EasyChair Preprint

№ 10266

Automatic Diagnosis of Autism Using Multilevel Wavelet Decomposition and Support Vector Machine

William Cancino and Said Pertuz

May 25, 2023

Automatic Diagnosis of Autism Using Multilevel Wavelet Decomposition and Support Vector Machine

William Cancino[0000-0002-6042-4345] and Said Pertuz*[0000-0001-8498-9917]

Connectivity and Signal Processing Group, Universidad Industrial de Santander, 680002, Bucaramanga, Colombia
* Corresponding Author: spertuz@uis.edu.co

Abstract. The current diagnosis of autism spectrum disorder (ASD) is very challenging due to the complex symptoms of this disease. Basically, this process is based on purely behavioral observations, which makes it a subjective method that could lead to incorrect diagnoses. To address the problem in question, in this study we propose an approach for the automatic diagnosis of autism based on Multilevel Discrete Wavelet Decomposition (MDWD) and Support Vector Machines (SVM). First, we use resting-state functional magnetic resonance imaging (rs-fMRI) from the Autism Brain Imaging Data Exchange I dataset. From these images, we extract time series of regions of interest defined by a brain atlas. Then, we apply MDWD to these time series and the resulting subseries are used for the construction of functional connectivity (FC) matrices. Finally, the FC feature vector serves as input to the SVM classifier. Our proposed method is evaluated on 175 rs-fMRI sequences. The results show that using MDWD to analyze signals provides a significant improvement in classifier performance. Our best model achieves an accuracy and F1-score of 72.5% and 63.8%, respectively.

Keywords: ASD, MDWD, rs-fMRI, SVM, wavelet

1 Introduction

Autism spectrum disorder (ASD) is a complex and highly heterogeneous neurodevelopmental disorder characterized by deficits in communication and social interaction, restricted activities and interests, and repetitive and stereotyped behavior patterns [1], [2], [3]. According to the World Health Organization, one in 100 children in the world has ASD.

Currently, the gold standard for the diagnosis of autism consists of a behavioral assessment that is supported by tools such as Autism Diagnostic Interview-Revised (ADI-R) [4] and Autism Diagnostic Observation Schedule-2 (ADOS-2) [5]. However, this procedure is subjective and susceptible to inaccuracies, as it depends largely on the specialist's experience and interpretation of the results obtained. In addition, the presence of psychiatric comorbidities in individuals with autism (e.g., depressive disorders, bipolar disorders, anxiety disorders or schizophrenia) can disguise or alter some symptoms [6], [7], which increases

the complexity of detection. In this way, diagnosis is a long and slow process, lacking neurobiological biomarkers [8], which may delay early intervention. In fact, accurate and early detection of ASD is fundamental to implement adequate treatments to improve the patient's condition and quality of life [9].

Recently, machine learning and deep learning methods trained on structural and functional imaging modalities have become attractive for the diagnosis of psychiatric disorders [10], [11], [12], [13]. Precisely, these studies are shown to be promising candidates that attempt to accelerate, improve and reduce subjectivity in the diagnostic process of autism. Within the different types of neuroimaging data, resting-state functional magnetic resonance imaging (rs-fMRI) is increasingly used and is of particular interest to researchers. This modality is a noninvasive technique that measures brain activity through changes in the blood oxygen level-dependent (BOLD) signal when the subject is in a resting state. In particular, rs-fMRI allows the study of functional connectivity, which examines the temporal correlation between BOLD signals from different brain regions. This concept is a valuable tool, as previous studies have reported alterations associated with autism that manifest as reductions or increases in functional connectivity [14].

