Accepted Manuscript

International Journal of Neural Systems

Article Title:	A Machine Learning Approach to Reveal the Neuro-Phenotypes of Autisms
Author(s):	Juan M. Gorriz, Javier Ramirez, F. Segovia, Francisco J. Martinez, Meng-Chuan Lai, Michael V. Lombardo, Simon Baron-Cohen, Mrc Aims Consortium, John Suckling
DOI:	10.1142/S0129065718500582
Accepted:	08 December 2018
To be cited as:	Juan M. Gorriz *et al.*, A Machine Learning Approach to Reveal the Neuro-Phenotypes of Autisms, *International Journal of Neural Systems*, doi: 10.1142/S0129065718500582
Link to final version:	https://doi.org/10.1142/S0129065718500582

World Scientific
www.worldscientific.com

International Journal of Neural Systems, Vol. 0, No. 0 (2018) 1–22

A MACHINE LEARNING APPROACH TO REVEAL THE NEURO-PHENOTYPES OF AUTISMS

JUAN M. GÓRRIZ*, JAVIER RAMÍREZ, F. SEGOVIA, FRANCISCO J. MARTÍNEZ
Dpt. of Signal Theory, Networking and Communications
University of Granada. Granada, 18071, Spain

MENG-CHUAN LAI
Centre for Addiction and Mental Health and The Hospital for Sick Children,
Department of Psychiatry, University of Toronto, Toronto, ON M6J 1H4, Canada

MICHAEL V. LOMBARDO
Department of Psychology, University of Cyprus, 2109 Aglantzia, Nicosia, Cyprus

SIMON BARON-COHEN, MRC AIMS Consortium†AND JOHN SUCKLING
Department of Psychiatry. University of Cambridge. Cambridge, CB2 0SZ, UK

Although much research has been undertaken, the spatial patterns, developmental course, and sexual dimorphism of brain structure associated with autism remains enigmatic. One of the difficulties in investigating differences between the sexes in autism is the small sample sizes of available imaging datasets with mixed sex. Thus, the majority of the investigations have involved male-samples, with females somewhat overlooked. This paper deploys machine learning on partial least squares feature extraction to reveal differences in regional brain structure between individuals with autism and typically developing participants. A four-class classification problem (sex and condition) is specified, with theoretical restrictions based on the evaluation of a novel upper bound in the resubstitution estimate. These conditions were imposed on the classifier complexity and feature space dimension to assure generalizable results from the training set to test samples. Accuracies above 80% on gray and white matter tissues estimated from voxel-based morphometry (VBM) features are obtained in a sample of equal-sized high-functioning male and female adults with and without autism ($N = 120$, $n = 30$/group). The proposed learning machine revealed how autism is modulated by biological sex using a low-dimensional feature space extracted from VBM. In addition, a spatial overlap analysis on reference maps partially corroborated predictions of the "extreme male brain" theory of autism, in sexual dimorphic areas.

Keywords: MRI; multiclass classification; upper bounds; Autism; spatial overlap analysis

1. Introduction

The discovery of a characteristic pattern of structural brain differences associated with autism spectrum conditions (ASC) would be a major advance in our understanding of this complex and highly variable developmental disorder. Not only would it be a substrate upon which to build a systems narrative of autism, but from this starting point it might be possible to run development "in reverse" to disentangle the roles of environmental and genetic risk factors in its aetiology. Unfortunately, "inconsistent" is perhaps the most common adjective associated with the

*Corresponding author: gorriz@ugr.es; jg825@cam.ac.uk
†See Appendix 2

extant literature in this area.

Initial MRI observations focused on increased total brain, total tissue, and total lateral ventricle volumes in adults with autism,[2] but subsequent meta-analyses, including MRI derived measurements as well as head circumference and post-mortem brain weights, were only able to detect a significant case-control difference amongst 4-5 year olds.[3] More recent meta-analyses based on fully automated voxel measures of grey and white matter volumes have confirmed the absence of any large difference in adults in either total grey matter[34] or total white matter,[4] or in cerebellar volume.? Thus, if there are differences in overall brain size then they occur very early in life[86] and are followed by a decrease in volume towards normative values during maturation in adolescence and early adulthood.[3,6]

1.1. *Grey and White Matter Distributions in the Autistic Brain*

Our knowledge of localized changes in brain anatomy has historically come from grey matter (GM) and white matter (WM) distributions estimated by voxel-based morphometry (VBM). Undergoing several iterations and improvements over time, VBM pipelines are automated, reliable[7] and statistically well-behaved.[8] The success of the VBM technique can be measured by the large number of wide-ranging longitudinally and cross-sectionally designed studies, as well as consistent cross-study patterns of tissue differences characterising disorders like schizophrenia,[9] Alzheimer's disease[83] and major depressive disorder.[10,11]

Whilst measures of global volume have generally pointed toward increases in the very early lives of those with autism, local comparisons of grey and white matter volume have variously implicated both regional increases and decreases. The greatest consistency that emerges from meta-analyses[12–14,16–19,34,85] is the large variance of both the primary literature and the outcomes of the derived meta-analyses, where the inclusion or exclusion of just a few primary sources can significantly alter the aggregated pattern of differences.

Explanations for the absence of a coherent narrative on the structural brain differences associated with autism have been suggested to arise from variety of origins. Whilst methodological differences vary between studies,[20] similar variability in VBM pipelines has not hindered the observation of characteristic patterns in other disorders.[9–11] Autism is also highly heterogeneous.[21] Indeed, under DSM-IV four categories were contained within Autism Spectrum Disorder: Asperger's disorder, childhood disintegrative disorder and pervasive developmental disorder. It has been argued that no reliable clinical diagnosis has been made with these sub-types[22] and that no consistent biological substrate has been discovered that differentiates between them leading to a single diagnostic category of Autism Spectrum Disorder in DSM-V[23] that subsumes the sub-types. Nevertheless, it is unlikely that there will be an observation that unites these disorders, and in fact greater diversity in phenotyping, perhaps down to the individual level, is argued to be more likely to identify an underlying neurobiology for autism.[24] Stratification does appears to yield sensitivity improvements; for example, children with regressive autism appear to have increases in brain size, whereas those with non-regressive autism are not associated with a significant size change;[25] and males and females with autism display different patterns of grey matter change.[28] What is almost certain is that should there be a true effect, it is spatially diffuse and generally of low effect size.

The main demographic features of ASC are its high prevalence, affecting 1% of the entire population,[30–32] and the significant skewing of the gender ratio towards male individuals to give values of 2-3:1 male:female,[31–33] and potentially contributing to the heterogeneous patterns of brain structure associated with the disorder. Previously, most studies of the biology of autism have focused predominantly on males[34] with male:female ratios in research samples in the range 8-15:1. Studies of functional imaging modalities that undertook analyses in each sex independently have found widespread significant cross-sectional differences.[35] Structural MRI studies of autism following similar stratification strategies have been successful in assessing the atypical brain areas that are shared and distinct across the sexes.[4] In a recent MRI study comprising high-functioning male and female adults with autism,[35] exploratory analyses based on a VBM univariate statistical framework uncovered substantial heterogeneity in the neurobiology of autism. Furthermore, stratification by sex can provide the empirical evidence to test predictions from the "extreme male brain" (EMB) theory

of autism[76] and other similar theories[79] that predict a relationship between autism and biological sexual differentiation.

1.2. *The machine learning approach*

A univariate analysis statistically processes single features independently. Features may be voxel-wise comparisons,[70,71] regions of interest (ROIs),[72] or models of brain features such as multiple hypothesis testing on VBM estimates[45] and intrinsic curvature.[50] However, detecting a putative small magnitude, diffuse spatial pattern of differences in autism with a univariate approach is difficult, even in large datasets,[26,27] due to its low sensitive to this type of effect.

Recently, effort has been expended on developing multivariate predictive models of biological sex based on univariate differential patterns of brain structure; i.e. cortical thickness.[39] In a sample of neurotypical males and females these models were subsequently applied to males and females with autism. Multivariate approaches have also been applied to other imaging measures that are hypothesized to be altered in autism, including brain networks,[40,51] texture features,[41,43,44] and other voxel and region-wise decomposition techniques,[52,53,74,84] In general, multivatriate cross-sectional studies of autism with VBM estimates of tissue volumes perform somewhat better than univariate methods.[26,29]

One of the major limitations of a multivariate computer-aided diagnosis (CAD) systems[87–89] is the need for a sample size (l) that is sufficiently large compared to the number of dimensions d (predictors);[58] that is, $l >> d$. Neuroimaging rarely satisfies this condition, thus the proposed learning machines must be designed and adapted to this hostile environment, otherwise control of the generalization ability of the learning process is arguably weak.[59,62] A solution to this problem may be achieved by the application of feature extraction/selection (FES) algorithms[58] with linear classifiers rather than defining a specific metric on the data,[90] or by evaluating the role that each dimension plays in the development of the machine learning architecture.[54] The aim is twofold: i) to reduce the number of input dimensions, retaining the relevant information by measuring the importance of each dimension, and ii) to control the complexity of the selected classifiers to avoid overfitting,[55,56] reducing the false positive rate (type I error). In this way, linear regressors endowed with regularization (e.g. Lasso) have been very useful in removing variables of reduced relevance from the model architecture, enhancing the contribution of the remaining variables.[57] Nevertheless, regularization approaches work on the input space enforcing an aggressive sparsity that results in a reduction of the set of non-filtered variables to a value comparable with the number of training instances,[57] which is particularly damaging in neuroimaging.

In this work, we propose a new multivariate methodology based on a one vs. one group classification scheme working on a *feature space*. The proposed system is designed under theoretical conditions imposed on the classifier complexity and the dimensionality of the input pattern. Briefly, the method comprises a subject-specific and region-wise partial least squares (PLS) FES that is then used to identify differences from patterns of GM and WM derived from MRI in a factorial design with factors of sex (male and female) and condition (autism or control), specifically: i) identification of potential sex-related risk and protective factors for autism; ii) characterization of the four classes described in terms of sex and condition; and iii) visualization of sexually dimorphic areas related to the probability of classification of autism in males and females.

This paper is organized as follows: Section 2 introduces the dataset and individual preprocessing steps implemented. In the following section 3, the overall methodology is presented, including a distribution-free upper bound for the true error rate of a classifier, using the resubstitution error estimate. These theoretical considerations permit the definition of reliable statistical learning and validation schemes using linear Support Vector Machine (SVM) and FES algorithms. In the context of the previous section, we demonstrate how the use of these techniques provides a meaningful overlap analysis between group-difference maps for VBM comparisons. Later, in Section 4, the experimental design is outlined, along with a description of a complete set of experiments, including the results obtained in each experiment, a t-test for significance in the classification experiments, and finally the spatial overlap analysis of the derived PLS-based maps with the same p-value as the t-test maps. In Section 5, these results are discussed and critiqued, and finally in Section 6 conclusions are drawn.

2. Materials: A heterogenous, multidimensional and multiclass MRI dataset

Participants ($l = 120$) included 30 right-handed premenopausal females and 30 males with autism, along with 30 neurotypical females and 30 neurotypical males (see Table 1 and[35]). All groups were matched for age (range: 18-49 years) and full-scale IQ. Participants with autism had a formal clinical diagnosis of International Classification of Diseases-10 (World Health Organization, 1992) childhood autism or Asperger's syndrome, or Diagnostic and Statistical Manual of Mental Disorders-IV text revision (American Psychiatric Association, 2000) autistic disorder or Asperger's disorder assessed by a psychiatrist or clinical psychologist in the National Health Service, UK. For further details on the inclusion criteria and the rationale in the diagnostic algorithm cut-offs please see.[35–38] Exclusion criteria for all groups are also outlined in[35] and included a current diagnosis or history of psychotic disorders, substance-use disorders, severe head injury, genetic disorders associated with autism (e.g. fragile × syndrome, tuberous sclerosis), intellectual disability (i.e. IQ <70), hyperkinetic disorder, Tourette's disorder, or any other medical condition significantly affecting brain function (e.g. epilepsy). The neurotypical groups did not have autism either themselves or in their family history.

The dataset was collected by the UK Medical Research Council Autism Imaging Multicentre Study (MRC AIMS) and were acquired at the Autism Research Centre, University of Cambridge. Further recruitment details can be found elsewhere.[35,38,45]

In addition, a larger multicentre male sample from the MRC AIMS project[38] was also used in this paper for spatial overlap analysis. It consisted of 84 neurotypical adult males and 84 males with autism matched for age and full-scale IQ (see[38] for further details).

