WAVELET FEATURES FOR IMAGE CLASSIFICATION IN NEUROSCIENCE RESEARCH

K. Maršálová[1,3], D. Schwarz[2,3]

1 Masaryk University, Faculty of Science

2 Masaryk University, Faculty of Medicine

3 Institute of Biostatistics and Analyses, Ltd.

Abstract

Recently, machine learning methods have been used for computer-aided diagnostics of neuropsychiatric disorders. The application of these methods on high-dimensional neuroimaging data may help in solving the problem of subjective diagnostics in psychiatric diseases, such as schizophrenia, where interviewing patients and their relatives is still considered as gold standard.

This paper presents a classification framework, which was designed to distinguish between two classes of subjects based on their imaging data from magnetic resonance. Features are extracted using two different automated brain morphometry methods: voxel-based (VBM) and deformation-based morphometry (DBM), and their representation in the wavelet domain. The number of features was reduced by thresholding in the wavelet domain and further by selecting only those features carrying the most discriminant information assessed with Fisher's discriminant ratio. The framework involves a support vector machine classifier (SVM) with linear kernel and an evaluation strategy based on leave-one-out cross-validation.

There are two basic parameters in the framework: (i) wavelet decomposition level and (ii) number of selected features. Their different setup and various types of features were tested on a dataset resulting from a clinical study focused on first-episode schizophrenia (FES) patients. Multiple experiments resulted in quantified quality of classification between 52 FES patients and 52 healthy controls. The highest classification accuracy – 73.08 % – was achieved with 1000 selected features extracted by VBM and four decomposition levels. In the case of DBM features, the classificator achieved the highest accuracy of 72.12 % for five decomposition levels and 5000 discriminating features.

The framework was deployed in the form of Matlab scripts and functions. Although the achieved classification quality is comparable with some other published results, there is still a space for improvements, as the accuracy in the real clinical practice is expected to be higher than 95 %.

1 Introduction

In recent years, medical imaging methods have been intensively developed to provide comprehensive and extensive data for further processing. This progress in the field of neuroscience allows, on the one hand, a thorough study of the brain structures, but also the finding of connections between the brain structure and its function. Hence, analysis of brain imaging data in psychiatric research has begun to be used with great potential, having discovered a link between neuropsychiatric disorders and structural change in the brain.

In the field of computational neuroanatomy, structural magnetic resonance imaging data (MRI) are often analysed using automatic brain morphometry, especially Voxel-based morphometry (VBM) and Deformation-based Morphometry (DBM). Unlike older morphometric methods (e.g. volumetry), it is not necessary to define arbitrary anatomical boundaries for these methods but spatial normalization of the image into the stereotactic must be performed [2].

VBM finds statistical differences in tissue volume at the level of individual voxels. The result of the analysis is the so-called Statistical Parametric Map (SPM). It represents voxels in which the concentration of grey or white matter is statistically significantly different among the studied groups of subjects [5]. DBM, similar to VBM, is another technique of computational neuroanatomy widely used to support diagnostics from image data. Unlike the aforementioned VBM, this method detects structural differences (changes in shape, position, size) of brain regions relative to a given template [14] without need to segment individual brain tissues.

Both of these methods are often used for feature extraction and subsequent comparative analysis to reveal anatomical or morphological abnormalities in a selected group of patients against healthy controls [6] [7] [14]. In order to differentiate those groups, the support vector algorithm (SVM - Support Vector Machine) is used very often [3] [4] [14].

This paper focuses on processing and analysis of magnetic resonance images of patients with the first episode of schizophrenia.

2 Methods

The proposed algorithm consists of 3 key steps – see Figure 1. The images are pre-processed with DBM and VBM in order to get features with clear biological-clinical interpretation. Thresholding in the wavelet domain is performed with the aim of basic dimensionality reduction. The reduced data enter into the classification algorithm and subsequently the quality of classification is evaluated quantitatively using the leave-one-out cross-validation approach.

2.1. Subjects and data

The data set contained a total of 104 subjects - 52 schizophrenic patients and 52 healthy subjects (NC - Normal Control). It consists of 104 T1-weighted MRI images of the entire subject's head. Images with a resolution of $160 \times 512 \times 512$ voxels were acquired by a 1.5 T magnetic resonance device. Patients, exclusively men with an average age of 24 years, were admitted to the Psychiatric Clinic of Masaryk University in Brno – upon admittance, they manifested features of this disease longer than one month for the first time. The diagnosis was based on an interview with senior psychiatrist. After follow-up examinations, some subjects were excluded from the study due to neurological diseases, drug addiction, and others [14]. 52 healthy male volunteers, were paired with patients based on their age and their writing hand preference.

