

# Learning from Enhanced Contextual Similarity in Brain Imaging Data for Classification of Schizophrenia

Tewodros Mulugeta Dagne<sup>1</sup>(✉), Letizia Squarcina<sup>2</sup>, Massimo W. Rivolta<sup>1</sup>, Paolo Brambilla<sup>3</sup>, and Roberto Sassi<sup>1</sup>

<sup>1</sup> Computer Science Department, Università degli Studi di Milano, Crema, Italy  
{tewodros.dagne, massimo.rivolta, roberto.sassi}@unimi.it

<sup>2</sup> Scientific Institute IRCCS “E. Medea”, Bosisio Parini, Italy  
letizia.squarcina@gmail.com

<sup>3</sup> Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy  
paolo.brambilla1@unimi.it

**Abstract.** In certain severe mental diseases, like schizophrenia, structural alterations of the brain are detectable by magnetic resonance imaging (MRI). In this work, we try to automatically distinguish, by using anatomical features obtained from MRI images, schizophrenia patients from healthy controls. We do so by exploiting contextual similarity of imaging data, enhanced with a distance metric learning strategy (DML - by providing “must-be-in-the-same-class” and “must-not-be-in-the-same-class” pairs of subjects). To learn from contextual similarity of the subjects brain anatomy, we use a graph-based semi-supervised label propagation algorithm (graph transduction, GT) and compare it to standard supervised techniques (SVM and K-nearest neighbor, KNN). We performed out tests on a population of 20 schizophrenia patients and 20 healthy controls. DML+GT achieved a statistically significant advantage in classification performance (Accuracy: 0.74, Sensitivity: 0.79, Specificity: 0.69, Ck: 0.48). Enhanced contextual similarity improved performance of GT, SVM and KNN offering promising perspectives for MRI images analysis.

## 1 Introduction

Schizophrenia (SCZ) is a severe, chronic and debilitating mental illness affecting around 0.4% of the population [1]. Magnetic resonance imaging (MRI) studies consistently observed alterations in cortical and subcortical brain areas, especially frontal [21] and temporal [14] regions. The capability of detecting these pathological alterations in brain images would be of high relevance in accelerating the diagnostic process, with clear benefits for both patients and psychiatrists. Given the complexity and multidimensionality of the problem, machine learning (ML) analysis of magnetic resonance (MR) images is recently becoming popular in the understanding of this domain.

ML algorithms have been used in SCZ studies [18] with the aim of detecting sets of features which could be discriminative in the diagnosis. In the literature, the majority of ML applications to psychiatric data are purely supervised methods that learn only from labeled data, with promising and interesting results [11, 13, 16]. However, while these findings have been received with great optimism within the neuropsychiatric community, a major criticism has been that these algorithms are ordinarily “trained” to categorize patients based on a symptom-based diagnosis. As such, there are inevitable uncertainty in the “gold standard”: learning from the unlabeled data seems a possibility to mitigate the problem. In these situations, classification performances might improve when the learning process incorporates unlabeled data. Moreover, semi-supervised and unsupervised schemes could provide a better phenotype identification and classification of diseases [20].

In this paper, we propose to exploit learning from both labeled and unlabeled MR images. The addition of learning from unlabeled data will decrease the risk of circular analysis, by exploiting similarities between data without prior information on the class. To do so, we applied graph transduction (GT), *i.e.*, a data-driven graph-based semi-supervised label propagation algorithm [4], which can learn from the contextual similarity (CS) of the imaging data. However, the problem with label propagation methods is that their performance heavily depends on the pre-existing CS of the input data. To deal with this problem, we applied a *distance metric learning* (DML) strategy, to enhance CS information of features obtained from MRI images, by providing “must-be-in-the-same-class” and “must-not-be-in-the-same-class” pairs of subjects (*i.e.*, healthy controls and SCZ patients), thus increasing the intra-cluster similarity and decreasing the inter-cluster similarity. The formalization of GT is inspired from game theoretic notions [4], where the final labeling corresponds to the Nash equilibrium of a non cooperative game. The players of the game correspond to data features (or nodes of the graph) and the class labels correspond to available strategies. In our case, we map the problem of classifying MRI images, where the brain imaging data of each subject correspond to a player who can choose a strategy to maximize its pay-off (the pair-wise similarity of the image features between subjects).

