



Article

Exploring Autism Spectrum Disorder: A Comparative Study of Traditional Classifiers and Deep Learning Classifiers to Analyze Functional Connectivity Measures from a Multicenter Dataset

Francesca Mainas, Bruno Golosio, Alessandra Retico and Piernicola Oliva

Special Issue

Editorial Board Members' Collection Series: Applied Affective and Cognitive Neuroscience

Edited by

Prof. Dr. Peter Walla and Prof. Dr. Alexander N. Pisarchik



Article

Exploring Autism Spectrum Disorder: A Comparative Study of Traditional Classifiers and Deep Learning Classifiers to Analyze Functional Connectivity Measures from a Multicenter Dataset

Francesca Mainas ^{1,2,3,*}, Bruno Golosio ^{1,2} , Alessandra Retico ⁴  and Piernicola Oliva ^{2,5} ¹ Department of Physics, University of Cagliari, 09042 Monserrato, Italy² National Institute for Nuclear Physics (INFN), Cagliari Division, 09042 Monserrato, Italy; oliva@uniss.it³ Department of Informatic, University of Pisa, 56127 Pisa, Italy⁴ National Institute for Nuclear Physics (INFN), Pisa Division, 56127 Pisa, Italy; alessandra.retico@pi.infn.it⁵ Department of Chemical, Physical, Mathematical and Natural Sciences, University of Sassari, 07100 Sassari, Italy

* Correspondence: francesca.mainas@ca.infn.it

Abstract: The investigation of functional magnetic resonance imaging (fMRI) data with traditional machine learning (ML) and deep learning (DL) classifiers has been widely used to study autism spectrum disorders (ASDs). This condition is characterized by symptoms that affect the individual's behavioral aspects and social relationships. Early diagnosis is crucial for intervention, but the complexity of ASD poses challenges for the development of effective treatments. This study compares traditional ML and DL classifiers in the analysis of tabular data, in particular, functional connectivity measures obtained from the time series of a public multicenter dataset, and evaluates whether the features that contribute most to the classification task vary depending on the classifier used. Specifically, Support Vector Machine (SVM) classifiers, with both linear and radial basis function (RBF) kernels, and Extreme Gradient Boosting (XGBoost) classifiers are compared against the TabNet classifier (a DL architecture customized for tabular data analysis) and a Multi Layer Perceptron (MLP). The findings suggest that DL classifiers may not be optimal for the type of data analyzed, as their performance trails behind that of standard classifiers. Among the latter, SVMs outperform the other classifiers with an AUC of around 75%, whereas the best performances of TabNet and MLP reach 65% and 71% at most, respectively. Furthermore, the analysis of the feature importance showed that the brain regions that contribute the most to the classification task are those primarily responsible for sensory and spatial perception, as well as attention modulation, which is known to be altered in ASDs.

Keywords: ABIDE; multi-site data; deep learning; machine learning; autism spectrum disorders



Citation: Mainas, F.; Golosio, B.; Retico, A.; Oliva, P. Exploring Autism Spectrum Disorder: A Comparative Study of Traditional Classifiers and Deep Learning Classifiers to Analyze Functional Connectivity Measures from a Multicenter Dataset. *Appl. Sci.* **2024**, *14*, 7632. <https://doi.org/10.3390/app14177632>

Academic Editors: Alexander N. Pisarchik, Qi-Huang Zheng and Peter Walla

Received: 4 July 2024

Revised: 5 August 2024

Accepted: 15 August 2024

Published: 29 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Autism spectrum disorders (ASDs) are a group of neurodevelopmental conditions characterized by repetitive and stereotyped behaviors as well as a deficit in social communication and interaction [1]. It affects approximately 1 child out of every 59, with a stronger prevalence among males (1 in 37) than females (1 in 51) [2]. Currently, the diagnosis of ASD is based on behavioral criteria, which requires a team of specialists, a process that can be time-consuming and may not always yield a definitive result due to factors such as comorbidity [3,4]. The heterogeneous nature of this condition requires continued study across various fields, leading to constant updates of the diagnostic criteria [5,6]. Early diagnosis and intervention are crucial for improving the quality of life and developing effective intervention strategies [7]. To date, numerous studies have focused on analyzing brain images acquired using resting-state functional magnetic resonance imaging (rs-fMRI). This non-invasive imaging technique involves acquiring functional magnetic resonance

images of the brain while the patient is at rest, without performing specific tasks. Rs-fMRI is often employed in the investigation of brain functional connectivity, which refers to the study of the correlation between the temporal signals of two anatomically distinct brain regions. By assuming that functional connectivity is a phenomenon involving interactions that occur on time scales shorter than acquisition times, it is possible to evaluate the correlation between the temporal signals of two anatomically distinct brain regions, taking into account the entire observation time interval, second by second. This process, repeated for all brain areas, allows us to quantify the functional connection between the different brain areas. These measures can be used to identify potential neurological distinctions between typically developing (TD) individuals and those with ASD. Given the abundance of data in neuroimaging, machine learning (ML) and deep learning (DL) classifiers have been widely employed to try to predict the ASD condition [8,9]. Neuroscientists typically employ both traditional machine learning classifiers, such as Support Vector Machines, [10] or random forests, [11,12], and DL classifiers, like convolutional neural networks [13,14], for classification purposes. Deep neural networks have achieved significant success in various fields, including image and text processing [15,16]. In practical applications, tabular data are the most common data type, particularly in medicine. Recent studies have shown that deep learning-based methods can have a crucial role in the diagnosis of ASD [17,18]. Over the past decade, traditional ML classifiers have remained dominant when dealing with tabular data and frequently achieved better performance than DL classifiers. Usually, ML classifiers are simpler than DL classifiers, and therefore, it is generally easier to understand and interpret their response. On the other hand, the complexity of DL and their lack of transparency and interpretability [19], limits their applicability in clinical contexts.

