JOINT ESTIMATION OF MULTIPLE CLINICAL VARIABLES OF NEUROLOGICAL DISEASES FROM IMAGING PATTERNS

Yong Fan[*1], *Daniel Kaufer*[2], *Dinggang Shen*[1]

Departments of Radiology[1] and Neurology[2], University of North Carolina, Chapel Hill

ABSTRACT

This paper presents a method to estimate multiple clinical variables associated with neurological pathologies from brain images, aiming to quantitatively evaluate continuous transition of neurological pathologies from the normal to diseased state. Built upon morphological measures derived from structural MR brain images, a Bayesian regression method is developed to jointly model multiple clinical variables for capturing their inherent correlations and suppressing noise. Coupled with a feature selection technique, the regression method is used to build a joint estimator of multiple clinical variables associated with Alzheimer's disease from structural MR brain images of elderly individuals. The cross-validation results demonstrate that the proposed method has superior performance over existing techniques.

***Index Terms*—**Structural MR brain image, Bayesian regression, Alzheimer's Disease, MMSE, ADAS-Cog

1 INTRODUCTION

Neuroimaging techniques have been increasingly used in studies of neurological diseases to complement clinical assessments, such as performance on standard neuropsychological tests. Compared with the clinical assessments that rely on symptomatic measures, neuroimaging measures are thought to more directly reflect the level of brain pathologies, such as Alzheimer's Disease (AD) [1]. AD displays a continuous transition from the normal to diseased state, which is often measured by clinical variables, like Mini Mental State Examination (MMSE) score and others. However, such measures are well known to be subject to a great deal of error and fluctuation from one evaluation to another, mostly because an individual's performance can be affected by a variety of psychophysical factors that are unrelated to the underlying pathology. Therefore, it would be important to quantitatively estimate clinical variables based on the relatively more objective MR scans of individuals. This is particularly important if we want to use such estimates to evaluate the stage of AD pathology, as well as to predict the likely progression of the individual in the future.

It is challenging to build robust and generalizable regression models to estimate clinic variables from imaging data due to the sheer dimensionality of imaging measures and the intrinsic noise in the clinically measured variables. The former problem may confound the feature extraction and feature selection and also limit the regression model's generalization performance. The large intrinsic noise in the measures of clinical variables is the major challenge, which hampers the construction of robust regression models. Recent efforts have been made to develop regression techniques to estimate clinical variables, like MMSE, from brain image data [2, 3]. In [2], normalized and modulated gray matter probability maps were directly used to build linear kernel regression estimators by relevance vector machine [4]. In [3], a linear regression model was used to estimate yearly MMSE changes from intensities of structural MRI brain images and their deformation fields. To reduce the high feature dimensionality, principal component analysis (PCA) techniques have been adopted in [3]. However, it remains a problem regarding how to deal with the intrinsic noise in the clinically measured data.

To overcome these problems, we propose a multivariate regression method consisting of the following two components: 1) a Bayesian regression method to jointly model multiple clinical variables in order to capture their inherent correlations and suppress noise; 2) a recursive feature elimination method to select the most informative features for regression, as well as to improve the estimation efficiency. The regression method and the feature selection technique are used in conjunction with a bagging (bootstrap aggregating) strategy to build an ensemble regression model for achieving robust regression as well as facilitating optimal selection of the number of features used in the regression. The proposed method has been applied to MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.loni.ucla.edu\ADNI), in order to obtain an estimate of scores of MMSE and the cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS-Cog). The superiority of the proposed method has been demonstrated in comparison with existing methods.

