



The “MS-ROM/IFAST” Model, a Novel Parallel Nonlinear EEG Analysis Technique, Distinguishes ASD Subjects From Children Affected With Other Neuropsychiatric Disorders With High Degree of Accuracy

Enzo Grossi¹ , Massimo Buscema^{2,3}, Francesca Della Torre², and Ronald J. Swatzyna⁴

Clinical EEG and Neuroscience
1–13
© EEG and Clinical Neuroscience
Society (ECNS) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1550059419861007
journals.sagepub.com/home/eeeg


Abstract

Background and Objective. In a previous study, we showed a new EEG processing methodology called Multi-Scale Ranked Organizing Map/Implicit Function As Squashing Time (MS-ROM/IFAST) performing an almost perfect distinction between computerized EEG of Italian children with autism spectrum disorder (ASD) and typically developing children. In this study, we assessed this system in distinguishing ASD subjects from children affected with other neuropsychiatric disorders (NPD). **Methods.** At a psychiatric practice in Texas, 20 children diagnosed with ASD and 20 children diagnosed with NPD were entered into the study. Continuous segments of artifact-free EEG data lasting 10 minutes were entered in MS-ROM/IFAST. From the new variables created by MS-ROM/IFAST, only 12 has been selected according to a correlation criterion. The selected features represent the input on which supervised machine learning systems (MLS) acted as blind classifiers. **Results.** The overall predictive capability in distinguishing ASD from other NPD cases ranged from 93% to 97.5%. The results were confirmed in further experiments in which Italian and US data have been combined. In this analysis, the best MLS reached 95.0% global accuracy in 1 out of 3 classes distinction (ASD, NPD, controls). This study demonstrates the value of EEG processing with advanced MLS in the differential diagnosis between ASD and NPD cases. The results were not affected by age, ethnicity and technicalities of EEG acquisition, confirming the existence of a specific EEG signature in ASD cases. To further support these findings, it was decided to test the behavior of already trained neural networks on 10 Italian very young ASD children (25–37 months). In this test, 9 out of 10 cases have been correctly recognized as ASD subjects in the best case. **Conclusions.** These results confirm the possibility of an early automatic autism detection based on standard EEG.

Keywords

children, electroencephalogram (EEG), artificial neural network, machine learning systems, autism spectrum disorder, differential diagnosis

Received September 25, 2018; revised March 4, 2019; accepted April 3, 2019.

Introduction

The cause of autism spectrum disorder (ASD) remains elusive despite a large amount of basic and clinical research performed over the past 10 years. There are reliable reasons to think that ASD is already present at birth, with many neurological changes that develop during fetal life in response to various heterogeneous factors.

As described in a recent article,¹ these include (a) alterations to columnar structure that have significant implications for the organization of cortical circuits and connectivity, (b) alterations to synaptic spines on individual cortical units that may underlie specific types of connectional changes, and (c) alterations within the cortical subplate—a region that plays a role in proper cortical development and in regulating interregional

communication in the mature brain. The relevant involvement of the cerebral cortex in substantially altering cortical circuitry explains the unique pattern of deficits and strengths that characterize cognitive functioning. However, this makes the

¹Villa Santa Maria Foundation, Neuropsychiatric Rehabilitation Center, Autism Unit, Tavernerio (Como), Italy

²Semeion Research Centre of Sciences of Communication, Rome, Italy

³Department of Mathematical and Statistical Sciences, University of Colorado at Denver, CO, USA

⁴Tarnow Center for Self-Management, Houston, TX, USA

Corresponding Author:

Enzo Grossi, Villa Santa Maria Foundation, Neuropsychiatric Rehabilitation Center, Autism Unit, IV Novembre street, Tavernerio (Como), 22038, Italy.
Email: enzo.grossi@bracco.com

potential usefulness of EEG recording plausible as a biomarker for these abnormalities.

ASD is associated with abnormal neural connectivity.²⁻⁷ Presently, neural connectivity is a theoretical construct that is hard to measure, but research in network science and time series analysis suggests that neural network structure, a marker of neural activity, is measurable with electroencephalography (EEG).⁸

The EEG can measure neural activity and may provide a useful tool to detect children at risk of developing ASD and, thus provide an opportunity for early intervention. In a recent review, the body of knowledge about the potential of EEG in ASD has been carefully described.⁹ The aim of this review was to examine evidence for the utility of 3 methods of EEG signal analysis in the ASD diagnosis and subtype delineation. Forty studies were identified and classified according to the EEG analysis method in 3 categories: (a) functional connectivity analysis, (b) spectral power analysis, and (c) information dynamics. All studies identified significant differences between ASD and non-ASD subjects, confirming the presence of specific EEG abnormalities. However, due to high heterogeneity in the results, generalizations were difficult to obtain and none of the methods alone have been proposed as a new diagnostic tool.

In a previous study, we have shown the ability of a novel kind of machine learning system (MLS) named MS-ROM/IFAST (Multi-Scale Ranked Organizing Map/Implicit Function As Squashing Time) developed by the Semeion Research Institute in Rome. This system can extract interesting features in computerized EEG that allow near perfect distinction of ASD children from those who are developing typically.¹⁰ This proof of concept study showed accuracy values near to 100% using a training-testing protocol and 84% to 92.8% using a leave-one-out protocol. The similarities among the MLS weight matrixes, measured with apposite algorithms, were not affected by the age of the subjects. This suggests the MLS does not read age-related EEG patterns, but rather invariant features related to the brain's underlying abnormalities.

The aim of this second study is to corroborate the findings obtained in the aforementioned pilot study where we distinguished ASD subjects from normally developing children. In this second study, we compared ASD subjects to children affected with other neuro-psychiatric disorders using EEG data collected with different equipment from another country (United States). In addition, to better circumstantiate the concept of EEG signature, experiments were done in which final classifiers have been trained on Italian data and tested on US data, and trained and tested on pooled Italian and US data. Finally, we trained the MLS on Italian and US data and then assessed how 10 completely unknown cases were classified.

Materials and Methods

Study Population

Diagnoses were made by board-certified psychiatrists and psychologists according to the DSM-V-TR (*Diagnostic and*

Statistical Manual of Mental Disorders, Fifth Edition, Text Revision) criteria. The data were collected over a 5-year period for those referred for an EEG assessment. The data was submitted to an institutional review board and granted a "waiver of approval," meeting the exemption categories set forth by federal regulation 45 CFR 46.101(b) [2] and [4].

