Supplementary Materials

Informative Biomarkers for Autism Spectrum Disorder Diagnosis in functional Magnetic Resonance Imaging Data on the Default Mode Network

Additional Analysis of the Extracted Features

As mentioned in the main manuscript, the features extracted included biological attributes (sFC, mean dFC and variance dFC) and non-physiological measurements (handedness, age, acquisition's protocol and movement parameters). Subsequently, the FS and classification approach utilized succeeded in producing a small feature subset employing 136 features, obtaining 76.63% accuracy and 82.74% AUC. Interestingly, the features selected included almost entirely FC characteristics (135 out of 136) with the exception of 1 non-biological feature (TE). In this regard, one could assume that the incorporation of only FC or FC along with acquisition features could enhance the obtained performance, if not in classification accuracy indubitably in computational cost (due to the smaller feature set that need to be fed into the FS process). On this premise, additional analysis was performed excluding the aforementioned unused features (one-by one and altogether), results of which are presented in Table S1 below.

Table S1. Additional analysis classification results

Features excluded	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
No exclusion	**76.63**	**78.42**	**74.27**	**82.74**
All unused features	75.81	80.43	70.12	83.63
Movement	75.47	77.13	73.86	82.71
Age	75.35	76.58	74.08	82.85
Handedness	74.65	74.18	75.36	79.79
Acquisition	73.37	76.85	69.62	78.69

Note: Results of the method employed in the main manuscript are presented in **bold**.

Interestingly, the exclusion of the different feature combinations provided lower classification performance from the framework presented in the main manuscript. However, all the additional analysis results illustrated better performance than previous related works (as indicated from Table 2 in the main manuscript). Of note is that all methodological approaches presented in Table S1, included 65-82% common features with the main method. This highlights the importance of static and dynamic FC as indicative biomarkers for ASD discrimination.

Furthermore, the unexpected deterioration in performance can be contributed to the nature of the FS process with regard to correlation bias. In fact, RFE-CBR estimates each feature and removes them one-by-one from the overall set to generate the ranked feature arrangement. Moreover, in every repetition, each feature is additionally estimated based on an additional correlation procedure. From this standpoint however, some less informative features (placed near the bottom of the ranked feature set) might be correlated to highly distinguishing ones. Thus, a smaller overall set could present higher chance of feature correlations. On the other hand, several additional approaches attained slightly higher AUC than the proposed approach, indicating a better model adoption, albeit with lower performance. Lastly, the exclusion of the hardware parameters during acquisition resulted in the larger deterioration of the overall additional analysis performance. This fact implies the effectiveness of the acquisition features to distinguish between ASD and TD individuals as it provides indices to better differentiate heterogeneous data from the various sites, although to a small degree.