NeuroImage: Clinical 19 (2018) 476–486

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Classification of multi-site MR images in the presence of heterogeneity using multi-task learning☆

Qiongmin Ma[a,b,c,*], Tianhao Zhang[b], Marcus V. Zanetti[d], Hui Shen[a], Theodore D. Satterthwaite[e], Daniel H. Wolf[e], Raquel E. Gur[e], Yong Fan[b], Dewen Hu[a,1], Geraldo F. Busatto[d,1], Christos Davatzikos[b,1]

[a] College of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan 410073, China
[b] Center for Biomedical Image Computing and Analytics, and Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, United States
[c] Beijing Institute of System Engineering, China
[d] Laboratory of Psychiatric Neuroimaging (LIM-21), Department and Institute of Psychiatry, Faculty of Medicine, University of São Paulo, São Paulo, Brazil
[e] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, United States

ARTICLE INFO

Keywords:
Multi-task learning
Multi-site classification
Imaging heterogeneity
Sparsity
Schizophrenia
MRI

ABSTRACT

With the advent of Big Data Imaging Analytics applied to neuroimaging, datasets from multiple sites need to be pooled into larger samples. However, heterogeneity across different scanners, protocols and populations, renders the task of finding underlying disease signatures challenging. The current work investigates the value of multi-task learning in finding disease signatures that generalize across studies and populations. Herein, we present a multi-task learning type of formulation, in which different tasks are from different studies and populations being pooled together. We test this approach in an MRI study of the neuroanatomy of schizophrenia (SCZ) by pooling data from 3 different sites and populations: Philadelphia, Sao Paulo and Tianjin (50 controls and 50 patients from each site), which posed integration challenges due to variability in disease chronicity, treatment exposure, and data collection. Some existing methods are also tested for comparison purposes. Experiments show that classification accuracy of multi-site data outperformed that of single-site data and pooled data using multi-task feature learning, and also outperformed other comparison methods. Several anatomical regions were identified to be common discriminant features across sites. These included prefrontal, superior temporal, insular, anterior cingulate cortex, temporo-limbic and striatal regions consistently implicated in the pathophysiology of schizophrenia, as well as the cerebellum, precuneus, and fusiform, middle temporal, inferior parietal, postcentral, angular, lingual and middle occipital gyri. These results indicate that the proposed multi-task learning method is robust in finding consistent and reliable structural brain abnormalities associated with SCZ across different sites, in the presence of multiple sources of heterogeneity.

1. Introduction

Neuroimaging studies have widely explored the clinical value of machine learning methods for differentiating psychiatric patients from healthy controls at the individual level. In addition to providing individualized indices for diagnostic purposes, machine learning methods may ultimately help identify brain regions affected by disease in subtle ways that can only be elucidated using multi-variate analysis. While most of these neuroimaging studies to date have been performed using single-site datasets, it is essential to integrate multi-site data for two reasons. First, multi-site data provide sufficient statistical power for detecting subtle, but informative patterns of brain structure and function (Brown et al., 2011; Friedman et al., 2006; Schnack et al., 2010), which may be difficult to unravel with the relative small sample sizes usually acquired in single centers (Pearlson, 2009; Segall et al., 2009). Second, large sample sizes enhance sample generalizability by pooling large patient populations with diverse demographic features and clinical characteristics including disease onset, symptom severity, and types and duration of treatment (Brown et al., 2011; Friedman et al., 2006; Glover et al., 2012; Pearlson, 2009; Sutton et al., 2008; Van Horn and Toga, 2009). Multi-site studies are therefore becoming increasingly the norm in neuroimaging research (Casey et al., 1998; Van Horn and Toga,

☆ There are no conflicts of interest.

* Corresponding author at: College of Mechatronics and Automation, National University of Defense Technology, 109 Deya Road, Changsha, Hunan 410073, China.
E-mail address: qiongmin.ma@nudt.edu.cn (Q. Ma).

[1] Authors contributed equally.

https://doi.org/10.1016/j.nicl.2018.04.037
Received 22 December 2017; Received in revised form 9 April 2018; Accepted 28 April 2018
Available online 09 May 2018

2009).

Multi-site data reflect a more comprehensive abnormal pattern of disease, and therefore may provide a richer understanding of disease signatures than single-site data. However, two recent studies found that simple pooling of multi-site data did not outperform single-site disease classification. Colby (Colby et al., 2012) and Nielsen (Nielsen et al., 2013) pooled multi-site data and trained a common classifier for all data, to identify attention deficit hyperactivity disorder (ADHD) and autism, respectively. Results in these two studies showed that the pooled dataset exhibited lower accuracy than each single-site datasets. Here we seek a new approach to synergistically integrating multi-site data, by emphasizing two points. First, since datasets collected in multiple imaging centers have a common disorder of interest (e.g. schizophrenia (SCZ), in our experiments herein), the abnormal patterns in each dataset are strongly related and thus, to some extent, may share a common imaging signature. The pattern reproducibility among multi-site data has been repeatedly demonstrated in several multi-site studies on functional MRI (fMRI) (Casey et al., 1998; Costafreda et al., 2007; Gee et al., 2015; Jovicich et al., 2015), morphometric MRI (Cannon et al., 2014; Schnack et al., 2010), and diffusion-tensor imaging (DTI) (Jovicich et al., 2014; Pfefferbaum et al., 2003). Most of these studies arrived at similar conclusions, namely that with appropriate multi-site data collection, different data sites shared highly consistent feature patterns. With this, the site-shared features reveal consistent brain abnormalities in multi-site data, which can lead to a more accurate neurobiological understanding of the psychiatric disorder under investigation. On the other hand, though the integration of multiple single-site data is advantageous, the unavoidable presence of site-specific features might decrease the accuracy of a classifier that merely pools data together across studies. Heterogeneity can emanate from multiple sources including scanner differences, differences in image acquisition protocols, or ethnic and treatment differences among participating patient populations (Jovicich et al., 2015; Schnack et al., 2010; Van Horn and Toga, 2009). Given such site-related heterogeneity, multi-site datasets should not simply be merged into larger cohorts for further machine learning investigation (Pan and Yang, 2010). In an attempt to eliminate or reduce the site-specific variability, studies have suggested same scanning protocols, consistent scanner parameters and etc. (Brown et al., 2011; Calhoun and Adali, 2009; Pearlson, 2009; You et al., 2011) in data collection, as well as the utilization of smoothness equalization (Friedman et al., 2006) and independent component analysis (Kim et al., 2009; Meda et al., 2008) in data preprocessing. Despite these efforts, site-specific heterogeneities still exist due to their complex causes (Pearlson, 2009; Segall et al., 2009).

