Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2015 April ; 2015: 980–984. doi:10.1109/ISBI.2015.7164035.

Information-Theoretic Characterization of Blood Panel Predictors for Brain Atrophy and Cognitive Decline in the Elderly

Sarah K. Madsen¹, Greg Ver Steeg², Adam Mezher¹, Neda Jahanshad¹, Talia M. Nir¹, Xue Hua¹, Boris A. Gutman¹, Aram Galstyan², Paul M. Thompson¹, and Alzheimer's Disease Neuroimaging Initiative (ADNI)

¹Imaging Genetics Center, USC, Marina Del Rey, CA, USA

²USC Information Sciences Institute, Marina Del Rey, CA, USA

Abstract

Cognitive decline in old age is tightly linked with brain atrophy, causing significant burden. It is critical to identify which biomarkers are most predictive of cognitive decline and brain atrophy in the elderly. In 566 older adults from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we used a novel unsupervised machine learning approach to evaluate an extensive list of more than 200 potential brain, blood and cerebrospinal fluid (CSF)-based predictors of cognitive decline. The method, called CorEx, discovers groups of variables with high multivariate mutual information and then constructs latent factors that explain these correlations. The approach produces a hierarchical structure and the predictive power of biological variables and latent factors are compared with regression. We found that a group of variables containing the well-known AD risk gene APOE and CSF tau and amyloid levels were highly correlated. This latent factor was the most predictive of cognitive decline and brain atrophy.

Keywords

Magnetic resonance imaging (MRI); Brain; Cells & molecules; Genes; Machine learning

1. INTRODUCTION

The search for biomarkers to predict cognitive decline and age-related brain atrophy [1] is at the forefront of Alzheimer's disease research. The most commonly used predictors include cognitive assessments, the AD risk gene APOE, and AD-related pathological compounds measured from CSF (amyloid beta 1-42, tau, and hyperphosphorylated tau 1-81)[2, 3].

Ideally, information collected in routine blood tests could help screen for individuals at higher risk of cognitive decline [4]. These markers are easily and routinely collected, are modifiable, and could form the basis for recommendations targeting cardiovascular, metabolic, hormonal, or other aspects of health. Such a strategy aims to improve overall health in at-risk individuals, but it is also critical to examine the relative predictive power of these various measures, compared to the more traditional biomarkers.

We used the recently developed CorEx [5, 6] method to discover hierarchical structure in over 200 biological measures in 566 older adults from the Alzheimer's Disease

Neuroimaging Initiative (ADNI). The extensive list of potential biomarkers for cognitive decline and brain atrophy included 200 laboratory tests of blood samples for cardiovascular health, oxidative stress, liver function, inflammation and immunology, hormones, growth factors, nutrition, metabolism and diabetes, cancer risk, and amyloid and tau levels. Theoretically, most of these blood test measures could be targeted with existing interventions. We also included the canonical genetic and cerebrospinal fluid measures to determine the best set of predictors for brain atrophy, defined by high dimensional morphometric and surface topology changes, as well as, clinical scores of cognition.

2. METHODS

2.1. Study Participants and Variables

We analyzed data from ADNI participants for whom at least half of the 206 laboratory test measures were available, resulting in a sample of $N=566$ older adults (age: 72.8 ± 7.4 years; 215 women, 351 men; education: 15.5 ± 3.1 years; 58 cognitively intact, 397 with mild cognitive impairment, 111 with probable AD). The minimum number of variables for any participant was 158.

Predictors in the model included 200 laboratory measures, standard demographic information (age, sex, years of education), height, weight, and APOE genotype. APOE genotype was coded as number of alleles of APOE4, which is the strongest known genetic risk factor for late-onset AD, increasing lifetime risk around 3-fold for each copy of APOE4 carried [7]. The laboratory measures included blood tests for metabolism, kidney function, immune system function, growth factor levels, tumor biomarkers, inflammation levels, nutrition, oxidative stress, cardiovascular function, blood components (including red and white blood cell counts), hormone levels, liver function, and amyloid and tau pathology. We also included a urine test of kidney function and a cerebrospinal fluid test for amyloid and tau pathology.