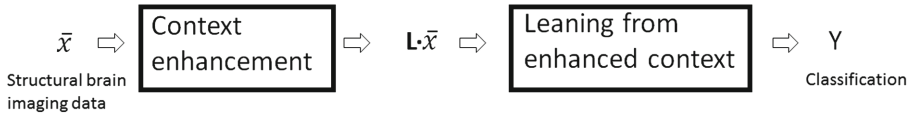
Authors of [3] showed a similar concept of what we present here, to solve a problem in object recognition and scene classification (a general computer vision problem), confirming the importance of enhancing CS to improve the performance of a label propagation algorithm. In our study, we implemented one of the latest and most robust metric learning [19] and label propagation algorithm [4], to be then applied to MRI data.

To the best of our knowledge, this is the first study to address classification of SCZ patients and healthy subjects applying a metric learning and graph-based semi-supervised learning strategy to structural MRI data.

## 2 Learning from Enhanced Contextual Similarity

In this section, we present a scheme of classification that exploits contextual brain anatomical similarities of subjects from MR images, so as to differentiate

healthy controls from SCZ patients. A set of features, characterizing the anatomy of the brain, was obtained from the MR images of every single subject. Then, we used a DML technique, specifically the one proposed in [19], to enhance the CS of the input MRI data and apply the GT algorithm [4] on top of this new metric space to learn from the newly enhanced context. The overall scheme is depicted in Fig. 1 and described step-by-step in the next sections.



**Fig. 1.** The proposed schizophrenia classification scheme using structural brain imaging data.

### 2.1 CS Enhancement Using DML

DML represents a useful technique widely exploited in pattern recognition, which aims to find a metric that maximizes the distance between features belonging to different classes (and viceversa, minimizes the distance between features belonging to the same class). With this aim, linear and non-linear metrics had been investigated. On one hand, linear metrics can be computationally less expensive, but often provide lower performances. On the other hand, non-linear algorithm might perform better but they are computationally expensive and application-dependent.

In the linear domain, DML remaps features using a linear combination carried out by the transformation matrix  $L$ , as follows:

$$\bar{x}' = L\bar{x},$$

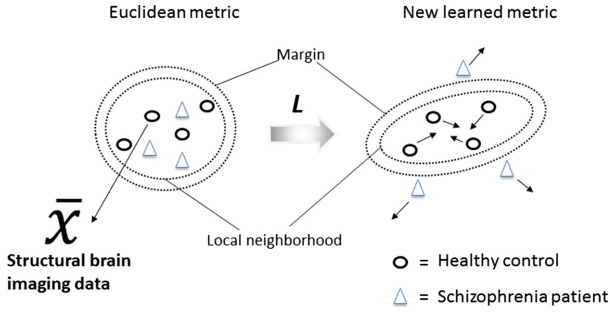
where  $\bar{x}$  is the input feature vector and  $\bar{x}'$  is the transformed feature vector. If the matrix  $L$  is full rank, it is possible to show that the Euclidean distance between two elements in the transformed space,

$$D(\bar{x}_i, \bar{x}_j) = \|L(\bar{x}_i - \bar{x}_j)\|_2,$$

represents a valid metric. Furthermore, the Euclidean distance can be rewritten using a matrix notation which becomes the so-called *Mahalanobis* distance. Such distance is defined as

$$D_M(x_i, x_j) = \sqrt{(\bar{x}_i - \bar{x}_j)^\top M (\bar{x}_i - \bar{x}_j)}$$

being  $M = L'L$  the Mahalanobis positive semidefinite matrix. The effect of such transformation is shown in Fig. 2. When  $L$  is the identity matrix, the Mahalanobis distance becomes the standard Euclidean distance.



**Fig. 2.** Illustration of feature context enhancement by means of large margin nearest neighbor (LMNN) distance metric learning. Before training (left) and after training (right).

