



Population-based detection of children ASD/ADHD comorbidity from atypical sensory processing

Manuel Fernández-Delgado¹ · Sara Cruz² · Eva Cernadas¹ · Heba Alateyat³ · María Tubío-Fungueiriño^{4,5,6} · Adriana Sampaio⁷ · Angel Carracedo^{4,6,8,9} · Montse Fernández-Prieto^{4,5,6,9}

Accepted: 25 June 2024 / Published online: 29 July 2024
© The Author(s) 2024

Abstract

Comorbidity between neurodevelopmental disorders is common, especially between autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD). This study aimed to detect overlapped sensory processing alterations in a sample of children and adolescents diagnosed with both ASD and ADHD. A collection of 42 standard and 8 proposed machine learning classifiers, 22 feature selection methods and 19 unbalanced classification strategies were applied on the 6 standard question groups of the Sensory Profile-2 questionnaire. The relatively low performance achieved by state-of-the-art classifiers led us to propose the feature population sum classifier, a probabilistic method based on class and feature value populations, designed for datasets where features are discrete numeric answers to questions in a questionnaire. The proposed method achieves the best kappa and accuracy, 60% and 82.5%, respectively, reaching 68% and 86.5% combined with backward sequential feature selection, with false positive and negative rates below 15%. Since the SP2 questionnaire can be filled by parents for children from three years, our prediction can alert the clinicians with an early diagnosis in order to apply early interventions.

Keywords Machine learning · Classification · ASD · ADHD · Sensory profile · Feature selection · Unbalanced classification

1 Introduction

Symptoms comorbidity between neurodevelopmental disorders has been increasing over the past years [1]. Autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD) show considerable symptomatology overlap [2] with rates ranging between 50 and 70% [3]. ASD is characterized by persistent deficits in socio-communication and socio-interaction across multiple contexts, alongside with restricted and repetitive patterns of behavior, interests, or activities [4], while ADHD is characterized by altered functioning in the areas of attention, hyperactivity, and impulsivity. Both disorders' prevalence has increased over the past decades [5], are more common in boys than girls and cause a significant disruption in children's life, such as difficulties in school performance or in social interactions with other children and adults [6].

Children and adolescents diagnosed with comorbid ASD/ADHD often present poorer behavioral, cognitive, and socio-emotional outcomes compared to children and adolescents with ASD alone, suggesting that ASD/ADHD comorbidity may constitute a distinctive phenotype characterized by marked difficulties compared to ASD alone, see [7] for a detailed review. Indeed, those with both ASD/ADHD present more severe autistic symptoms, particularly marked impairments in social interaction abilities, i.e., social responsiveness [8, 9], showing increased difficulties in adaptive skills [10] and poorer quality of life [11] than children and adolescents with ASD alone. In addition, compared to children with only ASD, children with both ASD and ADHD display lower cognitive functioning [12], especially problems in inhibitory control, attention abilities [13] and Theory of Mind [2]. Evidence also shows that children with ASD/ADHD have increased stereotypic and repetitive behaviors, such as mannerisms [12], tantrum behaviors and conduct problems [14] than those diagnosed with ASD alone. These increased difficulties frequently lead to anxiety, externalizing, and mood problems [15].

✉ Manuel Fernández-Delgado
manuel.fernandez.delgado@usc.es

Extended author information available on the last page of the article

The previous findings enclose relevant implications for clinical practice as they evidence the negative impact that the co-occurrence of ASD and ADHD symptoms have on children's multiple functioning outcomes. Thus, the early identification of ASD/ADHD comorbidity is important not only to help establishing earlier and more accurate diagnosis, but also to determine and optimize intervention programs for these children. The early identification of comorbidities facilitates the implementation of specific interventions for ASD and/or ADHD, positively influencing children's adhesion to treatment and functioning outcomes [16].

Altered sensory abilities are frequently co-observed in ASD and ADHD. Thus, a possible approach to early identify ASD/ADHD comorbidity may be related to sensory processing abnormalities. Difficulties in sensory processing are frequent and one of the earliest and common clinical observations in children with neurodevelopmental disorders, especially in those with ASD and behavioral problems [17], compared to typically developing children [18]. ASD is most often associated with impairments in sensory processing [19], but ADHD also presents atypical sensory processing [20]. Differences in the unusual responses to sensorial experiences exist between children with ASD alone, with ADHD alone, and with ASD/ADHD [21]. Children with ASD seem to have increased difficulties in oral processing (i.e., taste) compared to children with ADHD that, in turn, present more difficulties in visual processing [17] and proprioception [22]. Compared to children with ASD and ADHD alone, children with ASD/ADHD present increased overall sensory difficulties, e.g., visual, auditory or balance and motion [22]. However, a further characterization of sensory processing impairments in children with ASD/ADHD, ASD and ADHD is still lacking. Thus, clarifying the specificity of sensory processing difficulties in children with ASD and/or ADHD may enclose important information for an early differential diagnosis between these neurodevelopmental conditions with important implication for intervention programs.

Importantly, machine learning methods are a very promising avenue towards more successful clinical decision making and diagnosis [23]. The systematic review of Song et al [24] showed that deep learning improved the accuracy of the ASD and/or ADHD diagnosis by extracting clinical information from standardized diagnostic evaluations of neurodevelopmental disorders. The recent work of [25] evidenced good overall sensitivity and specificity obtained from machine learning methods in the diagnostic classification of ASD, particularly in children with less severe symptoms for whom establishing a diagnosis is frequently more difficult. Also, a previous study from our research group showed that machine learning techniques may be a useful method in predicting behavioral problems from altered sensory processing in children and adolescents with ASD [26].

Duda et al [27] performed the classification of 2,925 ASD and ADHD participants (150 with ADHD) using six machine learning classifiers (decision trees, linear discriminant analysis, random forest, support vector machines, logistic regression and categorical LASSO) and forward feature selection on social responsiveness scale score sheets, SRS [28], achieving 96% of area under curve with only 5 of 65 scores. Wall et al [29] use 16 classifiers to diagnose autism using the Autism Diagnostic Observation Schedule, ADOS [30] for 612 individuals, achieving a perfect classification with decision trees. Deep learning [31] achieves 70% of accuracy in the diagnosis of ASD using images of brain activation patterns for 1,035 patients. Abdullah et al [32] used the Autism Spectrum Quotient (AQ) questionnaire [33] to diagnose autism using random forest, k-nearest neighbors and logistic regression classifiers with three feature selection methods (chi-square, LASSO and random forest), achieving 97% of accuracy on a dataset of 704 persons, 189 of them with ASD.

Hence, applying machine learning methods to atypical sensory processing could be an effective approach to an early detection of ASD/ADHD comorbidity. The current study uses classification techniques to detect comorbid ADHD symptomatology in children and adolescents diagnosed with ASD. This is a binary classification problem with classes ASD and ASD/ADHD. The class prediction is performed using data about specific altered sensory processing abilities, that are described in section 2. It is expected that, in accordance with previous studies using machine learning, persons with comorbid ASD/ADHD will present increased sensory processing difficulties compared to those with ASD alone. Given the low performance achieved by the standard classifiers (section 3), we proposed alternative approaches specifically focused to this problem in subsections 4 and 5, that increased the classification performance. Since data are high-dimensional and exhibit class unbalance, we also tried a wide set of feature selection and unbalanced classification methods in subsections 6.1 and 6.2, respectively. The results are discussed in subsection 6.3. Section 7 compiles the main findings of this work.

2 Dataset description

We used a dataset that includes 52 children and adolescents, aged from 6 to 14 years (mean= 8.25; standard deviation=2.70), diagnosed with ASD. Of these, 14 children and adolescents also present comorbid ADHD (representing 26.9% of the total sample). All the diagnoses were made by qualified professionals of neurodevelopmental disorders and mental health based on: 1) ASD criteria established by the Diagnostic and Statistical Manual of Mental Disorders [4]

in its revised fourth version, DSM-IV-TR, or fifth version, DSM-5; or 2) a specific diagnosis instrument such as Autism Diagnostic Interview Revised, ADI-R [34] and the Autism Diagnostic Observation Schedule, ADOS [30]. The ADHD comorbidity was also carried out by clinicians and mental health professionals based on ADHD criteria established by the DSM-IV-TR or DSM-5. All parents who agreed to participate in this research gave the written informed consent. Informed consent was obtained in accordance with the Declaration of Helsinki. This study was approved in 2012 by the Galician ethic committee for clinical research with ethics board protocol number 2012/098.

Each participant in this study is defined by a vector of features, where each one is the numeric answer to a question in the Sensory Profile-2 questionnaire (SP2), Spanish version 3:0-14:11 [35]. The SP2 is a standardized tool to assess children and adolescents' sensory processing patterns in the context of everyday life. This questionnaire helps to determine how sensory processing difficulties in children and adolescents with neurodevelopmental disorders may be contributing to or interfering with their participation in activities and influencing their behavior. The questionnaire forms for children aged 3-14 years were completed by the caregivers. The SP2 measures children and adolescents' sensory processing patterns to different sensory modalities— Auditory Processing, Visual Processing, Touch Processing, Movement, Body Position and Oral Processing— in 86 questions scored on a five-point Likert-type scale (1=“Almost Never”, 2=“Occasionally”, 3=“Half of the time”, 4=“Frequently”, 5=“Almost Always”). Table 1 reports the question list of each group of questions in the SP2 questionnaire, with the number of questions, nature and list of the questions included. These scores characterize the sensory processing response into four different quadrants: *seeking*, *avoiding*, *sensitivity* and *registration*. In this study, these four quadrants, alongside with the *total* score (all the items in the questionnaire) and the *touch* processing sensory section score were used. The latter was included due to the evidence suggesting that tactile impairments are related to the onset of ASD, and because is the most common sensory alteration in this population [36, 37].

The machine learning classifier uses the feature vector, standardizing each feature to have zero mean and deviation one, as input data in order to predict the class for each participant: ASD alone, or ASD/ADHD. Each SP2 group defines a different dataset, e.g. *avoiding* has 20 questions listed in the corresponding row of Table 1, while *total* includes the 86 questions of SP2. The populations of both classes are unbalanced, because there are almost three times more participants with ASD alone (38 of 52) than participants with ASD/ADHD (only 14).

