FDR-HS: An Empirical Bayesian Identification of Heterogenous Features in Neuroimage Analysis

Xinwei Sun[1,6], Lingjing Hu[2](✉), Fandong Zhang[3,6], Yuan Yao[4](✉), and Yizhou Wang[5,6]

[1] School of Mathematical Science, Peking University, Beijing, 100871, China
[2] Yanjing Medical College, Capital Medical University, Beijing, 101300, China
[3] Key Laboratory of Machine Perception (Ministry of Education), Department of Machine Intelligence, School of Electronics Engineering and Computer Science,Peking University, Beijing 100871, China
[4] Hong Kong University of Science and Technology and Peking University, China
[5] National Engineering Laboratory for Video Technology, Key Laboratory of Machine Perception, School of EECS, Peking University, Beijing, 100871, China
[6] Deepwise Inc., Beijing, 100085, China

Abstract. Recent studies found that in voxel-based neuroimage analysis, detecting and differentiating "procedural bias" that are introduced during the preprocessing steps from lesion features, not only can help boost accuracy but also can improve interpretability. To the best of our knowledge, GSplit LBI is the first model proposed in the literature to simultaneously capture both procedural bias and lesion features. Despite the fact that it can improve prediction power by leveraging the procedural bias, it may select spurious features due to the multicollinearity in high dimensional space. Moreover, it does not take into account the heterogeneity of these two types of features. In fact, the procedural bias and lesion features differ in terms of volumetric change and spatial correlation pattern. To address these issues, we propose a "two-groups" Empirical-Bayes method called "FDR-HS" (False-Discovery-Rate Heterogenous Smoothing). Such method is able to not only avoid multicollinearity, but also exploit the heterogenous spatial patterns of features. In addition, it enjoys the simplicity in implementation by introducing hidden variables, which turns the problem into a convex optimization scheme and can be solved efficiently by the expectation-maximum (EM) algorithm. Empirical experiments have been evaluated on the Alzheimer's Disease Neuroimage Initiative (ADNI) database. The advantage of the proposed model is verified by improved interpretability and prediction power using selected features by FDR-HS.

Keywords: · Voxel-based Structural Magnetic Resonance Imaging · False Discovery Rate Heterogenous Smoothing · Procedural Bias · Lesion Voxel

Dedicated to Professor Bradley Efron on the occasion of his 80th birthday.

1 Introduction

In recent years, the issue of model interpretability attracts an increasing attention in voxel-based neuroimage analysis of disease prediction, e.g. [9,5]. Examples include, but not limited to, the preprocessed features on structural Magnetic Resonance Imaging (sMRI) images that usually contain the following voxel-wise features: (1) lesion features that are contributed to the disease (2) procedural bias introduced during the preprocessing steps and shown to be helpful in classification [12,3] (3) irrelevant or null features which are uncorrelated with disease label. Our goal is to stably select non-null features, i.e. lesion features and procedural bias with high power/recall and low false discovery rate (FDR).

The lesion features have been the main focus in disease prediction. In dementia disease such as Alzheimer's Disease (AD), such features are thought to be geometrically clustered in atrophied regions (hippocampus and medial temporal lobe etc.), as shown by the red voxels in Fig. 1 (A). To explore such spatial patterns, multivariate models with Total Variation [10] regularization can be applied by enforcing smoothness on the voxels in neighbor, e.g. the n^2GFL [15] can stably identify the early damaged regions in AD by harnessing the lesions.

Recently, another type of features called procedural bias, which are introduced during the preprocessing steps, are found to be helpful for disease prediction [12]. Again, taking AD as an example, the procedural bias refer to the mistakenly enlarged Gray Matter (GM) voxels surrounding locations with cerebral spinal fluid (CSF) spaces enlarged, e.g. lateral ventricle, as shown in Fig. 1 (A). This type of features has been ignored in the literature until recently, when the GSplit LBI [12] was targeted on capturing both types of features via a split of tasks of TV regularization (for lesions) and disease prediction with general linear model (with procedural bias). By leveraging such bias, it can outperform models which only focus on lesions in terms of prediction power and interpretability.

