Biomarker evaluation by Multiple Kernel Learning for Schizophrenia detection

Aydın Ulaş*, Umberto Castellani*, Vittorio Murino*†, Marcella Bellani‡, Michele Tansella‡ and Paolo Brambilla§¶
* *Departimento di Informatica, University of Verona, Verona, Italy*
† *Istituto Italiano di Tecnologia (IIT), Genova, Italy*
‡ *Section of Psychiatry and Clinical Psychology, University of Verona, Verona, Italy*
§ *IRCCS "E. Medea" Scientific Institute, Udine, Italy*
¶ *DISM, Inter-University Centre for Behavioural Neurosciences, University of Udine, Udine, Italy*

***Abstract*—In this paper, we use the promising paradigm of Multiple Kernel Learning (MKL) to challenge the problem of biomarker evaluation for schizophrenia detection. We use eight different Regions of Interest (ROIs) extracted from Magnetic Resonance Images (MRIs). For each region we evaluate both tissue and geometric properties. We show that with MKL we not only obtain more accurate classifiers than using single source support vector machines (SVMs), feature concatenation and kernel averaging but also we evaluate the relevance of the brain biomarkers in predicting this disease. On a data set of 50 patients and 50 healthy controls we can achieve an increase of 7 % accuracy compared to standard methods. Moreover, we are able to quantify the importance of each source of information by highlighting the synergies between the involved brain characteristics.**

***Keywords*-biomedical imaging; magnetic resonance imaging; support vector machines; multiple kernel learning**

I. INTRODUCTION

Computational neuroanatomy is a recent and promising research area aiming at quantifying morphological characteristics of different brains [1]. The final goal is to identify differences in brain structures between patients and controls. To this aim, advanced computer vision and pattern recognition techniques may deeply help the understanding of brain characteristics and functionalities and there are several studies where these techniques are applied [2], [3]. For instance, in schizophrenia, structural and functional brain abnormalities in patients have been demonstrated [4], [5]. For a detailed overview of MR for psychiatric disorders, we recommend the readers to refer to [6].

In this paper, we propose a new approach to integrate effectively different sources of information for Schizophrenia detection. We start from a wide set of brain scans acquired by 3D Morphological MRI (SMRI), which highlights morphological properties. For each brain, a set of Regions of Interest (ROIs) are available in order to concentrate the analysis only on brain subparts which are known to have structural abnormalities in people having the disease [7], [8]. In particular, in each ROI we compute both tissue and geometric properties. To integrate and select the contribution from different parts and different features of the brain we propose the Multiple Kernel Learning (MKL) approach. MKL algorithms learn a weighted combination of different kernel functions and are able to benefit from information coming from multiple sources. MKL methods have been recently proposed in the medical imaging community to detect Alzheimer's disease [9]–[11]. Here, we use this approach by focusing on Schizophrenia. Several kernels are computed for each ROI and feature, and the contributions of these kernels are combined using MKL. We observe that MKL algorithms have better accuracy then single SVMs, feature concatenation and kernel averaging. Moreover, differently from [9]–[11], we deeply evaluate the weights computed by the method in order to highlight the importance of each brain part and feature in the detection of the disease and their collaborative contribution.

In the following, Section II describes the Multiple Kernel Learning strategy. We describe the selected data set in Sect. III and the experimental set-up in Sect. IV with the results. Finally, last remarks are discussed in Sect. V.