The panel (a) of Fig. 1 shows the two *avoiding* feature vectors, one of class ASD (continuous squared line) and

other of class ASD/ADHD (dashed circled line), that have the lowest distance between them over all the training patterns of different classes. This distance is the number of features where they differ. In this panel (a), the minimum distance is 7. Considering only vectors of class ASD, the lowest distance is 9, see panel (b). With class ASD/ADHD, the lowest distance is 7 again, see panel (c). These distances are equal to or higher than the lowest distance between vectors of different classes, as in panel (a). Thus, similarities between feature vectors of participants with ASD and ASD/ADHD are comparable to differences between vectors of the same class. The panel (d) shows the histogram of distances between vectors of the same and different class (continuous and dashed lines, respectively), confirming that within- and between-class distances follow similar distributions and have similar ranges. This fact suggests that: 1) classifiers based on distances that consider equally all the features might be unable to adequately discriminate between classes; and 2) differences in certain features may be more important for classification than others. Therefore, methods that weight differently each feature might be able to detect differences between classes in just one or few features and, perhaps, to achieve better classification performance.

3 Standard classification methods

The classification of participants in ASD or ASD/ADHD according to their SP2 features was performed using a large collection of 42 standard state-of-the-art classifiers (Table 1 in the supplementary material) implemented in four programming languages: Matlab, Octave, Python (using the `scikit-learn` module) and R, using different packages listed in the table. Several versions of the most common classifiers were executed, such as classification trees (`ctree`) in Matlab, Octave (own implementation), Python and R. The recursive partitioning classification tree (`rpart`) was also executed in R. Multi-layer perceptron neural networks were executed in Matlab, Python and R with packages `nnet` and `neuralnet`. Other neural networks that were executed are extreme learning machine (`elm`) in Octave (own implementation) and R, and probabilistic neural network (`pnn`) in Matlab. Nearest neighbor classifiers were executed in Matlab, Python and R. Linear and diagonal linear, quadratic, flexible and mixture discriminant analysis were executed in Matlab, Python and R. The classifier ensembles executed in this work include: `adaboost` and `bagging`, both in Matlab and R; `gradient boosted machine (gbm)` in Python and R; and `random forest (rforest)` in Matlab, Python and R. The support vector machine was executed in Matlab, in Octave using the `LibSVM` library, in Python and in R. Other standard classifiers are: `naive Bayes (awnb, manb, nb, nbDiscrete and nbSearch)` in R; `k-nearest neighbors with Euclidean dis-`

Table 1 List of the 6 datasets created from the SP2 questionnaire: the four SP2 quadrants avoiding, registration, seeking and sensitivity, alongside with touch and total

Dataset	Size	Include questions about:
avoiding	20	retreats from unfamiliar situations, avoidings and/or limitations in sensory experiences: 1,2,5,15,18,58,59,61,63–68,70,71,72,74,75,81
registration	22	late reactions to sensory experiences or lack of response to sensory stimuli that other people usually respond to: 8,12,23,24,26,33–40,53,54,57,62,76,79,80,85,86
seeking	19	responses engaged in sensory experiences, increasings in sensory input: 14,21,22,25,27,28,30–32,41,48–51,55,56,60,82,83
sensitivity	19	intense reactions, or responses to sensory stimuli, that are not detected by others: 3,4,6,7,9,13,16,19,20,44–47,52,69,73,77,78,84
total	86	sensory processing response of the four previous quadrants: 1–86
touch	11	tactile processing: 16–26

The dataset size (number of questions) and the description and list of questions included in each dataset (e.g., question 1 means SP1) are also reported

tance (knn), in Matlab and R; multivariate adaptive regression spline (earth); Gaussian process regression (gpr) and learning vector quantization (lvq), both implemented in R. Table 2 of the supplementary material lists details about hyperparameter tuning.

The previous is a very diverse collection including many different classifiers, so there is no *a priori* reason to deem

them inadequate for automatic ASD/ADHD classification. However, their performance was poor for this problem, as reported in the results section below. In order to achieve better results, the following subsections propose alternative classification methods that exploit the characteristics of the problem of ASD/ADHD prediction using questions of the SP2 questionnaire.

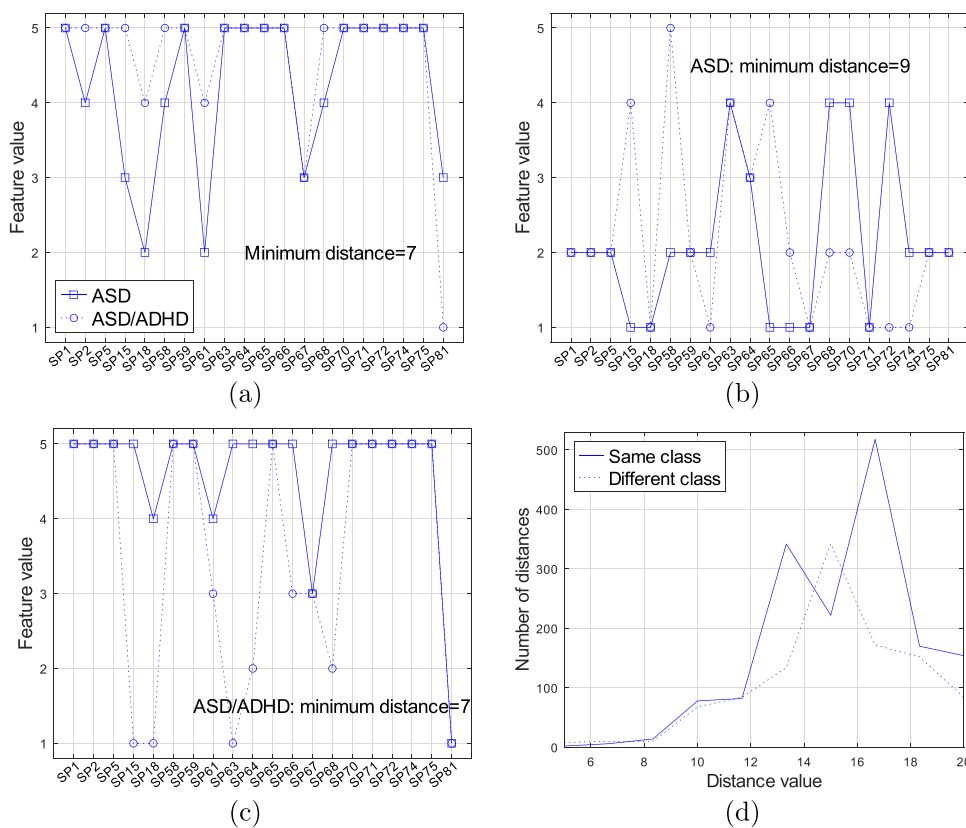


Fig. 1 (a) Two feature vectors of dataset avoiding of ASD and ASD/ADHD classes that have the lowest distance between them. The “Minimum distance” printed is the number of features that differ

between them. (b) Two nearest vectors of class ASD. (c) Two nearest vectors of class ASD/ADHD. (d) Histograms of distances between vectors of the same and different class

4 Proposed nearest neighbor classifiers

We also studied the use of other classifiers designed specifically for these data that take advantage from their singular properties. As explained above, the variability in the feature vectors is not higher between participants of different classes than between participants of the same class, so standard distance-based classifiers such as support vector machine and nearest neighbor methods might not achieve good performance. This suggested us to replace the standard Euclidean distance between vectors in neighbor classifiers by a distance measure specifically focused to this type of problems. A detailed analysis of vectors with the same and different class labels shows that the non-coincidence in a certain feature is often more relevant for classification than the specific value of the difference. Therefore, we used a distance measure that evaluates just the number of features that differ between vectors, as explained in the following paragraph.

Since features are answers to questions in the SP2 questionnaire, they share the same set of possible values $\mathcal{J} = \{1, \dots, J\}$, with $J=5$, so each feature vector can be defined as $\mathbf{x} = (j_1, \dots, j_I)$, where $j_i \in \mathcal{J}$ is the value of the i -th feature, for $i = 1, \dots, I$, being I the number of features (e.g. $I=20$ for dataset `avoiding`, see Table 1). Let N be the number of training vectors, and $\mathbf{x}_n = (j_{n1}, \dots, j_{nI})$ the n -th training vector, with $n = 1, \dots, N$. Let $c_n \in \{1, \dots, C\}$ be the class label of \mathbf{x}_n and $C=2$ the number of classes, where $c_n=1$ or $c_n=2$ for each ASD or ASD/ADHD participant, respectively. We propose to use the following distance measurement:

$$d_n(\mathbf{x}) = d(\mathbf{x}, \mathbf{x}_n) = I - \sum_{i=1}^I \delta(j_i, j_{ni}) \tag{1}$$

between a test vector \mathbf{x} and a training vector \mathbf{x}_n , where $\delta(i, j) = 1$ when $i = j$ and zero otherwise. Using this distance, we considered the following two nearest neighbor classifiers, alternatives to the standard k -nearest neighbor (knn) classifier:

1. **Nearest neighbor** (nn classifier in Table 1 of the supplementary material), the standard K -nearest neighbor classifier with distance $d_n(\mathbf{x})$ defined above, tuning the number K of neighbors with values from 1 to 19 with step 2.
2. **Nearest neighbors by class** (nnc). The high variability between ASD vectors, alongside with the population unbalance favouring this class, causes that ASD/ADHD vectors often find ASD nearest neighbors, biasing prediction to this class. The proposed nnc method tries to compensate this bias by scaling and translating the distances between a test vector and the training vectors of

both classes. We use the mean and deviations of distances for each class, in order to make a proper selection of nearest neighbors. For $n = 1, \dots, N$, let u_n be the lowest distance $d_n(\mathbf{x}_p)$ for $c_p = c_n$, with d_n defined above. For each class $k = 1, \dots, C$, let μ_k and σ_k be the mean and standard deviation of u_n for $c_n = k$. Then, for class k the lowest distance $D_k(\mathbf{x})$ between its training vectors $\{\mathbf{x}_n\}_{c_n=k}$ and the test vector \mathbf{x} is evaluated. This distance $D_k(\mathbf{x}) = \min_{c_n=k} \{d_n(\mathbf{x})\}$ is translated and scaled by

$$\mu_k \text{ and } \sigma_k, \text{ respectively, into } D'_k(\mathbf{x}) = \frac{D_k(\mathbf{x}) - \mu_k}{\sigma_k}$$

in order to be evaluated according to the typical distances between the training vectors of class k . The class prediction $Y(\mathbf{x}) = \arg \min_{k=1, \dots, C} \{D'_k(\mathbf{x})\}$ is the one with the lowest distance $D'_k(\mathbf{x})$.