For the study of rs-fMRI signals in ASD diagnosis, researchers have adopted different approaches, including deep learning (DL). DL algorithms can automatically extract features from the raw data through various levels of abstraction. Due to the multilayer representation of the information contained in the input data, where each successive layer extracts increasingly complex features, these methods have achieved revolutionary results [15], [16]. For this reason, in recent years, DL has been widely used in different fields of science. However, these models are seen as "black box" approaches due to the lack of interpretability, which derives from the inherent complexity in their processes. In this way, a uninterpretable model prevents the extraction of relevant knowledge from the relationships learned by the model itself. Thus, the analysis of data in different frequency bands, which could reveal information of interest, would not be evident by using DL. Precisely because of this problem, we prefer Multilevel Discrete Wavelet Decomposition (MDWD) over DL, since its multiresolution analysis provides an interpretable framework that allows understanding how an unwanted event affects the signal and, subsequently, if required, proposing strategies for its elimination. MDWD is a powerful time-frequency technique used in the analysis of non-stationary signals and allows decomposing an input signal into different levels. Each level contains low and high frequency subseries or contributions.

Thus, in this work we propose a classification framework based on MDWD and Support Vector Machines (SVM). Initially, MDWD is applied to the time series extracted using a 200-region brain atlas. Then, the resulting subseries or coefficients are used in the feature extraction stage, which is based on the construction of a functional connectivity matrix. Finally, the SVM classifier is trained with the acquired FC features. In this way, the main objective of the study is to test how the analysis using MDWD affects the performance of the model for

ASD diagnosis. Accordingly, for all experiments, we used 175 rs-fMRI sequences from the Autism Brain Imaging Data Exchange I dataset, which groups images from multiple international sites.

2 Previous related work

Previous studies have explored the use of the wavelet transform on rs-fMRI to improve ASD classification. For example, in [17] they take advantage of the temporal dynamics features present in scalogram images, which are constructed from Continuous Wavelet Transform (CWT). In the next stage, these images serve as input to 4 pre-trained deep learning frameworks for feature extraction. Finally, the extracted vectors are input to two different classifiers, SVM and K-Nearest Neighbors, reaching an accuracy of up to 85.9%. On the other hand, [13] extracts wavelet coherence maps from the time series of socio-executive resting-state networks. Subsequently, the maps are used to obtain a metric called time of in-phase coherence, which describes in-phase and coherent patterns (synchronicity) between pairs of networks. Precisely, this metric is used to train three classifiers based on SVM and Linear Discriminant Analysis, where they obtain an accuracy of up to 86.7%. Likewise, [18] introduces a multiclass classification oriented to the diagnosis of ASD subtypes. For this purpose, they determine dynamic FC between brain regions using a novel coherence metric. This metric quantifies the global variability of coherence on specific low-frequency scales of BOLD signals. In this way, they develop a classification algorithm based on convolutional neural networks and wavelet coherence maps of the pairwise regions. Thus, this configuration yields an average accuracy of 88.6%.

A point to highlight is that all the mentioned works are mainly based on continuous wavelet transforms (CWTs). However, unlike them, we use MDWD, since this methodology provides a minimally redundant representation of a signal. Therefore, the computational resources invested in the calculation and storage of its coefficients are significantly lower than the algorithms based on CWTs. Likewise, MDWD can capture the most important characteristics in a subset of coefficients much smaller than the original signal. This means that this method tends to concentrate the signals in a few large coefficients, while the noise is usually represented by several low magnitude coefficients. This allows the use of easy-to-implement techniques (e.g. statistical thresholding) to eliminate noise-associated components. Moreover, the coefficients obtained by MDWD tend to become decorrelated as the level of resolution increases [19]. Precisely, the decorrelated data are relevant in machine learning (ML) because it is desired that each coefficient provides unique information to the model. Thus, having many correlated coefficients can reduce the generalization and predictive ability of the ML model, since they would provide the same information of a common underlying pattern.

3 Materials and methods

3.1 Dataset and preprocessing

In our experiments we used rs-fMRI sequences available from the Autism Brain Imaging Data Exchange I (ABIDE I) [20]. ABIDE I is a consortium involving 17 international sites and shares neuroimaging data from 539 subjects with ASD and 573 typical development. Each of these 1112 samples consists of rs-fMRI and structural MRI data, and phenotypic information. In order to avoid confounding due to exogenous variables related to acquisition technologies and the sample population, for this study we considered data from only one site: NYU. This corresponds to 75 patients with ASD and 100 typical development.