However, GSplit LBI may suffer from inaccurate feature selection due to the following limitations in high dimensional feature space: [7]: (1) multicollinearity: high correlation among features in multivariate models [14]; (2) "heterogenous features": the procedural bias and lesion features differ in terms of volumetric change (enlarged v.s. atrophied) and particularly spatial pattern (surroundingly distributed v.s. spatially cohesive). Specifically, the multicollinearity could select spurious null features which are inter-correlated with non-nulls. Moreover, GSplit LBI fails to take into account the heterogeneity since it enforces correlation on features without differentiation. Such problems altogether may result in inaccurate selection of non-nulls, especially procedural bias. As shown in Fig. 1 (B) and Table 2, the procedural bias selected by GSplit LBI are unstably scattered on regions that are less informative than ventricle. Moreover, the collinearity among features tends to select a subset of features among correlated ones, as discussed in [16]. Such a limitation leads to the ignorance of many meaningful regions (such as medial temporal lobe, thalamus etc.) of GSplit LBI in selecting lesion features, as identified by the purple frames of FDR-HS in Fig. 1 (B).

[7] Please refer supplementary material for detailed and theoretical discussion

Moreover, the two problems above may get worse as dimensionality grows. In our experiments with a fine resolution ($4 \times 4 \times 4$ of 20,091 features), the prediction accuracy of GSplit LBI deteriorates to 89.77% (as shown in Table 3), lower than 90.91% reported in [12] with a coarse resolution ($8 \times 8 \times 8$ of 2,527 features).

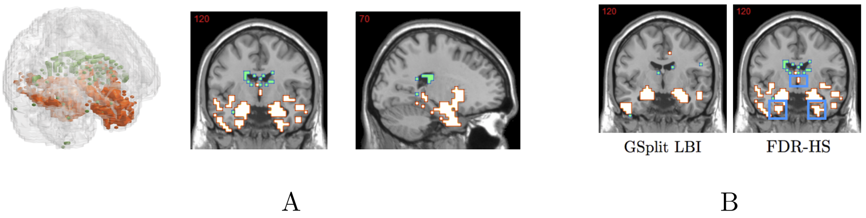


Fig. 1. A: the features selected by FDR-HS (green denotes procedural bias; red denotes lesion features which are geometrically clustered) B: comparison with GSplit LBI

To resolve the problems above, we propose a "two-groups" empirical Bayes method to identify heterogenous features, called FDR-HS standing for "FDR Heterogenous smoothing" in this paper. As a univariate FDR control method, it avoids the collinearity problem by proceeding voxel-by-voxel, as discussed in [7]. Moreover, it can deal with heterogeneity by regularizing on features with different levels of spatial coherence in different feature groups, which remedies the problem of losing spatial patterns that most conventional mass-univariate models suffer from, such as two sample T-test, BH$_q$ [4] and LocalFDR [7]. By introducing a binary latent variable, our problem turns into a convex optimization and can be solved efficiently via EM algorithm like [13]. The method is applied to a voxel-based sMRI analysis for AD with a fine resolution (4×4×4 of 20,091 features). As a result, our proposed method exhibits a much stabler feature extraction than GSplit LBI, and achieves much better classification accuracy at 91.48%.

2 Method

Our dataset consists of p voxels and N samples $\{x_i, y_i\}_1^N$ where x_{ij} denotes the intensity value of the j^{th} voxel of the i^{th} sample and $y_i = \{\pm 1\}$ indicates the disease status (-1 denotes AD). The FDR-HS method is proposed to select non-null features. Such method is the combination of "two-groups" model and heterogenous regularization, which is illustrated in Fig. 2 and discussed below. **Model Formulation.** Assuming for each voxel $i \in \{1, ..., p\}$, the statistic z_i is sampled from the following mixture:

$$z_i \sim \sum_{k=0}^{1} \mathrm{p}(s_i = k)\mathrm{p}(z_i | s_i = k) = c_i f_1(z_i) + (1 - c_i) f_0(z_i), \tag{2.1}$$

where s_i is a latent variable indicating if the voxel i belongs to the group of null features ($s_i = 0$) or the group of non-null ones ($s_i = 1$), $c_i = \mathrm{p}(s_i = 1) =$