5 Proposed feature population classifiers

The features are answers, with discrete values from 1 to 5, to questions in the SP2 questionnaire. To evaluate differences between participants over all the features jointly might not allow a reliable discrimination between both classes. On the contrary, it seems that some features might be much more relevant than others. The right prediction for each test vector might even depend on just one or few features, that however might not be the same for all the test vectors. This suggested us to evaluate separately each feature based on its value and on the number of training vectors of each class that had this value. The objectives are: 1) to perform a class prediction for each feature; 2) to evaluate the relevance of each feature for classification; and 3) to use these class predictions and feature relevances to make a global class prediction, i.e., over all the features. Following this idea, we propose several classifiers based on the population of training vectors of each class that have each value of each feature. In order to make the final prediction, these classifiers use different criteria that we will describe in the following paragraphs.

Let N_k be the number of training vectors of class k . For each feature (or input) $i = 1, \dots, I$, value $j = 1, \dots, J$, and class $k = 1, \dots, C$, let N_{ijk} be the number of training vectors $\mathbf{x}_n = (j_{n1}, \dots, j_{nI})$ with $c_n = k$ and $j_{ni} = j$. Let P_{ijk} be the relative population of training vectors of class k whose i -th relative has value j , calculated as the ratio between N_{ijk} and N_k :

$$P_{ijk} = \frac{N_{ijk}}{N_k}, \quad i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, C \tag{2}$$

For a test vector $\mathbf{x} = (j_1, \dots, j_I)$, let $P_{ik}(\mathbf{x})$ be the relative population of input i , value j_i and class k , so that $P_{ik}(\mathbf{x}) =$

P_{ijk} for $i = 1, \dots, I$ and $k = 1, \dots, C$. According to this distribution of the training vector population among feature values and classes, the prediction $y_i(\mathbf{x})$ issued by feature i should be the class k that maximizes $P_{ik}(\mathbf{x})$:

$$y_i(\mathbf{x}) = \arg \max_{k=1, \dots, C} \{P_{ik}(\mathbf{x})\}, \quad i = 1, \dots, I \tag{3}$$

Based on this, we can define the relevance measurement R_i of feature i as the average accuracy of this feature in the class prediction over the whole training set. For the training vector $\mathbf{x}_n = (j_{n1}, \dots, j_{nI})$, feature i predicts class $y_i(\mathbf{x}_n)$ achieved replacing \mathbf{x} by \mathbf{x}_n in (3), so that the relevance R_i of feature i can be defined as:

$$R_i = \frac{1}{N} \sum_{n=1}^N \delta(c_n, y_i(\mathbf{x}_n)) \tag{4}$$

In order to perform a class prediction $Y(\mathbf{x})$ for a test vector $\mathbf{x} = (j_1, \dots, j_I)$ based on the predictions $y_i(\mathbf{x})$ and relevances R_i of the I features, several criteria were considered, leading to the following population-based classifiers:

1. **Feature population voting (fpv)**. The predicted class $Y_1(\mathbf{x})$ is the most voted over the I features: $Y_1(\mathbf{x}) =$

$$\arg \max_{k=1, \dots, C} \left\{ \sum_{i=1}^I \delta(y_i(\mathbf{x}), k) \right\}, \quad \text{where } \delta(i, j) \text{ is}$$

defined in section 4. This approach was selected due to its simplicity, although the lack of an averaging procedure over features might lead to low classification performance. In case of tie, the predicted class k is the one with the largest sum of relevances over the features that

predict class k , so that $Y_1(\mathbf{x}) = \arg \max_{k=1, \dots, C} \left\{ \sum_{i \in \mathcal{A}_k(\mathbf{x})} R_i \right\}$,

where $\mathcal{A}_k(\mathbf{x}) = \{i = 1, \dots, I : y_i(\mathbf{x}) = k\}$.

2. **Feature population sum (fps)**. The predicted class $Y_2(\mathbf{x})$ is the one that maximizes the sum of relative populations $P_{ik}(\mathbf{x})$ for those features i with $y_i(\mathbf{x}) = k$, so that

$$Y_2(\mathbf{x}) = \arg \max_{k=1, \dots, C} \left\{ \sum_{i \in \mathcal{A}_k(\mathbf{x})} P_{ik}(\mathbf{x}) \right\}.$$

This sum over features that predict the same class performs averaging and might compensate some classification errors compared to the crisp decision used by fpv.

3. **Maximum relevance feature (mrf)**. The prediction $Y_3(\mathbf{x})$ is the class k predicted by the feature with the highest relevance R_i , so that $Y_3(\mathbf{x}) = y_p(\mathbf{x})$, where $p = \arg \max_{i=1, \dots, I} \{R_i\}$. The exclusive use of the most relevant feature p might lead to low performance due to the lack of averaging or smoothing over features.

4. **Voting among 5 most relevant features (mrf5)**. The predicted class $Y_4(\mathbf{x})$ is selected by voting among the

5 features with the highest relevances, so that $Y_4(\mathbf{x}) = \arg \max_{k=1, \dots, C} \left\{ \sum_{i \in \mathcal{B}} \delta(k, y_i(\mathbf{x})) \right\}$, where \mathcal{B} is the set of the 5 most relevant features. The voting over the 5 most relevant features might provide an averaging procedure to compensate errors that is missing in mrf.

5. **Feature relevance times population (frp)**. The predicted class $Y_5(\mathbf{x})$ is the one that maximizes the sum of $P_{ik}(\mathbf{x})$ multiplied by feature relevance R_i over the features $i \in \mathcal{A}_k(\mathbf{x})$, that predict class k . Thus, $Y_5(\mathbf{x}) =$

$$\arg \max_{k=1, \dots, C} \left\{ \sum_{i \in \mathcal{A}_k(\mathbf{x})} R_i P_{ik}(\mathbf{x}) \right\}.$$

The relevance weighted sum over features might also provide an averaging procedure that avoid errors caused by non-reliable features.

6. **Feature nearest neighbor (fnn)**. Instead of using the populations $P_{ik}(\mathbf{x})$ as the previous criteria, this method stores in set \mathcal{D}_{ijk} the indices n of the training vectors $\mathbf{x}_n = (j_{n1}, \dots, j_{nI})$ of each class k for each value j of each feature i , so that $\mathcal{D}_{ijk} = \{n = 1, \dots, N : c_n = k, j_{ni} = j\}$ for $i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, C$. For a test vector $\mathbf{x} = (j_1, \dots, j_I)$, let $d_{li}(\mathbf{x}) = \min_{n \in \mathcal{D}_{ijl}} \{d_n(\mathbf{x})\}$,

for $l = 1, \dots, C; i = 1, \dots, I$, be the lowest distance from \mathbf{x} to the training vectors of class l with the same value j_i of feature i as \mathbf{x} , whose indices are included in \mathcal{D}_{ijl} . The prediction $y_i(\mathbf{x})$ issued by fnn on feature i is the class l with the lowest distance $d_{li}(\mathbf{x})$ to training vectors in \mathcal{D}_{ijl} .

$$y_i(\mathbf{x}) = \arg \min_{l=1, \dots, C} \{d_{li}(\mathbf{x})\}, \quad i = 1, \dots, I \tag{5}$$

The global class prediction $Y_6(\mathbf{x})$ of fnn is selected by voting among features:

$$Y_6(\mathbf{x}) = \arg \max_{k=1, \dots, C} \left\{ \sum_{i=1}^I \delta(k, y_i(\mathbf{x})) \right\} \tag{6}$$

The pseudo-code of the proposed population-based classifiers is included in algorithms 1 and 2 of the supplementary material.

6 Results and discussion

We developed experiments comparing the standard and proposed classifiers (subsections 3 to 5) on the available datasets (subsection 2). Given the reduced number (52) of participants, or feature vectors, we used leave-one-out (LOO) cross validation [38]. The metrics of classification performance that we used are the Cohen kappa statistic [39] and the accu-

Table 2 Best kappa and accuracy (in %), best classifier and its language for each dataset

Dataset	Kappa	Accuracy	Classifier	Language
avoiding	60.5	82.7	fps	Octave
registration	41.4	76.9	mrf	Octave
seeking	47.4	78.8	nn	Octave
sensitivity	49.6	78.8	earth	R
total	37.9	75.0	fps	Octave
touch	25.0	71.2	lda	Python

racy, both in %. Since this is a binary classification problem, other measurements such as sensitivity, specificity, positive predictivity, F1 and area under curve, among others, were also calculated (see their definitions in Appendix E of the supplementary material).

Table 6 of Appendix H in the supplementary material reports the whole results for all the classifiers and datasets. Table 2 compiles the best kappa and accuracy achieved on each dataset by some of the 42 standard classifiers listed by Table 1 of the supplementary material. The best result (60.5%) is achieved by fps on dataset *avoiding*, that seems to be the most suited for classification. This kappa value can be considered, according to [40], as a “substantial agreement between true and predicted class labelings”. In the datasets *sensitivity*, *seeking* and *registration*, the per-

formance is about 15-20 points below *avoiding*, while results on *total* and especially on *touch* are fairly poor.

These results by datasets are confirmed by Fig. 2, that shows the box plot of the kappa values achieved by all the classifiers on each dataset. Besides achieving the highest kappa, dataset *avoiding* also achieves the largest upper box limit, much above the other datasets, and the largest median, although only slightly above *sensitivity* and *registration*. The *total*, *touch* and *seeking* datasets have even lower kappa medians (near 0%).