II. MULTIPLE KERNEL LEARNING

The assumption behind kernel methods is to transform linearly inseparable data into a higher dimensional (possibly with infinite dimension) space where it's possible to separate the classes linearly [12]. The SVM classifier in this sense converts the discriminant function into $f(\vec{x}) = \langle \vec{w}, \Phi(\vec{x}) \rangle + b$, where $\vec{w}$ and b are the parameters of the hyperplane which separates two classes, and $\Phi(\cdot)$ is the mapping function. With the so called "kernel trick", instead of defining the mapping function, the discriminant becomes

$$f(\vec{x}) = \sum_{i=1}^{N} \alpha_i y_i k(\vec{x}_i, \vec{x}) + b \,, \tag{1}$$

where $k(\vec{x}_i, \vec{x}_j) = \langle \Phi(\vec{x}_i), \Phi(\vec{x}_j) \rangle$ is the kernel function and for the linear case $k(\vec{x}_i, \vec{x}_j) = \langle \vec{x}_i, \vec{x}_j \rangle$. In this setup, there is only one kernel and one parameter set. MKL methods [13], [14], instead, learn a combination k_η of multiple kernels which allows one to do the selection/combination of different kernels or data sources automatically. The difference between most MKL algorithms is the optimization method which is applied to estimate the weights or the combination rule used [13]–[15] where the formulation for the linear case with P kernels is:

978-0-7695-4765-7/12 $26.00 © 2012 IEEE
DOI 10.1109/PRNI.2012.12

$$k\eta(\vec{x}_i, \vec{x}_j; \vec{\eta}) = \sum_{m=1}^{P} \eta_m k_m(\vec{x}_i^m, \vec{x}_j^m)\,, \qquad (2)$$

with $\eta_m \in \mathbb{R}$. The simplest approach is to use voting (mean-rule) which corresponds to $\eta_m = 1/P$. Generally, the weights η_m are automatically estimated from the data and to this aim, in the training phase, both MKL weights and SVM parameters are simultaneously estimated within the same optimization problem. In this paper, we apply `RBMKL`, `GLMKL` and `SMKL` where `RBMKL` is the rule-based MKL algorithm that trains an SVM with the mean of the combined kernels [16], `GLMKL` denotes the group Lasso based MKL algorithms [17], and `SMKL` is the iterative algorithm of [18] that uses projected gradient updates and trains single-kernel SVMs at each iteration.

III. DATA SET

In this work, we used a data set of 50 patients (30 male, 20 female) who were being treated for schizophrenia and 50 controls (30 male, 20 female) with no DSM-IV axis I disorders and had no psychiatric disorders among first-degree relatives. Diagnoses for schizophrenia were corroborated by the clinical consensus of two psychiatrists. Structural MRI scans were acquired with a 1.5 Tesla machine.

A. Tissue distribution

After the images were acquired, bias correction was applied using the SPM software [19][1]. After this step, images were segmented into specific brain regions called Regions of Interest (ROIs) manually by experts following a specific protocol for each ROI [20]. In this work, we use four ROIs from the two hemispheres of the brain summing up to a total of eight different brain regions: Amygdala (*lamyg* and *ramyg*), Entorhinal Cortex (*lec* and *rec*), Superior Temporal Gyrus (*lstg* and *rstg*), and Thalamus (*lthal* and *rthal*) which are found to be impaired in schizophrenic patients [8].

From these ROIs, we extract tissue distribution by computing histograms from intra-ROI MRI values, after MRI scale standardization based on landmark matching [21]. These histograms (119 bins) represent the first descriptor of the classification system (will be referred to as SMRI throughout the text).

B. Geometric shape descriptors

From the set of 2D ROIs of the shapes (slices) the 3D surface is computed as triangle mesh using marching cubes. A minimal smoothing operation is applied to remove noise and voxelization effect. We encode geometric properties of the surface using i) *Shape Index*, and ii) *Curvedness* [22]. Shape index is defined as:

$$si = -\frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right)\,,$$

[1]We thank Dr. Denis Peruzzo for the introduction to the SPM software

where $k_1 > k_2$ are the principal curvatures of a generic surface point. The Shape Index varies in $[-1, 1]$ and provides a local categorization of the shape into primitive forms such as spherical cap and cup, rut, ridge, trough, or saddle; is pose and scale invariant [22] and it has already been successfully employed in biomedical domain [23].

