

Multimodal Schizophrenia Detection by Multiclassification Analysis

Aydın Ulaş^{1,*}, Umberto Castellani¹, Pasquale Mirtuono¹, Manuele Bicego^{1,2},
Vittorio Murino^{1,2}, Stefania Cerruti³, Marcella Bellani³, Manfred Atzori⁴,
Gianluca Rambaldelli³, Michele Tansella³, and Paolo Brambilla^{4,5}

¹ University of Verona, Department of Computer Science, Verona, Italy

² Istituto Italiano di Tecnologia (IIT), Genova, Italy

³ Department of Public Health and Community Medicine,
Section of Psychiatry and Clinical Psychology, Inter-University Centre for
Behavioural Neurosciences, University of Verona, Verona, Italy

⁴ IRCCS “E. Medea” Scientific Institute, Udine, Italy

⁵ DISM, Inter-University Centre for Behavioural Neurosciences, University of Udine,
Udine, Italy

Abstract. We propose a multiclassification analysis to evaluate the relevance of different factors in schizophrenia detection. Several Magnetic Resonance Imaging (MRI) scans of brains are acquired from two sensors: morphological and diffusion MRI. Moreover, 14 Region Of Interests (ROIs) are available to focus the analysis on specific brain subparts. All information is combined to train three types of classifiers to distinguish between healthy and unhealthy subjects. Our contribution is threefold: (i) the classification accuracy improves when multiple factors are taken into account; (ii) proposed procedure allows the selection of a reduced subset of ROIs, and highlights the synergy between the two modalities; (iii) correlation analysis is performed for every ROI and modality to measure the information overlap using the correlation coefficient in the context of schizophrenia classification. We see that we achieve 85.96 % accuracy when we combine classifiers from both modalities, whereas the highest performance of a single modality is 78.95 %.

Keywords: Machine learning algorithms, Magnetic resonance imaging, Support vector machines, Correlation.

1 Introduction

Computational neuroanatomy using magnetic resonance imaging (MRI) is a growing research field that employs image analysis methods to quantify morphological characteristics of different brains [5]. The ultimate goal is to identify structural brain abnormalities by comparing normal subjects (controls) with patients affected by a certain disease. Advanced computer vision and pattern

* Corresponding author.

recognition techniques may deeply help the understanding of brain characteristics and functionalities and there are several studies where these techniques are applied [4,3]. In this paper we work on schizophrenia on which a substantial body of research demonstrates numerous structural and functional brain abnormalities in patients [11,9]. [1] is an excellent review on the use of MR for psychiatric diseases for the interested readers.

In this paper, we propose an image-based analysis starting from a rather wide set of brain scans acquired by two sensors: i) 3D Morphological (SMRI), which highlight morphological properties, and ii) Diffusion Weighted Imaging (DWI), which show the microstructure of the tissues. For each brain, a set of Region of Interests (ROIs) are available in order to concentrate the analysis only on brain subparts which are in relation with the disease [2].

The main contribution of the paper is the exploitation in a multiclassification scenario of both morphological and diffusion data for schizophrenia detection; moreover we show also the effect of combining different classifiers and ROIs. In particular, we evaluate several strategies for combining several aspects [6]: ROIs, data modalities, and type of classifiers. As a second contribution, we observe the correlations¹ between the classifiers trained on different ROIs and modalities to measure the level of overlap of information contained among the ROIs and modalities.

In the following, we describe the selected data set and the experimental setup in Sect. 2 and 3, respectively. Subsequently, we report the results combining several types of data and processing strategies in Sect. 4, and finally, last remarks are discussed in Sect. 5.

2 Data Set

We used a data set of 59 patients and 55 healthy subjects for both 3D SMRI and DWI modalities (Figure 1). SMRI data are more often used for human brain research. Data are quite accurate with respect to noise and the volume-data is represented with high resolution. Conversely, DWI data are more noisy and suitable for evidencing the microstructure of the tissues aiming at analyzing the integrity of the brain. It is worth noting that DWI data are less used for human brain research, and only few work have been done in schizophrenia. Several ROIs have been traced from SMRI data by drawing contours enclosing the intended region. For each ROI, the tracing has been carried out by a trained expert following a specific protocol [3]. In Fig. 2, we show a sample of the right superior temporal gyrus: the volume is composed of 35 slices of size 41×40 (ordered from left to right, top to bottom).

