyloid and hyperphosphorylated–τau, which can only be detected by analyzing the brain tissue under the microscope – an approach that is only possible during an autopsy. Diagnosis is usually done through clinical evaluations that include a battery of memory tests and interviews with patients and their caretakers. The diagnostic accuracy of these clinical assessments is about 90% when conducted by expert neurologists at specialized dementia clinics. Community hospitals, where most patients seek care, lack such expertise, and the diagnostic accuracy for AD is estimated to be 75%, with a sensitivity of 83%, and a specificity of 55%, even with the advantage of frequent follow-up [2]. On the other hand, pathologically targeted medications under current development can reduce the rate of the disease progression; however, their effectiveness is clearly dependent on an accurate diagnosis at the earliest stage of the disease. Hence, cost-effective biomarkers that can provide accurate diagnosis at a community clinic and hospital setting are needed.

New biomarkers that are under consideration typically fall into one of four categories: biochemical, anatomical, metabolic and physiological markers. The *biochemical* markers such as β-amyloid and tau are proteins found in the CSF that are linked to AD pathology. Accumulation of these proteins in AD effected brain interferes with signaling at the synapses, and eventually causes neuronal death. The levels of tau and β-amyloid can be detected in elevated levels in the CSF of patients with Alzheimer's through a lumbar puncture. While this is one of the most reliable predictors of pathology, it is also the most invasive and requires physicians skilled in performing the procedure. The *anatomical* marker for AD is the atrophy of certain regions of interest in the brain, which can be detected by MRI images. Similarly, a PET scan provides a *metabolic* marker, measuring the glucose metabolism in several regions of interest. Both the (gray matter) atrophy and loss of metabolism are directly linked to neuronal death.

Detection and measurement of event related potentials (ERPs) of the electroencephalogram (EEG), on the other hand, are *physiological* markers of the integrity of neuronal systems, which are shown to carry diagnostic information in previous studies, including our own [3]. A protocol, called *oddball paradigm* is used for the acquisition of ERPs, where the patient is asked to respond (by pressing a button) every time an infrequent target (oddball) stimulus (a tone at 2 kHz) is delivered in a series of frequent non-target stimuli (a tone at 1 kHz) and infrequent novel sounds (sounds clips from movies). Several changes in the structure of ERP (such as amplitude and latencies of certain components) are known to

T. Patel and R. Polikar are with Signal Processing and Pattern Recognition Laboratory of the Electrical and Computer Eng. Dept. at Rowan University, Glassboro, NJ 08028, USA. (e-mail: patelt99@students.rowan.edu).

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

C. Davatzikos is with Dept. of Radiology, University of Pennsylvania, Philadelphia, PA (e-mail: christos@rad.upenn.edu)

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

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

978-1-4244-1815-2/08/$25.00 ©2008 IEEE.

be altered by neurological disorders; providing a potential piece of diagnostic information [4,5]. ERP analysis has been used with a variety of signal processing techniques and automated classifiers for AD diagnosis with limited success [6-8], in part due to difficulty of early diagnosis, but also in part due to small cohort sizes, relying solely on specific components of the ERP, or on ERP obtained from certain electrodes only.

Our hypothesis is that ERP signals and MRI measures of regional brain atrophy carry complementary information, since ERP based features reflect the physiological and electrical changes affected by neuronal death and disrupted synaptic transmission, whereas MRI based features reflect the anatomical changes as loss of volume in certain regions of interest, a direct effect of neuronal death. If this hypothesis is true, then a strategic combination of such heterogeneous information should lead to an improvement in diagnostic accuracy compared to using either of these two sources alone.

An ensemble of classifiers based decision-fusion algorithm is proposed to combine ERP and MRI data. A separate classifier is trained using ERP data obtained from different electrode locations, along with another ensemble of classifiers trained on MRI data. These classifiers are then combined through weighted majority voting, where the voting weights are determined based on their performance on validation data.