3. Methodology

The main goal of this section is two-fold: i) to provide a statistical framework (see figure 1) under the conditions detailed in section 3.1 for a set of linear classifiers; and ii) to relieve the curse of dimensionality encountered in machine learning algorithms processing high-resolution images of the brain. The aim of using such filter methods[73] is to obtain a low-dimensional set of features and then to reduce the empirical risk without increasing the capacity of the set of classifiers (VC dimension,[61] or the number of separating dichotomies[65]) by means of a linear SVM.

3.1. *Upper bounds of error for machine learning in neuroimaging*

In a neuroimaging classification problem a d dimensional input pattern $\mathbf{x}$ is observed and the aim is to determine to which class or condition y belongs to, considered as a random variable $k = 1, \ldots, K$. In general, given a training sample of l pairs $(\mathbf{x}_i, y_i)$, the parametric set of functional dependencies $\{F(\mathbf{x}, \alpha)\}$, where α is a parameter defining the set of specific functions with cardinality equal to N, is required that minimizes the probability of misclassification. In other words, we minimize:

$$P(F(\mathbf{x}, \alpha)) \equiv P(F(\mathbf{x}, \alpha) \neq y) =$$

$$= \int_{y,X} (y - F(\mathbf{x}, \alpha))^2 P(x, y) dx dy \qquad (1)$$

on the basis of the empirical data. In this case, the functional is known as the *empirical risk* and can be computed without knowing $P(\mathbf{x}, y)$ as:

$$P_{emp}(F(\mathbf{x}, \alpha)) = 1/l \sum_{i=1}^{l} (y_i - F(\mathbf{x}_i, \alpha))^2 \qquad (2)$$

Let the minimum of equation 2 be attained for $F(\mathbf{x}, \alpha_{emp})$. The primary question is to establish when the decision rule $F(\mathbf{x}, \alpha_{emp})$ is close to $F(\mathbf{x}, \alpha_o)$, that which is obtained by minimizing equation 1. Following the derivation shown in the Appendix, the actual risk obtained by α_{emp} is bounded by probability $1 - \eta$:

$$P(\alpha_{emp}) \leq P_{emp}(\alpha_{emp}) + \sqrt{\frac{1}{2l} \ln \left(\frac{N}{\eta} \right)} \qquad (3)$$

In addition, by the use of homogeneous linear threshold functions as decision rules $F(\mathbf{x}, \alpha \equiv \mathbf{w}) : \mathbb{R}^d \longrightarrow \{-1, 1\}$, where

$$F(\mathbf{x}, \mathbf{w}) = \begin{cases} 1 & \text{if } \mathbf{x} \cdot \mathbf{w} > 0 \\ -1 & \text{if } \mathbf{x} \cdot \mathbf{w} < 0 \end{cases}$$

whenever the dimension (i.e. number of voxels/regions) of the input pattern $\mathbf{x}$ is far from the number of samples (i.e. scans), $d << l$, $P(\alpha_{emp})$ is

Table 1. Demographics details of the dataset, group means with their standard deviation (sd).

	Autism			Normal		
Sex	M	F	Total	M	F	Total
N	30	30	60	30	30	60
Age (SD)	27.2172 (7.2847)	27.8067 (7.6322)	27.5119 (7.4029)	28.1801 (5.6219)	27.4643 (6.4572)	27.8222 (6.0133)

seldom close to 0 and therefore the bound for the actual risk performs satisfactorily. In other words, for images of the brain where the number of scans is rather small compared to the number of voxels or regions, linear classifiers protect the system from overfitting, thus the variance of the actual risk is not large. Following the theory of homogeneous linear classifiers and applying the classical combinatorial geometry to develop the separating properties of decision surfaces,[59] we can take a step forward and refine the bound in equation 3 as:

$$P(\alpha_{emp}) \leq P_{emp}(\alpha_{emp}) + \sqrt{\frac{1}{2l} \ln \left(\frac{N(l,d)}{\eta} \right)} \quad (4)$$

where $N(l,d)$ is the number of linearly separable dichotomies of l samples in a d-dimensional space (see further details in the Appendix). Therefore, under the aforementioned conditions ($d << l$), the empirical risk $P_{emp}(\alpha_{emp})$ is an indicator of the generalization ability of the statistical classifier, and the maximum deviation of the frequencies from the corresponding probability of the empirical solution ($\Delta P = P - P_{emp}$) can be derived with probability $1 - \eta$ (see Appendix).

3.2. *PLS-based CAD System*

Providing a significant set of features is important to achieving a small empirical risk given a classifier of fixed complexity (capacity). The binary FES and classification stages within the multiclass classification problem are performed in a class pattern-specific manner, rather than using other strategies which combine different classes for subsequent fitting of binary statistical classifiers; i.e. the *one vs. rest*-based model. In this way, given K class-patterns a total number of $N_s = K(K-1)/2$ binary classifiers may be fitted based on the same number of training subsets. This one vs. one classification model (see figure 1) allows the assessment of which neuro-phenotypes defined in the N_s training subsets are dominant in the corresponding N_s test subsets. After that, the classification results can be merged to provide an overall classification result, defined as a decoding process in ternary error-correcting output code algorithms.[60] All of these analyses are achieved at a high confidence level given the upper bounds of the actual risk (see section 3.1), the conditions regarding the sample size and the capacity of the selected statistical classifier.

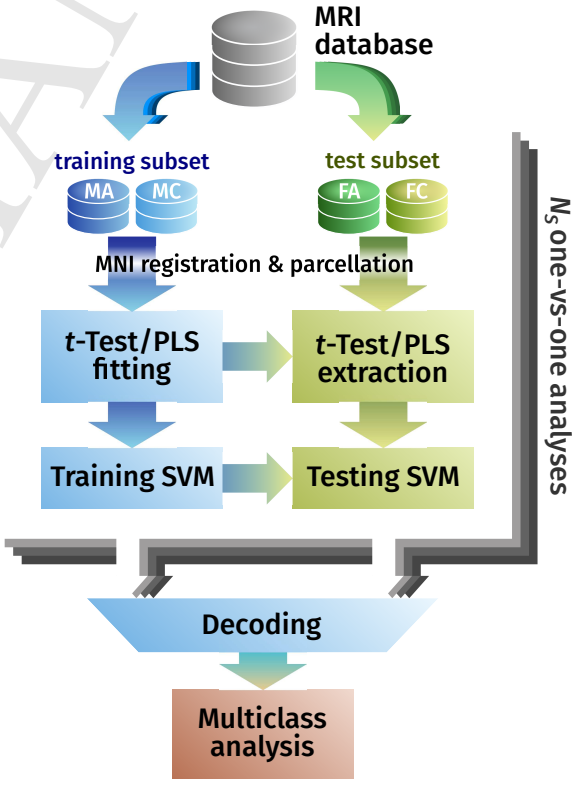


Fig. 1. Schematic representation of the proposed system. Given the K=4-class problem to be learned, $N_s = 6$ different bipartitions are formed, and 6 binary problems (dichotomizers) over the partitions are trained and tested. As a result, a codeword of the same length is obtained for each class that is decoded by a *Maximum A-Posteriori* probability principle.

3.2.1. *Structural magnetic resonance image acquisition and preprocessing*

All 120 participants were scanned using a contemporary 3 T MRI scanner (GE Medical Systems HDx) fitted with an 8-channel receive-only RT head-coil. Simulated T1-weighted inversion recovery images were segmented and normalized to the standard Montreal Neurological Institute (MNI) space using the SPM12 software (Wellcome Trust Centre for Neuroimaging, London, UK)[68] and the CAT12 toolbox.[69] Native space grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) images were obtained using standard automated segmentation routines. The native space GM and WM images were registered to a study-specific template using a high-dimensional non-linear diffeomorphic registration algorithm (DARTEL)[70] and then clustered into 116 standardized areas.[67] A modulation step was included to retain voxelwise information about local tissue volume. The modulated GM and WM maps were smoothed with a 4mm full-width at half-maximum Gaussian kernel. For the multicentre male sample all preprocessing steps were developed in the same way as described earlier for the latter dataset. In the statistical inference part, the general linear model for VBM considered the sites (i.e. scanning machines) as covariates (categorical fixed-effect factors).

GM and WM maps $\mathbf{X} = \{\mathbf{x}_i\}$, for $i = 1, \ldots, l$, were initially parcellated into $r = 1, \ldots, 116$ standardized regions of a brain anatomical atlas.[67] Thus, our multivariate analysis takes as its input mean grey and white matter volumes from within each atlas region, $\mathbf{x}_i(r)$, separately. This procedure is appropriate for our purpose of observing tissue-specific local variations, although partial results may be rearranged by aggregation in a complete volume, as shown in the following sections and figures 2 and 3.

3.2.2. *Two-sample t-test and PLS extraction*

A rank FES based on the standard two-sample t-test with pooled variance in combination with a PLS-based FES[74] were region-wise applied with the *one vs. one* classification model (see figure 1). Given a training subset comprising two classes $j = 1, 2$ with balanced samples sizes $l_1 = l_2 = l/2$, the $\mathbf{t}$ statistical vector is defined as:

$$\mathbf{t} = \frac{\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2}{\mathbf{s}_p \sqrt{\frac{4}{l}}} \tag{5}$$

where $\mathbf{s}_p = \sqrt{\frac{s_{\mathbf{x}^1} + s_{\mathbf{x}^2}}{2}}$ is the pooled standard deviation and $s_{\mathbf{x}^1}$ $s_{\mathbf{x}^2}$ are the unbiased estimators of the variances of the two classes (the assumption of the data in each group following a normal distribution was confirmed by a Kolmogorov-Smirnov test). The effect of this operation on the brain regions is to reduce the computational load prior to feature extraction, and to assure that regional differences in tissues volumes are considered when assessing overall patterns of tissue distribution.

Once the significant regions were identified by ranking the result of the t-test, we extracted the relevant patterns within those regions by a PLS regression between the $l \times d$ data matrix $\mathbf{X}$ and the $l \times 1$ vector labels $\mathbf{Y}$. Briefly, in the application of PLS to supervised classification we maximized:

$$\omega_o = \max_{\omega} \left(cov\left(\mathbf{X}\omega, \mathbf{Y}\right)\right)^2; \quad \text{s.t. } ||\omega|| = 1 \tag{6}$$

where the score vectors $\mathbf{s} = \mathbf{X}\omega$ were iteratively extracted and used to deflate the input matrix $\mathbf{X}$ by subtracting their rank-one approximations based on $\mathbf{s}$.[74] The deflation process was accomplished by the computation of the vector of loadings $\mathbf{p}$ as a coefficient of regressing $\mathbf{X}$ on $\mathbf{s}$:

$$\mathbf{p} = \frac{\mathbf{X}^T\mathbf{s}}{\mathbf{s}^T\mathbf{s}} = \mathbf{X}^T\hat{\mathbf{s}} \tag{7}$$

The vector of weights ω_f, where f indexes the number of extracted features, is a local property in the images; that is, the dimensional components are not mixed in its computation, whilst the score coefficients $\mathbf{s}(i) = \sum_{j=1}^{d} X_{ij}\omega_j$ (and the matrix of scores), and the deflation term $\mathbf{sp}^T|_{ij} = \frac{\sum_{k=1}^{d} s_k X_{ik} s_i}{\sum_{i=1}^{l} s_i^2}$, etc. are global. Therefore, the size of the input data d is crucial to the assessment of the relationship between GM (WM) volume and group membership with heterogeneous variances, where some statistical properties of the involved processes, such as the stationarity or the ergodicity in the correlation, must be assumed for the evaluated ROIs. In the present study the analysis was carried out on the set regions (see figures 2 and 3) selected by feature selection.