In this study, we used a linear DML to modify the pre-existing neighbouring structure of MRI data before feeding it to GT, aiming to achieve classification improvements. In order to determine the transformation matrix  $L$ , we used the *Large Margin Nearest Neighbor* DML method described in [19]. The algorithm makes use of the following equations

$$pullpush(L) = (1 - \mu) pull(L) + \mu push(L) \tag{1}$$

with

$$\begin{aligned}
 pull(L) &= \sum_{i,j \rightarrow i} \|L(\bar{x}_i - \bar{x}_j)\|^2 \\
 push(L) &= \sum_{i,j \rightarrow i} \sum_k (1 - \delta_{ik}) [1 + \|L(\bar{x}_i - \bar{x}_j)\|^2 \\
 &\quad - \|L(\bar{x}_i - \bar{x}_k)\|^2]_+
 \end{aligned} \tag{2}$$

where  $y_i$  is the class to which  $\bar{x}_i$  belongs and  $\delta_{ik} = 1$  if  $y_i = y_k$  or  $\delta_{ik} = 0$  otherwise.  $[f]_+$  implies a hinge-loss such that  $[f]_+ = \max(0, f)$ . The term  $j \rightarrow i$  in Eq. (2) implies that  $j$  belongs to the same class where  $i$  belongs too. Finally, the parameter  $\mu$  sets the trade-off between the pulling and pushing objectives and was set to 0.5 as suggested in [19].

The process of getting the transformation metric  $L$  involves minimizing the overall objective function in Eq. (1). The first term pulls subjects with the same class label closer in terms of the Mahalanobis distance. The second term pushes away differently-labeled instances by a large margin, so that they are located further apart in the transformed space (Fig. 2).

As stated in [19], it is worth noting that Eq. (1) does not define a convex optimization problem in terms of  $L$ . However, it can be rephrased in a convex fashion using a semi-definitive programming approach by determining  $M$  instead of  $L$ . Then,  $L$  can be computed using matrix factorization of  $M$ .

## 2.2 Learning from Enhanced CS of MR Images Using GT

The aim of GT is to address the problem of consistent labeling, with the aim of predicting or propagating class membership to unlabeled data by exploiting learning both from the labeled and unlabeled samples. Such methodology involves three different areas: (i) graph theory; (ii) evolutionary game theory; and (iii) dynamical systems and optimization.

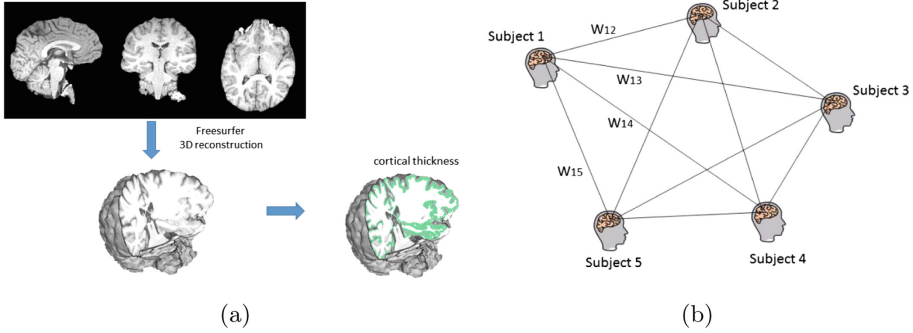
The main idea behind GT is to consider the samples of the dataset as nodes of a graph, and to propagate class labels to unlabeled nodes, by considering the CS among the samples. In particular, it exploits CS among data features to perform label propagation in a consistent way, relying on a common *a priori* assumption known as the “cluster assumption” (a reminiscent of the homophily principle used in social network analysis): nodes that are close to each other, in the same cluster or on the same manifold are expected to have the same label. Each node is then a feature vector  $\in \mathbb{R}^d$  (with  $d$  being the number of features). Moreover, each node can select a strategy, *i.e.*, class membership, that maximizes its CS. Finally, the output labeling corresponds to the Nash equilibrium of the game.

Input features are represented with graph nodes  $\mathcal{G} = (\mathcal{V})$ , where the vertex set  $\mathcal{V}$  is composed of  $n = l + u$  elements  $\in \mathbb{R}^d$  and consists of a first labeled set  $\{(x_1, y_1), \dots, (x_l, y_l)\}$  of  $l$  elements and a second unlabeled set  $\{(x_{l+1}, \dots, (x_{l+u})\}$  of  $u$  elements. Then, the similarity matrix  $\mathcal{E}$  between pairs of nodes is computed, after having selected a similarity metric. A simple and effective optimization algorithm to propagate the labels through the graph is given by the so-called replicator dynamics, developed and studied in evolutionary game theory, which has proven to be effective in many applications [7, 23].