In order to obtain the ROIs also on the DWI-space, we apply a nonrigid registration between diffusion and morphological images using the 3DSlicer (Available at <http://www.slicer.org/>) in accordance with a standard medical procedure. According with our previous work [3], we generate a histogram of properly normalized intensity values for each ROI and subject for SMRI. For DWI we extract

¹ The idea of using the correlations for information extraction is adapted from [13].

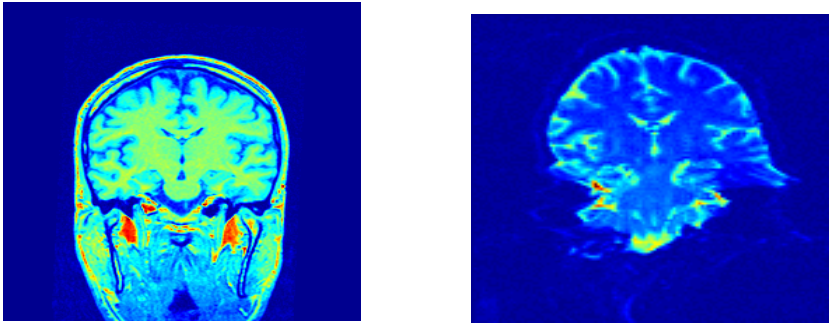


Fig. 1. Two slices acquired by 3D Morphological (SMRI) (left) and Diffusion Weighting Imaging DWI (right) techniques

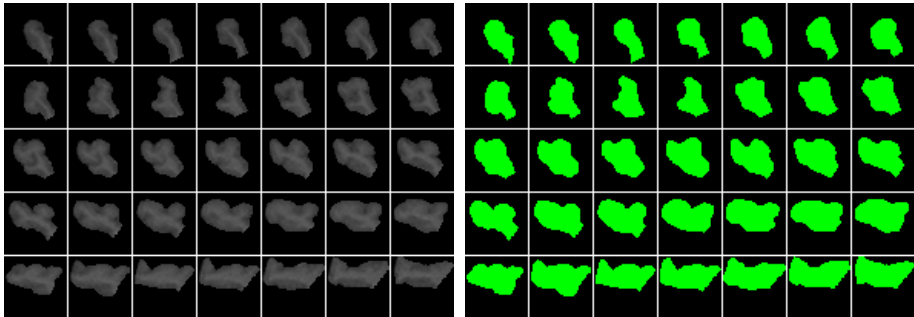


Fig. 2. Montage of the slices in the ROI volume ($41 \times 40 \times 35$) of *rstg* for subject 11. On the left, the MRI values; on the right, the corresponding binary masks.

the histograms of Apparent Diffusion Coefficient (ADC) values. Such histogram representations are the feature vectors which are evaluated by the classifiers.

3 Experimental Setup

In order to summarize the experimental setup, two issues should be highlighted, which are reported in the following.

3.1 Factor Combination

We propose to combine the following factors: i) three classifiers: k -nearest neighbor with $k = 1$ (*1nn*), support vector machine with a linear kernel (*svl*), and support vector machine with a radial basis kernel (*svr*), ii) seven ROIs \times two hemispheres: Amygdala (*lamyg* and *ramyg*), Dorso-lateral PreFrontal Cortex (*ldlpfc* and *rdlpfc*), Entorhinal Cortex (*lec* and *rec*), Heschl's Gyrus (*lhg* and *rhg*), Hippocampus (*lhippo* and *rhippo*), Superior Temporal Gyrus (*lstg* and *rstg*), and Thalamus (*lthal* and *rthal*), and iii) two modalities: SMRI data, and DWI data.

3.2 Validation Criteria and Decision Rule

For every factor combination, we use leave-one-out (LOO) methodology to train the models and assess the accuracy. Instead of using a binary decision rule, we record the posterior probability outputs of each model and use them for constructing ensembles using the mean combiner (i.e. SUM rule in classifier combination).