In this work, we investigated different ML and DL classifiers to show the differences in classification performances and in most important features involved in the classification of ASD subjects vs. TD ones, using tabular data derived from functional connectivity.

2. Materials and Methods

2.1. Data Selection

For this study, we used rs-fMRI data from the multicenter ABIDE dataset [20]. The dataset comprises two collections: ABIDE I and ABIDE II. ABIDE I includes scans of 1112 subjects, evenly distributed between ASD and TD, collected from 17 different sites. ABIDE II includes scans of 1114 subjects, evenly distributed between ASD and TD, collected from 19 different sites. In addition to the scan images, ABIDE also provides phenotypic information, such as age, sex, eye status at scan, and additional clinical information.

Although some sites in ABIDE II may be the same as those in ABIDE I, the pipeline and acquisition parameters may have been modified between the two datasets. For this reason, they will be considered as distinct acquisition sites. Furthermore, even within a single collection, such as ABIDE II, there are sites that have released two different data samples. For this reason, some of these samples are labeled with a subscription number (e.g., 1 or 2). Subjects from the ABIDE II collection will have the prefix “ABIDE II” before the site name. If this prefix is missing, the collection belongs to ABIDE I.

We selected 1001 male subjects, aged between 5 and 40 years, with eyes open during acquisition. The data of these subjects are collected from 23 different sites. Male subjects were chosen because the male sample is larger than the females one since males have a higher probability of being affected by ASD than females [21]. Moreover, the female dataset was insufficiently populated to allow for statistically significant studies. The choice of the condition with open eyes was made to avoid including subjects who were potentially sleeping during the examination. The dataset is evenly divided between ASD and TD, containing 506 TD and 495 ASD subjects. Figure 1 shows the distribution of ASD/TD for each site and Table 1 provides information on age distribution for each of the 23 sites.

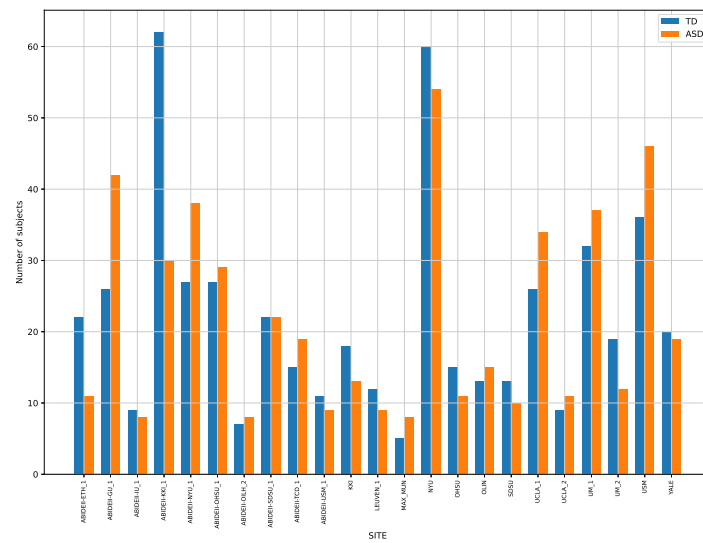


Figure 1. Dataset composition. Sites without a prefix belong to the ABIDE I collection.

Table 1. Information about the mean $\pm$ standard deviation, the minimum and the maximum value of age, in years, across each of the 23 considered sites.

SITE	Mean_Age $\pm$ Standard_Deviation	Min_Age	Max_Age
ABIDEII-TCD_1	15 $\pm$ 3	10	20
ABIDEII-SDSU_1	13 $\pm$ 3	8	18
ABIDEII-GU_1	11 $\pm$ 2	8	14
ABIDEII-NYU_1	9 $\pm$ 4	5	27
ABIDEII-OHSU_1	11 $\pm$ 2	7	15
ABIDEII-USM_1	21 $\pm$ 7	9	36
ABIDEII-IU_1	23 $\pm$ 5	17	34
ABIDEII-KKI_1	10 $\pm$ 1	8	13
ABIDEII-ETH_1	23 $\pm$ 4	14	31
ABIDEII-OILH_2	23 $\pm$ 3	18	28
YALE	12 $\pm$ 3	7	18
USM	21 $\pm$ 7	9	39
OLIN	16 $\pm$ 3	10	23
NYU	15 $\pm$ 6	7	39
UM_2	16 $\pm$ 3	13	29
UCLA_2	12 $\pm$ 1	10	15
UM_1	13 $\pm$ 3	8	19
SDSU	14 $\pm$ 1	12	17
KKI	10 $\pm$ 1	8	13
UCLA_1	13 $\pm$ 2	9	18
MAX_MUN	10 $\pm$ 2	7	13
LEUVEN_1	23 $\pm$ 3	18	32
OHSU	10 $\pm$ 2	8	14

2.2. Feature Generation

The data selected for this study were processed using the Configurable Pipeline for the Analysis of Connectome (CPAC) pipeline [22]. CPAC applies filters to remove noise from respiration, heart rate, movements of the subjects' heads, and other smoothing techniques. CPAC is among the most frequently used pipelines, and previous research has found that it leads to better ASD/TD classification compared to other pipelines, [23]. CPAC also provides the time series of brain areas of interest (ROIs) for patients. In this work, we used the Harvard–Oxford anatomical atlas [24], which consists of 110 ROIs. Figure 2 displays a horizontal section of the Harvard–Oxford atlas partition.