In practice, as explained in Sect. 2.1, labeled examples in the form of “must-be-in-the-same-class” and “must-not-be-in-the-same-class” pairs of subjects are provided to the DML framework, to learn the best feature space transformation matrix  $L$  using Eq. (1). Afterwards, the class label propagation occurs on such transformed feature space (*i.e.*,  $L\bar{x}$ ) by constructing the fully connected graph  $\mathcal{G} = (\mathcal{V})$ , where  $\mathcal{V}$  is now the set of graph nodes representing the transformed feature vectors, and  $\mathcal{E}$  encodes the brain anatomy similarity between subjects by means of the edge weights (similarity matrix) as depicted in Fig. 3b.  $\mathcal{E}$  is constructed in the following manner (for simplicity we show how an edge is constructed between two transformed feature vectors):

$$\mathcal{E}_{ij} = \exp \left[ -\frac{d(L\bar{x}_i, L\bar{x}_j)^2}{2\sigma^2} \right] \quad (3)$$

where  $d(L\bar{x}_i, L\bar{x}_j)$  is the Euclidean distance. For estimating  $\sigma$ , which is a critical parameter of the graph’s ability in representing the CS between data points, we adopted an automatic self tuning method as proposed in [22].



**Fig. 3.** (a) ROI and cortical thickness feature extraction from brain images. (b) Representation of brain anatomy similarity between subjects.

### 3 Experiments

#### 3.1 Dataset and Representation

The dataset consisted in T1-weighted MR images of 20 healthy control subjects ( $35.8 \pm 13$ , 8 males) and 20 SCZ patients ( $37.9 \pm 11$ , 13 males). The size of this dataset is in line with the dimensionality of datasets used in academic works aimed at medical applications [9, 10, 17]: in particular, it is not straightforward to obtain consistent MRI data of psychiatric patients, due to difficulties in recruitment and feasibility of MRI acquisitions in this kind of patients. The data were collected at the Psychiatric department of Ospedale di Verona (Verona, Italy). All involved subjects signed an informed consent, following the principle of the Helsinki’s declaration.

The T1-weighted images were preprocessed using the software FreeSurfer<sup>1</sup> as depicted in Fig. 3a. Based on prior knowledge on schizophrenia [14, 21], we considered the average cortical grey matter thickness of frontal and temporal regions (namely: *caudal middle frontal*, *inferior temporal*, *middle temporal*, *rostral middle frontal* and *superior frontal* of the left hemisphere) as features in the classification task. The ROI thickness measurement of the subjects is reported in Table 1. Also, in order to take into account the effect of age on the cortical thickness, we corrected all the data for age differences using a generalized linear model [8].

#### 3.2 Experimental Analysis

We performed two series of comparisons to assess the performances of the proposed classification scheme in differentiating healthy controls from SCZ subjects. First, we verified whether learning from CS (from both labeled and unlabeled data) might provide better classification results than just learning from labeled

<sup>1</sup> <http://surfer.nmr.mgh.harvard.edu/>.

**Table 1.** Grey matter cortical thickness of ROIs (in mm) of healthy controls and schizophrenia patients.

ROI	HC (mean $\pm$ std)	SCZ (mean $\pm$ std)
Caudal middle frontal	2.56 $\pm$ 0.21	2.44 $\pm$ 0.18
Inferior temporal	2.85 $\pm$ 0.16	2.64 $\pm$ 0.17
Middle temporal	2.89 $\pm$ 0.17	2.73 $\pm$ 0.18
Rostral middle frontal	2.47 $\pm$ 0.19	2.34 $\pm$ 0.17
Superior frontal	2.82 $\pm$ 0.18	2.65 $\pm$ 0.20

data. Second, we tested if the enhancement of CS by DML might provide further improvements. To do so, we compared the proposed classification scheme (DML+GT) with both GT [4] and KNN, with and without metric learning (KNN [6], DML+KNN [19]), linear SVM and DML+SVM.

We evaluated the classification performances by using accuracy (Acc), sensitivity (Se), specificity (Sp) and Cohen’s kappa (Ck) coefficients. Sensitivity refers to the true recognition of SCZ patients.

We considered first 70%, then 80% of the data from each class for training and input labeling of GT, while the rest of data was left to be predicted. In fact, GT was found to perform sufficiently well when the labeled data were just a small fraction of the dataset [4]. However, given the small size of our dataset, we considered labeling 70% and 80% of the data at disposal. We repeated this procedure by randomly sampling the dataset 100 times and computed the average performance. In all the experiments we avoided the risk of circular analysis [5]. For KNN, we chose  $K = 3$ , for limiting the possible overfitting due to the relatively small sample size.