3.2.3. *Statistical parametric PLS maps*

The analysis of functional and structural studies usually entails the construction of spatially extended statistical processes where each voxel of the novel image or map is the result of a statistical test. After that hypothesis testing process the main problem is to determine the significance of extrema by the use of several statistical field-based approximations,[15] i.e. a 3D t-statistical field. PLS methods have been demonstrated in the past to be very useful in describing the relation between brain activity and experimental design or behavior measures.[42,74] In this way PLS analysis of brain activations is able to reveal additional regions of salience that are not identified by typical univariate voxelwise methods such as SPM. Assume that the label vector $\mathbf{Y}$ contains the two global conditions in $\mathbf{X}$ ($\{1,-1\}$ indexed as $i \in C_1, C_2$) and $\mathbf{x}_{i_1} > \mathbf{x}_{i_2}$, for $i_1 \in C_1$ and $i_2 \in C_2$, without loss of generalization. The scores $\mathbf{s}$ will be ideally located around zero with different signs, thus from equation 7 for every loading component, we have:

$$p(k) = \sum_{i_1} x_i(k)\hat{s}(i) - \sum_{i_2} x_i(k)|\hat{s}(i)| \qquad (8)$$

for $k = 1, \ldots, d$, and a deviation from zero of this data-weighted score summation would suggest a region associated with a particular group membership (contrast). Remapped into image space the contents of the singular vector $\mathbf{p}$ indicate which pixels are most sensitive to those predefined contrasts and define the so-called PLS map. Comparing the latter expression to equation 5, the PLS-map can be seen as a multivariate two-sample test weighted by the scores of each subject with unknown distribution, except for the normalization term that depends on the pooled standard deviation. In fact, after some manipulations it yields:

$$\mathbf{p} = \bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2 \qquad (9)$$

where $\bar{\mathbf{x}}^j(k) \equiv \sum_{C_i} x_i(k)P(x_i(k))$ and $P(x_i(k)) = |\hat{s}(i)|$ is the frequency of the observation $x_i(k)$ that is assumed to be proportional to its score in the computation of the group mean.

The statistical significance of the PLS-maps can be assessed is many ways; e.g. by the use of a permutation test.[42] In this work we proposed the use of a parametric approach, that is very popular in neuroimaging, based on the Gaussian Random Field (GRF) theory. This theory models both the univariate probabilistic features of the resulting SPM and the non-stationary spatial covariance structure of that image.[68] This model can be applied as well to the proposed PLS-map as the selected quantity that characterized the PLS computation has shown to be very similar to the classical two-sample t-test mapping (except for the normalization factor). The methodology can be described as the following. Firstly, given a p-value and the $\mathbf{T}$ SPM extracted from the one tailed two-sample t-test with contrast matrix $[1,-1]$ for C_1 and C_2 conditions, we determine the statistically significant positive salience threshold t_{critic} using the inverse cumulative density function of the t-test distribution (the same applies for the negative one). The biased version of the PLS-based two-sample t-test in equation 9 makes up a novel $\mathbf{P}$ SPM that is linearly projected to a (bias corrected) distribution $\hat{\mathbf{P}}$ with the same parameters of the previous $\mathbf{T}$ SPM[a]. Finally, the positive saliences on this novel map are determined by evaluating the regions where $\hat{\mathbf{P}} > t_{critic}$ at the given p-value.

3.2.4. *A robust statistical classifier, SVM*

The use of SVMs was motivated by the minimization of the VC dimension and has been successfully shown to be a robust solution in classification learning,[62] which minimizes the separation margin between the binary-labeled training data, mapped into a (PLS-based) feature space $\mathfrak{F}$, by constructing a hyperplane $\mathbf{w}$ whose norm is minimum:[62]

$$||\mathbf{w}||^2 + C\sum_{i=1}^{l} \xi_i \qquad (10)$$

subject to

$$y_i(\mathbf{w} \cdot \mathbf{x}_i) \geq 1 - \xi_i; \quad \xi_i \geq 0; \quad i = 1, \ldots, l$$

where ξ_i are slack variables, C is a constant that allows a trade-off between training error and model complexity (C is usually optimized by several searching methods at the training stage, i.e. Bayesian optimization in a wide range $[1e-3, 1e3]$), and the decision rule is defined as $F(\mathbf{x}, \mathbf{w})$ in section 3.1. The

[a]Here we assume that the t-distribution is approximately normal for a high number of degrees of freedom. Note that if a variable $X \sim \mathcal{N}(\mu_1, \sigma_1)$ then $Z = (X - \mu_1)/\sigma_1 \sim \mathcal{N}(0, 1)$. Conversely, given Z then $Y = \mu_2 Z + \sigma_2 \sim \mathcal{N}(\mu_2, \sigma_2)$. Then we can connect any pair of distributions by $Y = \frac{\sigma_2}{\sigma_1}X + \left(\mu_2 - \mu_1\frac{\sigma_2}{\sigma_1}\right)$

solution is computed using $\mathbf{w} = \sum_{i=1}^{l} a_i y_i \mathbf{x_i}$, where the multipliers $0 \leq a_i \leq C$ were derived from the dual Lagrangian problem in equation 10.

4. Experiments and Results

The spatial representation of the binary classifier output and the corresponding "loading" of the PLS image as the new reference base for the analyzed input patterns is discussed and shown in this section. In this sense, we take a step forward with respect to the majority of exploratory analyses developed so far in the literature[35,45] and propose a specific cross-validation scheme with the purpose of discovering generalizable class-features in the GM or WM images. By training the system using binary group differences (gender and/or condition) and under the assumption that they are distributed across statistically significant regions, we aim to generalize these features in the remaining study groups which should share the same magnitude of differences across groups. The complete analysis included a class pattern-specific two-sample test for region-wise FES, an overlap analysis[35] across group-difference PLS maps with a VBM comparison to evaluate the Autism theories, the assessment of the sample Autism label probability with the determination of confidence intervals using the Clopper-Pearson method and finally, permutation tests to check the statistical significance of the classification results obtained by the machine learning-based system.

4.1. *Experimental Setup*

As discussed previously in section 3.2, the images were parcellated according to[67] and FES algorithms were applied to the resulting images obtained from a standardized preprocessing and image registration pipeline.[69] For the sake of clarity, we define $N_s = 6$ SVM classifiers acting on their corresponding group comparisons: Group 1 (G1): MC-MA (male controls vs male autistic individuals), G2: MC-FC (MC vs female controls), G3: MC-FA (MC vs female autistic individuals), G4: MA-FC, G5: MA-FA and G6: FC-FA.

4.1.1. *Preprocessing:*

The application of the FES approach detailed in section 3.2 results in several weight maps for the group comparisons as shown in figure 2.

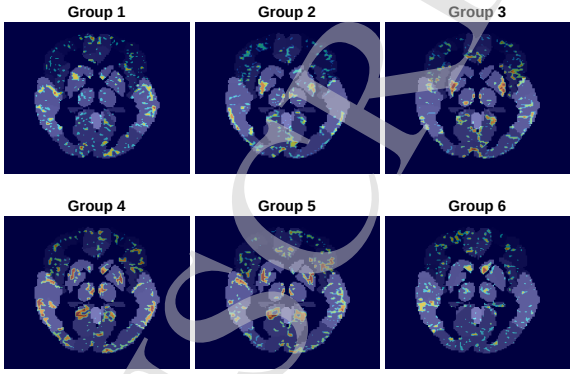


Fig. 2. Significant grey-matter differences (VBM) in the study groups ($p < 0.05$) plotted on the standardized atlas. Relative excesses in GM volume are displayed in orange/red, while deficits are displayed in blue (jet color map). From left to right, up to bottom: Male control (MC) vs. Male Autism (MA); MC vs. Female control (FC); MC vs. Female Autism (FA); MA vs. FC; MA vs. FA; FC vs. FA.

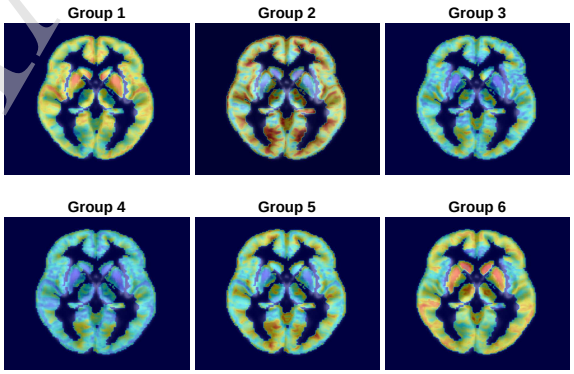


Fig. 3. Brain map (PLS) illustrating regions where GM volume was associated with group membership ($\mathbf{p}$ loading). Orange/red regions indicate areas with a positive association between GM volume and condition (i.e. ASD > control); blue regions indicate brain systems of decreased grey-matter volume in the condition group. The selected z-coordinate was 50 for the axial slice in the standard space of the MNI template.

With the aim of increasing the generalization ability of the proposed classification systems only the first PLS component was considered (a single dimension $d = 1$), thus the associated upper bound of the actual risk, as shown in equation 1, provides a strong connection between the risk and the empirical risk

(resubstitution error). It is worth mentioning what is actually shown in figures 3 and 4 is the first loading component **p** extracted from the brain regions, separately. Figure 3 highlights the relevant regions in terms of the PLS regression on the group differences. Group differences G1 and G6 illustrate the regions where the features associated with autism are different in males and females, respectively. In the remainder of the figures, features associated with gender are predominantly found in the PLS maps; e.g. the G2 pattern that refers to the VBM comparison between MC and FC, is modulated and found in group differences G3, G4 and G5. The statistical significance of these PLS maps is discussed in the following section 4.2, where the null distributions are modeled as a two-tailed Gaussian distribution or a t-distribution with large degrees of freedom ν.

In figure 4 the PLS global approach is shown on the left and is contrasted against the regionwise comparison for G1 (i.e. MC-MA). Observe that the regionwise comparison provides a noisy **p**-loading that is used as a reference base in the feature space for the extracted input scores. A one-sample t-test between the score data **s** obtained from the regionwise approach and the global approach reveals that, at a 95% significance level, 44 out of the 60 subjects present a difference in means in regions including the insula, hippocampus, and cingulum; figure 4.

4.1.2. *Classification*

The overall classification accuracy of each SVM model was estimated using the re-substitution error for the training subset and a $l/2$-holdout cross-validation error for the test subset, testing for significance by repeating the validation procedure $n = 500$ times, after randomly permuting the class labels. In this sense, each of $N_s = 6$ training subsets, under the one vs.one classification model, allows us to estimate the probability of the neuro-phenotype in terms of sex or condition and to evaluate it on the remaining N_s test subsets. Therefore, the four classes are considered as a set of participants with different proportions or combinations of these two types of effect (sex and condition).

The selection of other cross-validation procedures, i.e. K-fold, with the aim of "pure" classification leads to poor classification and generalization of results mainly due to the variability of the input patterns, the potentially preponderant role of one effect, and small sample size. Thus, we should consider the former $K = 4$ classes as tentative labels as they may contain spatially-dependent neurophenotypes with biases towards a particular effect that are non-uniformly distributed in the K classes.

4.2. *Visualizing the patterns of brain regions representative of normative sex/condition differences*

The aim of this section is to analyze the differences of the specific patterns representing sex or condition. Following a similar analysis as,[35] we undertook a two-sample t-test based on VBM comparisons between groups. Only voxelwise height thresholds with no spatial extent operation were applied under the two logical contrasts; i.e. $\text{group}_1 > \text{or} < \text{group}_2$. Three sets of VBM comparisons were performed to test EMB theory predictions;[76] namely, MC-FC, MC-MA and FC-FA comparisons. The aim of this procedure is to generate an overlap analysis across group-difference maps[35] and compare them with the PLS maps. The PLS maps can be interpreted as two-tailed statistical tests, unlike the classical one-tailed t-test maps obtained under the SPM framework,[69] and both are almost normally distributed with a high number of degrees of freedom (e.g. $\nu = 57$ for the condition comparison).[75] In order to make them comparable for the computation of group-difference maps, given a set of p values $\{0.0001, \ldots, 0.05\}$, we linearly transform the PLS maps, $P \sim \mathcal{N}(\mu_p, \sigma_p)$ to a new Gaussian distribution with mean and standard deviation of the corresponding comparison group T-maps, which are normally distributed $T \sim \mathcal{N}(\mu_t, \sigma_t)$. After this transformation, the PLS maps permit testing of the logical contrasts in a single map by $\text{group}_1 \neq \text{group}_2$. As an example, in those regions where there is a statistically significant sexual dimorphism $MC \neq FC$, we evaluated the significant differences in the neurophenotype of male and female individuals with autism, that is, $MA \neq MC$ and $FA \neq FC$ (see figure 5). As shown in the following section, the overlap analysis derived from these figures partially supports EMB theory predictions by considering both directions of effect at the same time ($> and <$).