In neuroimaging, Pearson correlation analysis is used to determine the potential correlation between the instantaneous variation in the activation state of different brain regions and their involvement in carrying out a specific function. The values, or coefficients, of Pearson correlation r_{xy} are defined as follows:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2)(\sum_{i=1}^n (y_i - \bar{y})^2)}} \quad (1)$$

where x and y represent the time series of two brain regions and n is their dimension (number of time points) [25]. Pearson coefficients were normalized using the Fisher transformation (2) to make them statistically more significant [26].

$$Z = \frac{1}{2} \sqrt{n-3} \ln \left(\frac{1+r}{1-r} \right) \quad (2)$$

In Equation (2), n represents the number of time points of the time series and r indicates the Pearson coefficient calculated with Equation (1). The Pearson coefficients were used as features for classification. The number of features depends on the atlas used. For instance, for N regions, $\frac{N(N-1)}{2}$ features are obtained. This is because calculating the Pearson correlation between the time series of each of the regions of an atlas generates a square and symmetric connectivity matrix (see Equation (1), which is invariant under the interchange of time series indices); thus, the upper triangle of the matrix contains the only elements of interest. The use of the Harvard–Oxford atlas, composed of 110 regions, leads to $\frac{110(110-1)}{2} = 5995$ features.

For this study, out of the initial 110 ROIs, 7 were excluded due to having null time series in a substantial number of patients. As a result, 103 ROIs for each patient were used. The correlation was then computed for each pair of brain areas, resulting in $N_{comb} = \frac{103(103-1)}{2} = 5253$ independent combinations of connectivity features for each subject.

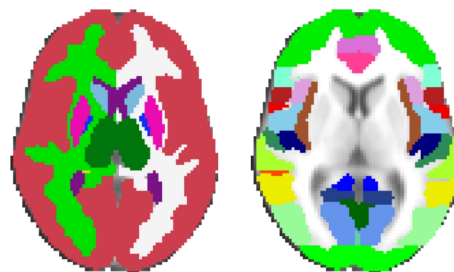


Figure 2. Horizontal section of Harvard–Oxford subcortical (**left**) and cortical (**right**) partitions.

2.3. Harmonization Procedure

Given that large datasets can be obtained by collecting images from different centers, this results in a heterogeneity of data due to differences in scanners and/or acquisition protocols that requires a harmonization technique to minimize these differences. In this work, we used the Neuroharmonize tool proposed by Pomponio et al. [27,28], which is

based on Fortin et al.'s ComBat [29,30]. Neuroharmonize aims to eliminate the site effect while preserving the dependence of the features on biologically significant covariates, such as age and sex.

According to Serra et al. [31], to avoid bias due to data leakage, the harmonization parameters were estimated using only the subjects belonging to the control group of the training set. Once the set of covariates is defined, the harmonization model is computed. In this work, we used age and site as covariates. Subsequently, the model is used to harmonize both the train and test sets. In a cross-validation scheme, the procedure is repeated for each fold separately.

2.4. Classification Strategy

For this work, traditional classifiers, such as Support Vector Machine with a linear kernel (L-SVM), Support Vector Machine with a Gaussian kernel (SVM-RBF) and Extreme Gradient Boosting (XGBoost) were chosen, as well as deep classifiers like Attentive Interpretable Tabular Learning (TabNet) and Multi Layer Perceptron (MLP). SVM classifiers are the most commonly used classifiers in these classification problems. They have demonstrated superior performance compared to other classifiers, particularly in scenarios with a small number of samples and a large number of features [32]. XGBoost belongs to the family of the three classifiers, which has proven to be particularly effective for tabular data classification problems [33]. Furthermore, this classifier has better generalization capability and is less susceptible to overfitting than other classifiers. TabNet is a deep learning classifier mostly used when dealing with tabular data; it employs sequential attention to select the most relevant features for reasoning at each decision step, enhancing interpretability and optimizing learning efficiency by focusing the learning capacity on the most significant features [34]. MLP consists of fully connected layers where every node of each adjacent layer is connected; it is a classifier that is easy to implement, fast and has shown performances that outperform other classifiers [35]. The L-SVM and SVM-RBF classifiers were implemented using the `sklearn.svm.SVC` module from the Python library `scikit-learn` [36,37].

For L-SVM and SVM-RBF, we specified only the kernel type, respectively, *linear* and *rbf*, leaving all the other parameters with the default values. For XGBoost, we used the `XGBoostClassifier` from the `xgboost` package in Python (V. 3.9.12) [38].

The hyperparameters used for the tuning were `max_depth`: [2,12], `min_child_weight`: [1,61], `eta`: [0.1,1]. All the other hyperparameters have been left at their default value. The TabNet classifier was implemented with the `TabNetClassifier` from the PyTorch library `pytorch_tabnet.tab_model` [39] and the MLP was implemented using the `MLPClassifier` from `sklearn.neural_network` package [40]. For TabNet's hyperparameter tuning, we used `n_d`: [8,64], `n_a`: [8,64], `n_steps`: [1,10], `gamma`: [1,2], `momentum`: [0.9,1] with `max_epochs` = 300, `batch_size` = 64, `patience` = 10, `learning-rate`: 2×10^{-2} and Adam optimizer. All the other hyperparameters have been left at their default values. For the MLP, we used `hidden_layer_size` = (128, 64, 64, 32), `activation` = `relu`, `batch_size` = 32, `max_iter` = 300, and `solver` = `adam`, leaving all the other parameters set at default values.