Twenty subjects diagnosed with ASD and 20 subjects diagnosed with other neuropsychiatric disorders, matched for age and gender distribution, were obtained from the institutional review board-approved data archive of a psychiatric practice in the United States. The 2 groups had the same age range (4-14 years) and male/female ratio (14/6). None of these children were affected by genetic conditions, cerebral malformations, or epilepsy. In the comparison group, the range of primary diagnoses was the following: attention deficit hyperactivity disorder (ADHD) (n = 16), mood disorders (n = 2), anxiety disorders (n = 2).

EEG Recordings

The EEG data were recorded at a psychiatric center in the United States, at resting state, eyes-closed condition. EEG acquisition was performed using Mitsar-EEG-10/70-201 equipment, with impedance maintained below 10 kohm. The patients were seated in a slightly reclining chair in a silent and low light environment. An Electrocap was used to collect the data according to the international 10-20 system with linked ears montage (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2). A minimum of 20 minutes of total data were recorded in both eyes open (10 minutes) and eyes closed (10 minutes) resting conditions. The order of these could vary among patients. This study used only the eyes closed data to be consistent with our pilot study. Data regarding quantitative EEG (qEEG) performed on these 40 cases are provided as supplementary information.

The EEG track was then saved in the database. Subsequently, 10 minutes of recording were exported ASCII files through the same acquisition program SystemPlus Evolution and saved, to make it possible to read in numerical format.

The EEG file in ASCII format was then included in the program's mathematical neural network for analysis. A continuous segment of artifact-free EEG data lasting 10 minutes was used to compute multi-scale entropy values and for subsequent analyses of each subject with artificial adaptive systems as described below.

EEG Analysis

All EEG data were evaluated and interpreted by the same neurophysiologist, a member of the American Board of Electroencephalography and Neurophysiology and the American Board of Clinical Neurophysiology. The neurophysiologist was blinded to the subject's diagnoses and medications. Visual inspection of the EEG was performed to search for paroxysms that are either focal or lateralizing.

Extra Cases

As an additional test of validity for our method, it was decided to consider 2 other groups of subjects.

The first dataset considered consists of 35 Italian subjects (15 ASD, 10 normal development). All the subjects with ASD received an independent diagnosis for autism according to DSM-5 criteria, which was then confirmed by a qualified psychiatrist at Villa Santa Maria, Tavernerio (Italy). The ADOS (Autism Diagnostic Observation Schedule) scale was used for diagnosis. The ages in this case ranged from 7 to 14 years. The EEG data were recorded at Villa Santa Maria Institute by C.O. from the recruited subjects at resting state (eyes-closed and eyes-open). The EEG recordings (0.3-70 Hz bandpass) were performed by the Micromed device equipped by SystemPlus Evolution software, using prewired headsets with cotton elastic inside from 19 electrodes in silver and chlorinated plastic positioned according to the International 10-20 system (ie, Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2, and common ground). Registration took about 3 minutes and 256-Hz sampling rate and was vigilantly carried with the eyes closed.

The second external control dataset consists of 10 children (8 males; 2 females) who received an independent diagnosis for autism according to DSM-5 criteria. The age of these subjects was clearly very low in comparison with the main study group, ranging from 25 to 37 months. No autistic subject was found to be affected by genetic conditions, cerebral malformations documented by neuroimaging or epilepsy. The EEG data were derived from 3 to 5 minutes of registration at 256-Hz sampling rate carried out in resting state with the eyes closed, according to standard methodology and with electrodes positioned according to the International 10-20 system.

Analysis of EEG Raw Data With MS-ROM IFAST Algorithm

The mathematical equations and computer programs described below have been created by M.B. at Semeion Research Center.

A description of the mathematical and computational characteristics of IFAST has recently been published.¹¹ This system uses unsupervised neural networks in the preprocessing phase.¹²⁻¹⁴

The structure of IFAST is composed of different phases: (1) The squashing phase, corresponding to the transformation of the EEG channels of each subject into a single vector of features. (2) The noise elimination phase, where the dynamic elimination of the noisy features from the vector representing each subject is performed. (3) The classification phase, that is, the intelligent classification of the features of each subject with the support of MLS.

Squashing Phase. In technical terms, this phase needs to transform a 2-dimensional matrix of each EEG into a 1-dimensional vector. The columns of the EEG matrix are the channels' values and the rows correspond to the EEG discrete time flow. The

squashing phase is designed to identify the spatial invariants in the EEG time flow. Two fundamental aspects should be noted: first, This phase is completely blinded: no information about the diagnosis of the subjects is known to the processing algorithm; second, after the squashing phase, even EEGs of different duration become comparable.

This task has been achieved with the Multi-Scale and Ranked Organizing Maps (MS-ROM).

The MS-ROM algorithm owes its name to the self-organizing map (SOM) neural network,¹⁵ the well-known unsupervised algorithm on which it is based. The MS-ROM is a processing system composed of 3 steps: sampling, projection and ranking.

During the sampling step, each EEG track is sampled many times with different time scales, according to the multiscale entropy (MSE) methodology.^{16,17} At the end of this step, a certain number of subsamples of the starting EEG track is available. Then, for each subsample generated through MSE, a self-organizing map is initialized and trained. In this way, at the end of this step, each one of the subsamples is projected into a 2-dimensional space, that is the $R \times C$ grid of the SOM (R = rows, C = columns). All the SOMs share the same dimensions, R and C , and the same random starting weights. The decision to produce P-scaling processing will result in P subsamples and P -trained SOM grids. After the training phase, each cell of the grid of the i th SOM includes all the rows of the relevant i th subsample considered similar from the 19 EEG channels' perspective and it is summarized by a codebook, that is, the vector of the trained weights connecting the input to that cell. After the projection step, each grid generated by the SOM is ranked according to its cell frequency, also called point-frequency. The X and Y coordinates of each ranked grid represent the invariant features of this algorithm. The final length of the input vector for each EEG track is equals to $2 \times (R \times C) \times P$, it depends on the number of scaling processing P for each EEG, and on the number of the cells of each SOM grid of which both the X and the Y coordinates are considered (see Figure 1).

Despite its simplicity, this algorithm has been demonstrated to be very efficient in detecting the complex similarities among channels in an EEG track. The key to its effectiveness is the ranking process of each trained grid which can transform a geometrical projection into a topological vector. We outline this difference for many reasons: SOM artificial neural networks (ANNs) can design a tessellation of the input space into a 2-dimensional grid. When 2 independent SOMs, with the same number of rows and columns, are trained with 2 different samples starting with the same random weights, each SOM will generate a codebook map influenced by the number of entries and by the input vector variance. When we rank the cells of each of the 2 SOM grids according to their frequency, we ignore the specific metric of each SOM and we enhance their topological interrelationship.