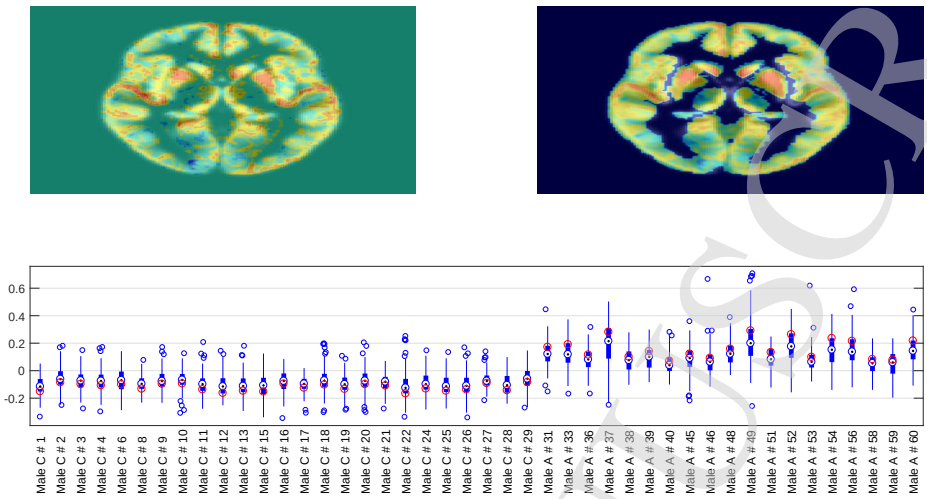


Fig. 4. Up: Detail of Slice #50 for the **p**-loading in Group 1 under global (on the left) and region-wise (on the right) PLS FE. Note the noisy PLS loading on the left image as a result of processing the whole image simultaneously. Bottom: one-sample t-test on the scores **s**, the rejection of the null hypothesis is obtained for 44 out of 60 subjects

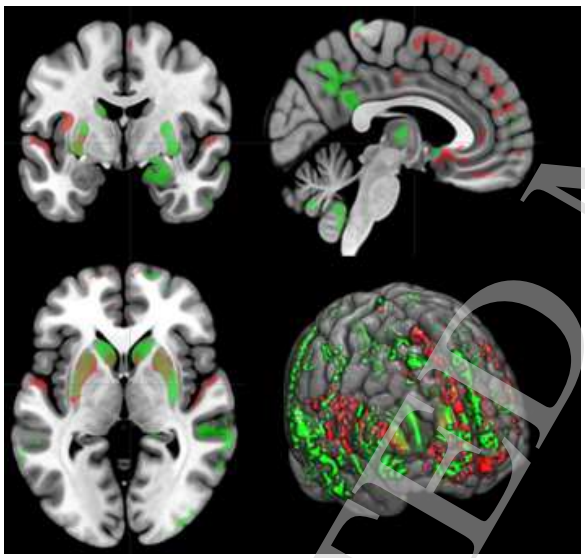


Fig. 5. Grey matter overlapping regions in sexually dimorphic areas of males (red) and female (green) individuals with autism at a voxel-level threshold of p-value=0.05.

4.2.1. *Spatial Overlap Analysis*

Following the discussion in section 4.2, three VBM comparisons, MC-FC, MA-MC and FA-FC, were evaluated on GM and WM volumes. We carried out the experiments on our dataset ($l = 120$) and the male sample ($l = 168$) described in,[38] using our PLS-based FE approach and the set of one-tailed contrasts obtained from the univariate 2×2 factorial design analysis provided by the SPM software.[68] For each comparison a conjunction analysis[35,78] consisting of logical AND masking, were assessed and tested for significance by running Montecarlo simulations (5000 iterations). The distribution of overlaps at the same p-value threshold were sampled; i.e. from 0.0001 to 0.05 in steps of 0.0001. This spatial overlap analysis, which considers synthetic GM (WM) map overlaps, allows us to assess the probability of whether the overlapping voxels of the comparisons occur by chance. In the baseline comparisons we included an additional spatial overlap analysis using a multicenter male sample as detailed in.[35,38] The purpose of this MA-MC comparison is to conduct the same analysis on a larger number of samples ($l = 168$) to provide higher power to detect the group differences, that were not found by the use of the current database ($l = 60$, males only).

The spatial overlap analysis between males and females with autism is small in both approaches (T and PLS maps) for GM and WM (e.g. up to 15% and 10% for GM and up to 23% and 10% for WM, respectively) for $p \leq 0.05$ as shown in the figure 6 in magenta color lines. This finding clearly identifies the differences between the neuroanatomy of autism across sexes that was previously demonstrated by the high precision of the four-class learning machine de-

tailed in the previous section, and in figure 5. The rest of the curves show a significant difference between approaches; that is, the one-tailed T maps versus the two tailed-PLS maps. In the latter approach, we find significant evidence for overlap between structures sensitive to diagnosis and sexual dimorphism in both sexes (red and blue solid lines). This evidence is only found for the one-tailed T-based approach in females (T-MC-FC&FC-FA, red dotted line in figure 5).

4.3. *Classification Results*

The output score pattern of the set of binary classifiers is shown in figure 7 for the training and test sets. From this set of figures what is interesting to note is the ability for generalization of the proposed system, under some constrained conditions, over the tests G2, G3, G4 and G5 (same rows on the figure). Another relevant feature found in the output score maps is the presence of vertical bars representing misclassification in training and test sets. This could be identification of the different neurophenotypes of specific individuals, e.g. participant 42 in the FC subset, as they are always located in the wrong hyperplane subspace at both training and test stages (see table 2). From this table it is clear that MA is the class with lower performance in classification accuracy in both stages. Surprisingly, the features associated with autism in the training of the SVM to classify G1 (males) cannot be generalized in the test set (females) although these features should be present in one class, and absent in the other (a real binary classification problem). The almost random pattern found in the test set could indicate the different nature of the features of autism in males and females. This situation is repeated in G6 where the females were used in the training stage and males made up the test set, confirming the latter hypothesis.

Table 2. GM (WM) Overlap and region averaged accuracy of the output score maps (%) for the selected subsets acting as Training and Test sets (Groups 1 and 6 are not considered). Note: the GM overlap of correctly classified patterns in the MC subset, used at the Training stage in the comparisons $MCvsFC$ and $MCvsFA$, is 90.98%.

	Subsets	Train Ov.	Test Ov.	Train Acc.	Test Acc.
GM	MC	90.98	89.68	77.07(7.68)	63.03(9.05)
	FC	92.16	90.43	80.60(6.65)	70.88(7.84)
	MA	92.27	89.34	73.09(7.96)	64.73(7.28)
	FA	90.55	90.00	81.16(6.33)	69.47(8.19)
WM	MC	90.92	86.95	81.38(9.16)	67.30(9.98)
	FC	91.18	88.97	80.00(8.58)	70.52(9.23)
	MA	93.22	87.64	77.63(9.00)	64.74(8.50)
	FA	91.24	87.82	80.45(8.88)	66.16(9.39)

The quantitative analysis of the output score pattern is found in figure 8 where we represent the distribution of accuracies, including the notch analysis to display the variability of the median between samples, for all the brain regions and the accompanying overlay histograms. From this figure we can observe how the learning systems generalize well with respect to the controlled upper bound (around 10%), excluding the G1 and G6 which behave as random classifiers on the test set. This effect indicates the varying nature in the differences between MC and MA (male autism features) which cannot be extrapolated to the FC-FA comparison, and vice versa. In the former groups, the gender feature is the most prevalent, providing a high overlap in the output of the systems for several comparisons; see Table 2.

Participants with a high or low probability of being classified as having autism may be determined by using a predictive mapping approach.[38] Given the output score map of a participant with sex S and condition C at the test&training stages, we first calculated the predictive class probability with respect to the male typical neurophenotype. For this purpose, the ratio was obtained of the number of regions associated with a classification of "male" to the overall number of brain regions. Then, we collected all these probabilities, one for each individual, into a discrete set of bins from 0 to 1 in steps of 0.25, and computed the sample Autism probability (that is, the probability of the male-neurophenotype, denoted in figure 9 as $P(ph = M|S, C = ASD)$) as the ratio of the number of individuals classified with autism to the total number of individuals within the bin. As shown in figure 9 for G2-5, where confidence intervals are determined using the Clopper-Pearson method in the binomial test,[77] the probability of an individual being classified as autistic in males (blue) and females (red) evolves differently across the male-

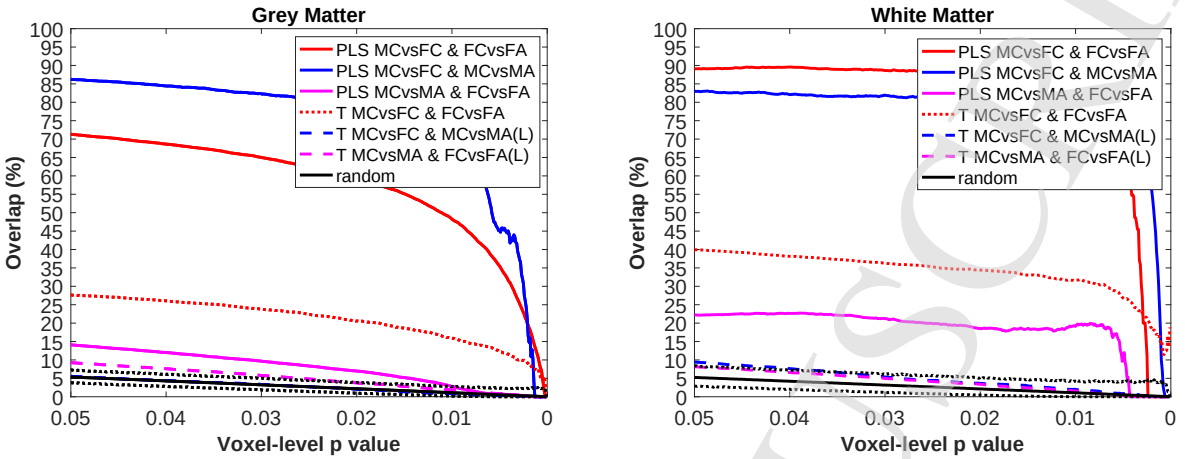


Fig. 6. Comparison of the testing brain-level predictions of the EMB theory of Autism. (L) indicates the use of the larger male database (i.e. using the DESPOT1 acquistions) for the VBM comparison, described in.[38]

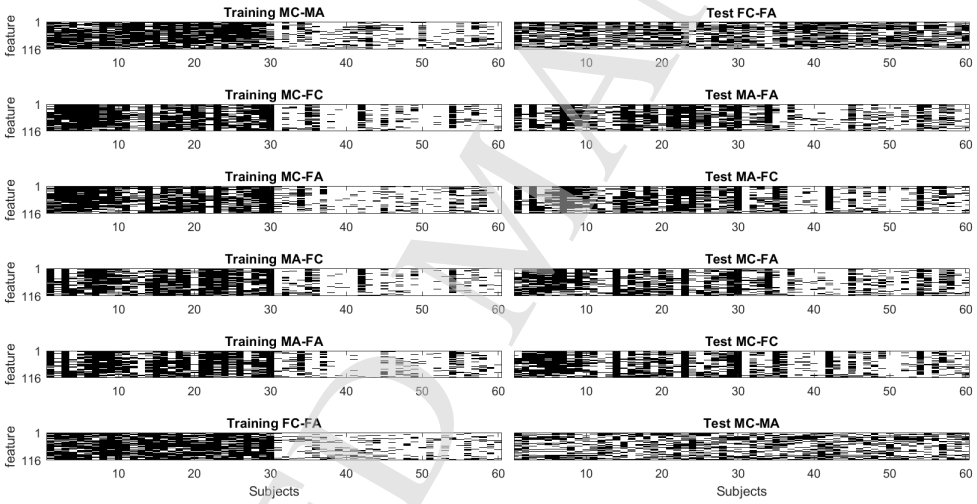


Fig. 7. Output score map of the linear SVM in the training&test configuration (GM). On the left column we plot the SVM-scores obtained from the training set while on the right column we show the consequent prediction on the test set. Each row contains one of the 6 binary problems to solve the multiclass framework, depicting all the 116 region-wise features (Y-axis) vs. each subject (X-axis). The hyperplane derived at each row using the training subsets on the left is employed in the prediction of the test subsets on the right. Note that for the ideal linearly-separable problem we would obtain two black (from 1 to 30) and white (from 31 to 60) columns in all cases. In this way, the fitted learner should generalize fairly-well on the remaining groups, obtaining two additional black and white columns on the test set, in case the features on the right were characterizing the groups on the left. This is clearly not the case in the 1 and 6 rows, where we are trying to predict male/female Autistic features from the opposite sex.

neurophenotypic axis. In addition, within gender, this probability increases or decreases depending on whether the group is processed at the test or training stage. This difference in behaviour is due to the presence (or absence) of the autism feature that adds to the gender-based difference between groups, during the training stage.