We also add another representation which is the concatenation of SMRI and DWI data, which we call CON. Therefore, each trained basic model consists of one modality, one ROI, and one classifier (i.e., SMRI-*lamyg-Inn*). The modality, the ROI or the classifier will be omitted throughout the text when the context is clear.

3.3 Correlation of Base Classifiers

We use the posterior probability outputs of the *svr* classifier to calculate the correlation coefficient between ROIs and modalities. It's known that for ensembles to have better accuracy, we need to have diverse base classifiers. Various measures of diversity exist [7], but in general, diversity can be defined as the classifiers responding differently to the same input. Correlations of base classifiers are also used when creating ensembles, and in this paper we used the correlation coefficient proposed by Petrakos *et al.* [8] which was studied empirically in the context of decision fusion. We apply the correlation computation to see how the ROIs and modalities are correlated. In principle, one would use a validation set to compute the correlation of the classification algorithms and then form the ensemble according to this information [7]. Instead, what we do in this study is to calculate the correlation of pairwise ROIs and modalities on the test set to *observe* the overlap of information contained in different ROIs [13].

The correlation between two classifiers is calculated as in [8]:

$$\rho_2 = \frac{2 \times N_{00}}{N_{01} + N_{10} + 2 \times N_{00}}$$

where N_{ij} shows the number of subjects (N_{00} : incorrectly classified by both classifiers, N_{11} : correctly classified by both classifiers, N_{01} incorrectly classified by the first and correctly classified by the second, and N_{10} : correctly classified by the first, and incorrectly classified by the second).

4 Results and Discussions

In this section, we evaluate different strategies to combine classifiers, ROIs, and data modalities.

4.1 Exp 1: Single Classifier per ROI

In the first experiment, we compare classification performances by training a single classifier per ROI and modality. In the left part of Table 1, one can see

the best LOO accuracies among the three different types of classifiers. Please, note that the best accuracy is 71.93 when we use SMRI-*ldlpfc*. The best accuracy of DWI is 62.28 when we use *lec* which is interesting because it is one of the only three ROIs where DWI is more accurate than SMRI. In half of the ROIs the accuracy improves when the feature concatenation approach is applied.

The right part of Table 1 shows instead the integration of the three types of classifiers using the mean combiner for each ROI and modality. Also this combination strategy does not yield better accuracies, sometimes makes it worse.

Table 1. Best and mean combiner (MC) accuracies using a single classifier per ROI and modality (in %)

	BEST			MC		
	SMRI	DWI	CON	SMRI	DWI	CON
<i>lamyg</i>	64.91	59.65	71.93	63.16	57.02	69.30
<i>ramyg</i>	64.91	53.51	52.63	63.16	47.37	56.14
<i>ldlpfc</i>	71.93	50.88	57.02	64.91	51.75	54.39
<i>rdlpfc</i>	61.40	51.75	64.91	63.16	48.25	59.65
<i>lec</i>	57.02	62.28	69.30	50.00	62.28	59.65
<i>rec</i>	63.16	54.39	63.16	65.79	53.51	55.26
<i>lhg</i>	55.26	58.77	62.28	52.63	57.89	58.77
<i>rhg</i>	57.02	57.02	58.77	51.75	51.75	54.39
<i>lhippo</i>	70.18	61.40	66.67	64.04	64.04	66.67
<i>rhippo</i>	57.89	60.53	57.02	55.26	46.49	47.37
<i>lstg</i>	62.28	57.89	64.04	59.65	57.02	62.28
<i>rstg</i>	62.28	51.75	64.91	58.77	49.12	66.67
<i>lthal</i>	63.16	52.63	57.02	60.53	50.00	57.89
<i>rthal</i>	61.40	57.02	58.77	58.77	53.51	57.02