In an effort to reduce the effects of unwanted artifacts and other noise, all 175 images were preprocessed using Configurable Pipeline for the Analysis of Connectomes (CPAC). CPAC is a strategy provided by the Preprocessed Connectomes Project (PCP) [21] and involves the following steps: slice time correction, motion realignment, intensity normalization, nuisance signal removal and band pass filtering (0.01 - 0.1 Hz). Furthermore, because the preprocessed data correspond to 4-dimensional sequences, 3 spatial and 1 temporal dimensions, it may be inefficient to analyze each voxel of the image. Instead, we divided the brain into 200 regions of interest (ROIs) defined by the Craddock 200 atlas (CC200). In this way, the time series of all voxels within each ROI are extracted and averaged, thus obtaining a single time series per ROI.

3.2 Multilevel Discrete Wavelet Decomposition

Multilevel Discrete Wavelet Decomposition (MDWD) is a versatile method used in the analysis of non-stationary signals, which allows extracting time-frequency characteristics from time series. MDWD decomposes a signal into low and high frequency subseries level by level. Therefore, to obtain the subseries of the (i+1)-th level, MDWD implements two fundamental steps: extraction of intermediate sequences and down-sampling.

Extraction of intermediate sequences. The sequences A^l_{i+1} and A^h_{i+1} are generated by convolving the low-frequency subband of the i-th level (X^l_i) with a low-pass filter $L = [l_0, l_1, ..., l_{K-1}]$ and a high-pass filter $H = [h_0, h_1, ..., h_{K-1}]$. Thus, in (1) and (2) the n-th element of A^l_{i+1} and A^h_{i+1} is computed. Additionally, the special case X^l_0 refers to the input time series (X).

$$A^l_{i+1}[n] = \sum_{k=0}^{K-1} X^l_i[k] \cdot L[n-k] \tag{1}$$

$$A^h_{i+1}[n] = \sum_{k=0}^{K-1} X^l_i[k] \cdot H[n-k] \tag{2}$$

Down-sampling. The low-frequency X_{i+1}^{l} and high-frequency X_{i+1}^{h} subseries are obtained by applying a decimation function to the A_{i+1}^{l} and A_{i+1}^{h} sequences, respectively. This process is described by (3) and (4), where $\downarrow 2(\cdot)$ is the operator that performs down-sampling by a factor of 2.

$$X_{i+1}^{l} = \downarrow 2(A_{i+1}^{l}) \tag{3}$$

$$X_{i+1}^{h} = \downarrow 2(A_{i+1}^{h}) \tag{4}$$

The result of MDWD is a set $S = [X_1^h, X_2^h, ..., X_j^h, X_j^l]$ of $j+1$ subseries, where j is the number of levels into which the time series is decomposed. Precisely, the levels used represent the signal viewed at different scales. Finally, in this work we select two decomposition levels ($j = 2$) and use Daubechies 2 to define the coefficients of the L and H filters. Furthermore, for each subject we obtain 200 subseries in the first level and 400 subseries in the second level (200 low frequency and 200 high frequency).

3.3 Feature extraction

In this work, we use two levels for the MDWD. As a result, 3 independent functional connectivity matrices are generated for each subject: one for the first level and two for the second level of the MDWD representation of the signal. The FC matrices are of size N x N (N = 200) and are constructed by calculating Pearson's correlation coefficients for each pair of time subseries. The resulting matrices are symmetric and contain coefficients ranging from 1 (two highly correlated time subseries) to -1 (two anti-correlated time subseries). Subsequently, the elements of the upper diagonal of the FC matrices are discarded, as they are repeated with those of the lower diagonal. Finally, the retained data are reduced to a one-dimensional feature vector of length L, given by $L = 0.5N(N-1)$. In the same way, the feature extraction process is presented in Figure 1.

3.4 Support Vector Machine Classifier

Support Vector Machines (SVM) are a set of supervised machine learning algorithms [22]. SVM develops the classification task by finding a hyperplane that optimally separates the data into two classes. This process is performed with the help of support vectors, which are the points in each class that are closest to the hyperplane and influence its orientation and position. Likewise, when the problems are complex, the solutions may require nonlinear hyperplanes. In that case, the original samples of the data set are mapped to a higher dimensional space by a kernel function so that they are now linearly separable. Therefore, in this work we consider the radial basis function (RBF) kernel.