The T1-weighted data was pre-processed using DBM or VBM. In addition to the images with morphometric features, the feature images were decomposed by discrete wavelet transformation at different levels. Morphometric imaging features in the original domain and in the wavelet domain were included in series of experiments aimed at classification performance assessment.

2.2 Sparsity and Wavelet Transform

In order to extract features, 3-D discrete wavelet transformation was used. It transforms the image data into a space where information contained therein is sparsely[1] represented. In order for this transformation to take place, two parameters have to be determined: maternal wavelet and image decomposition level. The levels of decomposition 3,4 and 5 and the sym5 wavelet were selected in consultation with an expert and based on literature [9].

After wavelet decomposition of the image data, each subject is described by means of a long wavelet coefficient vector The coefficients with the highest values, the greatest amount of the image energy is contained [15]. For this reason, values near to zero, i.e. values lower than the predetermined threshold were removed from the feature vector [1], and the dimensionality of the data was reduced. The threshold value was set to be 0.05 based on preliminary experiments.

[1] Discrete signal is sparse if most of his patterns or samples equal zero. If it is not sparse, it can be transformed into another domain and get the sparse characteristics, which might help with its analysis. [12].

In the DBM dataset after wavelet transformation to 3 levels of decomposition, only 90,000 of the original 8.5 million coefficients were left in this way, with 99% energy remaining. In the VBM after wavelet transformation, there were about 3 million features reduced to about 23 thousand, and 99% of the energy remaining. If the algorithm does not work with data that used the wavelet transformation to extract the features, they are selected in a different way from this data. Primary data range reduction is then performed using the binary brain mask instead of the above-mentioned thresholding. This step is followed by a metrics calculation in all data files to select discriminating features.

2.3 Feature Selection

The still too large dimensionality of the features space was reduced using Fisher's Discrimination Ratio (FDR), which seemed to be the most appropriate tool for selecting those features that divided the subjects into the desired NC and FES groups with the highest precision [17] [19].

$$FDR = \frac{(\mu_1 - \mu_2)^2}{{\sigma_1}^2 + {\sigma_2}^2}, \tag{2.1}$$

Where μ_1 and μ_2 are the mean feature values between first and second class subjects and σ_1^2 and σ_2^2 denote the variance of the feature values within the classes. After a particular criterion value is calculated for each feature, it is sorted down by size, and the first p features are then selected and entered into the algorithm. An important question is how many of these p best-discriminating features should be chosen. The cycled algorithm includes one more cycle that evaluates the classifier's success rate for selected p values. This number of features is set to {100, 500, 1000, 5000, 10,000} within the experiment. By using these number of features, the classifier is taught and results in a comparison of the overall classification accuracy for the given number of features.

2.4 Classification

At this stage, each subject is described by means of a long vector p of selected features with a high-discrimination score (FDR) and an identifier that determines to which group the subject is ranked. This data then serves as an input to the learning phase of the SVM classifier with a linear nucleus. The output of the classification algorithm is then an identifier that expresses the class to which the subject belongs according to the classifier. After teaching the classifier, its classification capability is verified on a test subject. Classification quality is determined here by sensitivity, specificity and overall accuracy:

$$SENS = \frac{TP}{TP+FN}, \tag{2.2}$$

$$\mathrm{SPEC} = \frac{\mathrm{TN}}{\mathrm{FP+TN}}, \tag{2.3}$$

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FN+TN+FP}}, \tag{2.4}$$

where TP expresses true positive results, TN denotes true negatives, FP indicates false positives, and FN indicates false negatives. These variables are calculated from the results obtained during leave-one-out cross-validation [10].