Table 3 reports the best kappa and accuracy achieved by each classifier over all the datasets, sorted by decreasing performance, and the dataset where the best kappa was achieved. Despite we tried 50 classifiers, 42 standard and 8 proposed, the table contains 33 rows because many classifiers have several implementations, but only the best of them is reported in the table. Eleven of the 20 highest kappa values are achieved using dataset *avoiding*, that is clearly the best one. The low kappa values achieved by the majority of the standard classifiers in Table 3 show that this is a hard classification problem. These values range from 4.6% to 60.5%, while accuracy ranges from 63.5% to 84.6%. Due to class unbalance, relatively high values of accuracy indeed correspond to very low kappa. In some cases, higher kappa correspond to lower accuracy, such as e.g. in positions 4 and 5, where nbDiscrete achieves kappa of 55.2% and accuracy of 80.8%, while fnn achieves lower kappa (54.8%) but higher accuracy (84.6%). Since accuracy is biased by class unbalance, kappa is a metric more significant than accuracy.

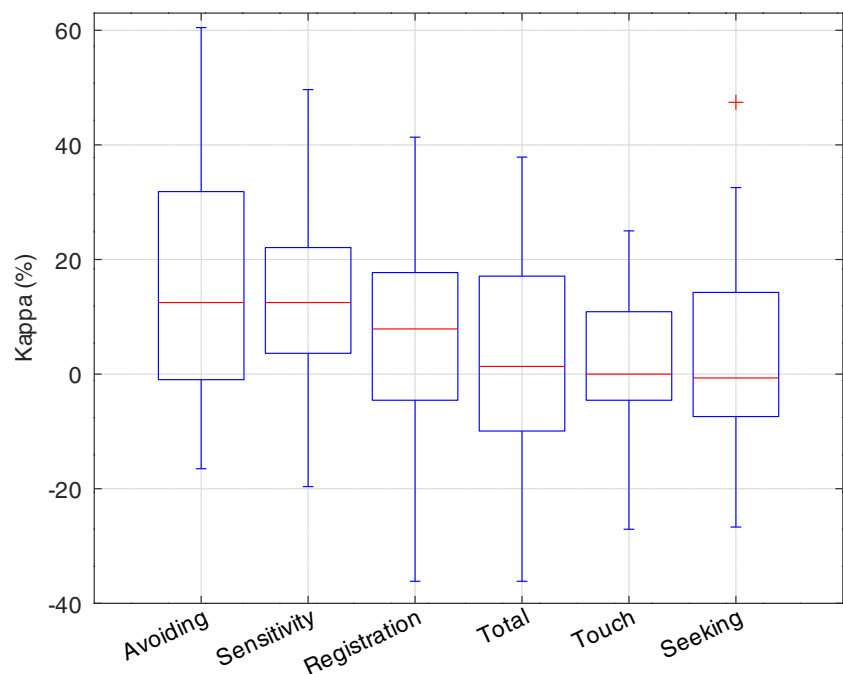
Fig. 2 Box plot of kappa (in %) for each dataset, sorted by decreasing medians

Table 3 Best kappa and accuracy (both in %) achieved by each classifier, sorted decreasingly, and dataset where the best result was achieved

Pos.	Classifier	Language	Kappa	Accuracy	Best dataset
1	fps	Octave	60.5	82.7	avoiding
2	fpv	Octave	57.0	80.8	avoiding
3	frp	Octave	55.2	80.8	avoiding
4	nbDiscrete	R	55.2	80.8	avoiding
5	fnn	Octave	54.8	84.6	avoiding
6	earth	R	49.6	78.8	sensitivity
7	rpart	R	49.6	78.8	sensitivity
8	nn	Octave	47.4	78.8	seeking
9	mrf	Octave	43.9	76.9	sensitivity
10	nb	R	41.4	76.9	avoiding
11	nnc	Octave	40.5	75.0	avoiding
12	mrf5	Octave	37.9	75.0	sensitivity
13	kfd	Python	34.5	73.1	avoiding
14	mlp	Python	32.5	78.8	seeking
15	rforest	Matlab	32.2	76.9	avoiding
16	elm	R	32.2	76.9	sensitivity
17	bagging	Matlab, R	31.9	75.0	avoiding
18	manb	R	31.9	75.0	sensitivity
19	dlda	Matlab	31.3	71.1	avoiding
20	knn	R	28.4	76.9	registration
21	ctree	Octave	25.2	69.2	seeking
22	fda	R	25.2	69.2	touch
23	lda	Python, R	25.2	69.2	touch
24	svm	Matlab	24.8	73.1	sensitivity
25	awnb	R	22.2	67.3	sensitivity
26	adaboost	Matlab	21.8	69.2	touch
27	nbSearch	R	21.4	71.1	avoiding
28	lvq	R	16.6	63.5	total
29	gbm	R	14.1	69.2	registration
30	mda	R	14.1	69.2	avoiding
31	qda	Matlab	11.7	73.1	sensitivity
32	gpr	R	6.2	73.1	sensitivity
33	pnn	Matlab	4.6	69.2	touch

The standard classifier that achieves the best kappa is discrete naive Bayes (nbDiscrete, 55.2%). This good result of Bayesian classification confirms the usefulness of probabilistic methods, as with fps, in this problem. Both the multivariate adaptive spline regression (earth) and rpart classification tree achieve kappa of 49.6%. The naive Bayes (nb) and kernel Fisher discriminant (kfd) also perform relatively well, with kappa of 41.4% and 34.5%, respectively. The remaining standard classifiers have lower kappa values. Neural networks: multi-layer perceptron (32.5%), extreme learning machine (32.2%) and probabilistic neural network (4.6%). Ensembles: random forest (32.2%), gradient boosted machine (14.2%), adaboost (21.8%) and bagging (31.9%).

Support vector machine (24.8%). Classification tree (25.2%). Discriminant analysis classifiers: dlda, fda, lda, mda and qda (between 11% and 31%). The remaining Bayes classifiers are: model averaged naive Bayes (manb, 31.9%), semi-naive structure learning wrapper (nbSearch, 21.4%) and attribute weighting naive Bayes (awnb, 22.2%). Other classical methods with low performance are learning vector quantization (lvq, 16.6%) and Gaussian process regression (gpr, 6.2%).

Considering the proposed methods fpv, fps, mrf, mrf5, frp, nn, nnc and fnn, they are all within the first 13 highest kappa values in Table 3, overcoming the vast majority of the standard classifiers excepting nbDiscrete, earth, rpart and nb. Specifically, fps, fpv and frp achieve the three highest kappa

values, and fnn is in the fifth position. Indeed, the 10 best classifiers include 6 proposed by us. The good result of fps (feature population sum, kappa 60.5%) reveals that classification based on populations of training vectors with each feature value is effective for this problem. The sum of populations in fps outperforms both the voting in fpv and the feature relevance in frp (feature relevance times population). This may happen because fps first sums the feature contributions (relative populations) of each class and then makes a prediction, while fpv first makes a class prediction for each feature and then sums the votes to make a final prediction. Thus, fps is able to compensate among features, being less prone to make mistakes than voting in fpv. The frp uses class populations and feature relevances, achieving less kappa than fps, so that relevances may be confusing classification. This is also confirmed by the lower positions, 9 and 12, achieved by mrf and mrf5, respectively. Besides, the voting among the 5 most reliable features in mfr5 reduces performance compared to mrf, that uses only the feature with the highest reliability. Appendix F in the supplementary material presents, for the avoiding dataset, the feature relevance values defined by section 5 and by other methods in the literature. These feature relevances are relatively coherent identifying the features that most influence classification.

The proposed fnn (feature nearest neighbor) also achieves a good (5th) position, with kappa 54.8%, outperforming nn and nnc. This fact suggests that information about class population helps the neighbor selection for this problem. The remaining proposed neighbor classifiers (nn and nnc, 47.4% and 40.5%, respectively) also outperform the standard neighbor methods (knn, 28.4%), so the distance $d_n(\mathbf{x})$ proposed in section 4 reveals more effective than standard Euclidean distance.

6.1 Feature selection

We also used 22 popular methods of feature selection (FS) listed in Table 3 of the supplementary material, including the feature relevance in (4), in order to find out whether some selected group of relevant features might achieve better performance than the whole set of features. To avoid an excessively large number of experiments, FS was applied with the best classifier, fps, and other 6 classifiers selected by their good performances: nbDiscrete, earth and rpart in R (positions 4, 6 and 7 in Table 3, respectively); kfd and mlp in Python (positions 13 and 14 in that table); and rforest in Matlab (position 15). In each LOO trial, the FS method is applied on the training set in order to rank features by decreasing importance. Backward and forward sequential feature selection (bsfs and fsfs, respectively) are well-known FS methods [41] that search sequentially a pre-specified number or features in order to maximize the performance of a classifier, that in the Matlab implementation used by us is a Bayesian classifier. The forward version (fsfs) starts from the scratch and iteratively adds features that raise the classifier performance until the number is reached. The backward version (bsfs) starts with all the I features and iteratively removes those when it raises the performance, stopping with the desired number of features. Both are executed for $i = 1, \dots, I$, selecting i features in iteration i . The importance of a feature is given by the number of times that it is selected.

Once the feature ranking is created, the classifier is used to classify the first i features in the ranking, for $i = 1, \dots, I$, and the performance is evaluated for each value of i . Some FS methods (e.g. sfprc2/sfprfc, sfdrc2/sfdrcf and sfwec2/sfwefc) do not accept a number of features to be selected. Instead, they require another parameter, such as a threshold α on the

Table 4 Best result (kappa and accuracy, in %) achieved by the best classifier (upper part) and by the best classifier with feature selection (lower part) for each dataset

	avoiding	registration	seeking	sensitivity	total	touch
Classification						
Kappa	60.5	41.4	32.5	49.6	37.9	25.0
Accuracy	82.7	76.9	78.8	78.8	75.0	71.2
Classifier	fps	mrf	mlp	earth	fps	lda
Classification and feature selection						
Kappa	68.0	57.0	43.5	53.2	53.2	42.0
Accuracy	86.5	82.7	80.8	80.7	80.8	73.1
Classifier	fps	fps	mlp	rf	fps	fps
Feature selection	bsfs	chi2	bsfs	skbfc	bsfs	bsfs
Language	Matlab	Python	Python	Python	Matlab	Matlab

Table 5 Confusion matrix achieved by fps-bsfs on dataset *avoiding*

Class True	Predicted	
	ASD	ASD/ADHD
ASD	33	5
ASD/ADHD	2	12

p -values of a chi-squared test or ANOVA F-value, so that only features with p -value $> \alpha$ are selected. The select scoring percentile (sp) method uses a percentile of features, with values from 10% to 100%, with step 10, of the features.