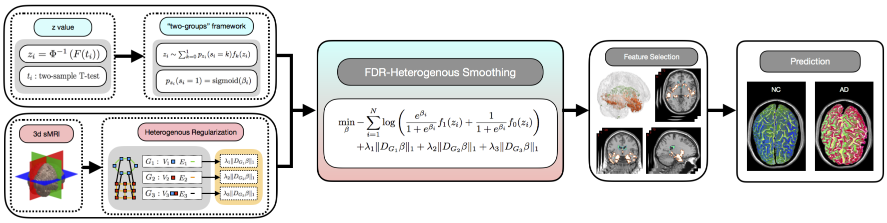

z value
$z_i = \Phi^{-1}(F(t_i))$
t_i : two-sample T-test
"two-groups" framework
$z_i \sim \sum_{k=0}^{1} p_{z_i}(s_i = k) f_k(z_i)$
$p_{z_i}(s_i = 1) = \text{sigmoid}(\beta_i)$
3d sMRI
Heterogenous Regularization
$G_1 : V_1$ E_1
$G_2 : V_2$ E_2
$G_3 : V_3$ E_3
$\lambda_1 \|D_{G_1}\beta\|_1$
$\lambda_2 \|D_{G_2}\beta\|_1$
$\lambda_3 \|D_{G_3}\beta\|_1$
FDR-Heterogenous Smoothing
$\min_{\beta} - \sum_{i=1}^{N} \log\left(\frac{e^{\beta_i}}{1+e^{\beta_i}} f_1(z_i) + \frac{1}{1+e^{\beta_i}} f_0(z_i)\right)$
$+\lambda_1 \|D_{G_1}\beta\|_1 + \lambda_2 \|D_{G_2}\beta\|_1 + \lambda_3 \|D_{G_3}\beta\|_1$
Feature Selection
Prediction
NC
AD

Fig. 2. Illustration of FDR-HS model.

$\text{sigmoid}(\beta_i) = e^{\beta_i} / \left(1 + e^{\beta_i}\right)$ and $z_i = \Phi^{-1}\left(F_{N-2}(t_i)\right)$ with t_i computed by two-sample t-test. Correspondingly, $f_0(\cdot)$ is density function of nulls, i.e. uncorrelated with AD and $f_1(\cdot)$ is that of non-nulls, i.e. procedural bias and lesions. The loss function can thus be defined as negative log-likelihood of z_i:

$$\ell(\beta) = -\sum_{i=1}^{N} \log\left(\frac{e^{\beta_i}}{1+e^{\beta_i}} f_1(z_i) + \frac{1}{1+e^{\beta_i}} f_0(z_i)\right) \tag{2.2}$$

which can be viewed as logistic regression (when f_0 and f_1 are replaced with binaries, as (2.6)) with identity design matrix since (2.1) proceeds voxel-by-voxel. Hence, it does not have the problem of multicollinearity.

Selecting Features. To select features, we compute the posterior distribution of s_i conditioned on z_i and $\widehat{\beta}_i$ (estimated β_i) and features with

$$\text{p}(s_i = 0 | z_i, \widehat{\beta}_i) = \frac{(1-\widehat{c}_i) f_0(z_i)}{\widehat{c}_i f_1(z_i) + (1-\widehat{c}_i) f_0(z_i)} < \gamma \quad \left(\widehat{c}_i = e^{\widehat{\beta}_i} / \left(1 + e^{\widehat{\beta}_i}\right)\right) \tag{2.3}$$

are selected. The $\gamma \in (0,1)$ is pre-setting threshold parameter.