In addition, we studied the binary classification problem using the group comparisons (MC-MA and FC-FA) using a classical cross-validation scheme

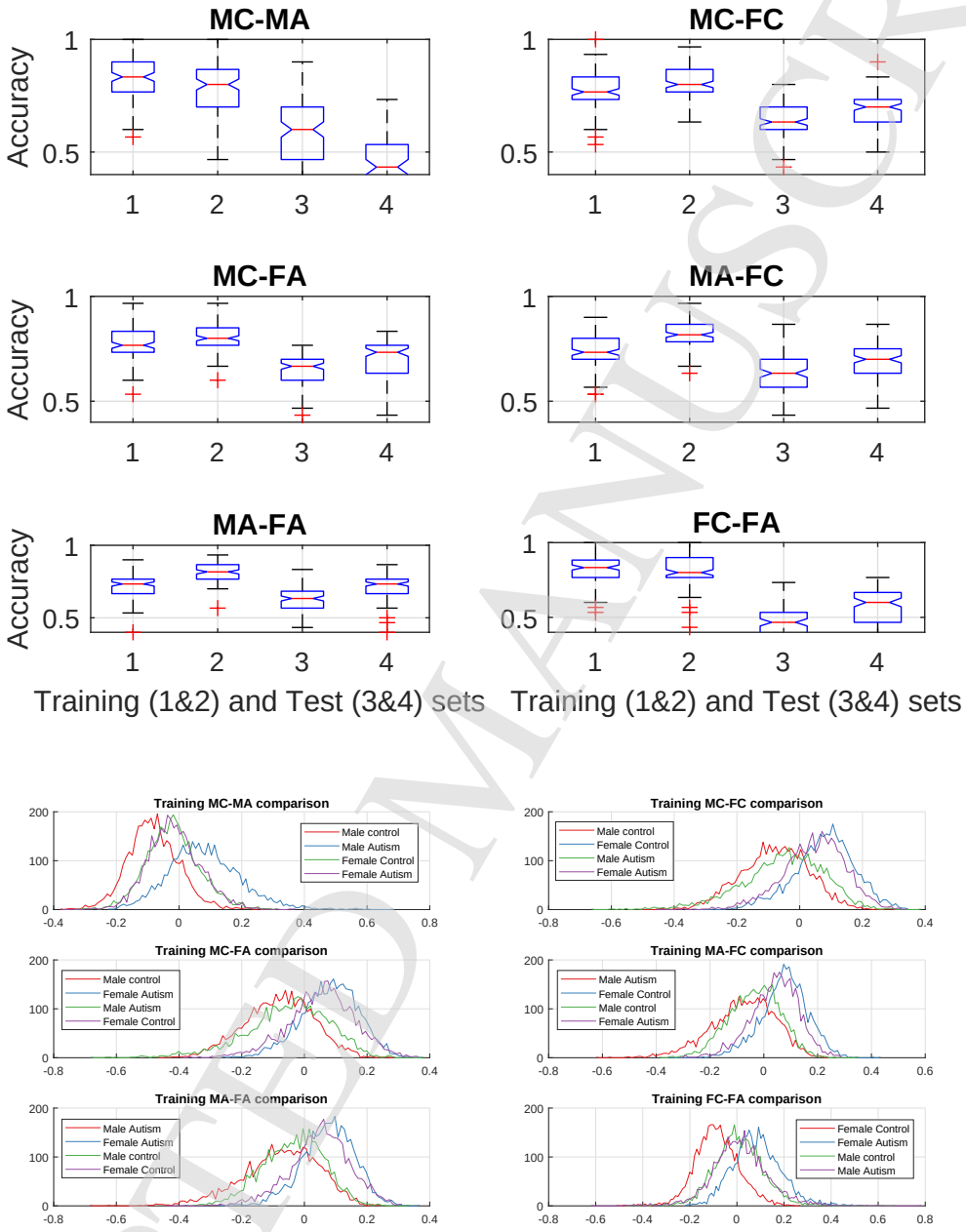


Fig. 8. Up: Distribution of accuracies in training and test subsets for Groups 1-6 (GM). Bottom: Overlay histogram Analysis for Groups 1-6 (GM).

used in small sample sizes; i.e. leave-one-out cross validation and $l = 60$. In this case, we employed the same feature selection and extraction methods with a similar parameter tuning prior to linear SVM-based classification. As an example, for the MC-MA comparison, the few regions extracted on GM by this more restrictive cross-validation scheme are coincident with the regions highlighted in figure 5, i.e. right middle frontal gyrus, right pallidum, etc. yielding classification rates from 70% up to 80% in the

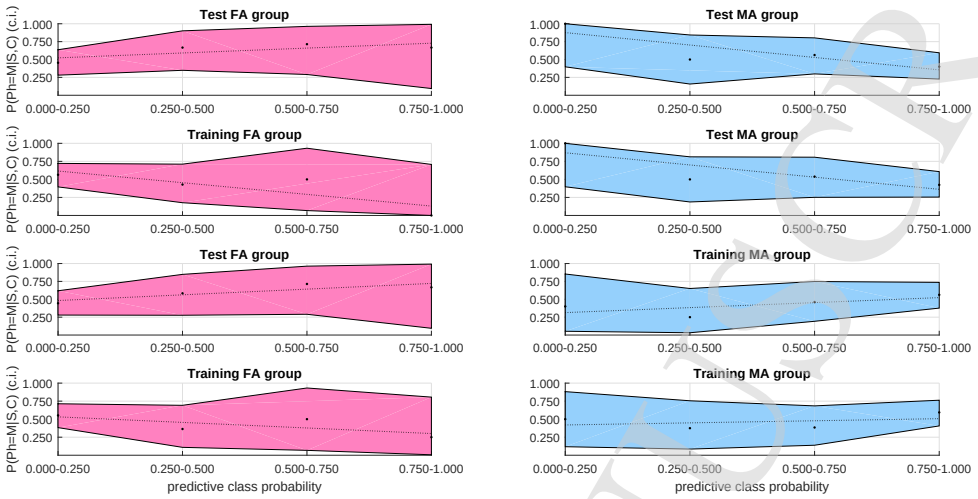


Fig. 9. Probability for autism as a Function of Normative Sex-Related Phenotypic Variability in Brain Morphology: Probability estimates for autism across four discrete bins along the axis of predictive class probabilities for sex are plotted for G2-5 and training (T) and test (t) stages. For male (blue) and female (red) models, the probability for autism behaves in a different manner with increasingly male-typical class predictions; that is, it increases (decreases) in females (males) at the test stage and the contrarily at the training stage.

study groups from GM and WM tissues.

Finally, using a *one vs. one* decoding scheme based on a MAP strategy, the multiclass problem may be undertaken on a region-wise basis and a output score map of the linear SVM may be derived with 4 class levels. Under the upper bound (~ 0.1) on the actual risk for $d = 1$ and using linear decision functions, the six binary classifiers can be combined obtaining the results shown in figure 10. Note in the upper subfigure that individuals with autism are ordered from male to female, and a deep (light) colour represents control (autism) participant sex. A region-wise analysis using this decoding strategy with GM reveals regions with a high accuracy (up to 83.33%); namely, left and right parietal, temporal, occipital lobes, calcarine sulcus, etc. and a mean across region of 62.61 ± 8.23 with confusion depicted in the bottom of the same figure. With WM, an improvement in performance is observed (up to 93.33%) mainly due to the improved classification of the MC group (averaged accuracy across regions of 67.50 ± 9.99), with relevant regions located at middle frontal gyrus, inferior parietal lobe, cerebellar cortical crus II, middle occipital and temporal lobes, etc. Additionally, in this multiclass problem we employed the output score maps depicted in figure 7 to decode the MAP output class for each region and then a majority voting scheme across the regions was applied to determine the final output class for each individual. Overall, the four groups are represented by different image patterns that can be classified with an accuracy up to 86.7% (91.70% on WM) using the information contained in the entirety of the image, as shown in the confusion matrix (fig 11).

4.4. *Permutation Tests*

Although the experimental conditions were rigorously established in the previous sections; that is, the results obtained by linear classifiers in low dimensional spaces are theoretically meaningful at a significance level η (see section 3.1), a permutation test to validate these results was applied to the whole dataset. For the complete set of binary classifiers the cross-validation procedure was repeated $n = 500$ times after randomly permuting the class labels to derive p-value maps. As expected, the classification obtained with this analysis for the set of classifiers is almost random and , at the 95% level of confidence, almost all the regions were considered as significant in this analysis as shown in table 3. The result of this analysis allows the construction of relevance maps for each classification model such as those displayed in

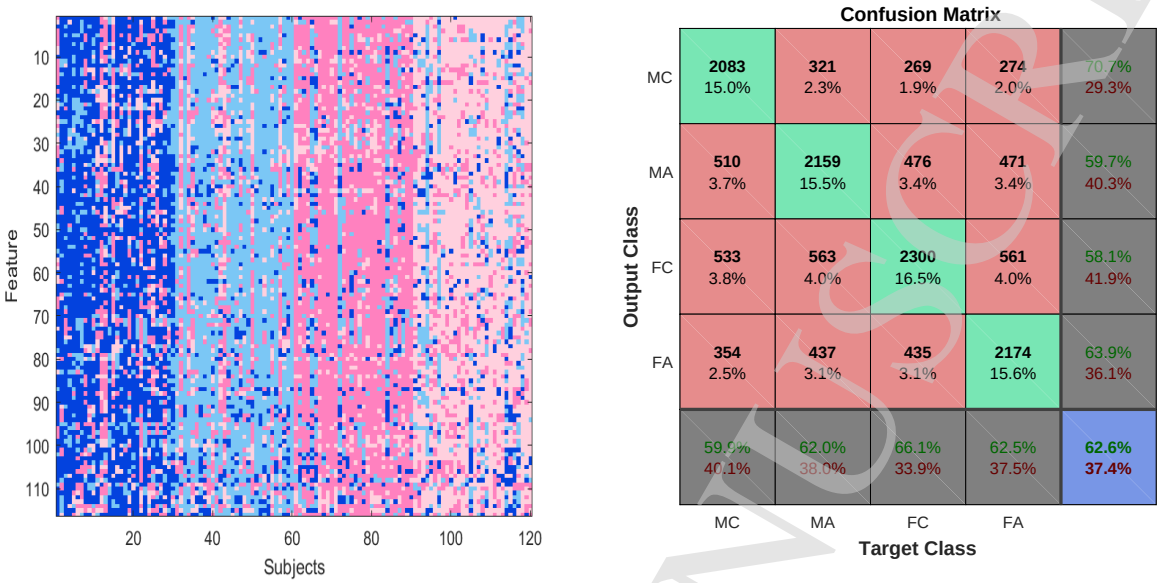


Fig. 10. Right: Output score map of the combination of six binary classifiers in a one vs. one classification scheme using GM-based features. Right: Confusion matrix of the region-wise classification scheme.

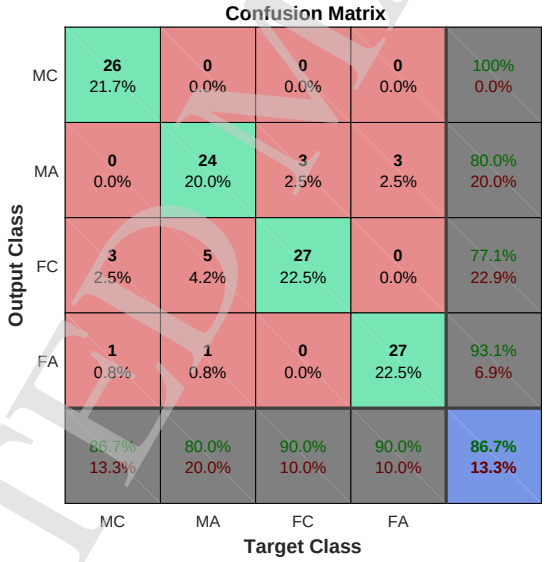


Fig. 11. Confusion matrix of the majority voting multiclass classification scheme corresponding to the output score map in figure 10. Observe how the overall classification result outperforms the region-wise approach, due to the presence of outliers in the four-class problem (patients containing a high ratio of misclassified regions).

the figure 12, where, as an example, we illustrate the prediction of sex based on VBM features using the sex classifier for G2.