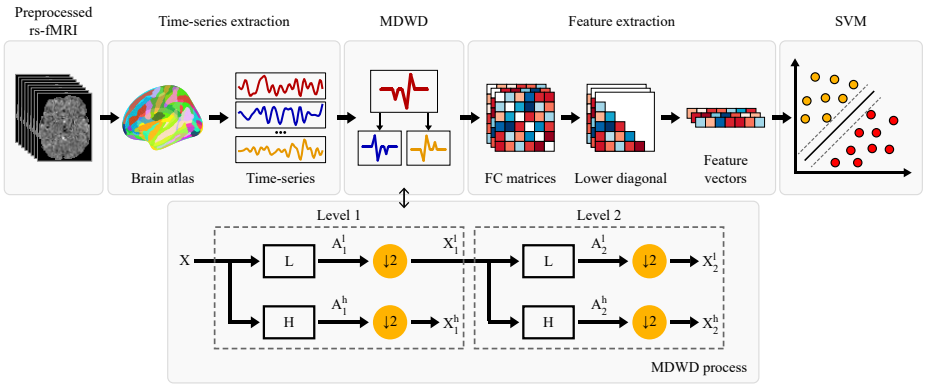

Fig. 1: In the illustration, the successive stages in the top line describe the proposed workflow for ASD classification. The overall process starts from the preprocessed rs-fMRI data and involves intermediate steps such as time series extraction, MDWD and feature extraction. Finally, the obtained vectors serve as input to the SVM classifier. Also, the MDWD process is presented in detail at the bottom, which decomposes the input signal into two levels.

3.5 Performance assessment

For the evaluation of the model we used 5-fold cross validation. Consequently, the data set is divided into 5 folds of the same size, i.e., each containing 35 subjects: 15 with ASD and 20 with typical development (TD). In this way, during the evaluation, a single fold is retained to serve as validation data, and the remaining 4 folds are used as training data. This process is repeated 5 times and each fold is used exactly once to evaluate the model. Similarly, the grid-search method is implemented to find the hyperparameters with which the model achieves the best performance. Therefore, the hyperparameters used for this process correspond to C and gamma with their values being established by random search within the intervals $[10^{-2}, 10]$ and $[10^{-5}, 10]$, respectively. The search is performed at each fold using only the training data and 10-fold cross validation.

Likewise, 5 metrics are used to measure classifier performance: accuracy (A), precision (P), F1-score ($F1$), recall (R) and specificity (S). These measures are calculated following the equations (5)-(9), where TP represents the true positives (patients correctly classified with ASD), TN the true negatives (patients correctly classified as TD or without ASD), FP the false positives (patients incorrectly classified with ASD), and FN the false negatives (patients incorrectly classified as TD).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = 2\left(\frac{PR}{P + R}\right) \tag{8}$$

$$S = \frac{TN}{TN + FP} \tag{9}$$

Accuracy measures the ability of a classifier to correctly identify all samples, regardless of whether they are positive or negative. Precision corresponds to the ratio of the number of positive samples correctly predicted to the total number of positive samples predicted. Recall or sensitivity represents the portion of ASD patients correctly classified as ASD. $F1$-score is the harmonic mean of precision and recall. Therefore, this measure considers the impact of false positives and false negatives on the classifier. Thus, the higher the precision and recall, the higher the $F1$-score. Specificity measures the ability of the classifier to correctly determine TD patients.