Cognitive decline was measured as change in the widely used Mini-Mental State Exam (MMSE) [8], one year after baseline. Brain atrophy was measured with a combined score based on one-year change in temporal lobe atrophy and lateral ventricle expansion.

ADNI was launched in 2004 by the National Institute of Health, the Food and Drug Administration, private pharmaceutical companies, and non-profit organizations to identify and evaluate biomarkers of AD for use in multisite studies. All ADNI data used in this analysis are publicly available online. The study was conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki, and the US 21 CFR Part 50—Protection of Human Subjects, and Part 56—Institutional Review Boards. Written informed consent was obtained from all participants.

2.2. Scan Acquisition

We analyzed brain MRIs from 566 ADNI participants. All participants received a high-resolution T1-weighted structural MRI brain scan [9]. Scans were acquired on 1.5T General Electric, Siemens, or Philips scanners using a 3D sagittal magnetization-prepared rapid gradient-echo (MP-RAGE) sequence (repetition time (2400ms), flip angle (8°), inversion

time (1000ms), 24-cm field of view, 192×192×166 acquisition matrix, voxel size 1.25×1.25×1.2 mm³, reconstructed to 1 mm isotropic voxels).

2.3. Image Analysis

For both baseline and one-year follow-up, we used one of the most powerful metrics from our lab to evaluate brain atrophy. This method combines two powerful approaches, tensor-based morphometry for the atrophy in the temporal lobe[10, 11] and shape analysis of the expansion of the lateral ventricles[12], described in further detail below.

2.3.1. Tensor Based Morphometry—We have previously developed methods for Tensor Based Morphometry (TBM) to quantify and localize atrophy by computing brain volume differences at the voxel level within a preprocessed MRI brain scan [10, 11].

After standard MRI preprocessing, each participant's brain scan from the one-year follow-up was registered to the same participant's baseline brain scan [13]. This alignment is performed by a nonlinear inverse consistent elastic intensity-based registration algorithm, which optimizes a joint cost function using mutual information and the elastic energy of the deformation.

The Jacobian determinant, a numeric value representing the degree of local expansion or contraction of the 3D elastic warping from one-year follow-up to baseline, is calculated at each voxel in every brain scan. The values in the Jacobian image are expressed as positive and negative percentages, denoting relative tissue volume difference.

2.3.2. Lateral Ventricle Segmentation—The ventricles were segmented with a validated, modified multi-atlas approach [12]. This segmentation approach applies a group-wise surface registration to existing templates in addition to surface-based template blending, which produces highly accurate results. Ventricular surfaces were then extracted from segmentations and an inverse-consistent fluid registration with a mutual information fidelity term aligned a set of hand-labeled ventricular templates to each scan. The template surfaces were registered as a group using medial-spherical registration and the template that best fits the new boundary was selected for each individual, allowing more flexible segmentation, particularly for outliers.

2.3.3. Combined Measure of Longitudinal Brain Change—We calculated a combined TBM score from the longitudinal change in the 3D Jacobian maps and the surface-based expansion rates of the ventricular boundary [14]. Following [12], we apply a one-class Linear Discriminant Analysis (LDA) model to learn a spatial signature of change to optimize the resulting LDA score for each variable's performance. Initially, the TBM-based score, and the surface expansion score, based on vertex-wise maps of change in radial distance, were computed separately. A combined score was then computed using hierarchical LDA model over the same training set. The combined score outperformed both the TBM- and the Ventricle-only scores as a biomarker for predicting anatomical change. This method also achieves better effect sizes for longitudinal changes in ADNI than virtually all prior methods.

2.4. Correlation Explanation (CorEx) for High-Dimensional Data

Many supervised learning methods have been used to combine features to improve prediction of biologically relevant outcomes. In this case, we learn latent factors in an unsupervised way (without using the outcomes at all), to capture latent causes of dependence in the data. Remarkably, this approach produces a latent factor, which happens to predict outcomes at a higher significance level than any feature

The X_i 's, represented by squares in **Figure 1**, denote observed quantities that may take values in an arbitrary domain and could represent different experimental modalities or heterogeneous data types. We assume that an observation, $x=(x_1, \dots, x_n)$, is drawn iid from some unknown joint distribution $p_X(x)$, often abbreviated $p(x)$. Groups of variables are indexed with superscripts as in **Figure 1**.

