

# DETECTION OF EARLY COGNITIVE LOSS FROM MEDICATION ADHERENCE BEHAVIOR

Todd Leen<sup>1</sup>, Zhengdong Lu<sup>1</sup>, Tamara Hayes<sup>1,2</sup>, Jeffrey Kaye<sup>2,3</sup>

<sup>1</sup>*OGI School of Science & Technology*, <sup>2</sup>*Oregon Center for Aging and Technology*,

<sup>3</sup>*Layton Aging & Alzheimer's Disease Center Oregon Health & Science University (OHSU)*

## INTRODUCTION

There is evidence that loss of memory contributes to poor medication adherence in the elderly [1,2]. We previously investigated this contribution in a group of independently-living seniors [3]. Here we describe the application of statistical pattern recognition techniques to medication adherence data, and demonstrate that patterns of adherence can reliably detect mild cognitive loss. We apply neural network classifiers to the task of discriminating between healthy individuals and those with early cognitive loss on the basis of medication adherence behavior. The results establish that data from relatively unobtrusive behavior monitoring can provide reliable inference for individuals.

## SUBJECTS AND ADHERENCE MONITORING

Forty independently-living elder subjects were recruited for the original study [3]. All had baseline mini mental state examination (MMSE) scores greater than 24, and Clinical Dementia Rating (CDR) of 0 or 0.5. Subjects were divided into two groups based on their memory function as assessed by Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) scores. Graham's normative ADAS-Cog data [4] were used to generate an age-adjusted 95th percentile interval for cognitively healthy individuals. We define the *Healthy Control* (HC) group as those subjects whose age-adjusted ADAS-Cog scores fell within this interval (N=19), and the *Cognitive Loss* (CL) group as those subjects whose scores fell outside this interval (N=21). After the data collection, results from two individuals were removed for

this study: Data from one HC subject was corrupted by equipment failure, and one CL subject met the clinical standard for mild cognitive impairment (MCI), greater cognitive loss than the intended target population. This left 38 subjects, N=18 in the HC group and N=20 in the CL group. The retained CL group corresponds to very mild cognitive loss.

Subjects were instructed to take a vitamin C supplement twice daily at agreed-upon times, one in the morning, and one in the evening. Their behavior was monitored by an instrumented seven-day pillbox, dubbed the MedTracker, developed in our group [5]. The device records the time of opening of the compartments. Event times are stored in an on-board buffer and transferred from the device by Bluetooth wireless every two hours. The design provides about eight weeks operation from a 9V battery.

Subjects in the pilot study [3] were monitored for approximately five weeks. A sample time series of events (from a healthy individual) recorded by the MedTracker is shown in Figure 1. Dots mark the compartment openings, and crosses mark missed doses. The two solid horizontal lines (at approximately 7:00am and 9:30pm) mark the *planned dose times* (explained below) for this subject. The dashed horizontal lines bound a window one hour before and two hours after the median dose times.

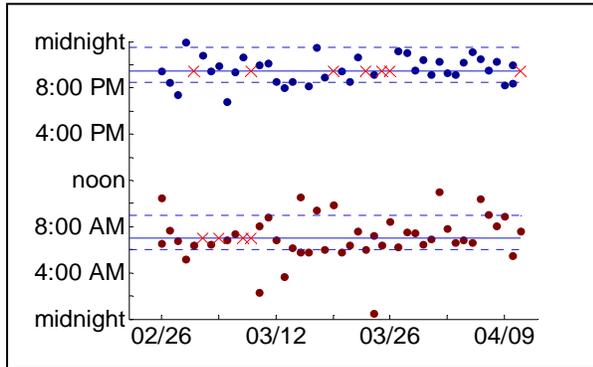


Fig.1: MedTracker event time series.

## CLASSIFIER CONSTRUCTION

### Data Preparation

Two aspects of the subjects' dose-taking behavior created the need for some care in the data analysis. First, although instructed to take one dose in the morning and one in the afternoon, the noon hour is not a reliable boundary between events for each subject. To determine an appropriate boundary between the two sets of events for each subject, we cluster the subject's events into morning and evening groups using the Matlab function `clusterdata`. The algorithm returned a good partition except on two subjects, which we clustered manually. Second, although each subject agreed on planned AM and PM times for their doses, their actual median dose times deviated considerably from their plan. For analysis we *define* a surrogate planned time for the AM dose as the median time for all morning events in a subject (rounded to the nearest half-hour). Planned time for the PM dose is similarly defined. The median times are shown by solid horizontal lines in the example time series in Figure 1.