We used the calculated features for each subject, along with the class labels +1 for ASD subjects and −1 for TD subjects, for the classification task. We applied the Scikit-learn's `RobustScaler` function for feature scaling for each classifier, within the classification step, and we conducted hyperparameter tuning for XGBoost and TabNet classifiers. The classification outcomes were obtained using the repeated stratified k-fold cross-validation, with 5 folds and 10 repetitions. To evaluate the classification performances, we used the area under the ROC curve (AUC) [41,42], calculating it for each fold and repetition. The final results are presented as the mean of the AUC with the associated standard deviation as error.

Considering the higher number of features compared to the number of samples, which increases the complexity of the analysis and the risk of overfitting, we evaluated the classifier

performances both with and without Principal Component Analysis (PCA). We varied the number of principal components (PCs) from 30 to 300 PCs (30, 50, 100, 200, 300).

2.5. Feature Importance

Understanding the most important features that contribute to the classification of ASD and TD individuals is crucial for advancing diagnostic and therapeutic strategies. To determine the most discriminative pairs of regions for distinguishing TD from ASD, the permutation importance [43] technique was used for each analyzed classifier, since it can be applied uniformly to all of the classifiers tested.

Permutation importance is generally useful for understanding data and interpreting classifier: by calculating the score for each feature, one can determine which features most influenced the utilized classifier. Permutation importance is considered as one of the global Explainable Artificial Intelligence (XAI) methods. It provides insights into the overall behavior of a classifier and offers a comprehensive view of feature contributions across the entire dataset. Using a global XAI approach, the interpretability and reliability of the classifier are increased. The basic concept of permutation importance involves observing how much a particular metric decreases when a feature is not available, with the score representing the importance of each feature. A higher score indicates that the feature in question has a greater impact on the classifier. In principle, one could remove features, retrain the classifier and check the score. However, this approach can be computationally complex because it would require retraining the classifier for each feature removal. Additionally, this method demonstrates which features might be important in the dataset rather than which features are important for the classifier. If a feature is replaced with noise, derived from the same distribution, as the original feature values during each permutation, it is possible to avoid retraining the classifier. The simplest way to derive this noise is by shuffling the values for a feature by using the values of the same feature across different examples.

In this study, we implemented the permutation importance as described above to determine whether the key features for the classification vary depending on the classifier used. We used the feature permutation importance implemented in the ELI5 python library [44]. This library offers a function that takes into account a trained classifier, a validation dataset, and a scoring metric, and it returns the importance score for each feature. The importance score reflects the decrease in the classifier's performance: the greater the drop in performance when a feature is shuffled, the more significant that feature will be considered. When a feature has a negative score, it means that the performance of the classifier has increased when this feature is replaced with noise, meaning that it is not important. We employed the AUC as a scoring metric and computed the permutation importance for each fold of the 5-fold cross-validation and repetition. The final results were obtained as the average importance score across the folds and repetitions.

3. Results

3.1. Classification Performances

In Figure 3, the classification performances in discriminating subjects with ASD from controls are illustrated. The results are reported for each classifier analyzed (TabNet, MLP, XGBoost, L-SVM, and SVM-RBF), and for different amounts of retained PCs. The best classification results were obtained by the SVM-RBF classifier, which achieved an AUC of 0.75 ± 0.03 (with 100 PCs), followed by L-SVM with an AUC of 0.74 ± 0.02 (with 50 and 100 PCs). As for the DL classifiers, the classification results fall behind for MLP, with AUC = 0.71 ± 0.02 (with 200 PCs and without PCA), and for TabNet, with AUC = 0.65 ± 0.02 (no PCA). These results are shown in Table 2.

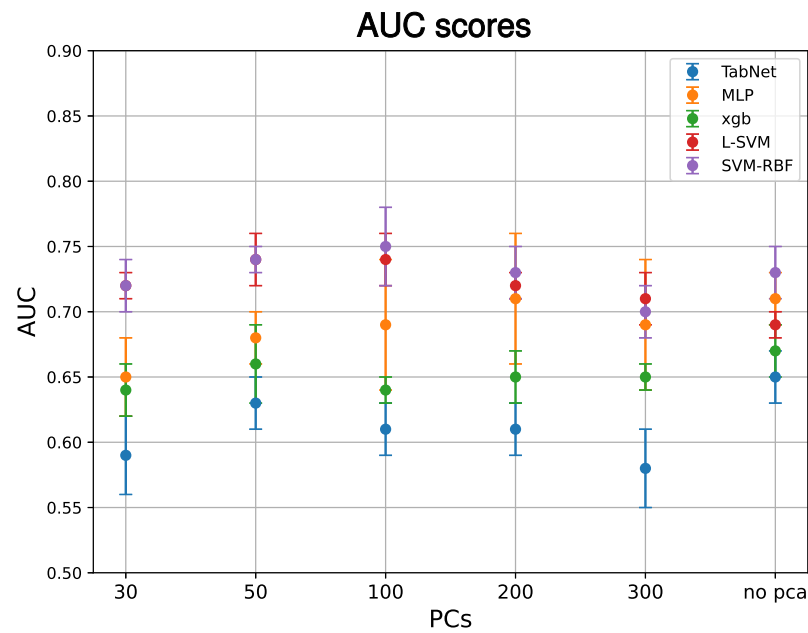


Figure 3. The ASD and TD classification results are reported for each classifier considered and for different values of PCs. The blue, orange, green, red and purple colors represent the different classifiers (TabNet, MLP, XGBoost, L-SVM and SVM-RBF, respectively). Dots and error bars represent the average AUC score and standard deviation, respectively. The average and standard deviation are calculated across the 5 folds and 10 repetitions of the repeated stratified k-fold cross-validation scheme.