We measure the relationships among the variables (in this case blood test measures) by a multivariate mutual information measure historically called “total correlation” (TC), although in modern terms it would be better described as measure of total dependence. TC is defined as:

$$TC(X) = D_{KL} \left(p(x) \parallel \prod_{i=1}^n p(x_i) \right) = \sum_i H(X_i) - H(X)$$

where H denotes the Shannon entropy, and D_{KL} is the Kullback-Leibler divergence.

The total correlation among a group of variables, X , after conditioning on some other variable, Y , can be simply defined in terms of standard conditional entropies as

$TC(X|Y) = \sum_i H(X_i|Y) - H(X|Y)$. Then we can measure the extent to which Y *explains* the correlations in X by looking at how much the total correlation is reduced by conditioning on Y :

$$TC(X;Y) = TC(X) - TC(X|Y) = \sum_i I(X_i;Y) - I(X;Y)$$

$TC(X|Y)$ is zero (and $TC(X;Y)$ maximized) if and only if the distribution of X 's conditioned on Y factorizes. This will be the case if Y contains full information about all of the common causes of the X_i 's (in other words we say that Y explains all of the correlation in X).

The principle behind Correlation Explanation (CorEx) is to search for latent factors, $Y_1, \dots, Y_m$, that maximize $TC(X;Y)$.

$$\max_{\forall j, p(y_j|x)} TC(X;Y) \quad (1)$$

This optimization searches *all possible functions of x* for the m representatives (shown as open circles in the middle row of **Figure 1**) that are most informative about the data. Directly optimizing this objective is intractable for large m , so we actually optimize a lower bound $TCL(X; Y)$ with two remarkable properties. First, we are able to optimize this lower bound efficiently (linear in the number of variables). Second, if we construct a hierarchy of

representations in which Y^1 explains correlations in X , and Y^2 explains correlations in Y^1 , etc., then to bound the information in X , we just add the contribution from each layer: $TC(X) = TC_L(X; Y^1) + TC_L(Y^1; Y^2) + \dots$. For a more detailed discussion of bounds and optimization procedure see [5, 6].

3. RESULTS

Figure 2 (next page) shows the hierarchical structure learned by CorEx including measured blood test measures and learned latent factors as numbered nodes. Variables are color-coded according to their type, as shown in the key. The significance of each variable for predicting delta MMSE (change in MMSE over one year) was determined with an F -test ($p < 0.01$) and significant factors are large and bolded in the graph. Latent factors that were significant predictors of atrophy are in red. The size of each node in **Figure 2** represents strength of correlation among connected variables and edge width is proportional to mutual information.

More factors were found to be significant predictors for brain atrophy than for cognitive decline. CorEx identified a group of clinical blood test measures with an associated latent factor, which outperformed any individual clinical measure in our dataset for predicting cognitive decline (highlighted in red in **Figure 2**). By combining correlated measures in an informative way, CorEx was able to produce a single factor that was a more robust predictor of both brain atrophy and cognitive decline. **Figure 3** demonstrates how the variables in latent factor 21 are related to each other.

Note that other structure discovered by CorEx may be biologically meaningful, even though the associated factors were not significant predictors. For example, the group of variables with the largest in-group correlation (node 0 in **Figure 2**) reflects sex-linked traits.

4. DISCUSSION

This study demonstrates how the latent factors we discovered with CorEx predict atrophy at a higher significance level and reveals relationships among a diverse set of blood test measures. The relationships were evaluated quantitatively, in terms of statistical significance, and qualitatively, in terms of the biological meaning of how measures were clustered. There were more significant predictors for brain atrophy, than for cognitive decline, perhaps because brain atrophy is a more stable measure [15]. These results also agree with existing literature [7] reporting that the commonly carried AD risk gene APOE, particularly the number of high-risk APOE4 alleles carried by an older adult, are associated with poorer longitudinal trajectory of cognitive decline and brain atrophy. Three proteins known to be associated with AD-related pathology in the aging human brain [2] were the other variables identified in the top group of predictors. The measures obtainable from routine clinical examinations for health conditions such as diabetes, heart disease, and hormone imbalance did not perform as well as the canonical measures in the group associated with latent factor #21. Future work will include longitudinal MRI data from ADNI.