4.2 Exp 2: Integration of ROIs

In the second experiment, we fix modalities and classifiers; and combine the ROIs (Table 2, upper part) using the mean combiner. One can see that the best accuracy of CON is improved from 71.93 % to 74.56 %. It's known that combining a subset of models may lead to better accuracies than using all the models [10,12]. Bearing this in mind, we have performed another experiment where we selected the best seven ROIs. We can see from Table 2 (middle) that the accuracies increase 5%, 14 % and 3% respectively when we use the *svr* classifiers and the best 7 ROIs. Using this selection strategy, we can have 76.32 accuracy when we use the morphological data. Even better is the result of the concatenation: 78.07 %. This experiment suggests that the best integration strategy is obtained when we combine a subset of different ROIs for any modality. Therefore, we design an exhaustive experiment to select automatically such subset of ROIs. We first have performed the analysis for the three modalities separately. Then, we allow the combination between all the ROIs of both the sensors at the same time. In these experiments, we use *svr*, because it has the best performance. The results can be seen in Table 2 (bottom). The optimum accuracy is 78.95 when a single modality is used, and 83.33 when we use the concatenation of both modalities whereas when both modalities are jointly combined the accuracy improves to 85.96. Another interesting fact is that optimal ROIs combination for

Table 2. Combination accuracies for combining all ROIs and a selection of ROIs (in %)

	<i>inn</i>	<i>svl</i>	<i>svr</i>
SMRI	60.53	64.91	71.05
DWI	63.16	56.14	57.02
CON	57.89	67.54	74.56
Best7-SMRI	59.65	68.42	76.32
Best7-DWI	53.51	51.75	71.05
Best7-CON	52.63	67.54	78.07
<i>opt</i> -SMRI	-	-	78.95
<i>opt</i> -DWI	-	-	74.56
<i>opt</i> -CON	-	-	83.33
<i>opt</i> -JOINT	-	-	85.96

SMRI includes five ROIs which are not the best five. The optimum combination for DWI selects four ROIs which are again in the best seven but not the best four. This shows us that different classifiers may perform well in classifying different parts of the input space, and even though they have low single accuracies, they may bring a relevant improvement when combined with other parts.

Another interesting aspect of this evaluation is that the optimum combination including both the modalities selects five ROIs from SMRI and three ROIs from DWI which again differ from those selected for the single modality evaluation. It is worth to note that this combination strategy not only provides the best performance in detecting schizophrenic subjects, but also allows us to localize the most discriminative brain subparts.

4.3 Exp 3: Correlations of Classifiers

In the above experiments we analyze that it is good to combine multiple ROIs in order to get better ensemble accuracy. The better method is to combine the outputs of classifiers coming from different modalities which our exhaustive analysis pointed out. One could also use the correlations to come up with such ensembles, but since we exhaustively searched all the solution space, instead of showing how to build an ensemble using correlation, we will observe the pairwise correlations among ROIs and modalities. Most of the correlations are below 0.50 (note that ρ^2 gets values in the range $[0, 1]$) and the correlations above this value are unevenly distributed amongst ROIs and the two modalities. The highest correlation is 0.61 between SMRI-*lthal* and SMRI-*ldlpfc*. This shows us that the correlations between these classifiers are low and combining them to construct ensembles would lead better ensemble accuracy which was also shown by our experiments above. This also has an important medical interpretation. It shows us that the information contained in each of these ROIs and across these two modalities are different (also complementary as we show above) and we should combine the information in these ROIs and modalities to get better ensemble accuracy. Table 3 shows the correlations across modalities and hemispheres. The first part of the table shows the correlations across the two modalities using the same ROI. We can see that the highest correlation is 0.56 which is observed on *rhippo*. The second part of the table shows the correlations across the brain hemispheres. This time we can see that the highest correlation is 0.63 and between *stg* of SMRI. We can see

that correlations inside ROIs are higher than cross-modality correlations which shows us that it may be a good idea to combine the two modalities which was confirmed by our combination experiments. Analyzing the table, we can also say the following: In the context of schizophrenia detection (using these features), the hemispheres of the brain are more similar in terms of morphology than function. Also we can see that the correlations between the right hemisphere of the two modalities are higher (except *stg*) than the correlations between the left hemispheres which shows that they are more similar in terms of discriminative power.

Table 3. Correlation of modalities and hemispheres. *l*- shows the left hemisphere and *r*- shows the right hemisphere.