Table 2. Best classification performances for each classifier.

Classifier	AUC	# of PCs
MLP	0.71 ± 0.02	no PCA
	0.71 ± 0.05	200 PCs
TabNet	0.65 ± 0.02	no PCA
XGBoost	0.67 ± 0.02	no PCA
L-SVM	0.74 ± 0.02	50 PCs
	0.74 ± 0.05	100 PCs
SVM-RBF	0.75 ± 0.03	100 PCs

3.2. Feature Importance

Identifying the key features that distinguish ASD from TD subjects is essential for understanding ASD. Given that these features assess the correlation between the temporal signals of ROIs, they provide valuable insight into which aspects most significantly impact the distinction between ASD and TD subjects.

As we only obtained 59 features with a positive score (greater than zero) using the XGBoost classifier, we selected the top 50 features with the highest scores for each classifier to compare which regions were most significant in discriminating ASD/TD across all the analyzed classifiers. Subsequently, we checked for common features among all the top 50 features. From this analysis, we did not find any common feature for all the classifiers but only some features that were in common between two or three of them.

However, despite the lack of global common features, we looked for brain region occurrences in the 50 most relevant correlations by counting the number of times these regions were present in all classifiers. This allowed us to observe which regions had the most significant effect on ASD/TD classification based on their connectivity to other regions. The results are shown in Table 3. Consistent regions can be identified in all classifiers. These regions are those whose correlation with other regions was most significant

in discriminating between ASD and TD. We also examined the location of these regions in the functional networks of the Mesulam [45] catalog. In this way, it was possible to highlight that the most significant areas to distinguish between ASD and TD belong to the heteromodal, unimodal, primary and paralimbic networks. The importance of these networks has also been found in the literature [46–48]. These highlighted areas are crucial for sensory perception, processing visual and auditory signals, spatial perception, and attention modulation. They are fundamental for understanding social signals that require the integration of complex sensory information such as facial expression, tone of voice, and gesture [49]. Therefore, they are important in understanding the mechanisms underlying autism spectrum disorder [50]. A heteromodal network, involving various cortical areas, is crucial for integrating complex sensory information and processing multisensory knowledge. In contrast, a unimodal network is specialized in a specific sensory modality. Neuroimaging studies have shown alterations in these areas in autism spectrum disorder, suggesting dysfunction in sensory integration and the processing of complex information in this disorder [51].

Table 3. ROIs whose connectivity with other regions had the most significant effect on ASD/TD classification. The *Occurrences* column includes the number of times an ROI appears in the five classifiers, while the numbers in the *ROI* column represent the identifiers of the ROIs in the HO atlas. The *Anatomical Part* column lists the corresponding anatomical parts of the brain (according to HO parcellation), while the *Mesulam* column identifies the associated functional networks.

Occurrences	ROI	Anatomical Part	Mesulam
18	3102	L-Precuneous Cortex	Heteromodal
15	1002	L-Superior Temporal Gyrus; posterior division	Unimodal
15	501	R-Inferior Frontal Gyrus; pars triangularis	Heteromodal
14	1302	L-Middle Temporal Gyrus; temporo-occipital	Heteromodal
11	1101	R-Middle Temporal Gyrus; anterior division	Heteromodal
10	1301	R-Middle Temporal Gyrus; temporo-occipital	Heteromodal
8	4301	R- Parietal Operculum Cortex	Unimodal
8	3301	R-Frontal Orbital Cortex	Paralimbic
8	2702	L-Subcallosal Cortex	Paralimbic
8	1102	L-Middle Temporal Gyrus; anterior division	Heteromodal
7	3401	R-Parahippocampal Gyrus; anterior division	Paralimbic
7	2801	R-Paracingulate Gyrus	Heteromodal
7	2302	L-Lateral Occipital Cortex; inferior division	Paralimbic
7	1702	L-Postcentral Gyrus	Primary
6	2201	R-Lateral Occipital Cortex; superior division	Unimodal
6	401	R-Middle Frontal Gyrus	Heteromodal
5	4402	L-Planum Polare	Unimodal

4. Discussion

The obtained results in the classification task are in agreement with the current literature, where performance typically hovers around 70% in the multicenter dataset [52–54]. This AUC percentage is also achieved using advanced techniques like in [55], where they used a multi-site clustering and nested feature extraction technique. When the analysis is limited to homogeneous and small datasets, the identification of ASD has a high accuracy [56–58], while classification results for heterogeneous and multicenter dataset have shown lower accuracy, like in Yang et al. [59]. With this type of analysis (multicenter dataset, classification task, and tabular data), we believe that traditional ML classifiers are

more suitable, but when dealing with multi-modal features, especially as the complexity increases with data, we believe that deep learning still has potential.