### 3.3 Experimental Results

The average and standard error of the classification performance for DML+GT (proposed scheme), DML+SVM, DML+KNN, GT, SVM and KNN (used for comparison), when 70% and 80% of the samples in each class are labeled are reported in Table 2 and Fig. 4.

As expected, performance got better when using higher percentage of labeled data on small datasets for DML+GT and GT. Moreover, in our proposed scheme, sensitivity was always lower than specificity (Fig. 4c and d), meaning that some subjects with schizophrenia were classified as healthy, regardless the labeled sample size. In addition, the increase of training data provided different relative improvements between sensitivity and specificity (Fig. 4c and d). This means that these methodologies, under the settings we considered, are capable of recognizing the healthy subjects more easily than the schizophrenia patients.

GT was more affected by the training set’s size (fourth bar in each plot of Fig. 4) than the other methods. However, when DML was applied before GT,

**Table 2.** Average test-set classification performance ( $\pm$  standard deviation across subjects) on brain sMRI data features using 70% and 80% of the data for training.

Methods		70%	80%
DML+GT	Acc	$0.70 \pm 0.01$	$0.74 \pm 0.01$
	Se	$0.66 \pm 0.02$	$0.69 \pm 0.02$
	Sp	$0.73 \pm 0.02$	$0.79 \pm 0.02$
	Ck	$0.39 \pm 0.02$	$0.48 \pm 0.03$
DML+SVM	Acc	$0.71 \pm 0.01$	$0.71 \pm 0.02$
	Se	$0.70 \pm 0.02$	$0.74 \pm 0.02$
	Sp	$0.73 \pm 0.02$	$0.68 \pm 0.02$
	Ck	$0.42 \pm 0.02$	$0.42 \pm 0.03$
DML+KNN	Acc	$0.68 \pm 0.01$	$0.70 \pm 0.02$
	Se	$0.64 \pm 0.02$	$0.66 \pm 0.02$
	Sp	$0.72 \pm 0.02$	$0.75 \pm 0.02$
	Ck	$0.36 \pm 0.02$	$0.41 \pm 0.03$
GT	Acc	$0.61 \pm 0.01$	$0.67 \pm 0.01$
	Se	$0.57 \pm 0.04$	$0.64 \pm 0.03$
	Sp	$0.65 \pm 0.03$	$0.71 \pm 0.03$
	Ck	$0.22 \pm 0.02$	$0.35 \pm 0.03$
SVM	Acc	$0.69 \pm 0.01$	$0.69 \pm 0.01$
	Se	$0.70 \pm 0.02$	$0.73 \pm 0.02$
	Sp	$0.69 \pm 0.02$	$0.66 \pm 0.02$
	Ck	$0.38 \pm 0.02$	$0.38 \pm 0.03$
KNN	Acc	$0.65 \pm 0.01$	$0.65 \pm 0.01$
	Se	$0.64 \pm 0.02$	$0.62 \pm 0.02$
	Sp	$0.65 \pm 0.02$	$0.69 \pm 0.02$
	Ck	$0.29 \pm 0.03$	$0.30 \pm 0.03$

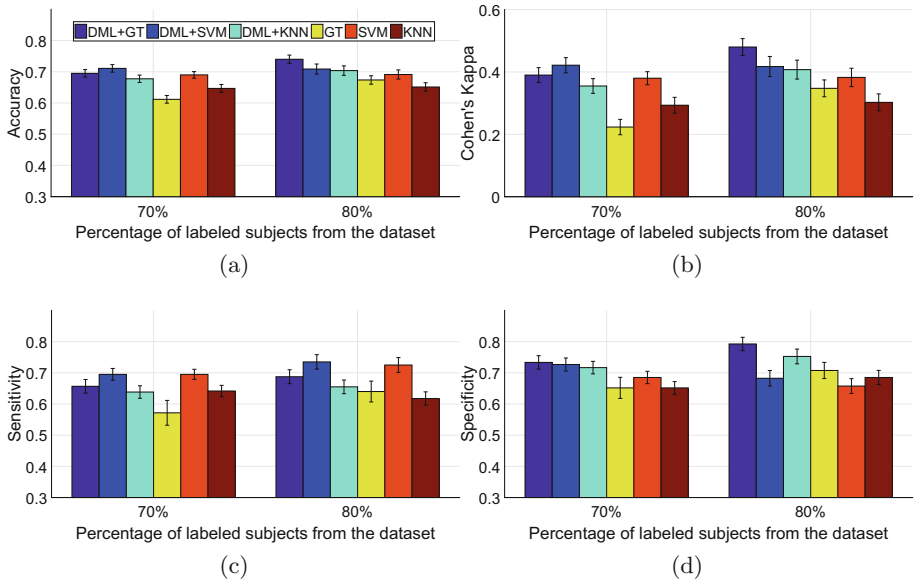
we obtained a drastic classification improvement of all measures except the sensitivity, even with a smaller training set. Furthermore, the use of DML resulted in a higher performance in all the cases except sensitivity (Fig. 4).