In this study, each EEG track has been divided in 5 multiscale samples and trained for 100 epochs. Each SOM sample used a 7×7 grid. At the end of the training phase, each EEG was thus represented by a vector of 490 features: $(49 \text{ x-coordinates} + 49 \text{ y-coordinates}) \times 5 \text{ samples}$. The MS-ROM system is implemented in the academic free Semeion Software Suite.

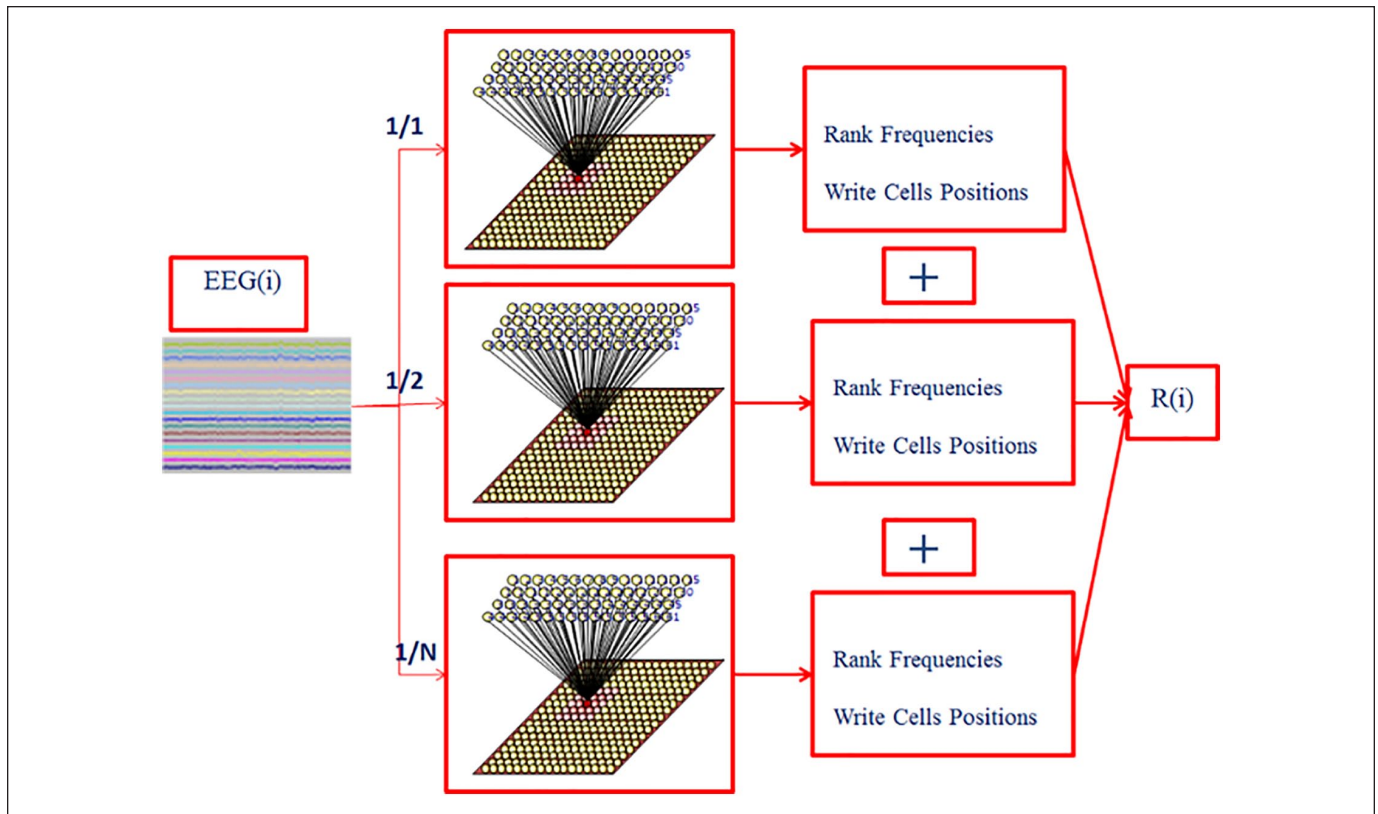


Figure 1. MS-ROM algorithm dynamics. Multiple SOM elaborations with a 7×7 grid are applied to the raw EEG data sampled with $1/1$, $1/2$, $1/3$, $1/4$, and $1/5$ pass according to the multiscale sampling procedure. The coordinates of cells are taken as input values for the final classifiers, ranked in order of the increasing number of records clustered in each cell. MS-ROM, Multi-Scale and Ranked Organizing Maps; SOM, self-organizing map.

Noise Elimination Phase. This fundamental phase serves to remove most of the noise introduced by the squashing phase. The starting 19 EEG variables (ie, the 19 electrodes traces) have been substituted with the new $2 \times (R \times C) \times P$ coming from the MS-ROM system but some of them can be redundant or even noisy and must be discarded. It should be noted that these variables do not correspond to measurable quantities but correspond to abstractions identified by the system. These virtual variables enclose and make more explicit all the information present in the EEG trace. In this study, a features selection criterion based on the correlation matrix has been used since if 2 features are independent, they are supposed to be even uncorrelated. From the 490 EEG features generated by the MS-ROM preprocessing algorithm, we select only those who have a linear correlation index with the diagnostic class higher than a certain value arbitrarily chosen. All the features having a linear correlation index below this threshold are eliminated while the remaining have been used as input data for supervised learning and testing.

It is important to note that, as we showed in our previous work,¹¹ the particular procedure employed in our method can avoid the need of traditional EEG preprocessing before characterization.

Classification Phase. The classification phase is executed through distinct types of learning machines. We have considered both the most well-known algorithms and new neural networks implemented in many freeware academic software packages to facilitate easy replication of our research.^{18,19}

Three different classification tests have been carried out.

Test 1. The main dataset (see “Study Population” paragraph in the “Materials and Methods” section) has been divided into 2 parts: a subset A consisting of 13 records (4 ASD and 9 affected by other pathologies) and a subset B consisting of 27 records (16 ASD and 11 Other pathologies) as shown in Table 1. This optimal subdivision has been carried out by means of an evolutionary algorithm named T&T.²⁰⁻²² that builds 2 sets, set A and set B, trying to approximate the same probability density function. Traditional techniques, such as cross-validation, leave-one-out variant and bootstrapping, do not guarantee good results when the global dataset is limited or complex, whose data is hyperpoint of an unknown nonlinear function: the sub-samples extracted are not always representative of the phenomenon and do not have the same probability density function. The division into training and testing sets on a random basis not only does not take into account the problem of outliers but has consequences in terms of variability of results.