The difference in the most important features between the studied classifiers can be attributed to the high number of features (5253) and the intrinsic multivariate nature of the problem. Hence, a large set of features appears to be relevant in the classification, while no small subset can be defined relevant in the classification. To provide an example, the most important features present a score that is in the range of 0.1–1% of AUC, depending on the classifier. This result could also be related to overfitting, highlighting the challenges posed by the dataset's complexity and the abundance of features in achieving robust classifier generalization and accurately identifying significant features.

One of the main advantages of DL classifiers is that it is possible to avoid the feature extraction procedure. This is useful, for example, when directly analyzing raw images. However, this approach brings an intrinsic challenge due to the high dimensionality of MRI data and limited size of the datasets. Despite the fact that there are more advanced DL classifiers to study brain images, as a consequence of the improvements in AI research, functional connectivity still remains one of the most used and reliable method when studying rs-fMRI. Hence, we chose to focus our work on comparing traditional ML classifiers with DL ones using functional connectivity tabular data and to determine whether common features emerged across different classifiers.

In this particular field, this work showed that, when dealing with tabular data, ML classifiers have better classification performance than DL ones.

5. Conclusions

In this work, we investigated the effectiveness of both traditional ML and DL classifiers in classifying individuals with ASD against TD controls. Our findings revealed that ML classifiers achieved state-of-the-art classification performance, outperforming the DL classifiers, TabNet and MLP. These results suggest that DL classifiers may not always provide optimal outcomes for this specific data domain. Moreover, our analysis emphasizes the need to pay attention when interpreting DL classifier performance, given that optimizing DL classifiers presents greater challenges compared to traditional ML classifiers. Additionally, the features that have the most significant impact in the classification task vary across different classifiers. This result indicates the need for great caution in determining the brain regions or features most involved in ASD when conducting a classification task.

Author Contributions: F.M. implemented, analyzed and interpreted the classification procedures and the feature importance and wrote the manuscript with input from all authors. P.O. contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. A.R. was involved in planning and supervised the work. B.G. was involved in planning and supervised the work. All authors have read and agreed to the published version of the manuscript.

Funding: Research partly supported by: Artificial Intelligence in Medicine (next AIM, <https://www.pi.infn.it/aim>) project, funded by INFN-CSN5; PNRR—M4C2—Partenariato Esteso “FAIR—Future Artificial Intelligence Research”—Spoke 8, and PNRR—M4C2—Centro Nazionale “ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing”—Spoke 8, funded by the European Commission under the NextGeneration EU programme.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for this work can be found in the ABIDE site for download. We provide all the additional and necessary information to interpret, replicate and build upon the findings reported in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABIDE	Autism Brain Imaging Data Exchange
ASD	Three-letter acronym
AUC	Area Under the Curve
BOLD	Blood Oxygenation Level Dependent
CPAC	Configurable Pipeline for the Analysis of Connectomes
CV	Cross-Validation
DL	Deep Learning
fMRI	Functional Magnetic Resonance Imaging
HO	Harvard–Oxford
L-SVM	Support Vector Machine with Linear Kernel
ML	Machine Learning
MLP	Multi Layer Perceptron
PCA	Principal Component Analysis
PCs	Principal Components
RBF-SVM	Support Vector Machine with Gaussian Radial Basis Function
ROC	Receiver Operating Characteristic
ROI	Region of Interest
rs-fMRI	Resting-state Functional Magnetic Resonance Imaging
SVM	Support Vector Machine
TabNet	Attentive Interpretable Tabular Learning
TD	Typically Developing
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

References

1. Isabelle, R.; Roberto, F.T. Autism: Definition, Neurobiology, Screening, Diagnosis. *Pediatr. Clin. N. Am.* **2008**, *55*, 1129–1146. [\[CrossRef\]](#)
2. Baio, J.; Wiggins, L.; Christensen, D.; Meanner, M.; Daniels, J.; Warren, Z.; Kurzius-Spencer, M.; Zahorodny, W.; Robinson, C.; Rosenberg, T.; et al. Prevalence of Autism Spectrum Disorder among Children Aged 8 Years- Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill. Summ.* **2018**, *67*, 1–23. [\[CrossRef\]](#)
3. Yarger, H.; Lee, L.C.; Kaufmann, C.; Zimmerman, A. Co-occurring Conditions and Change in Diagnosis in Autism Spectrum Disorders. *Pediatrics* **2012**, *129*, e305–e316. [\[CrossRef\]](#)
4. Falkmer, T.; Andeerson, K.; Falkmer, M.; Horlin, C. Diagnostic procedures in autism spectrum disorders: A systematic literature review. *Eur. Child Adolesc. Psychiatry* **2013**, *22*, 329–340. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Susan, E.L.; David S, M.; Robert T, S. Autism. *Lancet* **2009**, *374*, 1627–1638. [\[CrossRef\]](#)
6. Gallagher, S.; Varela, F.J. Redrawing the Map and Resetting the Time: Phenomenology and the Cognitive Sciences. *Can. J. Philos.* **2003**, *33*, 93–132. [\[CrossRef\]](#)
7. Barbara, R.; Ugis, S.; Gunter, S.; Antonio, M.P. Biomarkers in autism spectrum disorder: The old and the new. *Psychopharmacology* **2014**, *231*, 1201–1216. [\[CrossRef\]](#)
8. Büyükoflaz, F.N.; Öztürk, A. Early autism diagnosis of children with machine learning algorithms. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; pp. 1–4. [\[CrossRef\]](#)
9. Yousefian, A.; Shayegh, F.; Maleki, Z. Detection of autism spectrum disorder using graph representation learning algorithms and deep neural network, based on fMRI signals. *Front. Syst. Neurosci.* **2023**, *16*, 904770. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Koutsouleris, N.; Borgwardt, S.; Meisenzahl, E.M.; Bottlender, R.; Möller, H.J.; Riecher-Rössler, A. Disease Prediction in the At-Risk Mental State for Psychosis Using Neuroanatomical Biomarkers: Results from the FePsy Study. *Schizophr. Bull.* **2011**, *38*, 1234–1246. [\[CrossRef\]](#)
11. Ball, T.; Stein, M.; Ramsawh, H.; Campbell-Sills, L.; Paulus, M.P. Single-Subject Anxiety Treatment Outcome Prediction using Functional Neuroimaging. *Neuropsychopharmacology* **2014**, *39*, 1254–1261. [\[CrossRef\]](#)
12. Chen, T.; Chen, Y.; Yuan, M.; Gerstein, M.; Li, T.; Liang, H.; Froehlich, T.; Lu, L. The Development of a Practical Artificial Intelligence Tool for Diagnosing and Evaluating Autism Spectrum Disorder: Multicenter Study. *JMIR Med. Inform.* **2020**, *8*, e15767. [\[CrossRef\]](#)
13. Gao, J.; Chen, M.; Li, Y.; Gao, Y.; Li, Y.; Cai, S.; Wang, J. Multisite Autism Spectrum Disorder Classification Using Convolutional Neural Network Classifier and Individual Morphological Brain Networks. *Front. Neurosci.* **2021**, *14*, 629630. [\[CrossRef\]](#)