5. ACKNOWLEDGMENTS

This work was supported in part by a Consortium grant (U54 EB020403) from the NIH Institutes contributing to the Big Data to Knowledge (BD2K) Initiative, including the NIBIB and NCI and by AFOSR grants FA9550-12-1-0417 and FA9550-10-1-0569.

6. REFERENCES

- [1]. Rusinek H, De Santi S, Frid D, Tsui WH, Tarshish CY, Convit A, et al. Regional brain atrophy rate predicts future cognitive decline: 6-year longitudinal MR imaging study of normal aging. *Radiology*. Dec.2003 229:691–6. [PubMed: 14657306]
- [2]. Tapiola T, Pirttila T, Mehta PD, Alafuzoff I, Lehtovirta M, Soininen H. Relationship between apoE genotype and CSF beta-amyloid (1-42) and tau in patients with probable and definite Alzheimer's disease. *Neurobiology of Aging*. Sep-Oct;2000 21:735–740. [PubMed: 11016543]
- [3]. Chong MS, Sahadevan S. Preclinical Alzheimer's disease: diagnosis and prediction of progression. *Lancet Neurology*. Sep.2005 4:576–579. [PubMed: 16109364]
- [4]. Mapstone M, Cheema AK, Fiandaca MS, Zhong X, Mhyre TR, MacArthur LH, et al. Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med*. Apr.2014 20:415–8. [PubMed: 24608097]
- [5]. Ver Steeg G, Galstyan A. Discovering structure in high-dimensional data through correlation explanation. *Advances in Neural Information Processing Systems (NIPS)*. 2014
- [6]. Ver Steeg G, Galstyan A. Maximally informative representations for high-dimensional data. 2014 <http://arxiv.org/abs/1410.7404>.
- [7]. Sadigh-Eteghad S, Talebi M, Farhoudi M. Association of apolipoprotein E epsilon 4 allele with sporadic late onset Alzheimer's disease. A meta-analysis. *Neurosciences (Riyadh)*. Oct.2012 17:321–6. [PubMed: 23022896]
- [8]. Pangman VC, Sloan J, Guse L. An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. *Appl Nurs Res*. Nov.2000 13:209–13. [PubMed: 11078787]
- [9]. Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*. Apr.2008 27:685–91. [PubMed: 18302232]
- [10]. Hua X, Hibar DP, Ching CRK, Boyle CP, Rajagopalan P, Gutman BA, et al. Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials. *Neuroimage*. Feb 1.2013 66:648–661. [PubMed: 23153970]
- [11]. Leow AD, Klunder AD, Jack CR, Toga AW, Dale AM, Bernstein MA, et al. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage*. Jun. 2006 31:627–640. [PubMed: 16480900]
- [12]. Gutman BA, Hua X, Rajagopalan P, Chou YY, Wang YL, Yanovsky I, et al. Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features. *Neuroimage*. Apr 15.2013 70:386–401. [PubMed: 23296188]
- [13]. Leow AD, Yanovsky I, Chiang MC, Lee AD, Klunder AD, Lu A, et al. Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration. *IEEE Trans Med Imaging*. Jun.2007 26:822–32. [PubMed: 17679333]
- [14]. Gutman BA, Wang Y, Yanovsky I, Hua X, Toga AW, Jack CR Jr, et al. Empowering imaging biomarkers of Alzheimer's disease. *Neurobiol Aging*. Aug 27.2014
- [15]. Jack CR Jr, Shiung MM, Gunter JL, O'Brien PC, Weigand SD, Knopman DS, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*. Feb 24.2004 62:591–600. [PubMed: 14981176]

Figure 1.

The bottom row of variables (X_i 's) represents measured quantities. Variables in higher layers are learned latent factors that explain the correlations in the layer beneath.

Figure 3.

Each point represents an ADNI participant and point color represents the value of the latent factor #21.