Table 1. Summary of Data.

Subset	Autism Spectrum Disorder	Other Pathologies	Records, n	Chance Level, %
A	4	9	13	69.00
B	16	11	27	59.26
Total	20	20	40	

Table 2. Sample Division Into Sets A and B.

Subset	Autism Spectrum Disorder	Control	Other Pathologies	Records, n	Chance Level, %
Set A	18	4	8	30	60.00
Set B	17	6	12	35	48.57
Total	35	10	20	65	

Training and testing validation protocol has been used to compare the classification tasks. It consists in the execution of 2 independent procedure. The first, named a-b, uses as training set the previously built subset A and the subset B as testing set. The second, named b-a, reverses roles using subset B as training set and subset A as testing. The pair of predictions is then averaged to get a final value. Evaluating the results taking into account both classifications a-b and b-a prevents the possibility of selecting a particularly favorable sample.

The following different learning machines has been used for the final classification by processing both datasets: sine net neural networks (Sn)^{23,24}; back propagation algorithm (FF-Bp)²⁵; K-contractive map (K-CM)²⁶; kNN algorithm.²⁷

The objective of this trial is to establish whether, starting from the IFAST processing of EEGs, it is possible to distinguish between autistic subjects and subjects suffering from other pathologies, that is, to determine or not the existence of an ASD signature within the EEG.

Test 2. The second experiment has been carried out using the data coming from both the countries (ie, the main dataset and the first extra dataset). The new Italian–United States (IT-US) dataset is then composed of 65 cases: 10 typical (IT); 20 other pathologies (US); 35 ASD (15 IT + 20 US).

The EEG tracks have been subjected to the same procedure as the previous case: saved in a database and exported to ASCII for processing with MS-ROM/IFAST. The group of 15 ASDs consisted of 13 males and 2 females, leading to a male/female ratio of 27/8. The control group had a male/female ratio equals to 4/6 with and an age ranging from 7 to 12 years. The Other Pathologies group was the same previously described.

The cases have been divided into set a and set B, again by using T&T algorithm, to perform the same validation process previously used. Table 2 shows how the ASD, Control, and Other Pathologies subjects have been arranged in the subsets and the chance level for each subset.

The objective of this test is to determine whether the IFAST method can distinguish between ASD subjects, subjects with

Table 3. Summary of Data Relevant to Test 3.

Subset	Autism Spectrum Disorder	Other Pathologies	Records, n	Chance Level, %
A	20	20	40	50.00
B	15	10	25	60.00
Total	35	30	65	

other diseases and normal subjects. This type of analysis makes it possible to further highlight the presence or absence of a signature in the EEGs of autistic individuals. Moreover, since the EEGs come from different sources, it makes it possible to understand whether invariances determined by the MS-ROM/IFAST method have general validity.

Test 3. In the third experiment, the blind predictive power of the network was tested. Different methods of artificial intelligence have been trained through the modalities previously seen, dividing into training and testing set the 65 cases already used in Test 2 (see Table 3). Then, the second extra dataset, consisting of 10 subjects, all with ASD pathology, were used in the recall phase to see how much the algorithms could recognize completely extraneous pathological cases and to determine the MS-ROM/IFAST system's ability to extract invariant properties. Moreover, since the age of the subjects of the third dataset is decidedly lower than the ones used previously, we propose to understand if an early ASD diagnosis is possible.

Results

Test 1

The objective of this test is to determine whether it is possible to distinguish between autistic subjects and subjects suffering from other pathologies, using exclusively the IT data. Table 4 shows the results of the MS-ROM/IFAST protocol with learning machines using only 12 of the 490 features generated by MS-ROM system and selected by means of the linear correlation criterion, previously explained. To perform this kind of features extraction we chose the threshold $\theta = 0.35$ for all our tests. The 0.35 value was considered reasonable since the objective was to reduce the number of variables and eliminate redundancy; we wanted to consider only the informative variables. This value of θ has allowed us to have a 98% reduction in the number of variables and to obtain excellent results.

With the training-testing protocol, the overall predictive capability of the machine learning system employed in sorting out autistic cases from normal controls was consistently above 90%. Two of the machine learning systems reached an accuracy rate higher than 95%, as shown in Table 4. This confirms the superiority of this approach over the standard protocols available today as described in the review by Gurau et al.⁹

The ASD and Other Pathologies columns correspond to the sensitivities with regard to each group to be classified.

Table 4. Predictive Results From the EEG Features Detected by the MS-ROM Algorithm.

MLS	ASD, %	Other Pathologies, %	Overall Accuracy, %	Errors
FF_Bp(ab)	81.30	100.00	90.60	3
FF_Bp(ba)	100.00	88.90	94.40	1
Average	90.60	94.40	92.50	2
FF_Sn(ab)	87.50	100.00	93.80	2
FF_Sn(ba)	100.00	88.90	94.40	1
Average	93.80	94.40	94.10	1.5
K-CM(ab)	93.80	100.00	96.90	1
K-CM(ba)	100.00	100.00	100.00	0
Average	96.90	100.00	98.40	0.5
kNN(ab)	93.80	100.00	96.90	1
kNN(ba)	100.00	88.90	94.40	1
Average	96.90	94.40	95.70	1

Abbreviations: MS-ROM, Multi-Scale Ranked Organizing Map; MLS, machine learning system; FF_Bp, feed forward back-propagation neural network; FFSn, feed forward sine-net neural network; K-CM, K-contractive map neural network; kNN, k-nearest neighbors algorithm; ab, training on subset a and testing on subset b; ba, training on subset b and testing on subset a.

The best results were obtained using the K-CM algorithm, with an overall accuracy of 98.4%.

One of the prerogatives of K-CM is the possibility of visualizing the internal relationships among the test records as a semantic map, as shown in major detail in the mathematical appendix. Figure 2a and b show the clustering of the records in the 2 testing sets of the K-CM experiment according to their degree of membership to the 2 diagnostic classes in the training-testing sequence a-b (Figure 2a), where B was the testing set, and b-a (Figure 2b) where training and testing sets have been reversed. The unique misclassification is clearly visible (Figure 2a).