Table 3. Testing for significance the classification models on GM features (95% significance level). Accepted regions in %

out of the 116 analyzed BAs for the binary classifiers at the training stage.

%	MC	MA	FC	FA	
$G1$	95.69	83.62	-	-	GM
$G2$	89.66	-	96.55	-	
$G3$	90.52	-	-	98.28	
$G4$	-	82.76	98.28		
$G5$	-	80.17	-	98.28	
$G6$	-	-	91.38	94.83	

5. Discussion

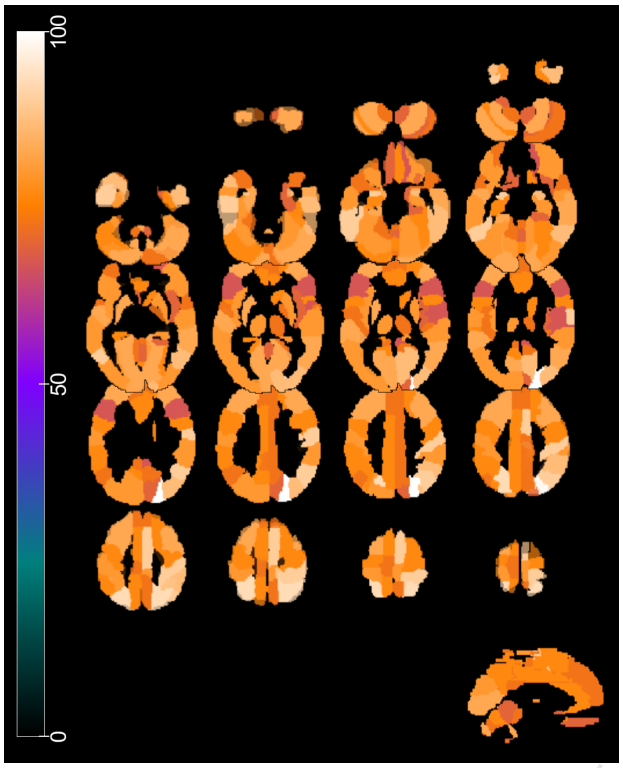
From the perspective of this machine learning-based approach, the neuroanatomy of autism in males and females is comprised of separate distinguishing patterns. A distinguishing pattern similarly identifies a male and female controls. These group-differences are observed even with a small sample set, although some theoretical bounds and suitable methodologies must be imposed to highlight such differences. The interactions between sex and diagnosis observed in this paper extend those previously found in prior work[35,38] that also showed sex×diagnosis group differences on GM and WM tissues. In particular, greater accuracy rates were obtained with the classification apparatus described herein.

In addition, based on this multivariate methodology, we also found that there was a significant overlap between the neuroanatomical features of autism in males and females, unlike the previous approaches, aligning with the predictions of the EMB theory in terms of the directionality of effect, and with the "gender incoherence" hypothesis[79]). These effects were observed by the use of a region-wise PLS-based activation map transformed into the same parameters as those used in the one-tailed t maps derived from the group differences. In fact, the approach proposed in this paper considers not only the EMB theory[76] directionality but also its inverse; that is, predicting that males with autism are feminized in terms of neuroanatomy.[79] This effect was clearly observed on GM and WM VBM comparisons and agrees with the results obtained by testing how the effect of autism overlaps with the effect of "femenization" using one-tailed contrasts.[35] Thus, on balance, these results may be a demonstration of gender incoherence of males with autism rather than "masculinization".

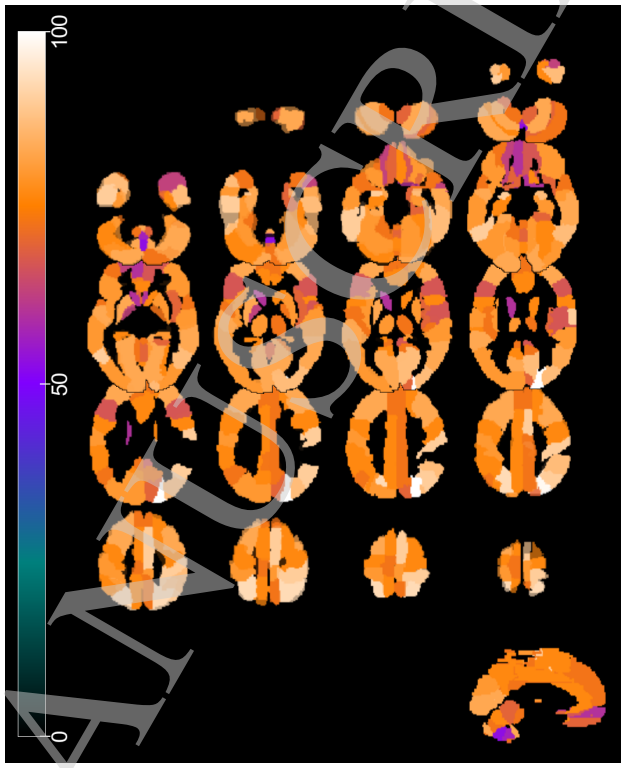
The heterogeneity of autism was detected by the assessment of the output score maps derived from the machine learning theory. The use of cross-validation groups further reduced the limited sample size and degraded the performance of the CAD system. Although some considerations and preventive measures regarding the curse of dimensionality were carried out to be conservative, the observed effects and group-differences could be partially sample-specific and the current analysis requires replication on larger datasets. These difficulties are detected in terms of the presence of outliers, that is, participants that are misclassified by resubstitution in almost all the analyzed regions. Those are participants that are located in the "wrong' feature subspace present an heterogeneous pattern that affects the performance of the SVM when they are considered in any validation/training fold. The output score maps revealed this misclassified sex/condition-specific neurophenotype on GM and WM tissues across regions and participants.

6. Conclusions

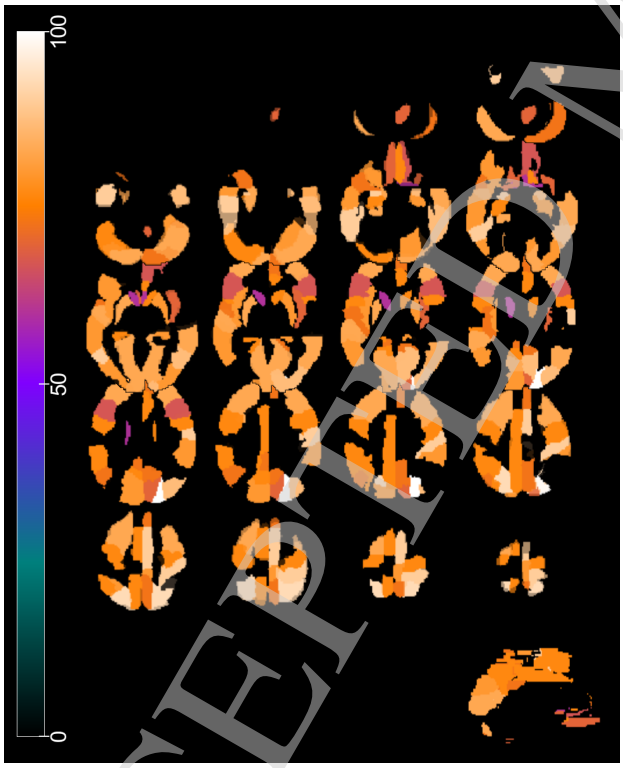
In summary, the research developed here is a first attempt to describe both sex-typical multivariate neurobiological phenotypes by the use of machine learning and a MRI sample of equal-sized high-functioning male and female adults with and without autism. Although the main limitation of the current equal-sized gender datasets is the samples sizes, we avoided this difficulty by imposing some restrictions on the learning parameters of the system by the use of a novel upper bound in the resubstitution estimate, obtaining a good trade-off between empirical risk and the variance of the actual risk estimation (upper bound). In addition, the system used a set of features extracted from PLS activation maps that are demonstrated to be statistically significant and in accordance with two theories of the neurobiology of autism. The complete classification system in a one vs. one configuration achieves, under these theoretical conditions, high classification results (up to 86%) in a four-label classification problem, thus outperforming the classification rate of previous published works on this field.[35,80–82]



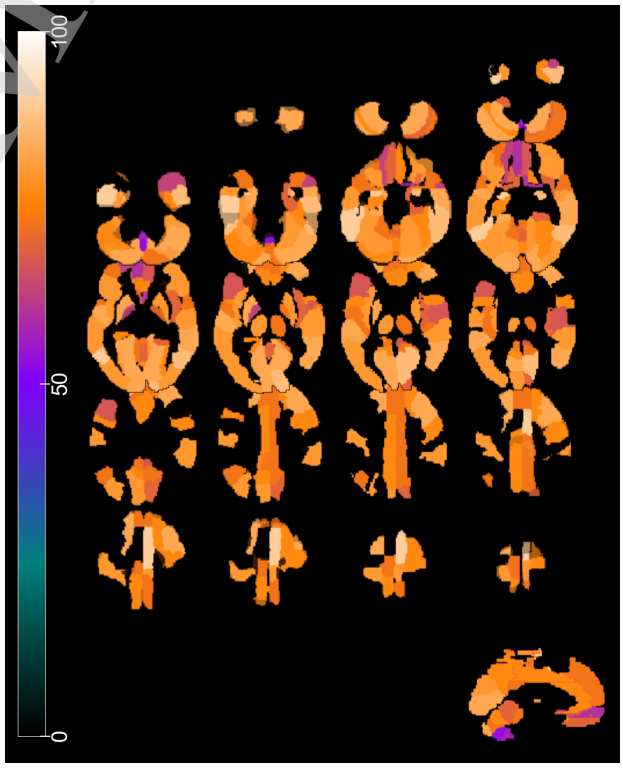
(a) MC relevant regions corrected by the permutation test



(b) FC relevant regions corrected by the permutation test



(c) MA relevant regions corrected by the permutation test



(d) FA relevant regions corrected by the permutation test

Fig. 12. Accuracy maps for G2: Training MC-FC comparison and test on MA, FA groups with GM features. The colorbar indicates the precision of each region in the binary classification problem of detecting sexual dimorphism in the training and test subsets.

7. Appendix 1

7.1. *On the upper bound of the actual risk*

In terms of the one-sided uniform convergence of the means we are interested in assessing for a given significance level η:

$$P\{\sup_i(P(\alpha_i) - P_{emp}(\alpha_i)) > \epsilon\} < \eta \qquad (11)$$

where $P(\alpha_i) = P(F(x, \alpha_i))$. Of course, with a sample size $l \to \infty$ the law of large numbers expressed in terms of the third Hoeffding inequality[63] for any functional α_i establishes that:

$$\lim_{l\to\infty} P\{\sup_i(P(\alpha_i) - P_{emp}(\alpha_i)) > \epsilon\} = 0 \qquad (12)$$

and the uniform convergence in equation 11 is achieved. In other cases $l < \infty$, the aforementioned inequality can be used to establish the bound of the actual risk as:[61]

$$P\{\Gamma_i > \epsilon\} \leq \sum_{i=1}^{N} P\{\gamma_i > \epsilon\} < \eta = N\exp(-2\epsilon^2 l) \qquad (13)$$

where $\Gamma_i = \sup_i(\gamma_i)$ and $\gamma_i = P(\alpha_i) - P_{emp}(\alpha_i)$ and N is the finite number of functional dependencies. Since the inequality is valid for all decision functions $F(x, \alpha_i)$, the actual risk obtained by α_{emp} is bounded with probability $1 - \eta$ by:

$$\gamma_{emp} \leq \sqrt{\frac{1}{2l}\ln\left(\frac{N}{\eta}\right)}$$

that is equivalent to equation 3. In general, these bounds could be further improved by considering the relative deviations[61,64] under scenarios where $P(\alpha_i)$ tends to the extremes $0, 1$, but that is far from our problem.

Definition: A set of l vectors is in general position in d-space if every subset of d or fewer vectors is linearly independent.