II. EXPERIMENTAL SETUP

A. *MRI Acquisition*

Magnetic resonance imaging (MRI) utilizes strong magnetic fields and specific radio frequency pulse sequences to produce cross-sectional images of the brain. MRI is based on manipulating the intrinsic spin of hydrogen nuclei by placing the hydrogen nuclei in a large magnetic field and exposing them to radio frequency (RF) pulses. MRI has proven to be an indispensible source for determining the anatomical configuration and tissue composition of various brain regions corresponding to different tissue types. Contrast between tissues is based primarily on the ratio of free to bound water in a given region. Since different types of brain tissue have different such ratios, a quantification of brain region volumes can be obtained.

In this application T_1 weighted MRI is utilized. This 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 T_1 durations. Brain tissue has a high amount of bound water compared to the surrounding CSF and therefore appears accentuated in a T1 based image. The image is segmented based on tissue type, followed by volume computations of anatomically defined regions of interest. These volumetric measurements can then be seen as anatomical markers of brain atrophy.

B. *The Oddball Paradigm and the ERP Acquisition*

Auditory oddball paradigm was used for ERP acquisition. The subjects were given a set of headphones and were presented with a series of tones and sounds occurring every 1-1.3 seconds. Most of the tones (65%) were low frequency (1 kHz) standard tones, and 20% were high frequency (2 kHz) target (oddball) tones. The remaining stimuli were novel sounds (15%). The subjects were instructed to press a button in response to a target tone only. The ERPs were acquired from 16 electrode locations, mounted to the scalp in accordance with the 10/20 International System of electrode placement (Figure 1). In this study, we focused on ERPs obtained in response to novel and target tones from the parietal regions, (P3, P4, P7, P8, and PZ) since these regions are known to generate strongest ERP signals. ERP signals were digitized at 256 samples/second, and notch filtered at 59-61 Hz. Artifactual recordings were rejected by an EEG technician. The ERP signals were then segmented to 1-second intervals, beginning 200ms pre stimulus and ending 800 ms post stimulus, followed by averaging of these synchronized signals. The resulting signals were 256-long average ERPs, one for each electrode and stimulus type, per patient.

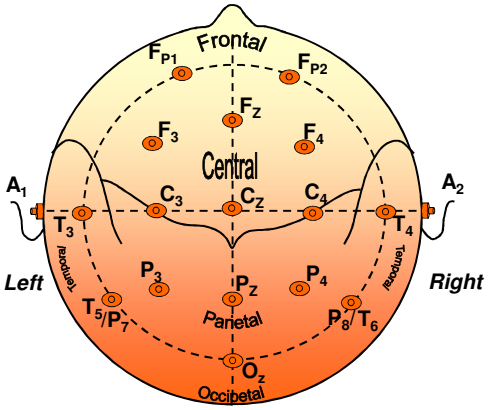

Fig. 1: Electrode placement.

C. *Patient Cohort*

ERP and MRI data were obtained from 83 subjects, 34 normal and 49 with AD. For cognitively normal subjects, the inclusion criteria was: age > 60; Clinical Dementia Rating (CDR) score = 0; Mini Mental State Exam (MMSE) score > 26; no cognitive decline within two years of testing, whereas for AD patients: age > 60; CDR score > 0.5; MMSE score ≤ 26; cognitive decline over the last 12 months; and meeting NINCDR-ADRDA criteria for probable AD [9-11].

III. METHODS

A. *Feature Extraction*

The T_1 weighted MRI data are topographic images of the brain taken in consecutive slices parallel to the transverse plane. In order to extract the anatomical biomarkers in the form of quantized volumetric data, the raw image is first segmented and a density map of the image is calculated. This analysis allows the determination of which tissues are experiencing neuronal atrophy. Once the density map is composed, an automated region of interest analysis then determines the brain regions visible in the image. This is repeated for each image slice and the slices are collectively processed using voxel analysis to quantize the volume of various brain images. The features used for the classification algorithm is quantization of volume from 12 brain regions such as hippocampus, temporal gray/white matter, parietal lobe, anterior lobe, etc. repeated for left and right sides for 24 total features.