### Feature Extraction

We summarize the time series by four features that describe the subject's dose behavior and will be used as input to the classifier: (1) `As_Prescribed` is the percentage of days with no less

than two compartment openings; (2) `As_Planned` is the percentage of the individual's events for which the dose was taken close to the planned time; no more than one hour before or two hours after the planned (that is, the median AM or PM) time; (3) `AM_STD` is the standard deviation of the time of the morning dose; and (4) `PM_STD` is the standard deviation of the time of the evening dose.

Not all of the features are equally useful for discriminating between the two groups. Figure 2 shows kernel density estimates [6] of the probability density of two of the features for the two groups. The plots suggest that the `As_Prescribed` feature has discriminatory power, and the `AM_STD` feature has much less. We are only using these density estimates for visualization; we do not suggest that they are accurate for constructing Bayesian classifiers, nor do we expect that they are valuable for ranking features according to their discriminatory capability. Below we discuss how we choose the optimal subset of features to obtain good classification.

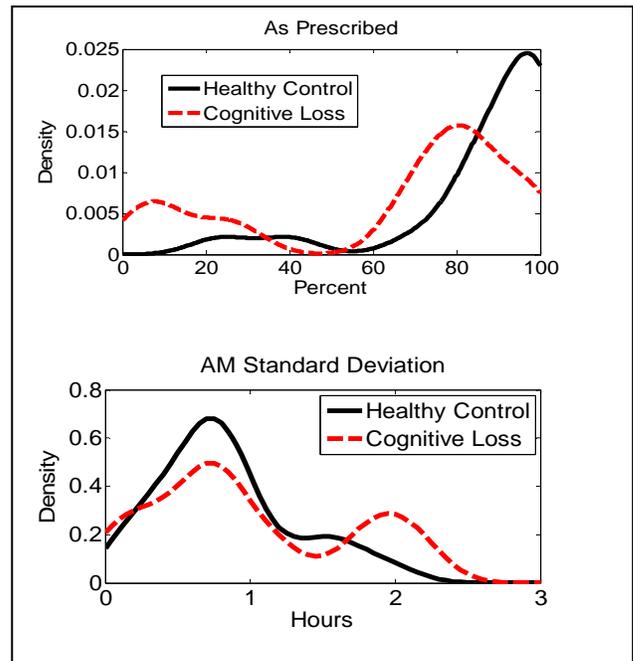


Figure 2: Density plots of the `As_Prescribed` and `AM_STD` summary features.

## Neural Network Classifier

Our goal is to build classifiers that use appropriate combinations of the four summary features as input, and accurately assign individuals to the *Healthy Control* or *Cognitive Loss* groups. We adopt multilayer perceptron (MLP) neural networks [7] for this study. We expect that similar results would be obtained with similarly flexible classifier technology such as support vector machines [8]. The classifier consists of three layers of nodes: the input layer (which receives the values of our summary features extracted from the MedTracker time series), the hidden layer, and the output layer whose single node activity identifies which of the two classes (HC or CL) the input belongs. A logistic function constrains the output to lie in the range (0,1). If the output is greater than 0.5, the subject is classified as CL, otherwise the subject is classified as HC. The numerous *weights* between the nodes comprise free parameters that are optimized to construct the final classifier.

The number of hidden nodes in the network is an architectural parameter determined by the user. Larger hidden layers provide more functional flexibility and hence power to construct complex decision boundaries when required. However the functional flexibility needs to be constrained to avoid overfitting to peculiarities of the training data. We discuss regularization below.

The network weights are trained to match the output to the class labels over the training data, which consists of sequences of pairs  $(F_i, y_i)$ , of the feature and class label (0 for HC and 1 for CL) respectively for the  $i^{\text{th}}$  subject. We use the mean square error between the network output and the class label as an error criteria, and optimize the weights using conjugate gradient descent. To avoid using solutions from poor local optima we use ten random-initializations and restarts of the optimizer. To avoid overfitting, we use weight decay regularization [7] and adjust the corresponding hyper-parameter  $\alpha$  by cross validation [6,7]. The hyper-parameter provides continuous control over the network's functional

flexibility. For our experiments, we limited the network size to 6 and 12 hidden units, and controlled the network complexity by regularization. Both network sizes produced very similar results.