Heterogenous Spatial Smoothing. However, (2.1) may lose spatial structure of non-nulls, especially lesion features. Besides, note that the procedural bias and lesion features are heterogenous in terms of volumetric change and level of spatial coherence. Hence, to capture the spatial structure of heterogenous features, we split the graph of voxels which denotes as $\boldsymbol{G}$ [8] into three subgraphs, i.e. $\boldsymbol{G} = \boldsymbol{G}_1 \cup \boldsymbol{G}_2 \cup \boldsymbol{G}_3$ with:

$$\boldsymbol{G}_1 = (\boldsymbol{V}_1, \boldsymbol{E}_1), \ \boldsymbol{V}_1 = \{i : z_i \leq 0\}, \ \boldsymbol{E}_1 = \{(i,j) \in \boldsymbol{E} : z_i \leq 0, z_j \leq 0\} \tag{2.4a}$$

$$\boldsymbol{G}_2 = (\boldsymbol{V}_2, \boldsymbol{E}_2), \ \boldsymbol{V}_2 = \{i : z_i > 0\}, \ \boldsymbol{E}_2 = \{(i,j) \in \boldsymbol{E} : z_i > 0, z_j > 0\} \tag{2.4b}$$

$$\boldsymbol{G}_3 = (\boldsymbol{V}_3, \boldsymbol{E}_3), \ \boldsymbol{V}_3 = \boldsymbol{V}_1 \cup \boldsymbol{V}_2, \ \boldsymbol{E}_3 = \{(i,j) \in \boldsymbol{E} : z_i > 0, z_j \leq 0\} \tag{2.4c}$$

where $\boldsymbol{G}_1$ denotes the subgraph restricted on enlarged voxels (procedural bias since -1 denotes AD); $\boldsymbol{G}_2$ denotes the subgraph restricted on degenerate voxels (lesion features); $\boldsymbol{G}_3$ denotes the bipartite graph with the edges connecting enlarged and degenerate voxels. The optimization function can be redefined as:

$$g(\beta) = \ell(\beta) + \lambda_{pro} \|D_{\boldsymbol{G}_1}\beta\|_1 + \lambda_{les} \|D_{\boldsymbol{G}_2}\beta\|_1 + \lambda_{pro\text{-}les} \|D_{\boldsymbol{G}_3}\beta\|_1 \tag{2.5}$$

[8] Here $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ is the node set of voxels, $\boldsymbol{E}$ is the edge set of voxel pairs in neighbor (e.g. 3-by-3-by-3).

where $D_{\boldsymbol{G}_k}\beta = \sum_{(i,j)\in \boldsymbol{E}_k} \beta_i - \beta_j$ for $k \in \{1,2,3\}$ denote graph difference operator on $\boldsymbol{G}_{k=1,2,3}$. By setting the group of regularization hyper-parameters $\{\lambda_{pro}, \lambda_{les}, \lambda_{pro\text{-}les}\}$ with different values, we can enforce spatial smoothness on three subgraphs at different level in a contrast to the traditional homogeneous regularization in [13]. The choice of each hyper-parameter, similar to [13], it is a trade-off between over-fitting and over-smoothing. Too small value tends to select features more than needed, while too large value will oversmooth hence the features are less clustered. Note that lesion features are more spatially coherent than procedural bias and they are located in different regions, the reasonable choice of regularization hyper-parameters tend to have $\lambda_{les} \leq \lambda_{pro} \leq \lambda_{pro\text{-}les}$.
Optimization. Note that the function (2.5) is not convex. Hence we adopted the same idea in [13] that introduced the latent variables s_i and $= 1$ if $z_i \sim f_1(z)$ and 0 if $z_i \sim f_0(z)$. The $\ell(\beta)$ and $g(\beta)$ are modified as:

$$\ell(\beta, s) = \sum_{i=1}^{N} \left\{ \log\left(1 + e^{\beta_i}\right) - s_i \beta_i \right\} \tag{2.6}$$

$$g(\beta, s) = \ell(\beta, s) + \lambda_{pro}\|D_{\boldsymbol{G}_1}\beta\|_1 + \lambda_{les}\|D_{\boldsymbol{G}_2}\beta\|_1 + \lambda_{pro\text{-}les}\|D_{\boldsymbol{G}_3}\beta\|_1 \tag{2.7}$$