The ERP signals are non-stationary time series data, which are composed of multiple functional components at different frequency bands. This type of data lends itself naturally to discrete wavelet transform analysis (DWT), which decomposes the signal into its constituent frequency bands [12].

DWT provides time localizations of the signal's spectral components, resulting in a time-frequency representation. This is achieved by decomposing the signal $\mathbf{x}[n]$ into its frequency subbands using a series of lowpass ($h[n]$) and highpass ($g[n]$) filters, followed by subsampling by 2, creating two signals of half the length and half the bandwidth at each level of decomposition. The outputs of the highpass filters at level i are the DWT (detail) coefficients ($\boldsymbol{d}_i$), whereas the outputs of the lowpass filters are approximation coefficients ($\boldsymbol{a}_i$), which are further decomposed by the next level of filters. In our implementation, the filters (length 8) were defined by the Daubechies wavelet with 4 vanishing moments. The complete decomposition is shown in Figure 2, which includes the frequency subband and number of coefficients at each level. Since the ERPs are known to be primarily in the 0 – 4 Hz, we chose $\boldsymbol{d}_5$, $\boldsymbol{d}_6$, $\boldsymbol{d}_7$ coefficients as our features.

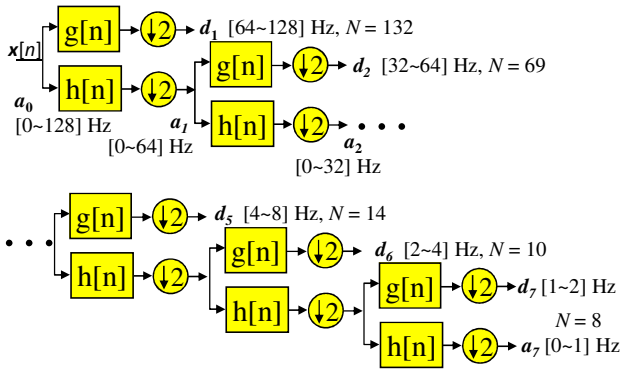

Fig. 2. DWT subband coding algorithm

B. *Ensemble Based Data Fusion Classification*

An ensembles of classifiers based data and decision fusion approach is used for automated classification. Ensemble systems combine several classifiers, each typically trained with a different subset of the training data to ensure the diversity of each classifier. Classifier diversity ensures that that each classifier creates a different decision boundary and hence makes different errors on each instance. A strategic combination of these classifier outputs can then reduce the total error. In this work, we use the ensemble of classifiers approach not only for reducing the error within each source, but also for combining heterogeneous data obtained from different sources.

First, an ensemble of classifiers was trained to form the "ERP expert." This expert consists of one classifier for each ERP source, that is, ERP signals obtained in response to two types of stimuli (target and novel sounds), from 5 electrode locations, and analyzed at three different frequency bands (1-2Hz, 2-4 Hz and 4-8 Hz), giving us 30 sources of ERP signals per patient. One support vector machine (SVM) with a Gaussian kernel was trained using the modified two-tier leave-one-out (LOO) cross validation (explained below) for each ERP based data source, giving us 30 ERP based SVMs. These classifiers were then combined using weighted majority voting (WMV), where the weights were determined based on the validation data performances (see below).

An "MRI expert" was then created. Because there was only one source of MRI information (24 volumetric measurements obtained from different regions of brain), the MRI classifiers were trained using *random subspace method* [13], which is similar to *bagging* [14] except applied to the features instead of training samples: a random subset of 18 (of 24) features were drawn to train each classifier, and 30 such classifiers were trained. The classifiers (SVMs with Gaussian kernels) were combined using a WMV, weights determined based on the validations performances.

The two-tier leave-one-out (LOO) cross validation, shown in Figure 3, was used for training, validation and testing, to ensure that the performance estimates reflect the true performances as closely as possible.