Similarly, the Curvedness is defined as:

$$curv = \frac{2}{\pi} \log \sqrt{\frac{k_1^2 + k_2^2}{2}}\,,$$

and it measures the size of curvature in a local area.

Both shape index and curvedness are computed at each vertex of the extracted mesh. Then, all the values are quantized and a histogram of occurrences is computed for both kinds of geometric properties. These histograms (150 bins each) represent the geometric descriptors of a given subject (will be referred as SHI and CURV respectively).

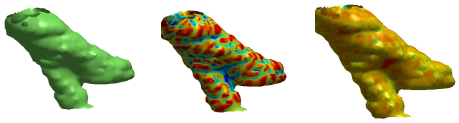

Figure 1. Geometric feature extraction: 3D surface of the thalamus (left), the surface colored according with Shape Index values (center), and Curvedness (right).

Fig. 1 shows the 3D surface of the left-Thalamus (left), the surface colored according with Shape Index values (center), and Curvedness (right).

IV. METHODOLOGY AND RESULTS

To assess the performance of the proposed methodology, we use a Leave One Out (LOO) cross validation strategy. For every subject $\vec{x}$ in the data set, we train all classifiers using all subjects but $\vec{x}$ and test the respective classifier on $\vec{x}$. We record if the classifier is able to detect if the subject is a patient or control, and calculate the overall accuracy on all subjects. We compare our results with single kernel SVM for every feature-ROI pair (will be called `SVM`) and an SVM on the concatenation of ROIs for every feature set (will be called `CONCAT`). The accuracy and the Area Under the Curve (AUC) values with the Receiver Operating Characteristic (ROC) curves clearly show the superiority of the MKL algorithms. To estimate fewer parameters, we applied linear kernels for all classifiers. The C values for the kernels and the regularization parameters of the `GLMKL` are selected using cross validation on the training set. For the MKL algorithms, all kernels are normalized to unit trace.

In Table I, we can see the accuracies of single kernel classification using the SMRI, SHI and CURV feature sets. We can see from the table that the best result obtained using single SVMs is 77.00 % when we use SMRI with *lamyg*. The closest accuracy to this value is again when we use SMRI, but this time with *rthal*. We can see from the table that tissue information always outperforms shape information.

Table I
SINGLE-KERNEL SVM ACCURACIES.

	SMRI	SHI	CURV
lamyg	**77**	51	58
ramyg	64	**57**	53
lec	74	**57**	56
rec	74	56	49
lstg	65	54	59
rstg	65	53	55
lthal	64	56	53
rthal	76	55	**64**

We then use MKL to combine different sources of information by combining multiple ROIs for each feature set. In Table II (columns 2-4), we see the accuracies of combining ROIs for each descriptor. We can see that MKL methods are better than single kernel SVMs and when we use the mean kernel, the accuracy may decrease which shows us that we need to do some kind of weighing or selection.

Table II
MKL RESULTS. COLUMNS 2-4: COMBINING ALL ROIS ON SEPARATE FEATURE SETS, COLUMN 5: COMBINING ALL FEATURE SETS AND ROIS.

	SMRI	SHI	CURV	ALL
SVM	77	57	**64**	N/A
CONCAT	79	60	55	76
RBMKL	79	56	59	80
SMKL	**81**	**65**	61	**84**
GLMKL	**81**	64	**64**	**84**

We can see the weights (average of LOO) of both MKL methods for combining multiple ROIs in the top part of Fig. 2 (We show weights from both algorithms to emphasize the consistency of different MKL methods). The figure shows how each ROI behaves when we combine all ROIs for each feature. When we analyze this figure we observe that the most important ROI for SMRI is *rec*. When we look at the shape based descriptors, we see that thalamus is the most important one and has the highest contribution to the MKL weights. These tables and figures show us that *rec* carries most of the information in the tissues whereas the shape of thalamus is the most important factor. Although MKL results using shape features alone are not promising, we will see that when we combine all information coming from all sources, we have the highest accuracy.