To solve (2.7), we can implement Expectation-Maximization (EM) algorithm to alternatively solve β and s. Suppose currently we are in the $(k+1)^{th}$ iteration. ***In the E-step***, we can estimate s_i by expectation value conditional on (β^k, z_i): $\tilde{s}_i = \mathrm{E}(s_i|\beta^k, z_i) = \frac{c_i^k f_1(z_i)}{c_i^k f_1(z_i) + (1 - c_i^k) f_0(z_i)}$.
In the M-step, we plug $\tilde{s}_i$ into (2.7), denote $\widetilde{D}_{\boldsymbol{G}} = \left[D_{\boldsymbol{G}_1^T}, \frac{\lambda_{les}}{\lambda_{pro}} D_{\boldsymbol{G}_2^T}, \frac{\lambda_{pro\text{-}les}}{\lambda_{pro}} D_{\boldsymbol{G}_3^T} \right]^T$ and expand $\ell(\beta|\tilde{s}^k)$ using a second-order Taylor approximation at the β^k. Then the M-step turns into a generalized lasso problem with square loss:

$$\min_{\beta} \frac{1}{2}\|\tilde{y} - \widetilde{X}\beta\|_2^2 + \lambda_{pro}\|\widetilde{D}_{\boldsymbol{G}}\beta\|_1 \tag{2.8}$$

where $\widetilde{X} = diag\{\sqrt{w_1}, ..., \sqrt{w_p}\}$ and $\tilde{y}_i = \sqrt{w_i}\left(\beta_i^k - \nabla_\beta \ell(\beta|\tilde{s}_i^k)_{|_{\beta^k}}/w_i\right)$ with $w_i = \nabla_\beta^2 \ell(\beta|\tilde{s}_i)_{|_{\beta^k}}$. Note that X and $\widetilde{D}_{\boldsymbol{G}}$ are sparse matrices, hence (2.8) can be efficiently solved by Alternating Direction Method of Multipliers (ADMM) [6] which has a complexity of $O(p \log p)$.
Estimation of f_0 and f_1. Before the iteration, we need to estimate $f_0(z)$ and $f_1(z)$. The marginal distribution of z can be regarded as mixture models with p components: $z \sim \frac{1}{p}\sum_{i=1}^{p} g_i(z),\ g_i(z) = p(s_i)p(z|s_i) = c_i f_1(z) + (1 - c_i) f_0(z)$ Hence, the marginal distribution of z is $f(z) = \bar{c} f_1(z) + (1 - \bar{c}) f_0(z)$, which is equivalent to LocalFDR [7]. We can therefore implement the CM (Central Matching) [7] method to estimate $\{f_0(z), \bar{c}\}$ and kernel density to estimate $f(z)$. The $f_1(z)$ can thus be given as $(f(z) - f_0(z)\bar{c})/(1 - \bar{c})$.

3 Experimental Results

In this section, we evaluate the proposed method by applying it on the ADNI database http://adni.loni.ucla.edu. The database is split into 1.5T and 3.0T (namely 15 and 30) MRI scanner magnetic field strength datasets. The 15 dataset contains 64 AD, 110 MCI (Mild Cognitive Impairment) and 90 NC, while the 30 dataset contains 66 AD and 110 NC. After applying DARTEL VBM [2] preprocessing pipeline on the data with scale of 4×4×4 mm^3 voxel size, there are in total 20,091 voxels with average values in GM population on template greater than 0.1 and they are served as input features. We designed experiments on 1.5T AD/NC, 1.5T MCI/NC and 3.0T AD/NC tasks, namely 15ADNC, 15MCINC and 30ADNC, respectively.