2 METHODS

To estimate the clinical variables from brain images, it is necessary to extract a compact set of robust and informative features. Since reliable prior knowledge about problems under study is not always available in real applications, whole brain voxel-wise measures have been good choices for multivariate analysis of brain images. In computational neuroanatomy, the typical voxel-wise measures include

[*]email: yfan@nlpr.ia.ac.cn; current affiliation: National Laboratory of Pattern Recognition, CASIA, Beijing, China

deformation fields, Jacobian determinants, and tissue density maps [5], which often come with hundred thousands of variables. To achieve effective and efficient regression, the voxel-wise measures need to be reduced to a small and optimal set of features. For feature dimensionality reduction, the PCA technique has been used to transform high dimensional voxel-wise features onto a linear subspace in [3] to build regression models of MMSE. Another strategy to reduce feature dimensionality is feature selection, which directly select features informative for learning tasks [6]. By considering that only parts of the brain are affected by a specific neurological pathology, the feature selection techniques might be more suitable than holistic PCA transformation. Therefore, we use a feature selection technique to select the most informative features for constructing regression models.

To evaluate the stage of neurological pathologies by quantitatively estimating clinical variables, similar to the clinical settings, multiple types of clinical variables should be estimated for achieving better accuracy. Therefore, we need to solve the problem of regression with multiple responses. The simplest way for regressing with multiple responses is to build a separate model for each target value, as done in [2]. Although this method is computational efficiency, it ignores the inherent correlation among clinical variables, which is useful to improve the estimation of noisy variables [7]. On the other hand, it is not helpful or even hurts the performance of regression if simply learning multiple variables together without properly modeling their relatedness [8]. Although inherent correlation exists among clinical variables, like MMSE and ADAS-Cog, it is a difficult task to explicitly model their relatedness. Instead, we model the relatedness of different clinical variables as sharing a common subset of informative basis functions and a common subset of relevant features within a kernel based regression framework, aiming to obtain solutions that are sparse in both sample space and feature space. Similar modeling strategies have been used separately to achieve sparseness in spaces of samples or features in multi-task learning problems [9]. We detail these techniques in following subsections.

2.1 Joint model of multiple clinical variables

Given a training set of brain scans with their corresponding clinical variables, represented by $\{(\boldsymbol{x}_j, \boldsymbol{y}_j)\}_{j=1}^{N}$, where $\boldsymbol{x}_j$ is a feature vector of imaging pattern computed from a brain scan, and $y_j = (y_j^1, y_j^2, \dots, y_j^m)'$ is an m-dimension vector of clinical variables, we model the clinical variables by a set of functions $f = (f^1, f^2, \dots, f^m)'$ linear in the parameters with additive noise $\omega = (\omega^1, \omega^2, \dots, \omega^m)'$, i.e.,

$$f^i(x; \beta^i) = \sum_{k=1}^{N} \beta_k^i K(x, x_k) + \beta_0^i \ , \tag{1}$$

where $\beta^i = (\beta_0^i, \beta_1^i, \dots, \beta_N^i)'$ are parameters to be learned from the training data, $K(\cdot,\cdot)$ is a kernel function, which allows us to easily handle problems with high feature dimensionality and model nonlinear functions with linear parameters. These regression functions model individual clinical variables with different sets of parameters, but share the same kernel basis functions.

With assumptions that $\{\omega^i\}_{i=1}^{m}$ are independent Gaussian noise, i.e., $\omega^i = \mathcal{N}(0, \sigma^i)$ and training data are generated independently, we can learn $f^i(x; \beta^i)$ separately by maximizing the likelihood of the training samples

$$p(\boldsymbol{y}^i | X, \beta^i, \sigma^i) = \prod_{j=1}^{N} (2\pi)^{-\frac{1}{2}} (\sigma^i)^{-1} exp\left\{-\frac{1}{2(\sigma^i)^2}\left(y_j^i - f^i(x_j; \beta^i)\right)^2\right\}, \tag{2}$$

where $X = \{x_j\}_{j=1}^{N}$ is the training set of image patterns, $\boldsymbol{y}^i = (y_1^i, y_2^i, \dots, y_N^i)'$ is the training samples of the i-th clinical variable to be estimated. In the Bayesian inference framework, one can get a sparse solution, robust to over-fitting, by defining automatic relevance determination (ARD) prior over the parameters of regression models [10]. The ARD prior over the model parameters is a zero-mean Gaussian distribution