In other words, Figure 2 shows a graph that is able to describe the relationships among all the test subjects. Each node represents a subject, two nodes are linked if their distances from each train record are similar. This mechanism allows a natural clustering between similar objects. In particular, it clarifies why the error occurred: The network judged the ASD_15 subject to be more similar to controls than to other ASDs. The value of each connection expresses how similar the 2 connected nodes are and are represented by the thickness of the link. We expect, as it happens, that elements belonging to the same cluster have medium-high similarity values (thick line) while the link of that allows the passage from one to the other has a low value (fine line). The graph has been built on the basis of the matrix of “meta-distances” produced by the K-CM algorithm (see Equation 2 in the mathematical appendix).

To get a better understanding of the potential of the system, the fitness of predictions, expressed as fuzzy membership to ASD class, and the age of the subjects have been plotted (Figure 3). Although Sine Net makes two errors in the

classification, its output for ASD subjects has been chosen to be plotted since the fuzzy nature of the results allows us to consider them as the membership to the output class. Thus, Figure 3 shows 18 subjects despite of 16 since also the 2 mistakes must be considered being a part of the actual network outputs. It is however interesting to observe how even if the errors are not considered the trend, although obviously with a minor slope, remains the same (Figure 4). It should be noted that the r value is low ($r = -0.21$), and more important, the trend is inversely proportionate to the age. This is an important finding because we should expect to obtain better results at a very early age. As an example, we add to the plot the regression line related to the ASD subset predictions.

Test 2

The objective of this test is to determine whether it is possible to distinguish between autistic subjects, subjects suffering from other pathologies and normal subjects, using both the IT and the US data. To better substantiate the concept of the EEG signature, the cases, both in the previous and in the current study, until now, were considered separately depending on the country of origin (IT or US). The procedure was the same as that applied in the US case though, in this case, the system had to distinguish between three classes. The results attained in this experimentation are shown in Table 5.

The average overall accuracy obtained with two of the four MLS employed (K-CM and KNN) is remarkably high for 1 out of 3 types of prediction, reaching 95%.

Figure 5a and b shows clustering of the records in the 2 testing sets according to their degree of membership to the 2 diagnostic classes in the training-testing sequence a-b (Figure 5a) and b-a (Figure 5b), according to K-CM semantic map. The 4 misclassifications are clearly visible in Figure 5a and none appears in Figure 5b.

Test 3

The objective of this test is to understand if an early diagnosis is possible by testing the model on data from the EEG of young children. Table 6 shows the behavior in testing of neural networks and leaning machines tuned on the data of Test 2, 40 subjects in training and 25 in testing. In this case, using the usual threshold criterion, linked to the linear correlation, 49 variables were selected corresponding to a linear correlation lower than 0.2.

Once neural networks and MLSs have been trained, the 10 new subjects have been placed as inputs. The relevant classification results are shown in Table 7. Observe that, while BP and SineNet respond in a fuzzy way, through a gradient of membership, K-CM responds in a crisp way, with 0 or 1 values.

Discussion

There is a paucity of literature available on the EEG changes associated with autistic spectrum disorders as witnessed by the

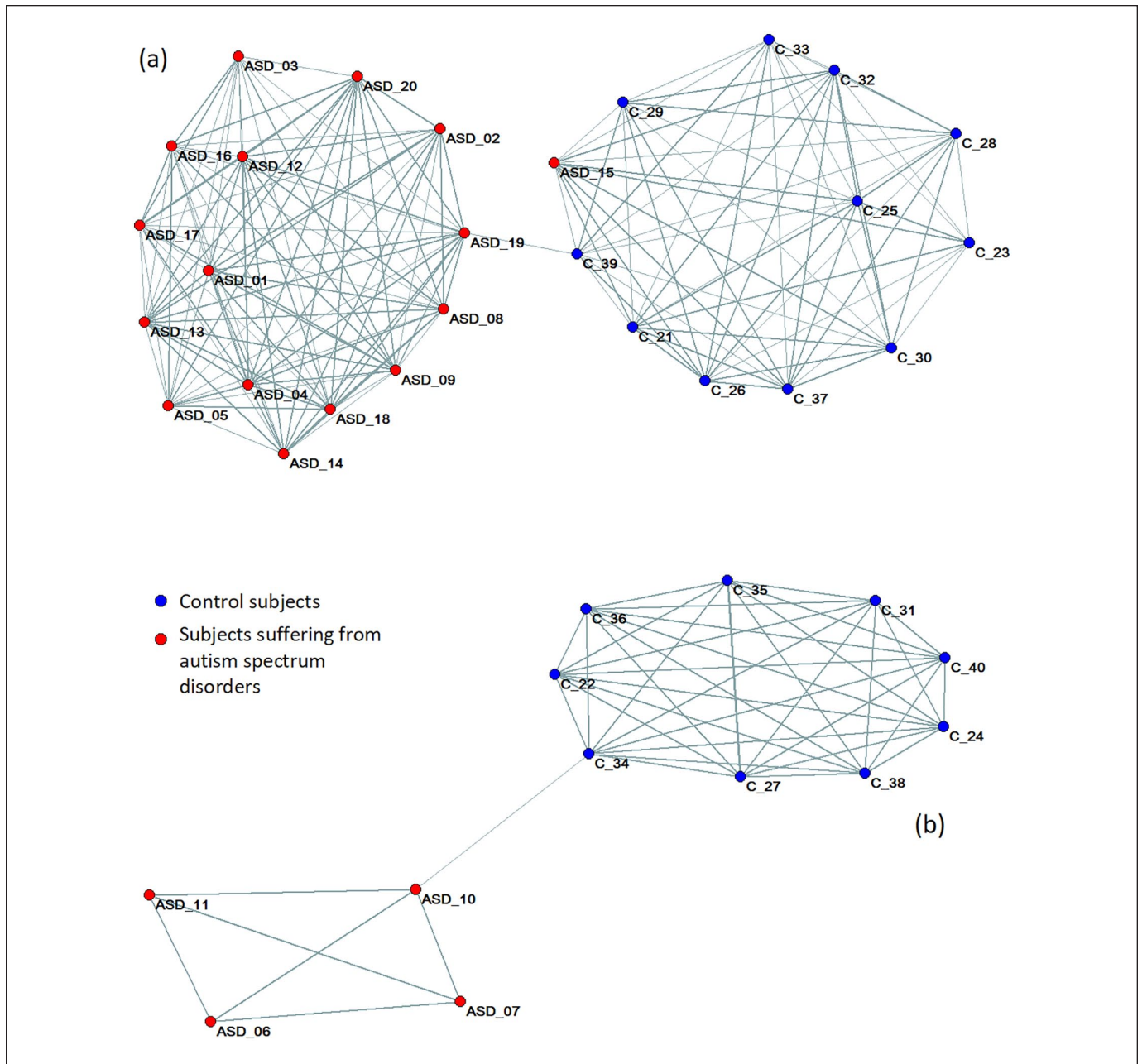


Figure 2. Clustering of the records in the 2 testing sets according to their degree of membership to the 2 diagnostic classes in the training-testing sequence a-b (figure a) and b-a (figure b). The line thickness indicates the strength of the connection, the thicker the connection, the greater the membership.

recent review by Gurau et al.⁹ However, most of the articles published refer to abnormalities in neural connectivity at both the global and local levels rather than to the potential diagnostic usage of this methodology. Our work appears to open a new frontier where mathematical optimization can emphasize the value of the data embedded in the EEG.