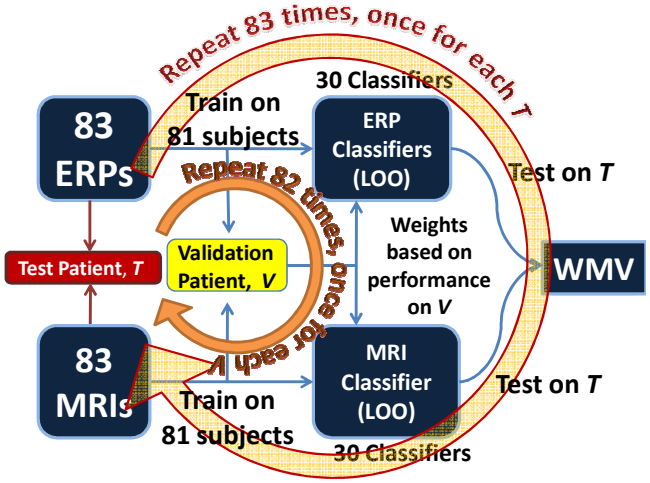

Fig. 3. Two-tier LOO for training / testing

Out of the 83 subjects, one subject was first removed for testing (T), and one was removed for validation (V). 30 ERP classifiers were then trained, one for each ERP source using the data from the remaining 81 patients. These classifiers were evaluated on the validation patient V, whose performance was noted. This procedure was repeated a total of 82 times, each using a different V. The average performance of each ERP classifier on its validation data was then used as the voting weight of that classifier for the subsequent weighted majority voting (WMV). Once all training was completed, the 30 classifiers were tested on the one test patient T that was left out. The entire procedure is repeated 83 times, each using a different T. The average of these 83 LOO trials was the overall classification performance of the "ERP expert."

A similar process was repeated for the MRI data, where 30 classifiers were generated for each LOO trial (each using 18 of the 24 features). The average performance of 82 LOO trials on the validation subject V was used to determine the voting weight, whereas the average performance of the 83 LOO trials on the test subject T was the overall performance of "MRI expert." Note that the LOO based validation scheme allows us to assess the relative diagnostic accuracy of different sources of information, and assign appropriate weights. In the absence of prior knowledge whether ERP or MRI provides more discriminatory diagnostic information, the num-

ber of classifiers (30) was kept the same, so that ERP and MRI experts have similar voting magnitudes.

IV. Results

The diagnostic performance figures obtained by using ERP based features alone, MRI based features alone, and with ERP + MRI data fusion are shown in Table 1. In addition to accuracy (average generalization performance on test data), we also provide sensitivity, specificity and positive predictive values (PPV). Sensitivity is the ratio of true positives (correctly classified as AD by the algorithm) to those clinically diagnosed as AD, specificity is the ratio of true negatives (correctly classified as normal by the algorithm), to those clinically identified as normal, and PPV (or precision) is the ratio of true positives to all subjects identified as AD (true positive + false positive).

TABLE 1. DIAGNOSTIC PERFORMANCES

	ERP	MRI	ERP+MRI
Accuracy	74.70%	89.16%	93.98%
PPV	80.48%	95.45%	95.83%
Sensitivity	75.51%	85.71%	93.88%
Specificity	73.53%	94.12%	94.12%

The overall diagnostic accuracy of the individual ERP data sources (not shown above) were in the upper 50% to mid 60% range, which improved to 74.7% when the decisions made by 30 such sources were combined through WMV. The average performance obtained by the MRI based features was 89.16%, which improved to 93.98% when combined with the ERP based decisions. Similar improvements can also be seen in all diagnostic metrics. These results, while preliminary, are clinically significant, since community clinic based diagnostic approaches consistently had lower specificity values (55% according to [2]).

V. Discussion & Conclusions

Our previous studies have shown that physiological (ERP) data from different sources provide complementary and valuable discriminatory information for diagnosis of AD, when combined with an ensemble of classifier based data and decision fusion algorithm. In this study, we report our preliminary results in extending this approach to include anatomical features measuring loss of brain volume as measured by T_1 weighted MRI images. The ERP data were obtained from five electrodes over the parietal region of the (P7, P3, PZ, P4, P8 electrodes), in response to target and novel stimuli, and their DWT coefficients in 0-8 Hz were used. MRI data were volumetric measurements from 24 regions of interest.