The above considerations highlight the need for a feature-learning framework in multi-site disease classification that can extract the site-shared features, while also accounting for the site-specific features; this approach generally seeks an overarching signature of disease, whereas it accommodates potential sub-cohort and other differences to be taken into consideration. In recent years, multi-task learning has been successful in learning task-shared and task-specific features simultaneously, which effectively improves generalization compared with traditional machine learning methods. For example, support vector machine (SVM) with single-task learning (Caruana, 1997) learns a distinct feature pattern and finds a maximum margin hyperplane to classify two groups, which extracts information within a single learning task. In contrast, multi-task learning extracts a subset of task-shared features to generate more accurate models on multiple tasks, with the task-specific features learned simultaneously (Marquand et al., 2014). The basic assumption of multi-task learning is that the feature weights of different tasks share similar sparse patterns (Chen et al., 2012), which can be learned by imposing sparsity regularization penalties on the task weight matrix (Kumar and Daume III, 2012). l_1-norm and $l_{2,1}$-norm are two commonly used sparsity regulating terms in multi-task learning, which enforces the weight matrix of different tasks to be sparse across all tasks. Particularly, an l_1-norm term highlights task-specific features by encouraging the weights of irrelevant features to be very small (Wang et al., 2015), while $l_{2,1}$-norm introduces group sparsity and enforces task-shared features to have larger weights (Watanabe et al., 2014; Yan et al., 2015).

The advantage of multi-task learning makes it suitable for multi-site data learning, considering the site as task, and the site-shared and site-specific features as task-shared and task-specific features. Neuroimaging studies have shown the effectiveness of performing multi-task learning in the brain decoding and disease classification (Marquand et al., 2014; Obozinski et al., 2010; Rao et al., 2013; Wang et al., 2015; Watanabe et al., 2014). Specifically, multi-site fMRI data of ADHD was demonstrated better than single-site classification by learning site-shared and site-specific features using multi-task scheme (Watanabe et al., 2014). In this work, though the multi-task learning scheme successfully extracted site-shared and site-specific features in multi-site data, the form of the objective function was rather complex and specific as it included *an* $l_{2,1}$-norm group sparsity regularization term and a 6-D spatial structure penalty (generated by the GraphNet, fused Lasso, or the isotropic total variation). In order to make the multi-task learning scheme more simple and applicable, an objective function including l_1-norm and $l_{2,1}$-norm penalty terms was used in the current study (Wang et al., 2011; Wang et al., 2015).

Building upon this emerging literature, we aim to distinguish SCZ patients from healthy controls across multiple-site MRI data using multi-task learning. We hypothesized that using multi-task learning framework on multi-site classification would not only have better performance than single-site data classification, but would also identify the abnormalities shared by all sites, and also specific to each site. These site-shared brain structural alterations should be consistent with the previously reported altered regions in SCZ, such as the brain regions involving prefrontal, superior temporal, insular, temporo-limbic regions, among others.

2. Materials and methods

2.1. Participants and MRI acquisitions

MRI data were collected by three academic centers, including locations in the United States (University of Pennsylvania; site A) (Davatzikos et al., 2005), Brazil (University of Sao Paulo; site B) (Schaufelberger et al., 2007; Zanetti et al., 2013), and China (Tianjin; site C). From each site, a balanced dataset was obtained with 50 normal controls (NCs) and 50 SCZ patients randomly selected from a larger pool of available subjects. In total, we had 150 NCs and 150 SCZs, which didn't differ by age and gender significantly ($p > 0.05$; see Table 1).

All SCZ patients met DSM-IV criteria. Written informed consent was obtained from all participants before MRI scanning. In site A, the 50 SCZ patients had chronic symptoms and were receiving treatment with antipsychotics (mean duration of illness 16.2 ± 12.3 years). In site B, all SCZ subjects were recruited shortly after they made their first contact with mental health services due to psychotic symptoms, and their duration of illness was 1.0 ± 1.3 years; 31 patients had been on antipsychotic treatment within 3 weeks of MRI, while the remaining 19 patients were free of antipsychotics at the time of MRI scanning. Site C contributed 5 first-episode, never-treated SCZ patients and 45 chronic SCZ patients under antipsychotic treatment (mean duration of illness 10.5 ± 7.2 years).

In site A, the imaging data were acquired using a Siemens Trio 3-T scanner (Siemens Medical Systems, Erlangen Germany), with the following protocol: slice thickness = 1 mm, TE = 3.51 ms, TR = 18.1 ms, flip angle = 9°, acquisition matrix = 240 × 180, and slice number = 160, no gaps, 1-mm isotropic voxels. In site B, the T1 images were acquired using two identical 1.5-T GE Signa scanners (GE Medical Systems, Milwaukee WI, USA) with the following protocol: T1-SPGR