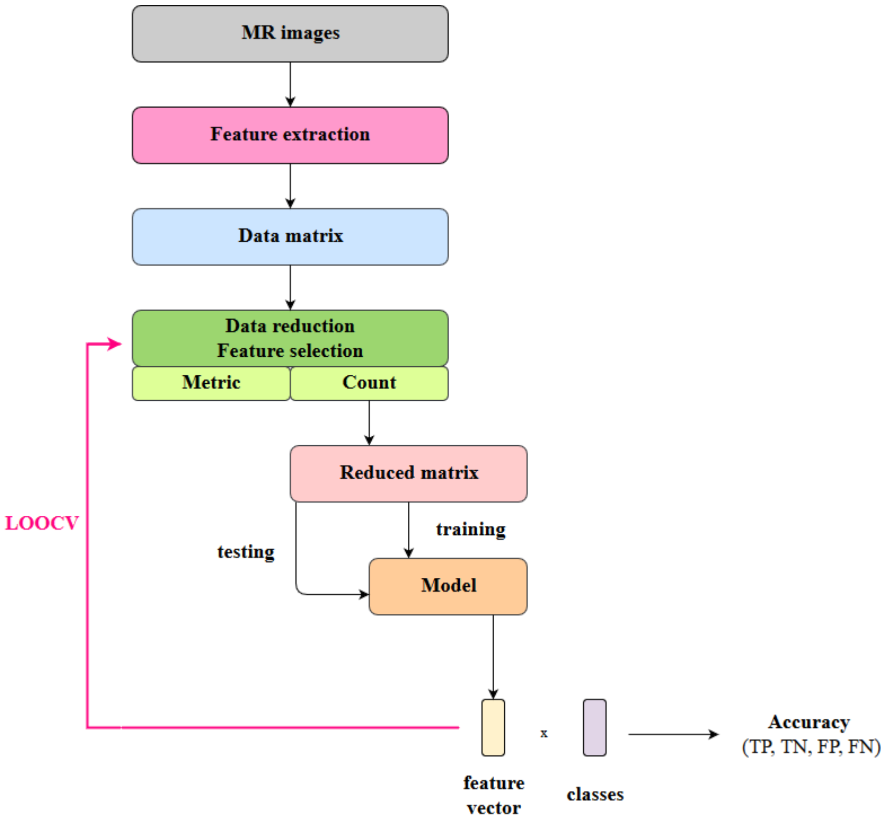

Figure 1 - a schematic diagram of the proposed classification algorithm. Classifier quality evaluation is done through LOOCV (leave-one-out cross-validation). TP indicates true positive classification results, FP indicates false positive classification results, TN indicates true negative classification results and FN indicates false negative classification results.

3 Experiments and Results

For the experiments in this paper, several optional parameters were selected in the classification algorithm: i) pre-processing of magnetic resonance imaging (DBM, VBM); ii) decomposition level by wavelet transformation (3, 4, 5); iii) selection of the best number of features (500, 1000, 5000, 10000). In the experiment, all of their combinations were tested, i.e. $2 \times 3 \times 5 = 30$ experiments with data that were processed by wavelet transformation and $2 \times 5 = 10$ with data that were not processed by the wavelet transformation. Altogether, 40 experiments were carried out. All of these experiments worked with different combinations of the above-mentioned parameters. On their basis, various features during the algorithm were selected which then entered its classification section.

The results of the experiments for different levels of decomposition by wavelet transformation and for the two best numbers of discriminating features are shown in Table 1. While VBM achieves the most accurate results for 4 decomposition levels by wavelet transformation when working with 1000 discriminating features, DBM achieves the highest accuracy for 5 levels of decomposition when selecting 5000 discriminating features.

Table 1: THE SUCCESS OF CLASSIFYING A DATA FILE DEPENDING ON THE LEVEL OF DECOMPOSITION BY WAVELET TRANSFORMATION AND THE NUMBER OF FEATURES. THE VALUES THAT GET THE BEST RESULTS ARE HIGHLIGHTED.

preprocess.	*decomposition levels*	*number of features*	*sensitivity [%]*	*specificity [%]*	*accuracy [%]*
VBM	0	1000	51,92	63,46	57,69
	0	5000	50,00	59,62	54,81
	3	1000	69,23	71,15	70,19
	3	5000	67,31	75,00	71,15
	4	**1000**	**71,15**	**75,00**	**73,08**
	4	5000	65,38	69,23	67,31
	5	1000	69,23	73,08	71,15
	5	5000	63,46	71,15	67,31
DBM	0	1000	51,92	42,31	47,12
	0	5000	67,31	59,62	63,43
	3	1000	59,62	61,54	60,58
	3	5000	69,23	67,31	68,27
	4	1000	55,77	57,69	56,73
	4	5000	67,31	71,15	69,23
	5	1000	57,69	59,62	58,65
	5	**5000**	**67,31**	**76,92**	**72,12**

4 Discussion

In this paper, a classification algorithm for image recognition was designed in the context of the upcoming era of computer-aided diagnostics of neuropsychiatric disorders. Four types of data entered into this algorithm depending on the pre-processing method: data from voxel-based morphometry and their decomposition by wavelet transformation and deformation-based morphometric data and their decomposition.