In Table II (last column), we can see the results combining all the 24 kernels (called ALL). We can see that we achieve the best classification accuracy when we combine everything. In this case, we have a 7 % increase in accuracy compared to the best single feature SVM result. This shows us the importance of combining different ROIs and we can achieve better results when we use multiple descriptors. When we check the weights of combining all kernels in Fig. 2 (bottom), we see that really the shape information of thalamus and the tissue information of entorhinal cortex are important factors in classifying schizophrenia from MRI images and give the most accurate classification result. We

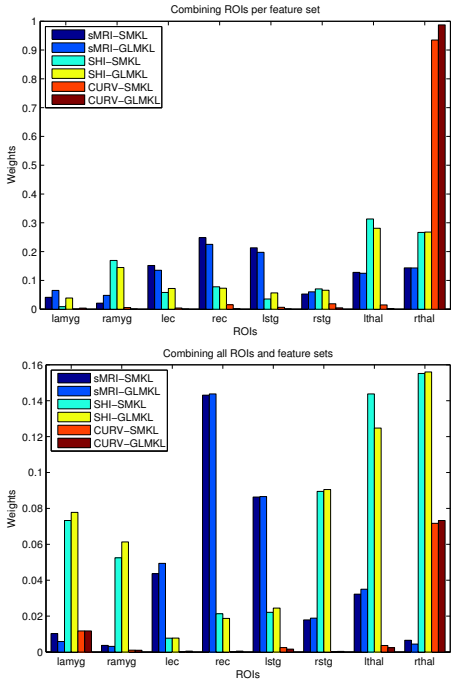

Figure 2. Weights of different descriptor-ROI combination i) top: when we combine ROIs for each descriptor and, ii) bottom: when we combine all descriptors and ROIs.

see that by combining multiple sources of information, we can achieve better accuracy than using a single source.

We can also observe the advantages of using MKL in Fig. 3 which shows the ROC curves for the best two MKLs which combine all 24 kernels (SMKL and GLMKL), the best MKL result for combining all ROIs using only one description (SMKL on SMRI) and the best single SVM (SMRI-*lamyg*). We can see a clear increase in AUC and True Positive Rate (TPR) when we use MKL.

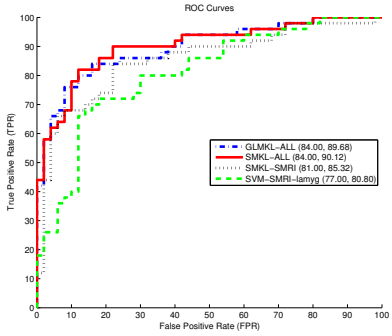

Figure 3. ROC curves of different methods. Values in parentheses show the accuracy and AUC respectively.

V. CONCLUSIONS

We have applied the Multiple Kernel Learning (MKL) paradigm to the combination of different sources of informa-

tion for schizophrenia detection using different Regions of Interest (ROIs) and feature descriptors. We have shown that MKL represents a natural and intuitive tool to evaluate the contributions of different sources of information. We have seen that the most important ROIs in the detection of this disease are Entorhinal Cortex and the Thalamus. Combining the tissue information of *ec* and shape information of *thal* gives us the most accurate discriminant. This analysis also reveals another structural information. Although by itself, *lstg* is not a good discriminant, when combined with structural differences in *rec* and shape differences in *thal*, the accuracy of detecting Schizophrenia increases. This shows us that there is an underlying relation between these ROIs for identifying the disease and one of them alone is not enough to get a good detection.

As a future work, we want to apply this paradigm to other modalities such as fMRI and DTI to achieve better classification accuracy and extract information on the functional aspects of Schizophrenia.

REFERENCES

[1] N. Giuliani, V. Calhouna, G. Pearlson, A. Francis, and R. Buchanan, "Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia," *Schizophr Res*, vol. 74, no. 2–3, pp. 135–147, 2005.