Autism spectrum disorder is characterized by a number of pathological changes that develop during fetal life in response to various heterogeneous toxic factors. These include (a) alterations to columnar structure that have significant implications

for the organization of cortical circuits and connectivity, (b) alterations to synaptic spines on individual cortical units that may underlie specific types of connectional changes, and (c) alterations within the cortical subplate—a region that plays a role in proper cortical development and in regulating interregional communication in the mature brain.¹ This pathophysiological substrate makes plausible the potential usefulness of EEG recording as a biomarker of loss of complexity as result of these abnormalities. The core of our algorithmic system relies in the mathematical optimization of the complex information

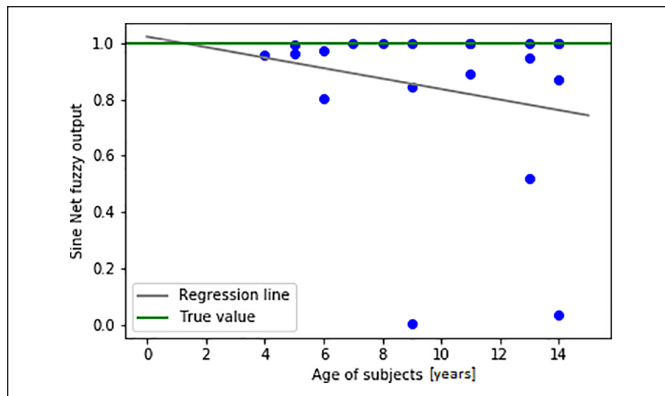


Figure 3. Correlation between neural network output in records classification in the autism spectrum disorder (ASD) subgroup and age of the subjects related to sine-net machine learning system (MLS). The interpolation straight line points toward a positive trend in the direction of early age. It seems promising to have the best results at the smallest ages.

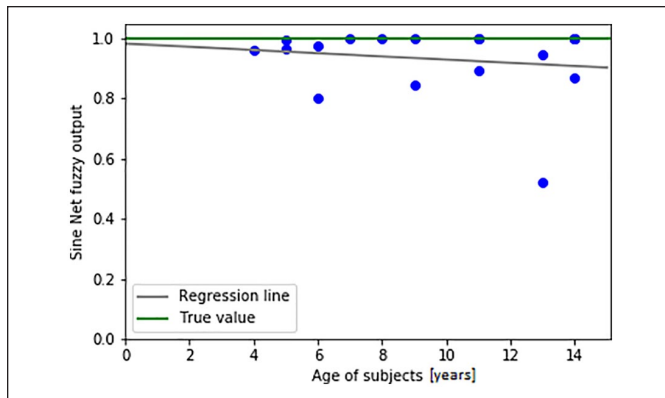


Figure 4. Analysis of the correlation without misclassifications. In this case, $r = -0.17$.

embedded in the EEG. The EEG complexity is completely blinded to human eyes even for the most expert neurophysiologist who is trained to observe single channels traces one at time, like most mathematical methods based on power spectra analysis. For this reason, we are not extracting particular features like those coming from quantified EEG, which express useful clinical insights but, on the contrary, we use the native information, that is, the unprocessed EEG, in a global way analyzing each time point of EEG referred to all 19 channels series. Each time point raw, made up by 19 numbers, becomes an individual. Our system groups together the most similar individuals through SOM. The clustering structure essentially reflects the EEG complexity. The final classifiers learn to distinguish ASD from control subject according the different complexity of their EEG.

Our original article addressed the use of artificial neural networks for an automatic classification of ASD. Useful data can be obtained by extracting spatial information contained in the resting EEG. The core of the procedure was that the ANNs did

Table 5. Predictive Results From the EEG Features Detected by the MS-ROM Algorithm After Training-Testing on Italian and US data.^a

MLS	ASD	Control	Other Path.	Overall accuracy	Errors
FF_Bp(ab)	0.78	I	0.92	0.9	7
FF_Bp(ba)	0.88	0.83	0.86	0.86	3
FF_Bp(Avrg)	0.83	0.92	0.89	0.88	5
FF_Sn(ab)	0.74	I	0.92	0.89	8
FF_Sn(ba)	0.88	0.83	0.71	0.81	4
FF_Sn(Avrg)	0.81	0.92	0.82	0.85	6
K-CM(ab)	0.85	I	I	0.95	4
K-CM(ba)	I	I	0.86	0.95	I
K-CM(Avrg)	0.93	I	0.93	0.95	2.5
kNN(ab)	0.85	I	I	0.95	4
kNN(ba)	I	I	0.86	0.95	I
kNN(Avrg)	0.93	I	0.93	0.95	2.5

Abbreviations: MS-ROM, Multi-Scale Ranked Organizing Map; ASD, autism spectrum disorder; MLS, machine learning system; FF_Bp, feed forward back-propagation neural network; FF_Sn, feed forward sine-net neural network; K-CM, K-contractive map neural network; kNN, k-nearest neighbors algorithm; ab, training on subset a and testing on subset b; ba, training on subset b and testing on subset a.

^aThe ASD, Control and Other Pathologies columns correspond to the sensitivities with regards to each group to be classified.

not classify individuals using EEG data as an input. Rather, the data inputs for the classification were the weights of the connections within an ANN trained to generate the recorded EEG data. The results were superior to those obtained with the more advanced currently available nonlinear techniques.

Now, the use of the same procedure based on EEG data from another country, collected with a different EEG machine, and with different protocols of registration and length (10 minutes vs 3 minutes in the original article) adds credibility to the goodness-of-fit results proving that the system is able to distinguish not only ASD from typically developing subjects but also ASD subjects from those with other pathologies (Test 1).