Table 7 of Appendix H in the supplementary material reports the whole results achieved with the selected classifiers and feature selection methods. Table 4 compiles, for each dataset, the kappa and accuracy achieved by the best classifiers without FS (upper part, same values as in Table 2 above). The lower part of the table reports kappa and accuracy achieved using the best combination of classifier and FS method. The last row specifies the implementation language. The FS increases kappa from 60.5% to 68% in dataset *avoiding* using bsfs (backward sequential feature selection) selecting the 17 most relevant features, and accuracy raises from 82.7% to 86.5%. The FS methods also overcome the best classifier's performances without FS in the remaining datasets. Figure 3 in the supplementary material shows the box plot comparing the the kappa achieved by the 7 selected classifiers without and with the best FS method over the 6 datasets. Note that feature selection increases the kappa values for all the classifiers, although only fps-bsfs overcomes 60%.

Table 5 reports the confusion matrix corresponding to the best result (kappa=68%, accuracy=86.5%), achieved by fps-bsfs on dataset *avoiding*. According to the scale defined by [40], this corresponds to a good agreement between true and predicted values. The ASD/ADHD class is considered as positive, so that TN (no. of true negatives)=33, FP (false positives)=5, FN (false negatives)=2 and TP (true positives)=12. This matrix reflects a reliable classification because the numbers outside the main diagonal are much below the numbers inside the main diagonal. Specifically, FP (with value 5) represents 41% of TP and 15% of TN, while FN (with value 2) represents 17% of TP and 6% of TN.

Table 6 Performance metrics (in %) achieved by fps-bsfs on dataset *avoiding* (see text for details)

Kappa	Accuracy	Recall/TPR/ Sensitivity	Precision/ PPV	Specificity	Balanced accuracy
68.0	86.5	86.5	70.6	86.8	86.3
FPR	FNR	F-score/F1	AUC (%)	Y-index	MCC (ϕ)
13.2	14.3	77.4	84.2	72.6	68.6

Table 6 reports other performance metrics for binary classification achieved by fps-bsfs: recall, a.k.a. true positive rate (TPR) and sensitivity; precision or positive predictive value (PPV); specificity; balanced accuracy; false positive rate (FPR); false negative rate (FNR); F1, equal to F-score for $\beta=1$; area under ROC curve (AUC); Youden index; and Mathews correlation coefficient (MCC), also known as Yule ϕ . Appendix E of the supplementary material defines these metrics.

The fps-bsfs correctly classifies 33+12=45 of 52 participants with an accuracy of 86.5% (Table 6) and a balanced accuracy of 86.3%. The prediction is wrong for the remaining 7 participants due to: 1) the inherent complexity of the classification problem, where similarity is large between feature vectors of both classes; and 2) machine learning classifiers can only learn approximately a classification problem, and can always make prediction errors. Recall is very high (86.5%) due to the low FN (2) compared to TP (12) in the confusion matrix (Table 5). Precision is lower (70.6%) because FP (5) is higher than FN. However, the specificity is higher (86.8%) because FP=5 is low compared to TN (33). This low FP also leads to a low FPR (13.2%). Since FN=2, only 2 participants of class ASD/ADHD are erroneously classified as ASD, giving a low FNR (14.3%). The F1, that combines recall and precision, is also high (77.4%). The Youden index and MCC are relatively high, 72.6% and 68.6%, respectively.

Figure 3 (left panel) plots the receiving operating curve (ROC) of fps-bsfs on dataset *avoiding*. The curve contains points fairly near to the upper left corner, that identifies the perfect classification. In fact, the area under curve (AUC) in Table 6 is also high (84.2%). The red square is the working point corresponding to the confusion matrix of Table 5, located fairly near to the upper left corner. The right panel of Fig. 3 shows the lift chart, see [42] p. 265, that plots the cumulative percentage of ASD/ADHD vectors that are correctly labeled versus the percentage of samples tested. The perfect (resp. random) classification corresponds to the left (resp. right) side of the shadowed triangle, whose upper side has length equal to the percentage of vectors of the minority class (ASD/ADHD). The classifier fps-bsfs is represented by the black line inside this triangle, that is fairly near to the left side, excepting in the higher percentages of tested samples because 5 ASD participants, that represent 9.6% of the total, are erroneously predicted as ASD/ADHD.

Fig. 3 Receiver operating curve (ROC, left panel, the red square is the working point) and lift chart (right panel) of classifier fps-bsfs on dataset avoiding

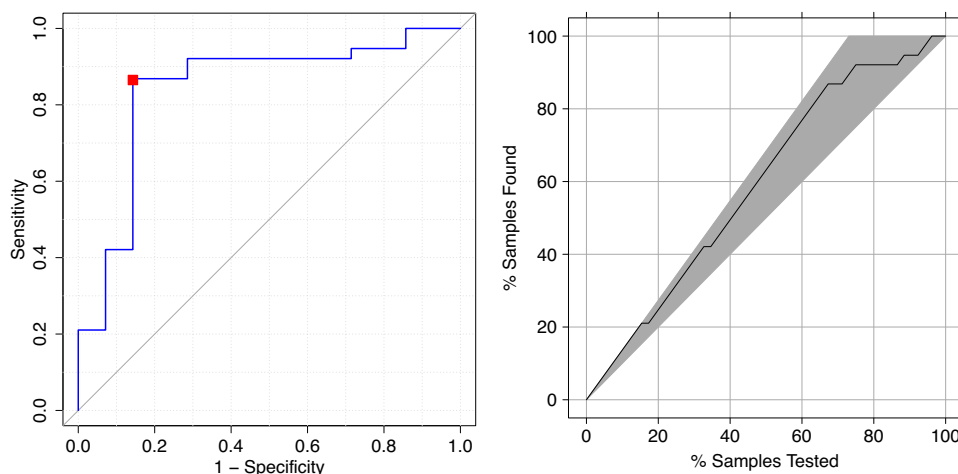


Table 7 shows a case example of prediction made by fps-bsfs for a test vector. The participant is identified by a vector \mathbf{x} composed of $I=17$ features selected by BSFS out of the 20 features in dataset `avoiding`. The names of the selected features are listed by the first column of Table 7. The feature values are listed in the second column. First, for each input the probabilities $P_{ik} = P_{i|j_i k}$ (2), calculated during training, are selected for $k=1$ for class ASD and $k=2$ for class ASD/ADHD. Here j_i is the i -th feature of \mathbf{x} . The values of

Table 7 Classification of FPS with 17 features selected by BSFS on `avoiding`

Feature	Value	Class probability		Prediction
		ASD	ASD/ADHD	
SP2	5	0.26	0.54	ASD/ADHD
SP5	5	0.16	0.46	ASD/ADHD
SP15	1	0.53	0.23	ASD
SP58	5	0.29	0.62	ASD/ADHD
SP61	3	0.21	0.15	ASD
SP63	1	0.16	0.00	ASD
SP64	2	0.13	0.08	ASD
SP65	5	0.13	0.54	ASD/ADHD
SP66	3	0.00	0.00	NONE
SP67	3	0.11	0.23	ASD/ADHD
SP68	2	0.26	0.08	ASD
SP70	5	0.24	0.62	ASD/ADHD
SP71	5	0.13	0.23	ASD/ADHD
SP72	5	0.29	0.38	ASD/ADHD
SP74	5	0.47	0.62	ASD/ADHD
SP75	5	0.50	0.69	ASD/ADHD
SP81	1	0.32	0.38	ASD/ADHD
Probability sum:		1.29	5.31	
Class probability:		0.19	0.81	ASD/ADHD

In bold the highest probability for each feature

P_{ik} for $i = 1 \dots, I$ and $k = 1, 2$ are reported by columns 3 and 4. Then, each feature i predicts the class $y_i(\mathbf{x}) \in \{1, 2\}$ with the highest P_{ik} as in (3). The prediction of each feature is listed in column 5, e.g. features SP2 and SP15 predict ASD/ADHD and ASD, respectively. Note that feature SP66 has no prediction because no training vector had the value 3, so for this test vector SP66 gives no information about class populations. For each class k , the probabilities P_{ik} are summed over the features i that predict class k . These sums are listed in the row “Probability sum” of the table, with values 1.29 and 5.31 for classes 1 and 2 (only bold values are summed for each class). The class probabilities P_k are $P_1/(P_1 + P_2)$ and $P_2/(P_1 + P_2)$ for $k=1,2$, with values 0.19 and 0.81. Finally, the predicted class is the one with the largest P_k , with probability or “certainty” 0.81 (last row of the table).

Subsequently, the participant with feature vector \mathbf{x} was predicted as ASD/ADHD because his SP2 questions 2, 5, 58, 65, 67, 70, 71, 72, 74, 75 and 81 had values that in the training are more probable for participants of class ASD/ADHD than of class ASD, while only SP2 questions 15, 61, 63, 64 and 68 had values more probable for participants of class ASD than of class ASD/ADHD. Summing the probabilities of the features that predict each class, ASD and ASD/ADHD achieve probabilities of 0.19 and 0.81 respectively, so the class with the highest sum (ASD/ADHD) was predicted with probability 0.81.