[2] C. Davatzikos, "Why voxel-based morphometric analysis should be used with great caution when characterizing group differences," *NeuroImage*, vol. 23, no. 1, pp. 17–20, 2004.

[3] A. Ulaş, R. Duin, U. Castellani, M. Loog, P. Mirtuono, M. Bicego, V. Murino, M. Bellani, S. Cerruti, M. Tansella, and P. Brambilla, "Dissimilarity-based detection of schizophrenia," *International Journal of Imaging Systems and Technology*, vol. 21, no. 2, pp. 179–192, 2011.

[4] M. Shenton, C. Dickey, M. Frumin, and R. McCarley, "A review of MRI findings in schizophrenia," *Schizophr Res*, vol. 49, no. 1–2, pp. 1–52, Apr 2001.

[5] D. Rujescu and D. Collier, "Dissecting the many genetic faces of schizophrenia," *Epidemiologia e Psichiatria Sociale*, vol. 18, no. 2, pp. 91–95, 2009.

[6] N. Agarwal, J. Port, M. Bazzocchi, and P. F. Renshaw, "Update on the use of MR for assessment and diagnosis of psychiatric diseases," *Radiology*, vol. 255, no. 1, pp. 23–41, 2010.

[7] M. Bellani and P. Brambilla, "The use and meaning of the continuous performance test in schizophrenia," *Epidemiologia e Psichiatria Sociale*, vol. 17, no. 3, pp. 188–191, 2008.

[8] C. Corradi-DellAcqua, L. Tomelleri, M. Bellani, G. Rambaldelli, R. Cerini, R. Pozzi-Mucelli, M. Balestrieri, M. Tansella, and P. Brambilla, "Thalamic-insular dysconnectivity in schizophrenia: Evidence from structural equation modeling," *Human Brain Mapping*, vol. 33, pp. 740–752, 2012.

[9] C. Hinrichs, V. Singh, G. Xu, and S. Johnson, "Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population," *NeuroImage*, vol. 55, no. 2, pp. 574–589, 2011.

[10] R. Filipovych, S. Resnick, and C. Davatzikos, "Multi-kernel classification for integration of clinical and imaging data: Application to prediction of cognitive decline in older adults," in *MLMI '11*, vol. 7009, 2011, pp. 26–34.

[11] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.

[12] V. N. Vapnik, *Statistical Learning Theory*. John Wiley and Sons, 1998.

[13] F. Bach, G. Lanckriet, and M. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *ICML '04*, 2004, pp. 41–48.

[14] G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. Jordan, "Learning the kernel matrix with semidefinite programming," *JMLR*, vol. 5, pp. 27–72, 2004.

[15] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *JMLR*, vol. 12, pp. 2181–2238, 2011.

[16] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[17] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien, "l_p-norm multiple kernel learning," *JMLR*, vol. 12, pp. 953–997, 2011.

[18] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *JMLR*, vol. 9, pp. 2491–2521, 2008.

[19] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, Eds., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007. [Online]. Available: http://store.elsevier.com/product.jsp?isbn=9780123725608

[20] M. Baiano, C. Perlini, G. Rambaldelli, R. Cerini, N. Dusi, M. Bellani, G. Spezzapria, A. Versace, M. Balestrieri, R. P. Mucelli, M. Tansella, and P. Brambilla, "Decreased entorhinal cortex volumes in schizophrenia," *Schizophr Res*, vol. 102, no. 1–3, pp. 171–180, 2008.

[21] L. Nyúl, J. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, 2000.

[22] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image and Vision Computing*, vol. 10, pp. 557–565, October 1992.

[23] S. P. Awate, P. Yushkevich, Z. Song, D. Licht, and J. C. Gee, "Multivariate high-dimensional cortical folding analysis, combining complexity and shape, in neonates with congenital heart disease," in *IPMI '09*, 2009, pp. 552–563.