This seems to highlight that a neural networks approach shows great potential to identify invariant features independent from technicalities related to EEG recording. The system in place is remarkable in a number of ways:

- Avoidance of preprocessing phase and filtering procedure of EEG data: We have shown that the new algorithms do not require the preprocessing of the EEG before applying artificial neural networks. On the contrary, they use the raw EEG information which can very efficiently accomplish our final objective. This improvement would resolve the problem of recording data with open or closed eyes and could represent an important practical simplification of the procedure.
- More robust classification phase: We have tested four distinct types of learning machines based on a vast array

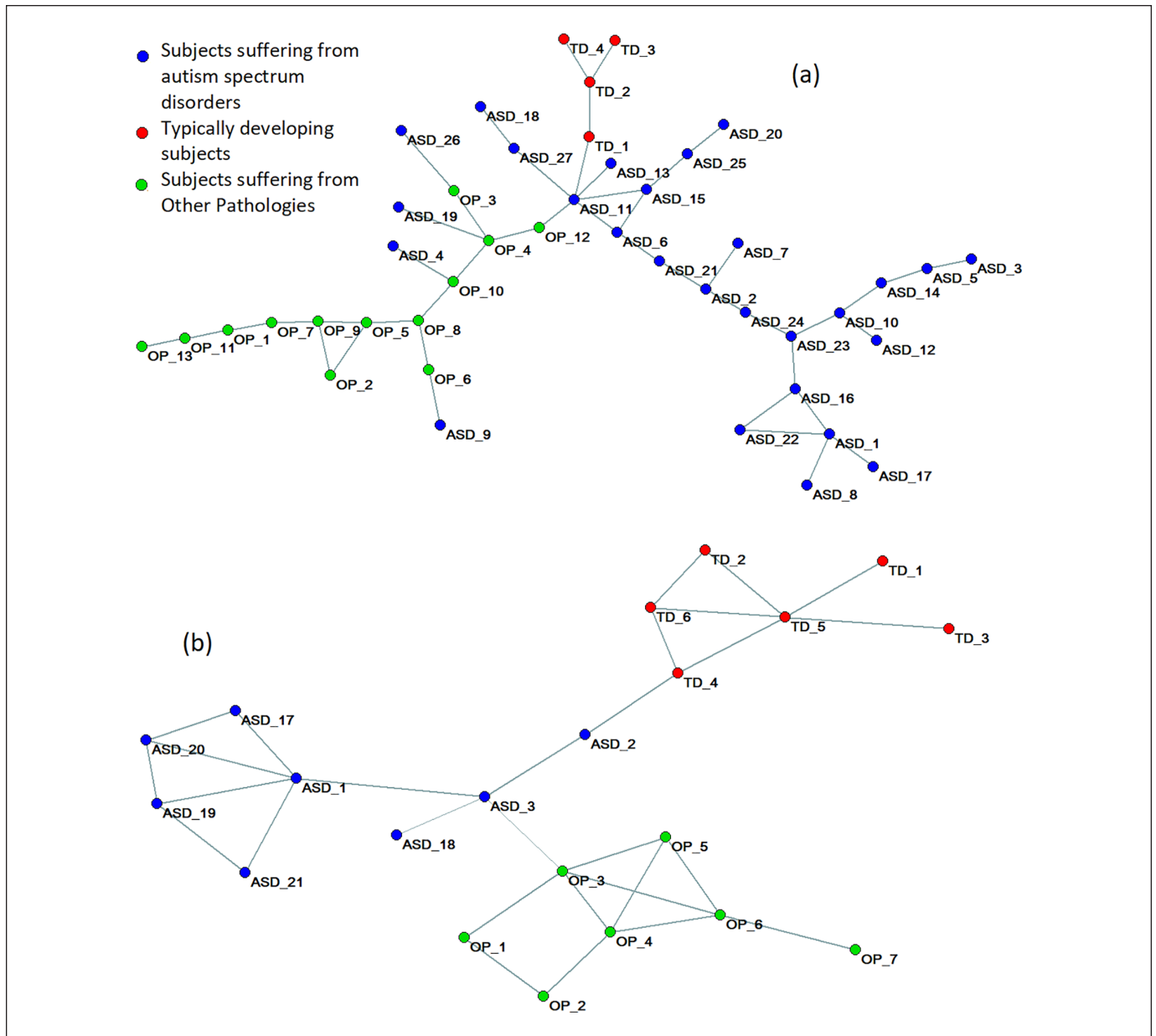


Figure 5. Clustering of the records in the 2 testing sets according to their degree of relationship to the 3 diagnostic classes in training-testing sequence a-b (figure a) and b-a (figure b).

of specific mathematics, showing the robustness of IF-AST similarly produces a very high accuracy rate. This gives consistency to the generalization of these findings in the real world.

(c) Extraction of spatial invariants of the EEG through MS-ROM.

To better circumstantiate the concept of an EEG signature, we carried out a second experiment in which final classifiers were trained and tested on combined Italian and US data (Test 2). After cross-validation on this mixed group of 10 typical IT; 20 other pathologies US; 35 ASD (15 IT + 20 US), the overall prediction, related to 1 out of 3 classes, with 2 of the 4 MLS

employed (K-CM and kNN) reached 92 % of global accuracy. This means that the results are not affected by age, ethnicity, and technicalities of EEG acquisition, confirming therefore, the existence of a specific EEG signature in ASD.

Finally, to be sure that the correctness of the predictions was absolutely free from conditioning related to the EEG used, we decided to train the neural networks on the dataset of Test 2 and to recall 10 ASD subjects completely unknown to the system (Test 3). The results obtained were encouraging as, in the best case, 10 out of 10 subjects were correctly classified. A further prerogative of the 10 new cases used lies in the very low age of the patients. The challenging objective that constitutes the end of this line of experimentations, in fact, lies precisely in the

Table 6. Results Attained by the MLS Before the Recall Phase.

MLS	ASD, %	Control, %	Accuracy, %	Errors
FF_Bp	100.00	100.00	100.00	0
k-NN	100.00	72.73	86.36	3
FF_Sn	91.67	100.00	95.83	1
K-CM	100.00	72.73	86.36	3

Abbreviations: ASD, autism spectrum disorder; MLS, machine learning system; FF_Bp, feed forward back-propagation neural network; kNN, k-nearest neighbors algorithm; FF_Sn, feed forward sine-net neural network; K-CM, K-contractive map neural network.

Table 7.