Finally, when 80% of the data is used as training, CS learning, *i.e.*, learning from unlabeled data as well, enhanced with DML outperformed both SVM and KNN with DML (first bar vs second and third bar).

### 3.4 Discussion of the Experimental Results

This work supports the finding that DML+KNN is better than KNN (*i.e.*, with respect to every evaluation metric considered), as found by other authors [19]. In particular, we showed that this finding holds true when applied to thickness features extracted from MRI data.





**Fig. 4.** Classification results for healthy controls vs. schizophrenia patients. Average performances and standard errors of the mean are reported.

Moreover, GT is consistently improved by the proposed scheme (DML+GT), which suggests that CS enhancement of MRI data coupled with learning from unlabeled samples, can result in a better performance of classifying schizophrenia. This result is also supported in [3], within the computer vision domain (object recognition and scene classification).

Finally, DML+GT performances are higher likely due to the additional information obtained from the unlabeled MRI data features. This confirms that DML and CS has the potential to improve schizophrenia classification.

The results obtained are comparable to the state-of-the-art in classification of schizophrenia. For example, in [12] using functional MRI (fMRI), they obtained an average classification accuracy of 0.59 and 0.84 using both static and dynamic resting-state functional network connectivity approach respectively and linear SVM. In [15] they obtained up to 0.75 accuracy (combining ROI thickness features) using 1.5 T sMRI and covariate multiple kernel learning approach using SVM. In [2], they achieved 0.75 accuracy considering the left hemisphere.

## 4 Conclusion

In this study, we designed a classification scheme to discriminate healthy controls from schizophrenia patients using MR images-derived data as features. We believe that learning from contextual anatomical similarity of subjects (*i.e.*, learning both from labeled and unlabeled MRI data features) has a great poten-

tial in dealing with schizophrenia, due to the nature and complexity of the disease and its associated diagnostic uncertainty.

Furthermore, we showed that enhancing the CS improved the classification performances of the label propagation algorithm (semi-supervised context learning). We demonstrated that the combination of metric learning and graph transduction (DML+GT) is useful to learn a meaningful underlying pattern from MRI data by exploiting contextual information, resulting in better classification performances.

In the future, we would like to test a non-linear metric for context enhancing to assess if it can further improve the classification results. Also, GT could be improved by using another anatomical feature (dis)similarity measurement instead of the symmetric Euclidean distance of Eq. (3) since it can handle asymmetric (dis)similarities also.