$$p(\beta^i | \boldsymbol{\alpha}^i) = \prod_{k=0}^{N} \mathcal{N}\left(0, (\alpha_k^i)^{-1}\right) = \prod_{k=0}^{N} (2\pi)^{-\frac{1}{2}} \alpha_k^i exp\{-(\alpha_k^i \beta_k^i)^2/2\}, \tag{3}$$

where $\boldsymbol{\alpha}^i = \left\{(\alpha_k^i)^{-1}\right\}_{k=0}^{N}$ is a vector of $N+1$ hyperparameters of the ARD prior over the parameters of the regression model f^i, which moderates the strength of the prior information. The parameters of regression models can be learned using the inference strategy proposed in [4]. In the inference process, some components of $\boldsymbol{\alpha}^i$ approach infinity, forcing their associated parameters of the regression model shrink to zero. We denote this method by S-modeling.

To benefit from inherent correlation in different clinical variables without implicit modeling of their relatedness, we adopt a method that incorporates relatedness of different variables by specifying the same hyperparameters for model parameters, i.e., different regression models of eqn. (1) use the same ARD prior of eqn. (3) for their model parameters which are different for models of different clinical variables. The joint modeling (J-modeling) problem can be solved by the method proposed in [9].

Given a testing set $\{(\tilde{\boldsymbol{x}}_j, \tilde{\boldsymbol{y}}_j), j = 1, \dots, n\}$, the performance of a regression model can be measured for each clinical variable separately by the correlation coefficient (CORR) between measured and estimated variables and the square root of mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left\| f^i(\tilde{\boldsymbol{x}}_j) - \tilde{\boldsymbol{y}}_j^i \right\|}. \tag{4}$$

2.2 Feature selection using a recursive feature elimination strategy

It is possible to simultaneously select features with the construction of the Bayesian regression model by putting a scaling parameter on each feature via ARD, however the inference task with a huge number of variables and a relatively small number of samples is too difficult if not impossible. Therefore, we propose to use a recursive feature elimination strategy to boost the regression performance. Such a feature selection strategy has been successful in classification studies.

The recursive feature elimination based feature selection strategy was proposed for gene selection based on linear support vector machines (SVM) [11]. Starting with all features, this feature selection method recursively removes the feature which causes the minimum variation in a generalization error bound derived from the weight vector of a SVM function [11]. This feature ranking criterion is equivalent to the norm of feature-wise partial derivative of the generalization error bound, so it is a feature ranking criterion based on a learning model. Since the regression model specified by eqn. (1) is in the same form of SVMs, the recursive feature selection strategy can be used with the regression models. However, unlike SVMs, the hyperparameters of Bayesian models are tuned systematically during the inference.

To build a joint regression model of multiple responses, the feature ranking criterion based on the regression model is computed separately for each response variable, and the same number of the top ranked features from different responses are combined to build the joint regression model.

To determine the number of features used for constructing a regression model, we use a v-fold cross-validation by dividing the training data set into v subsets with an approximately equal number of training samples. The cross-validation has v different runs, each with one different subset selected as the testing data set and others used as the training set. For feature elimination, the performance of regression models with different numbers of features is tested to select the optimal number of features that yield a model with the minimal RMSE.

Typically, different feature sets are selected in the v runs of cross-validation. Rather than combine the different feature sets somehow to build a single regression model, we build an ensemble regression model by averaging outputs of base models derived from different cross-validation runs. The ensemble regression tends to have more stable performance, as demonstrated empirically and theoretically [12].

3. RESULTS

The proposed methods were used to estimate the scores of MMSE and ADAS-Cog from structural MRI data. From ADNI database, structural MRI image data of 264 individuals were obtained (52 AD patients, 148 mild cognitive impairment patients: MCI, and 64 cognitive controls: CN), with demographic information shown in Table 1.

Table 1. Characteristics of the participants in this study.