6.2 Unbalanced classification

Given the class unbalance (ASD has 2.71 times more vectors than ASD/ADHD), we applied 19 methods for unbalanced classification (see Table 5 of the supplementary material) in order to explore whether they increase performance with respect to classifiers alone and to classifiers with feature selection. Some of the unbalanced methods are classifiers specifically devoted to unbalanced classification, while others are sampling methods that were used with the 7 selected

classifiers selected in the previous sections: fps in Octave; nbDiscrete, earth and rpart in R; kfd and mlp in Python; and rforest in Matlab. The three Octave classifiers in the upper part of Table 5 of the supplementary material are designed for unbalanced classification [43]. The Matlab methods include random undersampling boosting (RUSBoost), an ensemble that tunes the number of base classifiers with prior probabilities determined from the class populations [44], and synthetic minority over-sampling technique, SMOTE [45], a well-known sampling method for unbalanced classification. In Python, we used the 7 classifiers combined with three SMOTE versions implemented by package `imbalanced-learn` [46]. We also used the SVM with class weights (named `wsvm` in Table 5 of the supplementary material) inversely proportional to the class population, in order to take into account the class unbalance. In R, the 7 classifiers were combined with several sampling methods implemented by the `ROSE` package [47], including under- and oversampling, randomly oversampling examples, `ROSE` [48] and several SMOTE variants.

Table 8 in Appendix H of the supplementary material reports the kappa of all the unbalanced classifiers and datasets. The lower part of Table 8 compiles the best results achieved on each dataset by an unbalanced classifier or by the selected classifiers combined with a sampling method. For comparison, the middle part of the table reports the best values achieved by these classifiers with feature selection (already listed in Table 4), and the upper part reports the best classification without feature selection nor unbalanced classification (already listed in Table 2). The performance using unbalanced classification is slightly higher than only classification (upper part of the table), excepting dataset `seeking` where it is lower, and it is clearly lower than using feature

selection in all the datasets. Although unbalanced classification also achieves the best kappa (60.5%, equal to the upper part) for dataset `avoiding` and the same classifier (fps) as in upper and middle part, it does not overcome the globally best kappa (68%), achieved by fps with feature selection (bsfs). The reason might be the low degree of unbalance (class ASD only doubles population of class ASD/ADHD). The sampling methods `smote` and `rose` seem better than the others, achieving the best results in two datasets each one. Considering languages, four of six datasets achieve the best results with methods implemented in R. The classifiers designed for unbalanced classification, such as `adaboostM1`, `samme`, `adaboostNC`, `rusboost` and `wsvm`, did not outperform the original classification (upper part of Table 4) nor achieved the best result in any dataset.

6.3 Clinical discussion

Evidence supports an overlap of sensory processing difficulties in ASD and ADHD [17]. The early identification of the co-occurrence of these difficulties in these disorders may be useful to predict ADHD comorbidities in children and adolescents with ASD. This is even more relevant considering the increased negative impact of comorbid ASD/ADHD on children's functioning outcomes compared to ASD or ADHD alone, see [8–10]. Using distinct machine learning methods, the current study aimed at predicting ADHD comorbidities in children and adolescents diagnosed with ASD through atypical sensory processing. The experiments showed that machine learning methods can be an important and useful approach contributing to the early detection of ADHD comorbidity with ASD considering sensory processing. Comorbid ADHD in ASD tend to be detected after the

Table 8 Best kappa and accuracy (both in %) and classifier for each dataset: without FS nor unbalanced classification (upper part); with a FS method (middle part); with an unbalanced classification method (lower part)

	avoiding	registration	seeking	sensitivity	total	touch
Classification						
Kappa	60.5	41.4	32.5	49.6	37.9	25.0
Accuracy	82.7	76.9	78.8	78.8	75.0	71.2
Classifier	fps	mrf	mlp	earth	fps	lda
Classification and feature selection						
Kappa	68.0	57.0	38.6	49.6	53.2	42.0
Accuracy	86.5	82.7	76.9	78.8	80.8	73.1
Classifier	fps	fps	mlp	rf	fps	fps
FS method	bsfs	chi2	bsfs	skbfc	bsfs	bsfs
Unbalanced classification						
Kappa	60.5	42.9	20.3	49.6	40.6	35.5
Accuracy	82.7	71.1	75.0	78.8	69.2	76.9
Classifier	fps	rpart	fps	earth	earth	kfd
Unbalanced	ous	rose	smote	smote	rose	rsls
Language	R	R	R	Matlab	R	Python

ASD diagnosis [49]. As ASD is commonly detected around 3.5 years old [50] comorbid ADHD diagnosis may take years being diagnosed from 4.8 years old [50] to 8 years old [51]. In this study, we used the SP2 questionnaire that can be applied from the age of three-years old. We found that not only SP2 was able to predict ASD/ADHD, but also to do it several years before the diagnosis is made.

The results showed that sensory avoiding, registration and sensitivity quadrants presented more ability to predict ASD/ADHD. These findings are in line with previous studies demonstrating that children with ASD/ADHD present increased overall sensory processing difficulties compared to children with ASD or ADHD alone [21, 22]. Indeed, these three sensory quadrants reflect a child's increased difficulty in dealing with environment stimuli, which is in accordance with other findings [52, 53]. Particularly, sensory avoidance is indicative of children that retreat from situations and/or limit environmental sensory experiences (e.g., a child may refuse to being cuddled). Our results proved that selected questions from the avoiding quadrant of the SP2 questionnaire, filled by parents at low ages, selected by the backwards sequential feature selection, allow the feature population sum classifier to achieve an early ASD/ADHD comorbidity prediction with high accuracy (86.5%) and low false positive and negative rates (below 15%). The registration quadrant refers to difficulties related to later reactions to sensory experiences (e.g., taking longer to waking up in the morning). The sensitivity quadrant is associated with more intense reactions or responses to sensory stimuli that neurotypical children (e.g., easily startled by sounds).

Atypical touch processing and sensorial seeking quadrant were the two sensorial outcomes that were not able to discriminate between ASD/ADHD and ASD patients. This evidences that atypical tactile processing is one of the earliest and most commonly sensory alterations observed in ASD [52]. Besides, sensorial seeking (i.e., the need for additional sensory input from the environment, e.g., enjoying loud music) is more frequently reported in ADHD [54]. It may be possible that these two sensorial dimensions are more characteristic of ASD and ADHD alone, respectively, and accordingly not indicative of ADHD comorbidity in children and adolescents with ASD.

Our study suggests that atypical sensory processing is an important aspect to consider when characterizing and/or differentiating children with ASD, ADHD, and ASD/ADHD. Knowing that increased and overall sensory processing difficulties cascade into greater behavioral and adaptive problems, especially for children and adolescents with comorbid ASD/ADHD [7, 26], the distinct pattern of sensory processing difficulties in these disorders could assist an early and differential diagnosis. Because the SP2 questionnaire can be administered from early ages (i.e., starting from 3

years of age), atypical sensory processing can be used as an early indicator of comorbidity between these neurodevelopmental disorders as it encloses a possible predictive value of ADHD comorbidity with ASD. As a result, important implications for intervention programs can be draw. For example, altered sensory processing abilities could be considered in interventional approaches that include adapted sensory-based therapies for children with ASD, ADHD or ASD/ADHD. This can contribute to improve their adaptive functioning to environmental situations. The current study has limitations imposed by the low sample size. As well, eventual errors filling the answers to the SP2 questionnaire might bias the prediction. However, the impact of this bias on our experimentes are expected to be low because: 1) the experimental methodology employed, leave-one-out cross validation, is specifically designed for small-sample problems like this one, and it has been proven that it measures reliably the robustness of the classifiers; and 2) the errors in the answers, whose probability of occurrence is very low because the questionnaire has been previously validated and, if present, they will be both in the training and test data, so the evaluation of the classifier performance using leave-one-out cross validation already takes them into account.

7 Conclusion

In this study we have developed an algorithm to predict the presence of comorbid ADHD in youth and adolescents with ASD as young as 3 years old based on their sensory processing. To do so, we used the scores obtained in the questions of the Sensory Profile-2 (SP2) questionnaire. These questions are independent among them and have discrete integer values between 1 and 5. This kind of classification problems appear frequently in healthcare and other domains. Features were grouped into five standard combinations, of which four are sensory processing quadrants (avoiding, registration, seeking and sensitivity) and the other was a sensory modality (touch). Since the diagnosis of ASD/ADHD comorbidity is complex and usually performed at older ages, the number of participants enrolled in our study is small.

This prediction was performed using 42 standard classifiers of different families. We also proposed other 8 classifiers, that perform a prediction for a test data based on the class and feature values (supposed to be discrete and numeric) of the training data. Several proposed methods achieve the best performances on ADHD classification. Specifically, the feature population sum (fps) classifier achieved the highest kappa (60.5%) and accuracy (82.7%) using the avoiding quadrant. The best kappa achieved by standard classifiers is 55.2% (accuracy of 80.8%) using discrete naive Bayes, and 49.6% (78.8% of accuracy) using recursive partitioning trees (rpart) and multiadaptive regression splines (earth). Given the

high number (86) of features used, and the interest to know what are the most relevant ones, we also used a collection of 22 feature selection methods combined with the best classifier (fps) and other six standard classifiers selected by their good performance. Using the backwards sequential feature selection (bsfs), fps raised its performance until kappa of 68% and accuracy of 86.5%, very near to perfect classification with low ratios of false positives (13.2%) and negatives (14.3%). We also used 19 methods of unbalanced classification, but they no longer increased the performance, probably due to the low level of class unbalance (three to one). For a future full-scale practical application, the best classification model (fps) must be trained using only the 17 questions selected by the best feature selection method (bsfs) from the quadrant *avoiding* of the SP2 questionnaire. Once trained, the classifier is ready to make a prediction for new participants using their scores for these questions.

From a clinical perspective, the SP2 questionnaire is a useful tool as it is completed by caregivers, who are in the strongest position to indicate the child's response to sensory interactions that occur in everyday life, as early as three years of age. This information can assist early and more accurate ASD and/or ADHD diagnosis established by qualified professionals, and provides valuable information to a comprehensive assessment of the child and adolescent's sensory strengths and challenges and, consequently, to develop effective treatment plans, interventions, and everyday remediation strategies to support children's participation in all contexts (e.g., school). Future studies should validate the proposed models with a larger sample of participants. Equally, these models can be applied to other behavioral dimensions (e.g., social difficulties) where features are also discrete numeric answers to questionnaires, to further help more accurate diagnosis.