These two heterogeneous data types were chosen primarily because they can be obtained noninvasively, and using equipment that could be made available at community hospitals – unlike biochemical markers or the neuropsychological evaluations that are typically available only at dementia specialty clinics at major university or research hospitals.

Our preliminary results indicate that both ERP and MRI based features carry diagnostically useful information, with a diagnostic accuracy in the high 70% to 80% range. We have also shown that when the ensemble based decisions are combined using a weighted majority voting, the diagnostic accuracy increases to 90% range. We have also shown that other diagnostic metrics, such as sensitivity, specificity and PPV also show significant increases under the proposed data fusion approach.

We should mention, however, that our results represent the ability of the proposed approach in correctly matching the diagnoses of expert clinicians – and not predicting the true condition. Hence the true performance of the approach can be slightly worse or better, however, that would be impossible to assess without autopsy based confirmation of the diagnosis. With this distinction in mind, we can draw several conclusions from these preliminary results. First, ERPs and MRI carry complementary information – this is expected, since the features obtained from these two modalities are obtained by different underlying physical processes, that is, physiological vs. anatomical. Second, an ensemble of classifiers approach that provides a decision level fusion is an effective method to combine such complimentary information. Third, the diagnostic performance of the approach compares very favorably to that obtained by community clinics (of 75%). Hence, the approach promises to be a strong potential as a biomarker for early diagnosis.

References

[1] Alzheimer's Association, "Alzheimer's Disease Statistics," Available at: http://www.alz.org/AboutAD/statistics.asp.

[2] A. Lim, W. Kukull, D. Nochlin, J. Leverenz, et al., "Cliniconeuropathological correlation of Alzheimer's disease in a community-based case series," *J. of the American Geriatrics Soc*., 47, (1999), 564-569.

[3] Polikar R., Topalis A., Green D., Kounios J., Clark C.M., Ensemble based data fusion for early diagnosis of Alzheimer's disease, Information Fusion, vol. 9, no. 1, pp. 83-95, 2008.

[4] S. Yamaguchi, H. Tsuchiya, S. Yamagata, G. Toyoda, S.Kobayashi, "Event-related brain potentials in response to novel sounds in dementia," *Clinical Neurophysiology*, vol.112, no. 2, pp. 195-203, 2002.

[5] D. Linden, "The P300: Where in the brain is it produced and what does it tell us?" *The Neuroscientist*, vol. 11, no. 6, 563-576, 2005.

[6] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, et al., "Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)," Journal of Neuroscience Methods, vol. 161, no. 2, pp. 342-350, Apr. 2007.

[7] R. M. Chapman, G. H. Nowlis, J. W. McCrary, J. A. Chapman, et al., "Brain event-related potentials: Diagnosing early-stage Alzheimer's disease," Neurobiology of Aging, vol. 28, no. 2, pp. 194-201, 2007.

[8] 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

[9] G. McKhann, D. Drachman, M. Folstein, R. Katzman, and others, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Dept. of Health and Human Services Task Force on Alzheimer's Disease," *Neurology* ,vol.34**,** pp.939–944, 1984.

[10] L. Berg, "Clinical dementia rating (CDR),". *Psychopharmacol Bull*, vol. 24, pp.:637-639, 1988.

[11] R. M. Crum, Anthony J.C., Bassett S.S., and M. Folstein, "Population-based norms for the Mini-Mental State Examination by age and educational level," *Journal of American Medical Association*, vol. 269, no. 18, pp. 2386-2391, 1993.

[12] T. Demiralp, A. Ademoglu, "Decomposition of event related brain potentials into multiple functional components using wavelet transform," *Clinical Electroencephalography*, vol. 32, no. 3,pp. 122-138, 2001.

[13] Ho, T.K.: The random subspace method for constructing decision Forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 20, pp. 832-844, 1998.

[14] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.