NeuroImage 155 (2017) 10–24

Contents lists available at ScienceDirect

NeuroImage

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

Multi-center machine learning in imaging psychiatry: A meta-model approach

Petr Dluhoš[a,b,*], Daniel Schwarz[c], Wiepke Cahn[d], Neeltje van Haren[d], René Kahn[d], Filip Španiel[e], Jiří Horáček[e], Tomáš Kašpárek[a,b], Hugo Schnack[d]

[a] Behavioural and Social Neuroscience Group, CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic
[b] Department of Psychiatry, University Hospital Brno and Masaryk University, Brno, Czech Republic
[c] Institute of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic
[d] Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands
[e] National Institute of Mental Health, Klecany, Czech Republic

ARTICLE INFO

Keywords:
Machine learning
multi-center
meta-model
combining models
support vector machines (SVM)
first-episode schizophrenia
classification
prediction

ABSTRACT

One of the biggest problems in automated diagnosis of psychiatric disorders from medical images is the lack of sufficiently large samples for training. Sample size is especially important in the case of highly heterogeneous disorders such as schizophrenia, where machine learning models built on relatively low numbers of subjects may suffer from poor generalizability. Via multicenter studies and consortium initiatives researchers have tried to solve this problem by combining data sets from multiple sites. The necessary sharing of (raw) data is, however, often hindered by legal and ethical issues. Moreover, in the case of very large samples, the computational complexity might become too large. The solution to this problem could be distributed learning. In this paper we investigated the possibility to create a meta-model by combining support vector machines (SVM) classifiers trained on the local datasets, without the need for sharing medical images or any other personal data. Validation was done in a 4-center setup comprising of 480 first-episode schizophrenia patients and healthy controls in total. We built SVM models to separate patients from controls based on three different kinds of imaging features derived from structural MRI scans, and compared models built on the joint multicenter data to the meta-models. The results showed that the combined meta-model had high similarity to the model built on all data pooled together and comparable classification performance on all three imaging features. Both similarity and performance was superior to that of the local models. We conclude that combining models is thus a viable alternative that facilitates data sharing and creating bigger and more informative models.

Introduction

Schizophrenia (SZ) is a debilitating psychiatric disorder affecting almost one percent of the world population (McGrath et al., 2008). The manifestations range from misinterpretation of reality and delusions to disorganization of thinking and behavior and it has severe negative effects on the society on both social and economic levels (Knapp et al., 2004). Huge efforts to identify brain changes related to schizophrenia led to a large volume of magnetic resonance imaging (MRI) studies which reported a number of heterogeneous findings (Fusar-Poli et al., 2013; Haijma et al., 2013; Shepherd et al., 2012; Steen et al., 2006; van Erp et al., 2015). However, there is not yet clear consensus on what mechanism causes these changes (Insel, 2010) and the impact on schizophrenia diagnosis or treatment is minimal (Lawrie et al., 2011).

The contemporary diagnostic process is based on clinical interviews (International Statistical Classification of Disease and Related Health Problems ("ICD-10," 2016)), Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition ("DSM-5," 2016)) and lacks any objective tests or diagnostic tools (Lawrie et al., 2011). Such a tool in the form of, for example, a classification algorithm based on medical images, would provide multiple benefits: more objective and precise diagnosis, early detection of disease facilitating better prognosis (Riecher-Rössler et al., 2006), easier communication with the patient about their illness (better insight) and also deeper understanding of the disease itself.

Many promising results on automated schizophrenia classification from medical images have been published, with accuracies ranging from 71% to 98% (Ardekani et al., 2011; Demirci et al., 2008a; Ford et al., 2002; Liu et al., 2004; Shi et al., 2007; Wang and Verma, 2008). For an overview see (Kambeitz et al., 2015; Nieuwenhuis et al., 2012).

* Correspondence to: CEITEC, A35, Kamenice 5, 625 00 Brno, Czech Republic.
E-mail address: dluhos@mail.muni.cz (P. Dluhoš).

http://dx.doi.org/10.1016/j.neuroimage.2017.03.027
Received 22 December 2016; Received in revised form 6 March 2017; Accepted 14 March 2017
Available online 17 April 2017

However, these studies have two principal limitations in common. First, most of them were performed in chronic patients in which case long-term disease progress and medication could have altered the brain morphology (Puri, 2011; Vita et al., 2012) and thus have influenced the results. Studies performing automated classification on drug-naive, first-episode (FES) or at-risk schizophrenia patients achieved on average significantly lower classification accuracy (58–83%) (Zarogianni et al., 2013). Early stages of the disease are the stages which the automated diagnosis efforts should aim at, in order to utilize the developed methods and algorithms in future clinical practice. The second weakness is the potential lack of generalization of the prediction models, arising from relatively small samples and lack of testing the models on independent validation sets, which can both lead to a serious overestimation of classifier performance (Nieuwenhuis et al., 2012). Overestimated accuracy values and low generalizability to unseen data can also be a consequence of biases and errors made during the complex classification process (Demirci et al., 2008b). These errors can be minimized by proper utilization of cross-validation procedures while significance of the whole classification can be verified by permutation testing (Golland and Fischl, 2003). However, the cross-validation and permutation techniques do not eliminate the need for a sufficient number of subjects to build a robust model, and additional subjects to be used in an independent set for validation.