Group	AD	MCI	CN	All
No. of subjects	52	148	64	264
Percent of males	40%	64%	50%	56%
Age (year), mean± std	76.8±7.1	75.5±7.4	75.1±5.5	75.7±7.0
MMSE, mean± std	23.0±1.8	26.8±1.7	29.1±0.9	26.6±2.6
ADAS-Cog, mean± std	18.9±5.6	11.8±4.1	6.4±3.2	11.9±5.9

From structural MRI images, voxel-wise gray matter (GM) tissue density maps were computed by normalizing them to a standard template space within a mass-preserving shape transformation framework . Tissue density maps were normalized by total intra-cranial volume (ICV) in order to account for variations in head size, and smoothed using 8 mm full-width at half-maximum (FWHM) Gaussian smoothing kernel.

3.1 Regression performance and comparison with other methods

We applied the proposed method to estimate scores of MMSE and ADAS-Cog from GM tissue density maps for all the subjects. To get an unbiased estimation of the regression performance, a 10-fold cross-validation experiment was performed. The number of features used in regression was optimized based on the training data with another 10-fold cross-validation.

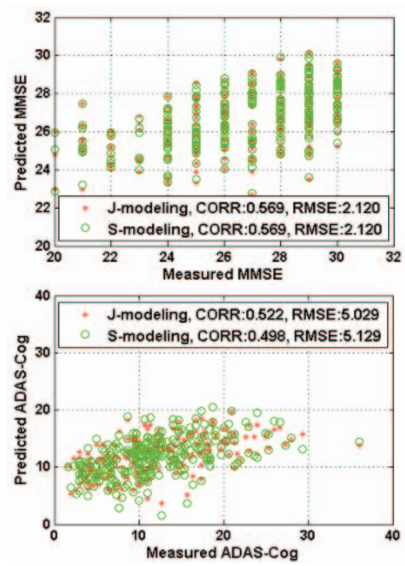

Fig.1. 10-fold cross-validation results of the regression modeling of MMSE and ADAD-Cog jointly and separately, with the clinically measured values shown in x-axes and the estimated values shown in y-axes. The performance of regression is measured by correlation coefficient (CORR) and square root of mean squared error (RMSE) between clinically measured and estimated clinical scores.

The testing results of both methods of J-modeling and S-modeling are shown in Fig. 1, demonstrating that the J-modeling achieved better performance than the S-modeling for ADAS-Cog w.r.t. both CORR and RMSE between the clinically measured and the estimated variables, although they have similar performance for MMSE. The regression method demonstrates a promising estimation performance in view of the fact that the measurement of MMSE and ADAS-Cog scores is subject to error. Actually, the differences between the clinically measured and the estimated variables are comparable to their fluctuation estimated from the longitudinal data of ADNI normal subjects between 6 months. Based on the data available on Jan, 2009, the square root of mean square difference between MMSE scores of baseline and 6-month follow-up for all available normal subjects is 1.28, and the same value for ADAS-Cog is 3.12.

In the same experiment setting, our method is compared with the regression model built using J-modeling based on PCA features [3]. The performances of these methods are summarized in Table 2, demonstrating the superiority of the proposed method.

Table 2. Comparison of regression performances.

	PCA		Feature selection	
	MMSE	ADAD-Cog	MMSE	ADAS-Cog
RMSE	2.55	5.86	2.12	5.03
CORR	0.18	0.11	0.57	0.52

3.2 Brain atrophy patterns detected by regression models

We construct a brain atrophy pattern as a spatial map of features that are selected to build regression models in J-modeling to show a visual picture of weights of voxel-wise features. Since a *v*-fold cross validation is used in the training stage to generate base estimators and optimize the number of features used in regression, the weights of each feature are computed and averaged from all regression models in the training stage. The spatial map is then normalized by its maximum value. So, the spatial map of features depicts the relative importance of these features in building estimators for scores of MMSE and ADAS-Cog.