Consider that the training sample $\{\mathbf{x}_i, y_i\}$ is distributed randomly in general position, then the number of linearly separable dichotomies of the set of input patterns is equal to N, that is, the number of decision functions $F(\mathbf{x}, \alpha)$ when the training sample is not a root of the set $(F(\mathbf{x}_i, \alpha) \neq 0)$. As shown in[59] for linear decision functions and based on the *Function Counting Theorem* this is equal to:

$$N(l, d) = 2\sum_{k=0}^{d-1}\binom{l-1}{k} \qquad (14)$$

Thus, the bound in equation 3 could be rewritten as shown in equation 4. As an example, if $d \simeq l$ then $N(l, d) \simeq 2^l$, thus the number of functions is such that it separates the sample size in all possible ways (non-falsifiable learning machine), the minimum of the empirical risk is zero and the upper bound of $P(\alpha_{emp})$ is trivial (> 0.5) independently of the sample size l. On the contrary, if $d << l$ the actual risk reaches its maximum value close to the empirical risk, i.e. for $d = \{1, 2, 3\}$ and $l = 120$, the maximum deviation of the frequencies ($\Delta P = P - P_{emp}$) are obtained with probability $1 - \eta(= 0.95)$ as:

$$\Delta P \leq \{0.1398, 0.1879, 0.2286\} \qquad (15)$$

8. Appendix 2

The Medical Research Council Autism Imaging Multicentre Study Consortium (MRC AIMS Consortium) is a UK collaboration between the Institute of Psychiatry, Psychology and Neuroscience (IoPPN) at King's College, London, the Autism Research Centre, University of Cambridge, and the Autism Research Group, University of Oxford. The Consortium members are in alphabetical order: Anthony J. Bailey (Oxford), Simon Baron-Cohen (Cambridge), Patrick F. Bolton (IoPPN), Edward T. Bullmore (Cambridge), Sarah Carrington (Oxford), Marco Catani (IoPPN), Bhismadev Chakrabarti (Cambridge), Michael C. Craig (IoPPN), Eileen M. Daly (IoPPN), Sean C. L. Deoni (IoPPN), Christine Ecker (IoPPN), Francesca Happé (IoPPN), Julian Henty (Cambridge), Peter Jezzard (Oxford), Patrick Johnston (IoPPN), Derek K. Jones (IoPPN), Meng-Chuan Lai (Cambridge), Michael V. Lombardo (Cambridge), Anya Madden (IoPPN), Diane Mullins (IoPPN), Clodagh M. Murphy (IoPPN), Declan G. M. Murphy (IoPPN), Greg Pasco (Cambridge), Amber N. V. Ruigrok (Cambridge), Susan A. Sadek (Cambridge), Debbie Spain (IoPPN), Rose Stewart (Oxford), John Suckling (Cambridge), Sally J. Wheelwright (Cambridge), and Steven C. Williams (IoPPN).

Acknowledgments

This work was partly supported by the MINECO under the TEC2015-64718-R project, the Salvador de Madariaga Mobility Grants 2017 and the Consejería de Economía, Innovación, Ciencia y Empleo (Junta

de Andalucía, Spain) under the Excellence Project P11-TIC-7103. The study was conducted in association with the National Institute for Health Research Collaborations for Leadership in Applied Health Research and Care (NIHR CLAHRC) East of England (EoE). The project was supported by the UK Medical Research Council (grant number GO 400061) and European Autism Interventions—a Multicentre Study for Developing New Medications (EU-AIMS); EU-AIMS has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115300, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007 - 2013) and EFPIA companies' in-kind contribution. During the period of this work M-CL was supported by the O'Brien Scholars Program in the Child and Youth Mental Health Collaborative at the Centre for Addiction and Mental Health (CAMH) and The Hospital for Sick Children, Toronto, the Academic Scholar Award from the Department of Psychiatry, University of Toronto, the Slaight Family Child and Youth Mental Health Innovation Fund, CAMH Foundation, and the Ontario Brain Institute via the Province of Ontario Neurodevelopmental Disorders (POND) Network; MVL was supported by the British Academy, Jesus College Cambridge, Wellcome Trust, and an ERC Starting Grant (ERC-2017-STG; 755816); SB-C was supported by the Autism Research Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health, UK.

References

1. Hauser SL, DeLong GR, Rosman NP. Pneumographic findings in the infantile autism syndrome. A correlation with temporal lobe disease. Brain. 1975 Dec;98(4):667-88.
2. Piven J, Arndt S, Bailey J, Havercamp S, Andreasen NC, Palmer P. Am J Psychiatry. 1995 Aug;152(8):1145-9. An MRI study of brain size in autism.
3. Redcay E, Courchesne E. When is the brain enlarged in autism? A meta-analysis of all brain size reports. Biol Psychiatry. 2005 Jul 1;58(1):1-9.
4. Radua J, Via E, Catani M, Mataix-Cols D. Voxel-based meta-analysis of regional white-matter volume differences in autism spectrum disorder versus healthy controls. Psychol Med. 2011 Jul;41(7):1539-50.
5. Traut N, Beggiato A, Bourgeron T, Delorme R, Rondi-Reig L, Paradis AL, Toro R. Cerebellar Volume in Autism: Literature Meta-analysis and Analysis of the Autism Brain Imaging Data Exchange Cohort. Biol Psychiatry. 2017 Oct 10.
6. Sacco R, Gabriele S, Persico AM. Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. Psychiatry Res. 2015 Nov 30;234(2):239-51.
7. Eggert LD, Sommer J, Jansen A, Kircher T, Konrad C. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. PLoS One. 2012;7(9):e45081.
8. Scarpazza C, Tognin S, Frisciata S, Sartori G, Mechelli A. False positive rates in Voxel-based Morphometry studies of the human brain: should we be worried? Neurosci Biobehav Rev. 2015 May;52:49-55.
9. Glahn DC, Laird AR, Ellison-Wright I, Thelen SM, Robinson JL, Lancaster JL, Bullmore E, Fox PT. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. Biol Psychiatry. 2008 Nov 1;64(9):774-81.
10. Lai CH. Gray matter volume in major depressive disorder: a meta-analysis of voxel-based morphometry studies Psychiatry Res. 2013 Jan 30;211(1):37-46.
11. Bora E, Fornito A, Pantelis C, Yucel M. Gray matter abnormalities in Major Depressive Disorder: a meta-analysis of voxel based morphometry studies. J Affect Disord. 2012 Apr;138(1-2):9-18.
12. Nickl-Jockschat T, Habel U, Michel TM, Manning J, Laird AR, Fox PT, Schneider F, Eickhoff SB. Brain structure anomalies in autism spectrum disorder–a meta-analysis of VBM studies using anatomic likelihood estimation. Hum Brain Mapp. 2012 Jun;33(6):1470-89.
13. Cauda F, Costa T, Palermo S, D'Agata F, Diano M, Bianco F, Duca S, Keller R. Concordance of white matter and gray matter abnormalities in autism spectrum disorders: a voxel-based meta-analysis study. Hum Brain Mapp. 2014 May;35(5):2073-98.
14. Yu KK, Cheung C, Chua SE, McAlonan GM. Can Asperger syndrome be distinguished from autism? An anatomic likelihood meta-analysis of MRI studies. J Psychiatry Neurosci. 2011 Nov;36(6):412-21.
15. Worsley, K. J. and Marrett, S. and Neelin, P. and Vandal, A. C. and Friston, K. J. and Evans, A. C. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping, vol 4, num 1 1996,58-73
16. Rojas DC, Peterson E, Winterrowd E, Reite ML, Rogers SJ, Tregellas JR. Regional gray matter volumetric changes in autism associated with social and repetitive behavior symptoms. BMC Psychiatry. 2006 Dec 13;6:56.
17. DeRamus TP, Kana RK1. Anatomical likelihood estimation meta-analysis of grey and white matter

anomalies in autism spectrum disorders. Neuroimage Clin. 2014 Nov 18;7:525-36.
18. Yang X, Si T, Gong Q, Qiu L, Jia Z, Zhou M, Zhao Y, Hu X, Wu M, Zhu H. Brain gray matter alterations and associated demographic profiles in adults with autism spectrum disorder: A meta-analysis of voxel-based morphometry studies.Aust N Z J Psychiatry. 2016 Aug;50(8):741-53.
19. Carlisi CO, Norman LJ, Lukito SS, Radua J, Mataix-Cols D, Rubia K. Comparative Multimodal Meta-analysis of Structural and Functional Brain Abnormalities in Autism Spectrum Disorder and Obsessive-Compulsive Disorder. Biol Psychiatry. 2017 Jul 15;82(2):83-102.
20. Chen R, Jiao Y, Herskovits EH. Structural MRI in autism spectrum disorder. Pediatr Res. 2011 May;69(5 Pt 2):63R-8R.
21. Lenroot RK and Yeung PK Heterogeneity within Autism Spectrum Disorders: What have We Learned from Neuroimaging Studies? Front Hum Neurosci. 2013; 7: 733.
22. Lord C, Jones RM. Annual research review: re-thinking the classification of autism spectrum disorders. J Child Psychol Psychiatry. 2012 May;53(5):490-509.
23. Diagnostic and Statistical Manual of Mental Disorders , Fifth Edition Edited by American Psychiatric Association, 2013.
24. Waterhouse L, Gillberg C. Why autism must be taken apart. J Autism Dev Disord. 2014 Jul;44(7):1788-92.
25. Nordahl CW, Lange N, Li DD, Barnett LA, Lee A, Buonocore MH, Simon TJ, Rogers S, Ozonoff S and Amarala DG Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders Proc Natl Acad Sci U S A. 2011 Dec 13; 108(50): 20195–20200.
26. Haar S, Berman S, Behrmann M, Dinstein I. Anatomical Abnormalities in Autism? Cereb Cortex. 2016 Apr;26(4):1440-52.
27. Riddle K, Cascio CJ, Woodward ND. Brain structure in autism: a voxel-based morphometry analysis of the Autism Brain Imaging Database Exchange (ABIDE). Brain Imaging Behav. 2017 Apr;11(2):541-551.
28. Beacher FD, Minati L, Baron-Cohen S, Lombardo MV, Lai MC, Gray MA, Harrison NA, Critchley HD. Autism attenuates sex differences in brain structure: a combined voxel-based morphometry and diffusion tensor imaging study. AJNR Am J Neuroradiol. 2012 Jan;33(1):83-9.
29. Uddin LQ, Menon V, Young CB, Ryali S, Chen T, Khouzam A, Minshew NJ, Hardan AY. Multivariate searchlight classification of structural magnetic resonance imaging in children and adolescents with autism. Biol Psychiatry. 2011 Nov 1;70(9):833-41.
30. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE, *et al.* Prevalence of autism-spectrum conditions: UK school-based population study. Br J Psychiatry 2009; 194: 500-9.
31. Mattila ML, Kielinen M, Linna SL, Jussila K, Ebeling H, Bloigu R, et al. Autism spectrum disorders according to DSM-IV-TR and comparison with DSM-5 draft criteria: an epidemiological study. J Am Acad Child Adolesc Psychiatry 2011; 50: 583-92 e11.
32. Idring S, Rai D, Dal H, Dalman C, Sturm H, Zander E, et al. Autism spectrum disorders in the stockholm youth cohort: design, prevalence and validity. PLoS One 2012; 7: e41280.
33. Kim YS, Leventhal BL, Koh YJ, Fombonne E, Laska E, Lim EC, et al. Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry 2011; 168: 904-12
34. Via E, Radua J, Cardoner N, Happe F, Mataix-Cols D. Meta-analysis of grey matter abnormalities in autism spectrum disorder: should Asperger disorder be subsumed under a broader umbrella of autistic spectrum disorder? Arch Gen Psychiatry 2011; 68: 409-18.
35. Lai, M.C., Lombardo, M. V., Suckling, J., Ruigrok, A. N. V., Chakrabarti, B., Ecker, C., Deoni, S. C. L., Craig, M. C., Murphy, D. G. M., Bullmore, E. T. and Baron-Cohen, S. (2013) Biological sex affects the neurobiology of autism. Brain : a journal of neurology, 136 (9). pp. 2799-2815. ISSN 1460-2156
36. Lai MC, Lombardo MV, Chakrabarti B, Sadek SA, Pasco G, Wheelwright SJ, Bullmore ET, Baron-Cohen S; MRC AIMS Consortium, Suckling J. A shift to randomness of brain oscillations in people with autism. Biol Psychiatry. 2010 Dec 15;68(12):1092-9.
37. Lombardo MV, Chakrabarti B, Bullmore ET, Sadek SA, Pasco G, Wheelwright SJ, et al. Atypical neural self-representation in autism. Brain 2010; 133(Pt 2): 611-24.
38. Ecker C, Ginestet C, Feng Y, Johnston P, Lombardo MV, Lai MC, Suckling J, Palaniyappan L, Daly E, Murphy CM, Williams SC, Bullmore ET, Baron-Cohen S, Brammer M, Murphy DG. Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. JAMA Psychiatry 2013; 70: 59-70.
39. Ecker C, Andrews DS, Gudbrandsen CM, et al. Association between the probability of autism spectrum disorder and normative sex-related phenotypic diversity in brain structure. JAMA Psychiatry. 2017; 74(4):329-338.
40. Ortiz A, Munilla J, Álvarez-Illán I, Górriz JM, Ramírez J Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's Disease diagnosis. Front Comput Neurosci. 2015 Nov 3;9:132.
41. U. Acharya, R. Yanti, J. Zheng, M. Krishnan, J. Tan, R. Martis and C. Lim, Automated diagnosis of epilepsy using cwt, hos and texture parameters., *International Journal of Neural Systems* **23**(03) (2013)

p. 1350009.