The problem of the low sample size is even intensified in case of the FES patients: Larger datasets are needed in order to detect the subtler brain abnormalities, but it is more difficult to acquire patients in the first stage of the disease. A multicenter study is usually the natural solution of this problem. However, analysis of data acquired at multiple study sites brings three additional problems:

1) **Technical issues** – dealing with different scanners, imaging sequences and preprocessing at each site (Kostro et al., 2014; Schnack et al., 2010; van Erp et al., 2015)
2) **Legal or ethical issues** – sharing sensitive information about patients and controls, such as the images themselves (Rubinstein et al., 2009; Sarwate et al., 2014)
3) **Computational complexity** – increasing with large number of subjects, especially in cases when particular classifiers exhibit higher-order time or memory complexity (Chu et al., 2007).

In the case of group-level analyses, a clever way how to overcome or at least mitigate these problems was adopted by consortia such as ENIGMA (Thompson et al., 2014). Instead of pooling all images together and performing one analysis on them, each site analyzes their own image data and shares only their (small-sized) results. All local results are then combined by a meta-analysis at the level of statistical representations. This solves the problem number 2, because instead of individual images, only (local) group-level data are shared. It also solves the problem number 3, because it is easier to compute M multiple smaller analyses and then combine them than to perform one large analysis on all data. Suppose the time complexity of analyzing N subjects is N^k. Then dividing the problem into M independent subproblems and combining the resulting models (this combination, for example some sort of averaging, can be typically performed in linear time) would require $M(\frac{N}{M})^k = \frac{N^k}{M^{k-1}}$, which is M^{k-1} less than for solving the problem as a whole. The technical issues (problem number 1) can be mitigated partly by establishing a common standard which would be used by each site, thus reducing the unwanted variability originating from different preprocessing methods.

Inspired by the ENIGMA concept which requires no sharing of personal imaging data in a multicenter study, we investigated here the possibility of building "local" Support Vector Machines (SVM) models for classification and combine them to a larger meta-model. For all models we used three types of imaging features extracted from structural MR images of FES patients and healthy controls acquired at four different sites. A cross-validation methodology was used for evaluation of our meta-model approach: For every combination of three out of four sites, we compared the meta-models to a joint model built on data from all subjects from the three sites pooled together. The performance of the classification models was evaluated using various techniques, including classification of the fourth independent dataset not used for training of any of the models.

Material and methods

Datasets

In this study we used four datasets totaling 258 first-episode schizophrenia patients and 222 healthy controls. Item-by-item, each dataset was set aside as an independent validation set and classification models for discriminating between patients and controls were built on the remaining three datasets.

Dataset 1

Fifty-two patients (age 24 ± 5.1 years) were recruited from those admitted to the all-male unit of the Department of Psychiatry, University Hospital Brno (UHB), Czech Republic for first episode of schizophrenia. Their symptoms fulfilled the criteria for schizophrenia for the first time when admitted to the department, including the time criterion – duration of symptoms longer than 1 month. The patients were matched for age and handedness by 52 healthy males (age 24 ± 3.7 years) recruited from the community, the local staff and medical students. Diagnosis was established during clinical interviews held in compliance with the International Statistical Classification of Disease and Related Health Problems (ICD-10) research criteria. The patients were briefly rehospitalized one year after the first episode to confirm the diagnosis. All patients were taking atypical antipsychotics for 1–4 weeks by the time of MRI scanning (79% risperidone or olanzapine, 21% other). More detailed information about this dataset can be found in Kasparek et al. (2011).

Dataset contains T1-weighted images of the entire head scanned by 1.5 T Siemens Symphony MR device, IR/GR sequence, repetition time (TR)=1700 ms, echo time (TE)=3.93 ms, inversion time (TI)=1100 ms, flip angle (FA)=15°, sagittal tomographic plane thickness 1.17 mm, the in-plane resolution 0.48 mm×0.48 mm, 3-D field of view (FOW) =160×512×512 voxels.

Dataset 2

A sample of ninety-five first-episode patients (age 24.1 ± 5.4 years) and fifty-three healthy controls (age 25.0 ± 6.0 years) were selected from a study carried out at the University Medical Center Utrecht (UMC), Utrecht, the Netherlands. Eligible patients and healthy controls in this study had to fulfill the following criteria: age between 16 and 50 years, be fluent in Dutch, and be able and willing to give written informed consent. Patients met the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria for a nonaffective psychotic disorder (including schizophrenia, schizophreniform disorder, and schizoaffective disorder), subjects with substance dependence/abuse and a major medical or neurological illness were excluded. By the time of MRI scanning, 24 patients were using typical medication, 13 atypical medication, 3 both, 15 was not using any antipsychotics and for 40 patients is the information about medication not available. Details of the recruitment criteria and diagnosis are described in Cahn et al. (2002) and Hulshoff Pol et al. (2001).

T1-weighted structural MRI scans were obtained on a 1.5-T Achieva scanner (Philips, Best, the Netherlands), coronal spoiled-gradient echo scan of the whole head, 256×256 matrix, TE=4.6 ms, TR=30 ms, FA=30°, 160–180 contiguous slices, 1×1×1.2 mm^3 voxels, FOV=256 mm/70%.