## Feature Selection and Network Regularization

Due the very small number of data samples, we adopted leave-one-out cross validation (jackknife) [6,7] for feature selection and for determining the regularization hyper-parameter  $\alpha$ . For each of the 15 possible combinations of input features we trained networks, choosing as the optimal weight decay parameter  $\alpha$  (from a discrete set) that which gave the best validation performance. (Validation performance is the average of the classification rate of 38 classifiers, each of which is trained on a different 37-sample training set, and evaluated on a one-sample validation set.) The discriminative ability of each feature combination is measured using the validation error achieved with the optimal weight decay. In our experiment, the combination of `As_Planned` + `As_Prescribed` features returns the best performance.

## **CLASSIFIER PERFORMANCE**

Having selected the optimal feature set, we want to estimate the classifier performance on *an independent test set* not used for training or adjusting regularization. Again, the scarcity of samples suggests we adopt leave-one-out cross-validation. With a regularization parameter to choose, this is a two-loop process ( $O(N^2)$  complexity in the number of samples). We sequentially select one subject as the *test set*, and use the remaining 37 subjects as the *development set*. On the development set, we do another leave-one-out cross-validation to determine the optimal weight decay parameter  $\alpha$ . Then we train a classifier on the complete development set and apply it to the single-sample test set. We repeat this over all 38 development–test partitions and report the total number of misclassified test samples as test set error.

In our experiments, the combination of As\_Planned + As\_Prescribed features provides the best performance. This configuration misclassifies 10 of the 38 examples (error rate 0.263), or equivalently yields a correct classification rate of 0.737. (The 95% confidence limit computed using binomial statistics on the classification rate is [0.58, 0.85].) In contrast, guessing based on the class probabilities would return a correct classification rate of 0.53. Table 1 gives error rates for several other feature combinations; the ranking of feature combinations according to their classification performance parallels that returned by the feature selection study discussed in the last section, lending confidence to the selection process.

Feature Combination	Error
As_Pres.+As_Planned	0.263
As_Pres.+As_Planned+PM_STD	0.342
All Four	0.385

Table 1: Test error rates

## DISCUSSION

Our results show that relatively unobtrusive automated monitoring of medication adherence behavior provides a feasible detector for mild cognitive loss at the level of individuals. This study was based on approximately five weeks of adherence data for each subject. Classifiers trained on features derived from shorter data streams performed less well and the curve of performance versus observation time suggests that longer data streams could provide better discrimination. The study could be extended by appeal to alternative classifiers (such as the SVM), and resampling and boosting techniques.

Acknowledgements The authors thank Catherine Huang for help with figures, and Deniz Erdogmus for helpful discussion. This work was supported by Intel under the OHSU BAIC project. Data was made available by OHSU ORCATECH under NIA grant P30 AG024978. Dr. Hayes has a

significant financial interest in Intel Corporation. Results of this study and others like it could result in financial benefit for Dr. Hayes and Intel. Dr. Kaye’s time was partially supported by the Merit Review Grant, Office of Research and Development, Department of Veterans Affairs.

## REFERENCES

- [1] L.J. Fitten, L. Coleman, D.W. Siembieda, M. Yu, and S. Ganzell (1995). “Assessment of capacity to comply with medication regimens in older patients,” *J. Am. Geriatr. Soc.*, 43, 361-367.
- [2] H.K. Edelberg, E. Shallenberger, and J.Y. Wei (1999). “Medication management capacity in highly functioning community-living older adults: Detection of early deficits,” *J. Am. Geriatr. Soc.*, 47, 592-596.
- [3] T.L. Hayes, J.A. Kaye, M. Pavel, J. Lundell, and J. Kimel (2006). “Effect of early cognitive changes on medication adherence,” *59<sup>th</sup> Annual Scientific Meeting of the Gerontological Society of America*, Dallas, TX.
- [4] D.P. Graham, J.A. Cully, A.L. Snow, P. Massman, and R. Doody (2004). “The Alzheimer’s Disease Assessment Scale — Cognitive subscale: Normative data for older adult controls,” *Alzheimer Dis. Assoc. Discov.*, 18, 236-240.
- [5] T.L. Hayes, J.M. Hunt, A. Adami, and J.A. Kaye (2006). “An electronic pillbox for continuous monitoring of medication adherence,” *28<sup>th</sup> Annual Int. Conf. IEEE EMBS*, 6400-6403.
- [6] Keinosuke Fukunaga (1990). *Introduction to Statistical Pattern Recognition*, 2<sup>nd</sup> Edition, Academic Press Inc., Boston.
- [7] Richard O. Duda, Peter E. Hart, and David G. Stork (2001). *Pattern Classification*, Second Edition, John Wiley & Sons, Inc., New York.
- [8] Bernhard Schölkopf and Alexander Smola (2002). *Learning with Kernels*, The MIT Press, Cambridge, Massachusetts.