42. McIntosh AR1, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage. 1996 Jun;3(3 Pt 1):143-57.
43. A. Ortiz, J. M. Górriz, J. Ramírez and F. Martínez-Murcia, Lvq-SVM based CAD tool applied to structural MRI for the diagnosis of the alzheimer's disease, *Pattern Recognition Letters* **34** (Oct 2013) p. 1725-1733.
44. F. Martinez-Murcia, J. Górriz, J. Ramírez, M. Moreno-Caballero, M. Gómez-Río, P. P. M. Initiative *et al.*, Parametrization of textural patterns in 123i-ioflupane imaging for the automatic detection of parkinsonism, *Medical physics* **41**(1) (2014) p. 012502.
45. Ecker C, Suckling J, Deoni SC, Lombardo MV, Bullmore ET, Baron-Cohen S, Catani M, Jezzard P, Barnes A, Bailey AJ, Williams SC, Murphy DGM, MRC AIMS Consortium FT. Brain Anatomy and Its Relationship to Behavior in Adults With Autism Spectrum Disorder A Multicenter Magnetic Resonance Imaging Study. Arch Gen Psychiatry. 2012;69(2):195-209. doi:10.1001/archgenpsychiatry.2011.1251
46. P. Piaggi, D. Menicucci, C. Gentili, G. Handjaras, A. Gemignani and A. Landi, Singular spectrum analysis and adaptive filtering enhance the functional connectivity analysis of resting state fmri data., *Int J Neural Syst* **24** (May 2014) p. 1450010.
47. K. Michalopoulos and N. Bourbakis, Combining eeg microstates with fmri structural features for modeling brain activity., *Int J Neural Syst* **25** (Dec 2015) p. 1550041.
48. S. Ferdowsi, S. Sanei and V. Abolghasemi, A predictive modeling approach to analyze data in eeg-fmri experiments., *Int J Neural Syst* **25** (Feb 2015) p. 1440008.
49. D. Chyzhyk, M. Graña, D. Öngür and A. K. Shinn, Discrimination of schizophrenia auditory hallucinators by machine learning of resting-state functional mri., *Int J Neural Syst* **25** (May 2015) p. 1550007.
50. L. Ronan, R. Pienaar, G. Williams, E. Bullmore, T. J. Crow, N. Roberts, P. B. Jones, J. Suckling and P. C. Fletcher, Intrinsic curvature: a marker of milimeter-scale tangential cortico-cortical connectivity?, *International Journal of Neural Systems* **21**(05) (2011) 351–366.
51. M. Ahmadlou and H. Adeli, Functional community analysis of brain: a new approach for eeg-based investigation of the brain pathology., *Neuroimage* **58** (Sep 2011) 401–408.
52. I. A. Illan, J. M. Górriz, J. Ramírez and A. Meyer-Base, Spatial component analysis of mri data for alzheimer's disease diagnosis: a bayesian network approach., *Front Comput Neurosci* **8** (2014) p. 156.
53. F Segovia, JM Górriz, J Ramírez, I Alvarez, JM Jiménez-Hoyuela, Improved parkinsonism diagnosis using a partial least squares based approach, Medical physics 39 (7), 4395-4403. 2012.
54. B. S. Khundrakpam, J. Tohka, and A. C. Evans, Prediction of brain maturity based on cortical thickness at different spatial resolutions," NeuroImage, vol. 111, pp. 350-359, 2015.
55. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning, vol. 46, no. 1, pp. 389-422, 2002.
56. J. Bi, K. Bennett, M. Embrechts, C. Breneman, and M. Song, Dimensionality reduction via sparse support vector machines, JMLR, vol. 3, pp. 1229-1243, 2003.
57. E. Parrado-Hernandez, V. Gomez-Verdejo, M. Martinez-Ramon, J. Shawe-Taylor, P. Alonso, J. Pujol, J. M. Menchon, N. Cardoner, and C. Soriano-Mas, Discovering brain regions relevant to obsessive compulsive disorder identification through bagging and transduction, Medical image analysis, vol. 18, no. 3, pp. 435-448, 2014.
58. Juan Manuel Górriz, Javier Ramírez, John Suckling, Ignacio Álvarez Illán, Andrés Ortiz García, Francisco Jesús Martínez-Murcia, Fermín Segovia, Diego Salas-Gonzalez, Shuihua Wang: Case-Based Statistical Learning: A Non-Parametric Implementation With a Conditional-Error Rate SVM. IEEE Access 5: 11468-11478 (2017)
59. Juan M. Gorriz, Javier Ramirez, John Suckling, F.J. Martinez-Murcia, I.A. Illán, F. Segovia, A. Ortiz, D. Salas-González, D. Castillo-Barnés, C.G. Puntonet, A semi-supervised learning approach for model selection based on class-hypothesis testing, Expert Systems with Applications, Volume 90, 2017, Pages 40-49, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2017.08.006.
60. Sergio Escalera, Oriol Pujol, and Petia Radeva, On the Decoding Process in Ternary Error-Correcting Output Codes, IEEE Trans. on pattern analysis and machine learning intelligence, Vol. 32, No. 1, Jan 2010
61. Vladimir Vapnik. Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics). 1982. Springer-Verlag New York, Inc.
62. Vladimir N. Vapnik. The nature of statistical learning theory 1995. ISBN 0-387-94559-8 Springer-Verlag New York, Inc.
63. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. 58, 13-30.
64. Olivier Gascuel, Gilles Caraux, Distribution-free performance bounds with the resubstitution error estimate, Pattern Recognition Letters, Volume 13, Issue 11, 1992, Pages 757-764, ISSN 0167-8655,
65. Thomas M. Cover, Geometrical Statistical and applications in inequalities linear of pattern properties recognition systems with applications in pattern

recognition. IEEE Transactions on Electronic Computers. 3, EC-14. (1965) 326-334
66. Tohka J, Dinov ID, Shattuck DW, Toga AW. Brain MRI Tissue Classification Based on Local Markov Random Fields. Magnetic resonance imaging. 2010;28(4):557-573. doi:10.1016/j.mri.2009.12.012.
67. N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain", NeuroImage, vol. 15, no. 1, pp. 273-279, Jan. 2002.
68. K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, Statistical Parametric Mapping, Academic Press, London, 2007, ISBN 9780123725608.
69. C. Gaser, R. Dahnke (2016). CAT - A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. HBM 2016. http://www.neuro.uni-jena.de/hbm2016/GaserHBM2016.pdf
70. John Ashburner, A fast diffeomorphic image registration algorithm, In NeuroImage, Volume 38, Issue 1, 2007, Pages 95-113, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2007.07.007.
71. A. Shiino, T. Watanabe, K. Maeda, E. Kotani, I. Akiguchi and M. Matsuda, Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease, *NeuroImage* **33** (October 2006) 17–26.
72. N. Ayache, Analyzing 3D Images of the Brain, *NeuroImage* **4**(3) (1996) S34–S35.
73. I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds., Feature Extraction: Foundations and Applications. Berlin, Germany: Springer, 2006.
74. R. Rosipal, N. Kramer, Overview and Recent Advances in Partial Least Squares, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 34-51.
75. Hazewinkel, Michiel, ed. (2001) [1994], "Student distribution´´, Encyclopedia of Mathematics, Springer Science+Business Media B.V. / Kluwer Academic Publishers, ISBN 978-1-55608-010-4
76. Baron-Cohen S. The extreme male brain theory of autism. Trends Cogn Sci 2002; 6: 248-54.
77. Johnson, N. L., S. Kotz, and A. W. Kemp. Univariate Discrete Distributions. Hoboken, NJ: Wiley-Interscience, 1993.
78. Lombardo MV, Ashwin E, Auyeung B, Chakrabarti B, Taylor K, Hackett G, et al. Fetal testosterone influences sexually dimorphic grey matter in the human brain. J Neurosci 2012; 32: 674-80.
79. Bejerot S, Eriksson JM, Bonde S, Carlstrom K, Humble MB, Eriksson E. The extreme male brain revisited: gender coherence in adults with autism spectrum disorder. Br J Psychiatry 2012; 201: 116-23.
80. Martinez-Murcia FJ, Lai MC, Górriz JM, Ramírez J, Young AM, Deoni SC, Ecker C, Lombardo MV; MRC AIMS Consortium,, Baron-Cohen S, Murphy DG, Bullmore ET, Suckling J. On the brain structure heterogeneity of autism: Parsing out acquisition site effects with significance-weighted principal component analysis. Hum Brain Mapp. 2017 Mar;38(3):1208-1223. doi: 10.1002/hbm.23449. Epub 2016 Oct 24.
81. Fermín Segovia, Rosemary Holt, Michael Spencer, Juan M. Górriz, Javier Ramírez, Carlos G. Puntonet, Christophe Phillips,Lindsay Chura, Simon Baron-Cohen and John Suckling. Identifying endophenotypes of autism: a multivariate approach Front Comput Neurosci. 2014; 8: 60. Published online 2014 Jun 6. doi: 10.3389/fncom.2014.00060
82. Anibal Solon Heinsfelda, Alexandre Rosa Francob, R. Cameron Craddockf, Augusto Buchweitz, Felipe Meneguzzia. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical 17 (2018) 16–23.
83. Mirzaei, G., Adeli, A., and Adeli, H., "Imaging and Machine Learning Techniques for Diagnosis of Alzheimer Disease," Reviews in the Neurosciences, 27:8, 2016, pp. 857-870.
84. Ahmadlou, M., Adeli, H., Adeli, (2012) A., "Fuzzy Synchronization Likelihood-Wavelet Methodology for Diagnosis of Autism Spectrum Disorder," Journal of Neuroscience Methods, 211:2 pp. 203-209.
85. Bhat, S., Acharya, U.R., Adeli, H. (2014), Muralidhar Bairy, G.M., and Adeli, A., "Automated Diagnosis of Autism: In Search of a Mathematical Marker," Reviews in the Neurosciences, 25:6, 851-861.
86. Bhat, S., Acharya, U.R., Adeli, Muralidhar Bairy, G.M., and Adeli, A. (2014), "Autism: Cause Factors, Early Diagnosis and Therapies," Reviews in the Neurosciences, 25:6, 841-850.
87. F.C. Morabito, M. Campolo, N. Mammone, M. Versaci, S. Franceschetti, F. Tagliavini, V. Sofia, D. Fatuzzo, A. Gambardella, A. Labate, L. Mumoli, G.G. Tripodi, S. Gasparini, V. Cianci, C. Sueri, E. Ferlazzo, U. Aguglia, "Deep Learning Representation from Electroencephalography of Early-stage Creutzfeld-Jakob Disease and Features for Differentiation from Rapidly Progressive Dementia" International Journal of Neural Systems, 27:2, 2017, 1650039
88. A. Antoniades, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei, and C.C. Took (2018), "Deep neural architectures for mapping scalp to intracranial EEG" International Journal of Neural Systems, 28:8,
89. S. Wang, Y. Hu, and Y. Shen (2018), "Classification of Diffusion tensor metrics for the Diagnosis of a Myelopathic Cord Using Machine Learning" International Journal of Neural Systems, 28:2, 1750036
90. X. Li, Y. Bai and Y. Peng, S. Du, and S. Ying (2018), "Nonlinear Semi-supervised Metric Learning via Multiple Kernel and Local Topology", International Journal of Neural Systems, 28:2, 1750040