3.1 Prediction Results

To test the efficacy of selected features by FDR-HS and compare it with other univariate models (as listed in Table 1), we feed them into elastic net classifier, which has been one of the state-of-the-arts in the prediction of neuroimage data [11]. The hyper-parameters are determined by grid-search. In details, the threshold hyper-parameter of p-value in T-test and q-value in BH$_q$ are optimized through $\{0.001, 0.01, 0.02, 0.05, 0.1\}$; the threshold hyper-parameter for choosing non-nulls, i.e. γ for FDR-HS (2.3) and the counterpart of LocalFDR [7], are chosen from $\{0.1, 0.2, ..., 0.5\}$. Besides, the regularization parameters λ_{pro}, λ_{les} and $\lambda_{pro\text{-}les}$ of FDR-HS are ranged in $\{0.1, 0.2, ..., 2\}$. For elastic net, the regularization parameter is chosen from $\{0.1, 0.2, ..., 2, 5, 10\}$; the mixture parameter α is from $\{0, 0.01, ..., 1\}$. Moreover, we compare our model to GSplit LBI and elastic net, adopting the same optimized strategy for hyper-parameters in [12] (the top 300 negative voxels are identified as procedural bias [12]) and those of elastic net following after the univariate models, as mentioned above.

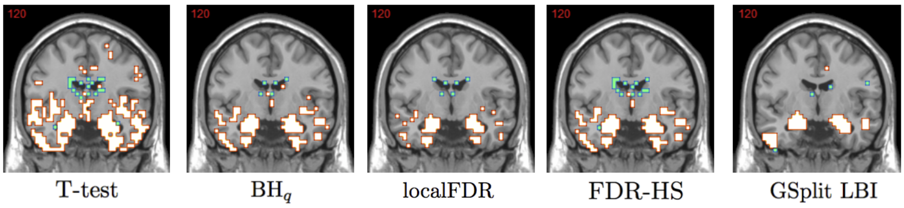
A 10-fold cross-validation strategy is applied and the classification results for all tasks are summarized in Table 1. As shown, our method yields better results than others in all cases, that includes: (1) FDR-HS can select features with more prediction power than other univariate models due to the ability to capture heterogenous spatial patterns; (2) FDR-HS can achieve better classification results than multivariate methods in high dimensional settings, in which the non-nulls may be represented by other nulls that are highly correlated with them.

Table 1. Comparison between FDR-HS and others on 10-fold classification result

	Univariate + ElasticNet				Multivariate	
	T-test	BH$_q$ [4]	LocalFDR [7]	**FDR-HS**	GSplit LBI [12]	Elastic Net [16]
15ADNC	89.61%	89.61%	87.01%	**90.26%**	85.06%	87.01%
15MCINC	70.50%	71.00%	73.50%	**75.00%**	72.50%	72.00%
30ADNC	88.64%	89.77%	89.77%	**91.48%**	89.77%	88.07%

3.2 Feature Selection Analysis

We used 2-d images of 30ADNC to visualize the features of all methods under the hyper-parameters that give the best accuracy. As shown in Fig. 3,



T-test BH$_q$ localFDR FDR-HS GSplit LBI

Fig. 3. The comparison of FDR-HS between others in terms of feature selection (30ADNC). Red denotes lesions; blue denotes procedural bias.

the lesion features selected by FDR-HS are located clustered in early damaged regions; while procedural bias are surrounding around lateral ventricle. Besides, such a result is given by $\lambda_{les} < \lambda_{pro} < \lambda_{pro\text{-}les}$, which agrees with that the larger value results in features with lower level of spatial coherence. In contrast, the lesions selected by T-test and BH$_q$ are scattered and redundant; some procedural bias around lateral ventricle are missed by BH$_q$ and LocalFDR. Moreover, GSplit LBI selected procedural bias on regions with CSF space less enlarged than lateral ventricle; besides, it ignored lesions located in medial temporal lobe, Thalamus and Fusiform etc., which are believed to be the early damaged regions [1,8].