4 Results

In this work, we want to assess whether the discrimination of ASD from fMRI data has relevant information at different frequency bands. For this reason, we compare the behavior of the model when using or not using the MDWD block. Specifically, for the case where MDWD is added, we explore two options. The first option consists of using data from all decomposition levels. This involves training three different SVM classifiers and then implementing voting classifier (majority voting). The second option considers only the data associated with the low frequency subband of the last level (X_2^l). Therefore, Table 1 shows the results of our experiments, where the highest values obtained in each measure are highlighted in bold. We can observe that, for all measurements, the best performances are obtained when using MDWD. This suggests the importance of applying such an algorithm as a method of time series analysis. Also, adding MDWD to our pipeline yields an improvement of up to 5.1% in terms of accuracy. Similarly, the classifier trained with all levels (All) achieves a precision of up to 79.6%. On the other hand, evaluating the performance differences between using and not using MDWD, we find that improvements range from 2.7% to 12.1%. Finally, we consider that the best model is the one that uses the data associated with X_2^l, since it has the best results in terms of accuracy, $F1$-score and recall.

5 Conclusions

In this study, we proposed an approach based on Multilevel Discrete Wavelet Decomposition and SVM for ASD classification. We used 175 rs-fMRI sequences

Table 1: Results obtained in terms of different performance measures when using and not using MDWD.

MDWD	**Levels**	**Measures**				
		A	***P***	***F1***	***R***	***S***
No	-	0.674	0.691	0.536	0.439	0.850
Yes	All	0.708	**0.796**	0.580	0.466	**0.890**
	X_2^l	**0.725**	0.747	**0.638**	**0.560**	0.850

from the ABIDE I dataset and evaluated the model by 5-fold cross validation. Experimental results evidence a significant improvement in different performance measures when applying MDWD on time series. Moreover, in our work, the best performing model was the one that considered only the low frequency subband of the last decomposition level. This suggests that analyzing the signal in different frequency bands may unveil relevant information that improves classification performance. Similarly, our approach may be of particular interest for frameworks that rely on traditional machine learning algorithms, as in many occasions the performance of these methods is limited by the complexity of the raw data. The implemented code is available at https://github.com/WilliamCancino/wavelet-autism-classification.git

References

[1] *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, USA: American Psychiatric Association, 2013.

[2] Tyler C. McFayden et al. "Sex Differences in an Autism Spectrum Disorder Diagnosis: Are Restricted Repetitive Behaviors and Interests the Key?" In: *Review Journal of Autism and Developmental Disorders* 7.2 (Aug. 2019), pp. 119–126. DOI: 10.1007/s40489-019-00183-w. URL: https://doi.org/10.1007/s40489-019-00183-w.

[3] Suma Jacob et al. "Neurodevelopmental heterogeneity and computational approaches for understanding autism." In: *Translational Psychiatry* 9.1 (Feb. 2019). DOI: 10.1038/s41398-019-0390-0. URL: https://doi.org/10.1038/s41398-019-0390-0.

[4] Michael Rutter, A Le Couteur, and Catherine Lord. "Autism diagnostic interview-revised." In: *Los Angeles, CA: Western Psychological Services* 29.2003 (2003), p. 30.

[5] Catherine Lord et al. "Autism diagnostic observation schedule, second edition (ADOS–2)." In: *Los Angeles, CA: Western Psychological Corporation* (2012).

[6] Yvette Hus and Osnat Segal. "Challenges Surrounding the Diagnosis of Autism in Children." In: *Neuropsychiatric Disease and Treatment* Volume 17 (Dec. 2021), pp. 3509–3529. DOI: 10.2147/ndt.s282569. URL: https://doi.org/10.2147/ndt.s282569.

[7] Laura Fusar-Poli et al. "Missed diagnoses and misdiagnoses of adults with autism spectrum disorder." In: *European Archives of Psychiatry and Clinical Neuroscience* 272.2 (Sept. 2020), pp. 187–198. DOI: 10.1007/s00406-020-01189-w. URL: https://doi.org/10.1007/s00406-020-01189-w.

[8] Xingdan Liu and Huifang Huang. "Alterations of functional connectivities associated with autism spectrum disorder symptom severity: a multi-site study using multivariate pattern analysis." In: *Scientific Reports* 10.1 (Mar. 2020). DOI: 10.1038/s41598-020-60702-2. URL: https://doi.org/10.1038/s41598-020-60702-2.

[9] Stephen N. James and Christopher J. Smith. "Early Autism Diagnosis in the Primary Care Setting." In: *Seminars in Pediatric Neurology* 35 (2020), p. 100827. ISSN: 1071-9091. DOI: https://doi.org/10.1016/j.spen.2020.100827. URL: https://www.sciencedirect.com/science/article/pii/S1071909120300383.