## References

1. Bhugra, D.: The global prevalence of schizophrenia. *PLoS Med.* **2**(5), e151 (2005)
2. Castellani, U., Rossato, E., Murino, V., Bellani, M., Rambaldelli, G., Perlini, C., Tomelleri, L., Tansella, M., Brambilla, P.: Classification of schizophrenia using feature-based morphometry. *J. Neural Transm. (Vienna)* **119**(3), 395–404 (2012)
3. Ebert, S., Fritz, M., Schiele, B.: Pick your neighborhood-improving labels and neighborhood structure for label propagation. In: *ICPR*, pp. 152–162 (2011)
4. Erdem, A., Pelillo, M.: Graph transduction as a noncooperative game. *Neural Comput.* **24**(3), 700–723 (2012)
5. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I.: Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**(5), 535–540 (2009)
6. Larose, D.T.: K-nearest neighbor algorithm. In: *Discovering Knowledge in Data: An Introduction to Data Mining*, pp. 90–106. Wiley (2005)
7. Zemene, E., Bulò, S.R., Pelillo, M.: Dominant-set clustering using multiple affinity matrices. In: Feragen, A., Pelillo, M., Loog, M. (eds.) *SIMBAD 2015. LNCS*, vol. 9370, pp. 186–198. Springer, Cham (2015). doi:[10.1007/978-3-319-24261-3\\_15](https://doi.org/10.1007/978-3-319-24261-3_15)
8. Montgomery, D.C., Peck, E.A., Vining, G.G.: *Introduction to Linear Regression Analysis*. Wiley, Hoboken (2015)
9. Mwangi, B., Spiker, D., Zunta-Soares, G.B., Soares, J.C.: Prediction of pediatric bipolar disorder using neuroanatomical signatures of the amygdala. *Bipolar Disord.* **16**(7), 713–721 (2014)
10. Mwangi, B., Wu, M.J., Bauer, I.E., Modi, H., Zeni, C.P., Zunta-Soares, G.B., Hasan, K.M., Soares, J.C.: Predictive classification of pediatric bipolar disorder using atlas-based diffusion weighted imaging and support vector machines. *Psychiatry Res.* **234**(2), 265–271 (2015)
11. Peruzzo, D., Castellani, U., Perlini, C., Bellani, M., Marinelli, V., Rambaldelli, G., Lasalvia, A., Tosato, S., De Santi, K., Murino, V., et al.: Classification of first-episode psychosis: a multi-modal multi-feature approach integrating structural and diffusion imaging. *J. Neural Transm.* **122**(6), 897–905 (2015)
12. Rashid, B., Arbabshirani, M.R., Damaraaju, E., Cetin, M.S., Miller, R., Pearson, G.D., Calhoun, V.D.: Classification of schizophrenia and bipolar patients using static and dynamic resting-state fmri brain connectivity. *Neuroimage* **134**, 645–657 (2016)

13. Schnack, H.G., Van Haren, N.E., Nieuwenhuis, M., Hulshoff Pol, H.E., Cahn, W., Kahn, R.S.: Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study. *Am. J. Psychiatry* **173**(6), 607–616 (2016)
14. Shenton, M.E., Kikinis, R., Jolesz, F.A., Pollak, S.D., LeMay, M., Wible, C.G., Hokama, H., Martin, J., Metcalf, D., Coleman, M., et al.: Abnormalities of the left temporal lobe and thought disorder in schizophrenia: a quantitative magnetic resonance imaging study. *N. Engl. J. Med.* **327**(9), 604–612 (1992)
15. Squarcina, L., Castellani, U., Bellani, M., Perlini, C., Lasalvia, A., Dusi, N., Bonetto, C., Cristofalo, D., Tosato, S., Rambaldelli, G., et al.: Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques. *Neuroimage* **145**(Part B), 238–245 (2015)
16. Squarcina, L., Perlini, C., Bellani, M., Lasalvia, A., Ruggeri, M., Brambilla, P., Castellani, U.: Learning with heterogeneous data for longitudinal studies. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 535–542. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4\\_64](https://doi.org/10.1007/978-3-319-24574-4_64)
17. Valli, I., Marquand, A.F., Mechelli, A., Raffin, M., Allen, P., Seal, M.L., McGuire, P.: Identifying individuals at high risk of psychosis: predictive utility of support vector machine using structural and functional MRI data. *Front. Psychiatry* **7**, 52 (2016)
18. Veronese, E., Castellani, U., Peruzzo, D., Bellani, M., Brambilla, P.: Machine learning approaches: from theory to application in schizophrenia. *Comput. Math. Methods Med.* **2013**, 867924 (2013)
19. Weinberger, K., Saul, L.: Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* **10**, 207–244 (2009)
20. Wu, M.J., Mwangi, B., Bauer, I.E., Passos, I.C., Sanches, M., Zunta-Soares, G.B., Meyer, T.D., Hasan, K.M., Soares, J.C.: Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. *Neuroimage* **145**, 254–264 (2017)
21. Xiao, Y., Lui, S., Deng, W., Yao, L., Zhang, W., Li, S., Wu, M., Xie, T., He, Y., Huang, X., et al.: Altered cortical thickness related to clinical severity but not the untreated disease duration in schizophrenia. *Schizophr. Bull.* **41**(1), 201–210 (2015)
22. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: *NIPS*, pp. 1601–1608 (2004)
23. Zemene, E., Tariku, Y., Idrees, H., Prati, A., Pelillo, M., Shah, M.: Large-scale image geo-localization using dominant sets. *CoRR* abs/1702.01238 (2017)