The voxel-wise correlation between GM tissue density maps and the clinical variables is also estimated using SPM5 (www.fil.ion.ucl.ac.uk/spm/software/spm5). The resulting t-statistic maps are visually compared with those generated from the regression models. As shown in the top panel of Fig. 2, the brain atrophy patterns identified (thresholded with 0.2) by the regression method are different w.r.t. different clinical variables, however they are largely in agreement with the brain atrophy highly correlated with clinical variables detected via voxel-wise correlation analysis, shown in the bottom panels of Fig. 2 ($p<0.05$, FDR corrected for multiple comparison). The difference of these two techniques is also reflected by the difference of brain atrophy patterns. However, all of these figures reveal that the brain atrophy pattern highly correlated with clinical measures is complex, including the hippocampus, consistent with the known contribution of these regions to memory processes.

4 CONCLUSION

A general regression method is proposed in this paper to estimate clinical variables based on neuroimages. Instead of constructing regression models separately for different tasks, the proposed method builds a joint estimator, aiming to improve the prediction robustness. The regression method of jointly modeling clinical variables demonstrated a promising estimation performance. The differences between the clinically measured and the estimated variables are comparable to their fluctuation estimated from the longitudinal data of ADNI normal subjects between 6 months. We expect better results can be obtained if advanced feature extraction techniques and more sophisticated modeling of relatedness among clinical variables are used in conjunction with the regression algorithms.

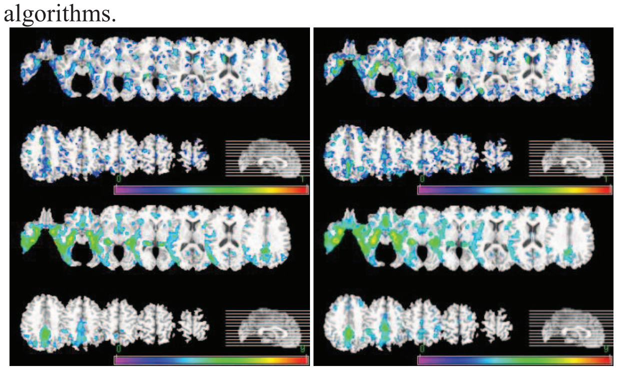

Fig. 2. Spatial patterns of brain atrophy that are significantly correlated with MMSE (left) and ADAS-Cog (right), measured by regression analysis (top) and by voxel-wise correlation analysis ($p<0.05$, FDR corrected) (bottom). The color scales indicate, respectively, regression weight for regression analysis (top) and t-value for voxel-wise correlation analysis (bottom). Images are displayed in radiological convention.

REFERENCES

[1] P. Thompson, K. Hayashi, et al., "Tracking Alzheimer's disease," *Annals of New York academy of Sciences,* vol. 1097, pp. 183-214, 2007.

[2] C. Chu, S. Klöppel, et al., , "Regression analysis for clinical scores of Alzheimer’s Disease using multivariate machine learning method," in *Human Brain Mapping* Chicago, IL, USA, 2007.

[3] S. Duchesne, A. Caroli, et al., "Predicting clinical variable from MRI features: application to MMSE in MCI," in *MICCAI*, 2005, pp. 392-399.

[4] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research,* vol. 1, pp. 211-244, 2001.

[5] J. Ashburner, J. G. Csernansky, et al., "Computer-assisted imaging to assess brain structure in healthy and diseased brains," *The Lancet (Neurology),* vol. 2, pp. 79-88, 2003.

[6] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[7] P. Boyle and M. Frean, "Dependent Gaussian processes," in *Advances in Neural Information Processing Systems*, 2005, pp. 217-224.

[8] E. V. Bonilla, K. M. A. Chai, and C. K. I. Williams, "Multi-task Gaussian Process Prediction," in *Advances in Neural Information Processing Systems* 2008.

[9] A. Thayananthan, R. Navaratnam, et al., "Pose estimation and tracking using multivariate regression," *Pattern Recognition Letters,* vol. 29, pp. 1302-1310, 2008.

[10] R. M. Neal, *Bayesian learning for neural networks*: Springer, 1996.

[11] I. Guyon, J. Weston, et al., "Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning,* vol. 46, pp. 389-422, January 2002.

[12] L. Breiman, "Bagging predictors," *Machine Learning,* vol. 24, pp. 123-140, 1996.