The experiments mainly dealt with data that were processed by wavelet transformation. This data were decomposed to three, four and five levels, based on expert consultation and literature [9]. There has been some limitation of the classification algorithm parameters. Another limiting parameter was a predetermined sym5 mother wavelet from the symlets family, which was also determined based on a study [9] that states their good performance for the decomposition of natural images.

All data decomposed by wavelet transformation were classified after eliminating coefficients at an absolute value smaller than the pre-selected threshold T = 0.05. This value was determined on the basis of several experiments and calculations of the total image energy, which decreased with each higher value of the threshold.

The discriminating power of the coefficients determined by the Fisher Discrimination Ratio was calculated independently of the other features. The average values of these coefficients increase slightly depending on the decomposition level (Table 2).

Table 2: AVERAGE VALUES OF THE DISCRIMINATING POWER OF THE WAVELET COEFFICIENTS AT DIFFERENT DECOMPOSITION LEVELS.

decomposition levels	*preprocessing*	
	DBM	VBM
3	0,1385	0,0825
4	0,1438	0,0889
5	0,1451	0,0956

As the final results suggest (Table 1), the classification algorithm, which recognizes the images from which the discriminating features were extracted by wavelet transformation compared to the case without transformation, produces better results. In the case of data that describes local volume changes, especially in the recognition of images decomposed into 5 levels of decomposition, while in grey matter density data, the algorithm achieves the best accuracy at 4 levels of decomposition. The fact that higher classification performance is achieved by data processed by wavelet transformation could be caused by the fact that a significant amount of information is contained in a smaller number of coefficients thanks to wavelet transformation.

Another factor influencing the resulting classification accuracy is, in addition to the above mentioned decomposition level and the data pre-processing method which is no less important, is the number of discriminating features with which the algorithm works. The best results, i.e. 73.08% accuracy, 71.15% sensitivity, and 75.00% specificity were obtained from the data that describes the grey matter density, that were processed by wavelet decomposition at 4 levels and the classifier worked with 1000 features. In the case of data describing local volumetric changes, it also achieved the best results at 5 levels of decomposition for 5000 features with 72.12% accuracy, 67.31% sensitivity and 76.92% specificity.

For a fair comparison of this paper with literature, it is necessary to compare it only in the context of works that did not work with the image data of patients who suffer from the chronic form of schizophrenia. This is due to the fact that individuals with chronic schizophrenia have more marked morphological differences in the brain [11] [13].

The results of this work are comparable to scientific papers dealing with the design of algorithms for the objective method of diagnosing the first episode of schizophrenia [8] [14] [16] [18]. However, it is very important to note that the success rate of about 70% is still insufficient to use these procedures for diagnostic purposes in clinical practice.

5 Conclussion

This paper dealt with the application of wavelet transformation to image data, which was further used for computer-aided diagnosing the first episode of schizophrenia. Schizophrenia is currently diagnosed only through neuropsychological tests and interviews, therefore these procedures could lead to a more effective and accurate diagnosis of this neuropsychiatric disorder. In such case, it would be possible to refine the medication, or set individual treatment for each patient, thereby increasing the number of healed patients.

Differently pre-processed data entered into the proposed algorithm; their dimensionality was subsequently reduced by reducing and selecting the features suitable for subsequent classification with the teacher. Success rate of the proposed classifier has been tested too.

The obtained results can be compared using precision indicators to the results of similarly focused works, which also focused on designing an algorithm for recognizing images of schizophrenic and healthy controls. Although the results are comparable to those of other authors, the resulting accuracy is still too low to make the proposed algorithm serve as an objective method for diagnosing this neuropsychiatric disorder.

References

[1] Anutam, Rajni, 2014. Performance Analysis of Image Denoising with Wavelet Thresholding Methods for Different Levels of Decomposition. *The International Journal of Multimedia and Its Applications*, vol. 6(3).