The co-occurrence of ASD and ADHD is well-documented. The overlap of symptoms between these disorders (e.g., inattention) poses challenges for an early and accurate diagnosis. Children and adolescents with comorbid ASD/ADHD experience greater difficulties in social skills [8], adaptive functioning [10] and quality of life [11] compared to children and adolescents with only one of these disorders, which enclosed long-term negative consequences [15]. This study contributes to the literature by showing that the comorbidity between ASD and ADHD involves greater sensory processing difficulties. In addition, it showed that atypical sensory responses may be an important differentiating factor between these disorders at early ages.

Supplementary information The supplementary material includes the classifier list (Appendix A), hyper-parameter tuning of standard classifiers (Appendix B), pseudo-code of population-based methods (Appendix C), list of feature selection methods used (Appendix D), list of performance

metrics for binary classification (Appendix E), feature ranking (Appendix F), list of unbalanced classification methods (Appendix G) and detailed results of the experimental work (Appendix H).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10489-024-05655-z>.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work has received financial support from the Consellería de Educación, Universidade e Formación Profesional (accreditation 2019-2022 ED431G-2019/04) and the European Regional Development Fund (ERDF), which acknowledges the CiTIUS (Centro Singular de Investigación en Tecnoloxías Intelixentes da Universidade de Santiago de Compostela) as a Research Center of the Galician University System. Sara Cruz is supported by the Psychology for Development Research Center, Lusíada University, Portugal, supported by FCT - Fundação para a Ciência e Tecnologia, I.P., by project reference UIDB/04375/2020 and DOI identifier <10.54499/UIDB/04375/2020 (<https://doi.org/10.54499/UIDB/04375/2020>)>. Adriana Sampaio is supported by the Psychology Research Center (PSI/01662), School of Psychology, University of Minho, through the Foundation for Science and Technology (FCT) and the Portuguese State Budget (Ref. UIDB/PSI/01662/2020). María Tubío-Fungueiriño, Angel Carracedo, and Montse Fernández-Prieto were funded by Instituto de Salud Carlos III (projects PI19/00809 and PI22/00208 to Angel Carracedo) and co-funded by European Union (ERDF) "A way of making Europe", and by Fundación María José Jove.

Data availability statement Data used on this study are of medical nature and therefore private, so they can not be publicly available.

Declarations

Conflict of interest There are no potential competing interests in our paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Cleaton M, Kirby A (2018) Why do we find it so hard to calculate the burden of neurodevelopmental disorders? *J Childhood Development Disorders* 4(10). <https://doi.org/10.4172/2472-1786.100073>

2. Berenguer C, Roselló B, Colomer C, Baixauli I, Miranda A (2018) Children with autism and attention deficit hyperactivity disorder. Relationships between symptoms and executive function, theory of mind, and behavioral problems. *Res Dev Disabil* 83:260–269. <https://doi.org/10.1016/j.ridd.2018.10.001>
3. Rong Y, Yang CJ, Jin Y, Wang Y (2021) Prevalence of attention-deficit/hyperactivity disorder in individuals with autism spectrum disorder: a meta-analysis. *Res Autism Spectr Disord* 83:101759. <https://doi.org/10.1016/j.rasd.2021.101759>
4. APA (2022) Diagnostic and statistical manual of mental disorders. Am Psychiat Assoc, Washington DC. <https://doi.org/10.1176/appi.books.9780890425596>
5. Hansen BH, Oerbeck B, Skirbekk B, Petrovskian BE, Kristensen H (2018) Neurodevelopmental disorders: prevalence and comorbidity in children referred to mental health services. *Nord J Psychiat* 72(4):285–291. <https://doi.org/10.1080/08039488.2018.1444087>
6. Harkins CM, Handen BL, Mazurek MO (2021) The impact of the comorbidity of ASD and ADHD on social impairment. *J Autism Dev Disord* 52:2512–2522. <https://doi.org/10.1007/s10803-021-05150-1>
7. Roselló R, Martínez-Raga J, Mira A, Pastor JC, Solmi M, Cortese S (2022) Cognitive, social, and behavioral manifestations of the co-occurrence of autism spectrum disorder and attention-deficit/hyperactivity disorder: a systematic review. *Autism* 26(4):743–760. <https://doi.org/10.1177/13623613211065545>
8. Avni A, Ben-Itzhak E, Zachor DA (2018) The presence of comorbid ADHD and anxiety symptoms in autism spectrum disorder: clinical presentation and predictors. *Front Psychiat* 9. <https://doi.org/10.3389/fpsy.2018.00717>
9. Sprenger L, Bühler E, Poustka L, Bach C, Heinzel-Gutenbrunner M, Kamp-Becker I, Bachmann C (2013) Impact of ADHD symptoms on autism spectrum disorder symptom severity. *Res Devel Disabil* 34(10):3545–3552. <https://doi.org/10.1016/j.ridd.2013.07.028>
10. Scandurra V, Emberti-Gialloreti L, Barbanera F, Scordo MR, Pierini A, Canitano R (2019) Neurodevelopmental disorders and adaptive functions: a study of children with autism spectrum disorders (ASD) and/or attention deficit and hyperactivity disorder (ADHD). *Front Psychiat* 10. <https://doi.org/10.3389/fpsy.2019.00673>
11. Sikora DM, Vora P, Coury DL, Rosenberg D (2012) Attention deficit/hyperactivity disorder symptoms, adaptive functioning, and quality of life in children with autism spectrum disorder. *Pediatrics* 130(Suppl. 2):91–97. <https://doi.org/10.1542/peds.2012-0900G>
12. Rao PA, Landa RJ (2014) Association between severity of behavioral phenotype and comorbid attention deficit hyperactivity disorder symptoms in children with autism spectrum disorders. *Autism* 18(3):272–280. <https://doi.org/10.1177/1362361312470494>
13. Rowena N, Kimberley H, Elise H (2021) Associations between ADHD subtype symptomatology and social functioning in children with ADHD, autism spectrum disorder, and comorbid diagnosis: utility of diagnostic tools in treatment considerations. *J Atten Disord* 25(6):820–828. <https://doi.org/10.1177/1087054719855680>
14. Jang J, Matson JL, Williams LW, Tureck K, Goldin RL, Cervantes PE (2013) Rates of comorbid symptoms in children with ASD, ADHD, and comorbid ASD and ADHD. *Res Devel Disabil* 34(8):2369–2378. <https://doi.org/10.1016/j.ridd.2013.04.021>
15. Houghton S, Alsalmi N, Tan C, Taylor M, Durkin K (2017) Treating comorbid anxiety in adolescents with ADHD using a cognitive behavior therapy program approach. *J Atten Disord* 21(13):1094–1104. <https://doi.org/10.1177/1087054712473182>
16. Manohar H, Kuppili PP, Kandasamy P, Chandrasekaran V, RP R (2018) Implications of comorbid ADHD in ASD interventions and outcome: results from a naturalistic follow up study from South India. *Asian J Psychiatr* 33:68–73. <https://doi.org/10.1016/j.ajp.2018.03.009>
17. Little LM, Dean E, Tomchek S, Dunn W (2018) Sensory processing patterns in autism, attention deficit hyperactivity disorder, and typical development. *Phys Occup Ther Pediatr* 38(3):243–254. <https://doi.org/10.1080/01942638.2017.1390809>
18. Dellapiazza F, Michelon C, Vernhet C, Muratori F, Blanc N, Picot MC, Baghdadli A (2020) Sensory processing related to attention in children with ASD, ADHD, or typical development: results from the ELENA cohort. *Eur Child Adolesc Psychiat* 30:283–291. <https://doi.org/10.1007/s00787-020-01516-5>
19. Thye MD, Bednarz HM, Herringshaw AJ, Sartin EB, Kana RK (2018) The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Dev Cogn Neurosc* 29:151–167. <https://doi.org/10.1016/j.dcn.2017.04.010>
20. Ghanizadeh A (2011) Sensory processing problems in children with ADHD, a systematic review. *Psychiat Invest* 8(2):89–94. <https://doi.org/10.4306/pi.2011.8.2.89>
21. Rommelse N, Visser J, Hartman C (2018) Differentiating between ADHD and ASD in childhood: some directions for practitioners. *Adolesc Psychiat* 27:679–681. <https://doi.org/10.1007/s00787-018-1165-5>
22. Sanz-Cervera P, Pastor-Cerezuola G, González-Sala F, Tárraga-Mínguez R, Fernández-Andrés MI (2017) Sensory processing in children with autism spectrum disorder and/or attention deficit hyperactivity disorder in the home and classroom contexts. *Front Psychol* 11:1–12. <https://doi.org/10.3389/fpsyg.2017.01772>
23. Richens JG, Lee CM, Johri S (2020) Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* 11:3923. <https://doi.org/10.1038/s41467-020-17419-7>
24. Song JW, Yoon NR, Jang SM, Lee GY, Kim BN (2020) Neuroimaging-based deep learning in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Soa Chongsongyon Chongsin Uihak* 31(3):97–104. <https://doi.org/10.5765/jkacap.200021>
25. Schulte-Rüther M, Kulvicus J, Stroth S, Wolff N, Roessner V, Marschik PB, Kamp-Becker I, Poustka L (2022) Using machine learning to improve diagnostic assessment of ASD in the light of specific differential and co-occurring diagnoses. *J Child Psychol Psychiat*. <https://doi.org/10.1111/jcpp.13650>
26. Alateyat H, Cruz S, Cernadas E, Tubío-Fungueiriño M, Sampaio A, González-Villar A, Carracedo A, Fernández-Delgado M, Fernández-Prieto M (2022) A machine learning approach in autism spectrum disorders: from sensory processing to behavior problems. *Front Mol Neurosci* 15. <https://doi.org/10.3389/fnmol.2022.889641>
27. Duda M, Ma R, Haber N (2016) Use of machine learning for behavioral distinction of autism and ADHD. *Transl Psychiat* 6(e732):1–5. <https://doi.org/10.1038/tp.2015.221>
28. Constantino JN (2002) The social responsiveness scale. Western Psychological Services, Los Angeles, CA
29. Wall DP, Kosmicki JA, DeLuca TF, Harstad EB, Fusaro VA (2012) Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl Psychiat* 2(e100):1–9. <https://doi.org/10.1038/tp.2012.10>
30. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL, Luyster RJ, Guthrie W (2008) ADOS. Escala de Observación Para el Diagnóstico del Autismo, TEA Editions, Madrid (ES)
31. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F (2017) Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical* 17:16–23. <https://doi.org/10.1016/j.nicl.2017.08.017>
32. Abdullah A, Rijal S, Dash SR (2019) Evaluation on machine learning algorithms for classification of autism spectrum disorder (ASD). *J Phys: Conf Ser* 1372. <https://doi.org/10.1088/1742-6596/1372/1/012052>
33. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E (2001) The autism-spectrum quotient (AQ): evidence from

- Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* 31(1):5–17. <https://doi.org/10.1023/a:1005653411471>
34. Rutter M, Couteur AL, Lord C (2006) ADI-R: Entrevista Para el Diagnóstico del Autismo-revisada. TEA Editions, Madrid (ES)
 35. Dunn W (2014) *Sensory Profile 2 Manual*. Pearson, San Antonio, TX
 36. Mikkelsen M, Wodka EL, Mostofsky SH, Puts NAJ (2018) Autism spectrum disorder in the scope of tactile processing. *Dev Cogn Neurosci* 29:140–150. <https://doi.org/10.1016/j.dcn.2016.12.005>
 37. Piccardi ES, Begum AJ, Jonesm EJH, Mason L, Charman T, al MHJ (2021) Behavioural and neural markers of tactile sensory processing in infants at elevated likelihood of autism spectrum disorder and/or attention deficit hyperactivity disorder. *J Neurodev Disord* 13(1):1–18. <https://doi.org/10.1186/s11689-020-09334-1>
 38. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Intl J Conf Artif Intel (IJCAI)*, Montreal (Canadá), pp 1137–1143
 39. Landis J, Koch G (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174. <https://doi.org/10.2307/2529310>
 40. Blackman N, Koval J (2000) Interval estimation for Cohen's kappa as a measure of agreement. *Stat Med* 19(5):723–741. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000315\)19:5<723::AID-SIM379>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0258(20000315)19:5<723::AID-SIM379>3.0.CO;2-A)
 41. Ferri F, Pudil P, Hatef M, Kittler J (1994) Comparative study of techniques for large-scale feature selection. *Mach Intel Patt Rec* 16:403–413. <https://doi.org/10.1016/B978-0-444-81892-8.50040-7>
 42. Kuhn M, Johnson K (2013) *Applied predictive modeling*. Springer, New York (NJ)
 43. Zhang C, Bi J, Xu S, Ramentol E, Fan G, Qiao B, Fujita H (2019) Multi-imbalance: an open-source software for multi-class imbalance learning. *Knowl-Based Syst* 174:137–143. <https://doi.org/10.1016/j.knsys.2019.03.001>
 44. Seiffert C, Khoshgoftaar T, Hulse JV, Napolitano A (2008) RUSBoost: improving classification performance when training data is skewed. In: *19th Intl Conf Pattern Recogn*, pp 1–4. <https://doi.org/10.1109/ICPR.2008.4761297>
 45. Chawla W, Bowyer K, Hall L, Kegelmeyer W (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intel Res* 16:321–357. <https://doi.org/10.1613/jair.953>
 46. Lemaitre G, Nogueira F, Aridas CK (2016) Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *arXiv*. <https://doi.org/10.48550/ARXIV.1609.06570>
 47. Lunardon N, Menardi G, Torelli N (2014) ROSE: a package for binary imbalanced learning. *R Journal* 6(1):82–92. <https://doi.org/10.32614/RJ-2014-008>
 48. Menardi G, Torelli N (2014) Training and assessing classification rules with imbalanced data. *Data Min Knowl Disc* 28:92–122. <https://doi.org/10.1007/s10618-012-0295-5>
 49. Sainsbury WJ, Carrasco K, Whitehouse AJO, McNeil L, Waddington H (2022) Age of diagnosis for co-occurring autism and attention deficit hyperactivity disorder during childhood and adolescence: a systematic review. *Rev J Autism Dev Disord*. <https://doi.org/10.1007/s40489-022-00309-7>
 50. Hatakenaka Y, Fernell E, Sakaguchi M, Ninomiya H, Fukunaga I, Gillberg C (2016) ESSENCE-Q - a first clinical validation study of a new screening questionnaire for young children with suspected neurodevelopmental problems in South Japan. *Neuropsych Dis Treat* 12:1739–1746. <https://doi.org/10.2147/ndt.s108411>
 51. Jensen CM, Steinhausen H-C (2015) Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *ADHD Atten Def Hyp Disord* 7:27–38. <https://doi.org/10.1007/s12402-014-0142-1>
 52. Dellapiazza F, Audras-Torrent L, Michelon C, Baghdadli A (2021) Clinical characteristics of children with ASD and comorbid ADHD: association with social impairment and externalizing and internalizing behaviours. *Res Dev Disabil* 113:103930. <https://doi.org/10.1016/j.ridd.2021.103930>
 53. Mattard-Labrecque C, Amor LB, Couture MM (2013) Children with autism and attention difficulties: a pilot study of the association between sensory, motor, and adaptive behaviors. *J Can Acad Child Adolesc Psychiatr* 22(2):139–146. <https://doi.org/10.1007/s00787-020-01516-5>
 54. Rani I, Agarwal V, A A, P M (2023) Sensory processing in children and adolescents with attention deficit hyperactivity disorder. *J Atten Disord* 27(2):145–151. <https://doi.org/10.1177/10870547221129306>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Manuel Fernández-Delgado was born in Vimianzo, A Coruña, Spain, in 1971. He studied Physics and became doctor in Computer Science at the University of Santiago de Compostela in 1994 and 1999, respectively. He is lecturer of Computer Science and researcher in machine learning and artificial intelligence in this university.



Sara Cruz PhD, develops research develops research activities in the field of developmental Psychology, with a particular interest in the interplay between brain, cognition and behaviour in typical and atypical development, in infancy and childhood.



Eva Cernadas was born in A Coruña, Spain, in 1969. She studied Physics and became doctor in Computer Science at the University of Santiago de Compostela, in 1992 and 1997, respectively. Nowadays she is a lecturer in Computer Science and researcher of this university about image processing and pattern recognition, mainly applied in the food technology, biological and medical domains.



Adriana Sampaio MD, PhD, her main research is in the field of developmental cognitive neuroscience, namely in the structural neurofunctional and genetic basis of (ab)normal sensorial-perceptual processing as a scaffold for cognitive and emotional processes.



Heba Alateyat achieved its BS (2001) in Balqa Applied University (Jordan) and PhD and Computer Science (2023) in University of Santiago de Compostela (Spain). She is a lecturer on Computer Science at Balqa Univ. Her research interests are about machine learning for classification and regression.



Angel Carracedo was born in Santa Comba, Spain, in 1955. He is Professor of Forensic Medicine at the University of Santiago (USC). He also director of Fundación Pública Galega de Medicina Xenómica and the Genotyping National Centre (ISCIII), and coordinator of the Genomic Medicine Group. His main current lines of research include the genetics of neurodevelopmental diseases, cancer, rare diseases and pharmacogenomics.



María Tubío-Fungueiriño was born in Boiro, Spain, in 1992. She holds a PhD in Neuroscience and Clinical Psychology from the University of Santiago de Compostela. She is currently a researcher at the Instituto de Investigación Sanitaria de Santiago de Compostela (Spain), in the Genomic Medicine Group. Her research interests include neuropsychology of neurodevelopmental disorders.



Montse Fernández-Prieto was born in Mondoñedo, Spain, in 1969. She holds a PhD in Psychology from the University of Granada. She is a senior researcher at the Instituto de Investigación Sanitaria de Santiago de Compostela (Spain). Her main line of expertise focuses on the neurodevelopmental area, in which she has several international publications. MFP directs the cognitive part of a line of research linked to the genetic-phenotypic association in autism

spectrum disorder.

Authors and Affiliations

Manuel Fernández-Delgado¹  · Sara Cruz² · Eva Cernadas¹ · Heba Alateyat³ · María Tubío-Fungueiriño^{4,5,6} · Adriana Sampaio⁷ · Angel Carracedo^{4,6,8,9} · Montse Fernández-Prieto^{4,5,6,9}

Sara Cruz
saracruz@por.ulusiada.pt

Eva Cernadas
eva.cernadas@usc.es

Heba Alateyat
hebaateyat@gmail.com

María Tubío-Fungueiriño
maria.tubio.fungueirino@usc.es

Adriana Sampaio
adriana.sampaio@psi.uminho.pt

Angel Carracedo
angel.carracedo@usc.es

Montse Fernández-Prieto
montse.fernandez.prieto@usc.es

¹ Singular Research Center on Intelligent Technologies of the USC (CiTIUS), Univ. Santiago de Compostela (USC), Xenaro de la Fuente Domínguez, Santiago de Compostela 15782, A Coruña, Spain

² Psychology for Development Research Centre, Lusíada University, Porto, Portugal

³ Computer Dept., Al-Balqa Applied Univ., Al-Salt 19117, Jordan

⁴ Genomics and Bioinformatics Group, Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), University of Santiago de Compostela (USC), Santiago de Compostela, Spain

⁵ Fundación Pública Galega Instituto de Investigación Sanitaria de Santiago de Compostela (FIDIS), Santiago de Compostela, Spain

⁶ Genetics Group, Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain

⁷ Psychological Neuroscience Lab, CIPsi, School of Psychology, University of Minho, Campus de Gualtar, Braga 4710-057, Portugal

⁸ Fundación Pública Galega de Medicina Xenómica, Servicio Galego de Saúde (SERGAS), Santiago de Compostela, Spain

⁹ Centro de Investigación Biomédica en Red de Enfermedades Raras, Instituto de Salud Carlos III, Madrid, Spain