14. Yang, X.; Islam, M.S.; Khaled, A.M.A. Functional connectivity magnetic resonance imaging classification of autism spectrum disorder using the multisite ABIDE dataset. In Proceedings of the 2019 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Chicago, IL, USA, 19–22 May 2019; pp. 1–4.
15. Jacob, D.; Ming-Wei, C.; Kenton, L.; Kristina, T. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
16. Kaiming, H.; Xiangyu, Z.; Shaoqing, R.; Jian, S. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
17. Marjane, K.; Afshin, S.; Delaram, S.; Navid, G.; Mahboobeh, J.; Parisa, M.; Ali, K.; Roohallah, A.; Assef, Z.; Yinan, K.; et al. Deep learning for neuroimaging-based diagnosis and rehabilitation of Autism Spectrum Disorder: A review. *Comput. Biol. Med.* **2021**, *139*, 104949. [\[CrossRef\]](#)
18. Yang, X.; Sarraf, S.; Zhang, N. Deep Learning-based framework for Autism functional MRI Image Classification. *J. Ark. Acad. Sci.* **2018**, *72*, 47–52. [\[CrossRef\]](#)
19. Shwartz-Ziv, R.; Tishby, N. Opening the Black Box of Deep Neural Networks via Information. *arXiv* **2017**, arXiv:1703.00810.
20. Autism Brain Imaging Data Exchange. Available online: <http://preprocessed-connectomes-project.org/abide/index.html> (accessed on 15 September 2023).
21. Rachel, L.; Laura, H.; William, P.L.M. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **2017**, *56*, 466–474. [\[CrossRef\]](#)
22. Configurable Pipeline for the Analysis of Connectomes. Available online: <https://fcp-indi.github.io/> (accessed on 10 March 2024).
23. Yang, X.; Schrader, P.T.; Zhang, N. A Deep Neural Network Study of the ABIDE Repository on Autism Spectrum Classification. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*. [\[CrossRef\]](#)
24. Atlases. Available online: <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases> (accessed on 30 April 2024).
25. Ross, S.M. *Introduzione Alla Statistica*; Maggioli Editore: Rimini, Italy, 2014.
26. Chen, H.; Nomi, J.; Uddin, L.; Duan, X.; Chen, H. Intrinsic functional connectivity variance and state-specific under-connectivity in autism. *Hum. Brain Mapp.* **2017**, *38*, 5740–5755. [\[CrossRef\]](#)
27. NeuroHarmonize. Available online: <https://github.com/rpomponio/neuroHarmonize> (accessed on 4 March 2024).
28. Pomponio, R.; Erus, G.; Habes, M.; Doshi, J.; Srinivasan, D.; Mamourian, E.; Bashyam, V.; Nasrallah, I.M.; Satterthwaite, T.D.; Fan, Y.; et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage* **2020**, *208*, 116450. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2006**, *8*, 118–127. [\[CrossRef\]](#)
30. Fortin, J.P.; Cullen, N.; Sheline, Y.I.; Taylor, W.D.; Aselcioglu, I.; Cook, P.A.; Adams, P.; Cooper, C.; Fava, M.; McGrath, P.J.; et al. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* **2018**, *167*, 104–120. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Serra, G.; Mainas, F.; Golosio, B.; Retico, A.; Oliva, P. Effect of data harmonization of multicentric dataset in ASD/TD classification. *Brain Inform.* **2023**, *10*, 32. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Kassraian-Fard, P.; Matthis, C.; Balsters, J.H.; Maathuis, M.H.; Wenderoth, N. Promises, Pitfalls, and Basic Guidelines for Applying Machine Learning Classifiers to Psychiatric Imaging Data, with Autism as an Example. *Front. Psychiatry* **2016**, *7*, 177. [\[CrossRef\]](#)
33. Shwartz-Ziv, R.; Armon, A. Tabular Data: Deep Learning is Not All You Need. *arXiv* **2021**. [\[CrossRef\]](#)
34. Arik, S.O.; Pfister, T. TabNet: Attentive Interpretable Tabular Learning. *arXiv* **2020**, arXiv:1908.07442.
35. Hossain, M.; Kabir, M.; Anwar, A.; Islam, M.Z. Detecting autism spectrum disorder using machine learning techniques. *Health Inf. Sci. Syst.* **2021**, *9*, 17. [\[CrossRef\]](#)
36. sklearn svm. Available online: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC> (accessed on 14 March 2024).
37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
38. xgboost. Available online: https://xgboost.readthedocs.io/en/stable/python/python_intro.html (accessed on 14 March 2024).
39. tabnet. Available online: <https://pypi.org/project/pytorch-tabnet/> (accessed on 14 March 2024).
40. MLPClassifier. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html#sklearn.neural_network.MLPClassifier (accessed on 14 March 2024).
41. Hanley, J.; Mcneil, B. The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology* **1982**, *143*, 29–36. [\[CrossRef\]](#)
42. Metz, C.E. Receiver Operating Characteristic Analysis: A Tool for the Quantitative Evaluation of Observer Performance and Imaging Systems. *J. Am. Coll. Radiol.* **2006**, *3*, 413–422. Special Issue: Image Perception. [\[CrossRef\]](#) [\[PubMed\]](#)
43. ELI5's Documentation: Permutation Importance. Available online: https://eli5.readthedocs.io/en/latest/blackbox/permutation_importance.html (accessed on 15 April 2024).
44. ELI5's Documentation. Available online: <https://eli5.readthedocs.io/en/latest/index.html> (accessed on 15 April 2024).
45. Mesulam, M.M. Form sensation to cognition. *Brain* **1998**, *121*, 1013–1052. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Martínez, K.; Martínez-García, M.; Marcos-Vidal, L.; Janssen, J.; Castellanos, F.X.; Pretus, C.; Villarroja, Ó.; Pina-Camacho, L.; Díaz-Caneja, C.M.; Parellada, M.; et al. Sensory-to-Cognitive Systems Integration Is Associated with Clinical Severity in Autism Spectrum Disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **2020**, *59*, 422–433. [\[CrossRef\]](#)