[2] Ashburner, J., Friston, K. J., 2000. Voxel-based morphometry – the methods. *NeuroImage*, vol. 11, p. 805–821.

[3] Castellani, U., Rossato, E., Murino, V., Bellani, M., Rambaldelli, G., Perlini, C., Brambilla, P., 2012. Classification of schizophrenia using feature-based morphometry. *Journal of Neural Transmission*, vol. 119(3), p. 395–404.

[4] Fan, Y., Shen, D., Gur, R. C., Gur, R. E., Davatzikos, C., 2007. COMPARE: Classification of morphological patterns using adaptive regional elements. *IEEE Transactions on Medical Imaging*, vol. 26(1), p. 93–105.

[5] Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., Frackowiak, R. S. J., 1994. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, vol. 2 (4), p. 189–210.

[6] Glahn, D. C., Laird, A. R., Ellison-Wright, I., Thelen, S. M., Robinson, J. L., Lancaster, J. L., Bullmore, E., Fox, P. T., 2008. Meta-Analysis of Gray Matter Anomalies in Schizophrenia: Application of Anatomic Likelihood Estimation and Network Analysis. *Biological Psychiatry*, vol. 64(9), p. 774–781.

[7] Honea, R., Crow, T. J., Passingham, D., Mackay, C. E., 2005. Regional deficits in brain volume in schizophrenia: A meta-analysis of voxel-based morphometry studies. *American Journal of Psychiatry*, vol. 162, p. 2233–2245.

[8] Kašpárek, T., Thomaz, C. E., Sato, J. R., Schwarz, D., Janousova, E., Marecek, R., Prikryl, R., Vanicek, J., Fujita, A., Ceskova, E., 2011. Maximum-uncertainty linear discrimination analysis of first-episode schizophrenia subjects. *Psychiatry Research: Neuroimaging*, vol. 191(3), p. 174–181.

[9] Kumari, S., 2012. Effect of Symlet Filter Order on Denoising of Still Images. *Advanced Computing: An International Journal*, vol. 3(1), p. 137–143.

[10] Kuncheva, L.I., 2004. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons. ISBN 0471210781.

[11] Lieberman, J. A., Tollefson, G. D., Charles, C., Zipursky, R., Sharma, T., Kahn, R. S., Group, H. S., 2005. Antipsychotic drug effects on brain morphology in first-episode psychosis. *Arch Gen Psychiatry*, vol. 62(4), p. 361–370.

[12] Mallat, S., 2007. *A wavelet tour of signal processing: the sparse way*. Elsevier, Boston. ISBN 9780123743701.

[13] Pierson, R., Magnotta, V., 2012. Long-term Antipsychotic Treatment and Brain Volumes: A Longitudinal Study of First-Episode Schizophrenia. *Archives of General Psychiatry*, vol. 68(2), p. 128–137.

[14] Schwarz, D., Kašpárek, T., Provazník, I., Jarkovský, J., 2007. A deformable registration method for automated morphometry of MRI brain images in neuropsychiatric research. *IEEE Transactions on Medical Imaging*, vol. 26, p. 452–461.

[15] Starck, J.-L., Murtagh, F., Fadili, J. M., 2010. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press. ISBN 9780521119139.

[16] Vyškovský, R., Schwarz, D., Janoušová, E., Kašpárek, T., 2016. Random Subspace Ensemble Artificial Neural Networks for First-episode Schizophrenia Classification. *Annals of Computer Science and Information Systems*, vol. 8, p. 317 – 321.

[17] Wang, S., Li, D., Song, X., Wei, Y., Li, H., 2011. A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. *Expert Systems with Applications*, vol. 38(7), p. 8696–8702.

[18] Yoon, J. H., Nguyen, D. V, McVay, L. M., Deramo, P., Michael, J., Ragland, J. D., Niendham, T., Solomon, M., Carter, C. S., 2013. Automated classification of fMRI during cognitive control identifies more severely disorganized subjects with schizophrenia. *Schizophrenia Research*, vol. 135, p. 28–33.

[19] Zhang, H., Ho, T. B., Lin, M., Liang, X., 2006. Feature Extraction for Time Series Classification Using Discriminating Wavelet Coefficients. *Advances in Neural Networks*, vol. 3971, p. 1394–1399.

Kateřina Maršálová
+420 605 766 523
katerina.marsalova299@gmail.com