[10] Jihoon Oh et al. "Identifying Schizophrenia Using Structural MRI With a Deep Learning Algorithm." In: *Frontiers in Psychiatry* 11 (Feb. 2020). DOI: 10.3389/fpsyt.2020.00016. URL: https://doi.org/10.3389/fpsyt.2020.00016.

[11] Atif Riaz et al. "DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI." In: *Journal of Neuroscience Methods* 335 (Apr. 2020), p. 108506. DOI: 10.1016/j.jneumeth.2019.108506. URL: https://doi.org/10.1016/j.jneumeth.2019.108506.

[12] Md Rishad Ahmed et al. "Single Volume Image Generator and Deep Learning-Based ASD Classification." In: *IEEE Journal of Biomedical and Health Informatics* 24.11 (2020), pp. 3044–3054. DOI: 10.1109/JBHI.2020.2998603.

[13] Antoine Bernas, Albert P. Aldenkamp, and Svitlana Zinger. "Wavelet coherence-based classifier: A resting-state functional MRI study on neurodynamics in adolescents with high-functioning autism." In: *Computer Methods and Programs in Biomedicine* 154 (2018), pp. 143–151. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2017.11.017. URL: https://www.sciencedirect.com/science/article/pii/S0169260717303097.

[14] Oualid Benkarim et al. "Connectivity alterations in autism reflect functional idiosyncrasy." In: *Communications Biology* 4.1 (Sept. 2021). DOI: 10.1038/s42003-021-02572-6. URL: https://doi.org/10.1038/s42003-021-02572-6.

[15] Ahmed Sedik et al. "Efficient deep learning approach for augmented detection of Coronavirus disease." In: *Neural Computing and Applications* (Jan. 2021). DOI: 10.1007/s00521-020-05410-8. URL: https://doi.org/10.1007/s00521-020-05410-8.

[16] Xiao-Ling Zou et al. "A promising approach for screening pulmonary hypertension based on frontal chest radiographs using deep learning: A

retrospective study." In: *PLOS ONE* 15.7 (July 2020). Ed. by Jie Zhang, e0236378. DOI: 10.1371/journal.pone.0236378. URL: https://doi.org/10.1371/journal.pone.0236378.

[17] Mohammed I. Al-Hiyali et al. "Classification of BOLD FMRI Signals using Wavelet Transform and Transfer Learning for Detection of Autism Spectrum Disorder." In: *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (2021), pp. 94–98. DOI: 10.1109/IECBES48179.2021.9398803.

[18] Mohammed Isam Al-Hiyali et al. "Classification of ASD Subtypes Based on Coherence Features of BOLD Resting-state fMRI Signals." In: *2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)* (2021), pp. 17–22. DOI: 10.1109/ICICyTA53712.2021.9689092.

[19] Abdourrahmane M. Atto, Dominique Pastor, and Alexandru Isar. "On the statistical decorrelation of the wavelet packet coefficients of a band-limited wide-sense stationary random process." In: *Signal Processing* 87.10 (Oct. 2007), pp. 2320–2335. DOI: 10.1016/j.sigpro.2007.03.014. URL: https://doi.org/10.1016/j.sigpro.2007.03.014.

[20] A. Di Martino and S. Mostofsky. *ABIDE I.* [Online]. Available: http://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html. 2016.

[21] Craddock Cameron et al. "The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives." In: *Frontiers in Neuroinformatics* 7 (2013). DOI: 10.3389/conf.fninf.2013.09.00041. URL: https://doi.org/10.3389/conf.fninf.2013.09.00041.

[22] Robert Gove and Jorge Faytong. "Machine Learning and Event-Based Software Testing: Classifiers for Identifying Infeasible GUI Event Sequences." In: *Advances in Computers* (2012), pp. 109–135. DOI: 10.1016/b978-0-12-396535-6.00004-1. URL: https://doi.org/10.1016/b978-0-12-396535-6.00004-1.