	<i>amyg</i>	<i>dlpfc</i>	<i>ec</i>	<i>hg</i>	<i>hippo</i>	<i>stg</i>	<i>thal</i>
<i>l</i> -SMRI vs. <i>l</i> -DWI	0.37	0.41	0.39	0.39	0.44	0.49	0.44
<i>r</i> -SMRI vs. <i>r</i> -DWI	0.44	0.55	0.46	0.45	0.56	0.48	0.47
<i>l</i> -SMRI vs. <i>r</i> -SMRI	0.53	0.47	0.49	0.46	0.49	0.63	0.56
<i>l</i> -DWI vs. <i>r</i> -DWI	0.45	0.58	0.46	0.42	0.43	0.51	0.52

5 Conclusions

In this work, we evaluate the effectiveness of different classifier combination strategies in the context of schizophrenia detection. Even if the question “Can schizophrenia be detected just by analyzing MRI images?” is still unsolved, this study provides novel useful insights on the combined effect of two data modalities: in particular we have considered morphological and diffusion data, extracted from 14 ROIs associated to brain subparts, and classified using three types of classifiers. We have seen that neither concatenating the two modalities, nor combining different types of classifiers on the same ROI, provides the expected effect of increasing ensemble accuracy. Such effect may be found when we use a carefully selected subset of ROIs, which are combined using the mean combiner strategy. The best increase in accuracy occurs when we combine also the two data modalities (JOINT). This highlights that the information encoded in the morphological and diffusion data are different, and their contributions to the classification are complementary. We also showed this using correlation coefficients between classifiers. Using the correlation coefficients, we have seen that the classifiers trained using different modalities are more diverse. Also we have seen that in terms of schizophrenia detection, the two hemispheres are more similar in terms of morphology than in terms of function. As a future work, we will exploit the use of other classifier combination algorithms to improve the ensemble accuracies.

Acknowledgements. We acknowledge financial support from the FET programme (EU FP7), under the SIMBAD project (contract 213250).

References

1. Agarwal, N., Port, J.D., Bazzocchi, M., Renshaw, P.F.: Update on the use of MR for assessment and diagnosis of psychiatric diseases. *Radiology* 255(1), 23–41 (2010)
2. Bellani, M., Brambilla, P.: The use and meaning of the continuous performance test in schizophrenia. *Epidemiologia e Psichiatria Sociale* 17(3), 188–191 (2008)
3. Cheng, D.S., Bicego, M., Castellani, U., Cerruti, S., Bellani, M., Rambaldelli, G., Atzori, M., Brambilla, P., Murino, V.: Schizophrenia classification using regions of interest in brain mri. In: *Proceedings of Intelligent Data Analysis in Biomedicine and Pharmacology, IDAMAP 2009*, pp. 47–52 (2009)
4. Davatzikos, C.: Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* 23(1), 17–20 (2004)
5. Giuliani, N.R., Calhoun, V.D., Pearson, G.D., Francis, A., Buchanan, R.W.: Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophrenia Research* 74(2-3), 135–147 (2005)
6. Kuncheva, L.I.: *Combining pattern classifiers: methods and algorithms*. Wiley-Interscience (2004)
7. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning* 51(2), 181–207 (2003)
8. Petrakos, M., Kannelopoulos, I., Benediktsson, J.A., Pesaresi, M.: The effect of correlation on the accuracy of the combined classifier in decision level fusion. In: *Proceedings of the IEEE International Geo-science and Remote Sensing Symposium, IGARSS 2000*, vol. 6, pp. 2623–2625 (2000)
9. Rujescu, D., Collier, D.A.: Dissecting the many genetic faces of schizophrenia. *Epidemiologia e Psichiatria Sociale* 18(2), 91–95 (2009)
10. Ruta, D., Gabrys, B.: Classifier selection for majority voting. *Information Fusion* 6(1), 63–81 (2005)
11. Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W.: A review of mri findings in schizophrenia. *Schizophrenia Research* 49(1-2), 1–52 (2001)
12. Ulaş, A., Semerci, M., Yıldız, O.T., Alpaydm, E.: Incremental construction of classifier and discriminant ensembles. *Information Sciences* 179(9), 1298–1318 (2009)
13. Ulaş, A., Yıldız, O.T., Alpaydm, E.: Eigenclassifiers for combining correlated classifiers. *Information Sciences* (accepted, 2011)