Besides, we also evaluated the stability of selected features using multi-set Dice Coefficient (mDC) measurement defined in [15]. Larger mDC implies more stable feature selection. As shown in Table 2, our model can obtain more stable results than GSplit LBI which suffer the "collinearity" problem.

Table 2. Comparison between FDR-HS and others on stability (measured by mDC)

	T-test	BH$_q$	LocalFDR	FDR-HS	GSplit LBI
mDC$^{(+)}$ (Lesion features)	0.6705	0.6248	0.6698	**0.6842**	0.4598
mDC$^{(-)}$ (Procedural Bias)	0.6267	0.5541	0.5127	**0.6540**	0.3033

4 Conclusions

In this paper, a "two-groups" Empirical-Bayes model is proposed to stably and efficiently select interpretable heterogenous features in voxel-based neuroimage analysis. By modeling prior probability voxel-by-voxel and using a heterogenous regularization, the model can avoid multicollinearity and exploit spatial patterns of features. With experiments on ADNI database, the features selected by our models have better interpretability and prediction power than others.

Acknowledgements. This work was supported in part by 973-2015CB351800, NSFC-61625201, 61527804, National Basic Research Program of China (Nos.

2015CB85600, 2012CB825501), NNSF of China (Nos. 61370004, 11421110001), HKRGC grant 16303817, Scientific Research Common Program of Beijing Municipal Commission of Education (No. KM201610025013) and grants from Tencent AI Lab, Si Family Foundation, Baidu BDI and Microsoft Research-Asia.

References

1. Aggleton, J.P., Pralus, A., Nelson, A.J., Hornberger, M.: Thalamic pathology and memory loss in early alzheimers disease: moving the focus from the medial temporal lobe to papez circuit. Brain 139(7), 1877–1890 (2016)
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage 38(1), 95–113 (2007)
3. Ashburner, J., Friston, K.J.: Why voxel-based morphometry should be used. Neuroimage 14(6), 1238–1243 (2001)
4. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological) pp. 289–300 (1995)
5. Bießmann, F., Dähne, S., Meinecke, F.C., Blankertz, B., Görgen, K., Müller, K.R., Haufe, S.: On the interpretability of linear multivariate neuroimaging analyses: filters, patterns and their relationship. Citeseer
6. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning 3(1), 1–122 (2011)
7. Efron, B., Hastie, T.: Computer age statistical inference: Algorithms. Evidence and Data Science, Institute of Mathematical Statistics Monographs (2016)
8. Galton, C.J., Patterson, K., Graham, K., Lambon-Ralph, M., Williams, G., Antoun, N., Sahakian, B., Hodges, J.: Differing patterns of temporal atrophy in alzheimers disease and semantic dementia. Neurology 57(2), 216–225 (2001)
9. Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F.: On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110 (2014)
10. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(1-4), 259–268 (1992)
11. Shen, L., Kim, S., Qi, Y., Inlow, M., Swaminathan, S., Nho, K., Wan, J., Risacher, S.L., Shaw, L.M., Trojanowski, J.Q., et al.: Identifying neuroimaging and proteomic biomarkers for mci and ad via the elastic net. In: International Workshop on Multimodal Brain Image Analysis. pp. 27–34. Springer (2011)
12. Sun, X., Hu, L., Yao, Y., Wang, Y.: Gsplit lbi: Taming the procedural bias in neuroimaging for disease prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 107–115. Springer (2017)
13. Tansey, W., Koyejo, O., Poldrack, R.A., Scott, J.G.: False discovery rate smoothing. Journal of the American Statistical Association (just-accepted) (2017)
14. Tu, Y.K., Kellett, M., Clerehugh, V., Gilthorpe, M.S.: Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. British dental journal 199(7), 457 (2005)
15. Xin, B., Hu, L., Wang, Y., Gao, W.: Stable feature selection from brain smri. AAAI pp. 1910–1916 (2014)
16. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301–320 (2005)