47. Martineau, J.; Roux, S.; Garreau, B.; Adrien, J.; Lelord, G. Unimodal and crossmodal reactivity in autism: Presence of auditory evoked responses and effect of the repetition of auditory stimuli. *Biol. Psychiatry* **1992**, *31*, 1190–1203. [\[CrossRef\]](#)
48. d’Albis, M.A.; Guevara, P.; Guevara, M.; Laidi, C.; Boisgontier, J.; Sarrazin, S.; Duclap, D.; Delorme, R.; Bolognani, F.; Czech, C.; et al. Local structural connectivity is associated with social cognition in autism spectrum disorder. *Brain* **2018**, *141*, 3472–3481. [\[CrossRef\]](#)
49. Maximo, J.O.; Kana, R.K. Aberrant “deep connectivity” in autism: A cortico–Subcortical functional connectivity magnetic resonance imaging study. *Autism Res.* **2019**, *12*, 384–400. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Neuroanatomia dell’autismo. Available online: <https://www.igorvitale.org/neuroanatomia-autismo-cervello-caratteristiche/> (accessed on 15 April 2024).
51. Gotts, S.J.; Simmons, W.K.; Milbury, L.A.; Wallace, G.L.; Cox, R.W.; Martin, A. Fractionation of social brain circuits in autism spectrum disorders. *Brain* **2012**, *135*, 2711–2725. [\[CrossRef\]](#)
52. Plitt, M.; Barnes, K.A.; Martin, A. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage Clin.* **2015**, *7*, 359–366. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Nielsen, J.; Zielinski, B.; Fletcher, P.; Alexander, A.; Lange, N.; Bigler, E.; Lainhart, J.; Anderson, J. Multisite functional connectivity MRI classification of autism: ABIDE results. *Front. Hum. Neurosci.* **2013**, *7*, 599. [\[CrossRef\]](#)
54. Jain, V.; Rakshe, C.; Sengar, S.; Murugappan, M.; Ronickom, J.F.A. Age-and Severity-Specific Deep Learning Models for Autism Spectrum Disorder Classification Using Functional Connectivity Measures. *Arab. J. Sci. Eng.* **2024**, *49*, 6847–6865. [\[CrossRef\]](#)
55. Nan, W.; Dongren, Y.; Lizhuang, M.; Mingxia, L. Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with resting-state fMRI. *Med. Image Anal.* **2022**, *75*, 102279. [\[CrossRef\]](#)
56. Yin, W.; Mostafa, S.; Wu, F.x. Diagnosis of Autism Spectrum Disorder Based on Functional Brain Networks with Deep Learning. *J. Comput. Biol.* **2021**, *28*, 146–165. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Salim, I.; Hamza, A. Classification of Developmental and Brain Disorders via Graph Convolutional Aggregation. *Cogn. Comput.* **2024**, *16*, 701–716. [\[CrossRef\]](#)
58. Nogay, H.; Adeli, H. Multiple Classification of Brain MRI Autism Spectrum Disorder by Age and Gender Using Deep Learning. *J. Med. Syst.* **2024**, *48*, 15. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Xin, Y.; Ning, Z.; Paul, S. A study of brain networks for autism spectrum disorder classification using resting-state functional connectivity. *Mach. Learn. Appl.* **2022**, *8*, 100290. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.