Subject	FF_Bp	FF_Sn	k-NN	K-CM
Case_1	0.74	0.46	0	0
Case_2	0.77	0.93	1	1
Case_3	0.77	0.98	1	1
Case_4	0.91	1	1	1
Case_5	0.92	1	1	1
Case_6	0.77	0.89	1	1
Case_7	0.77	0.92	1	1
Case_8	0.39	0	1	1
Case_9	0.85	1	1	1
Case_10	0.95	1	1	1

Abbreviations: FF_Bp feed forward back-propagation neural network; kNN, k-nearest neighbors algorithm; FF_Sn, feed forward sine-net neural network; K-CM, K-contractive map neural network.

construction of a model that allows the earliest possible diagnosis of autism spectrum disorders. Both Tests 1 and 2 seem to encourage this ambition.

K-CM besides having obtained the best results allows to build a semantic map able to represent the relationships detected within the test set. The K-CM semantic map is useful for evaluating results and capturing further information, such as the major similarities between testing records. This neural network, as underlined in the appendix, is able to combine the kNN technique, effective but sometimes not precise, with the most sophisticated artificial intelligence techniques able to explain to the maximum the information hidden in the data. This combination of different but compatible techniques used as a pipeline has allowed to get the best from each one and to obtain, in spite of this delicate problem, positive and encouraging results for possible future applications. Moreover, contrary to the classical neural networks, K-CM lets the input variables interact in an “all against all” mode, in auto-association, before being used to determine the output (fuzzy profiling, see mathematical appendix). This freedom of interaction makes it possible for highly nonlinear connections between variables and targets to emerge spontaneously.

The implementation of this network, resulting from the combination of methods already known and published, is not problematic even if, obviously, as the amount of data increases, the speed of learning may be affected. However, it should be

considered that, basically, the actual learning phase is carried out at intervals that are rather distant from each other. The diagnosis phase of a new subject (recall phase), on the other hand, is extremely rapid, so the hypothesis of a system that works almost in real time is absolutely plausible and to be hoped. Over the past decade, a growing body of evidence points out that to effectively treat ASD, the earliest possible detection is directly proportional to successful therapy.²⁸⁻³⁰ This clearly implies that the availability of an accurate and relatively inexpensive diagnostic method for early diagnosis should be one of the medical community's highest priorities.

Conclusions

This study points out that standard EEGs contain the delicate information that can substantially differentiate typically developing brains from brains belonging to children with ASD, provided that this information is processed with very sophisticated analytical systems like those employed by our group. Our study demonstrates how advanced artificial adaptive systems are able to distinguish people with autism from both normal and other types of neuropsychiatric disease. The average accuracy obtained in the ASD versus control subjects classification reaches 98.4%, while in the 1 of 3 experiments between ASD, control, and other pathologies subjects it reaches 95.0%.

The intelligent systems used seem to be more effective in the case of younger children. This is crucial as it would be a major step toward early diagnosis.

In addition, although MLS were trained on data from older children, it was possible to correctly classify 10 out of 10 cases in a much lower age group. This seems to suggest that autistic spectrum disorders carry a fingerprint present in EEGs and that early diagnosis may be possible.

This procedure has proved effective both on the data recorded in Italy and in the United States, confirming that the results do not depend on the acquisition mode or on the specific devices but only on the type of analysis: the EEG.

We are fully aware of the limitation on relatively small sample size. Further work is needed before deriving definitive conclusions. A larger sample is needed to confirm our results. If the relevant results obtained in this limited-sample study were to eventually be confirmed in a larger study, and then the open question would become: Is this signature already present during the first months of a child's life? Answering this is the aim of our future research, which we hope will make a significant step forward in the treatment of this impairing disease.

Appendix

K-Contractive Map

To allow a better understanding of the results, in this section, a brief explanation about K-CM and its peculiarities will be given.

K-CM²⁶ is a methodology consisting of different steps, able to solve supervised pattern recognition problems. At first, the

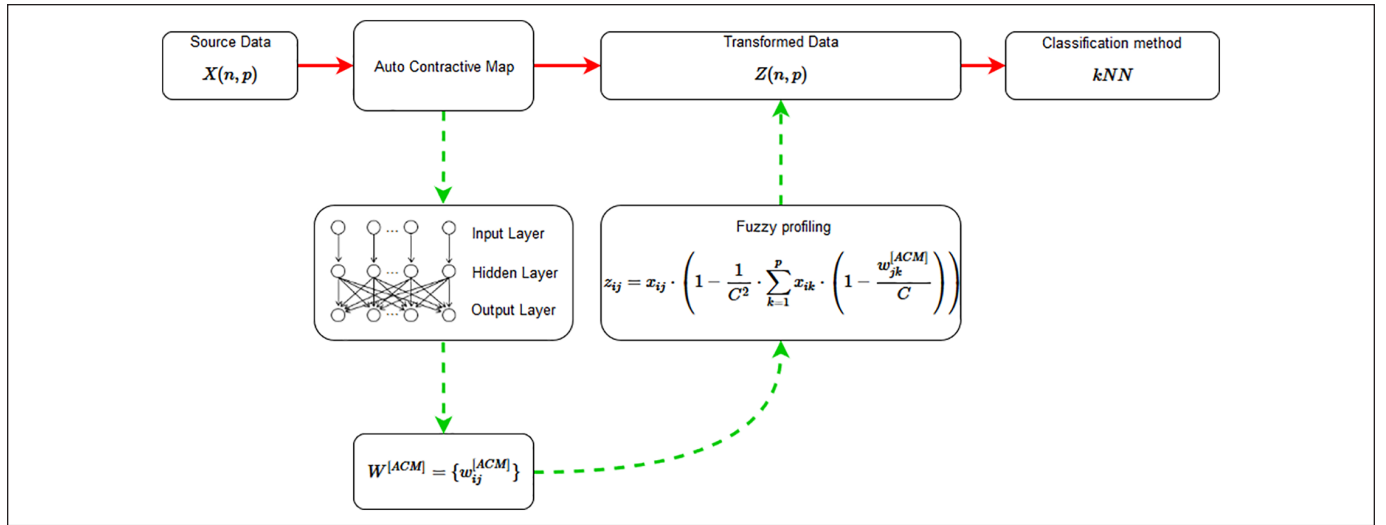


Figure 6. The K-Contractive Map (K-CM) flow, from the source data to the classification.

source dataset $X(n, p)$, consisting of p variables and n records, is used as input for auto contractive map (AutoCM),³¹ an unsupervised neural network particularly effective in making explicit the hidden relationships in the data. The values of the relationships found by AutoCM are stored in $\mathbf{W}^{[ACM]}$, the weights matrix. Each record of $X(n, p)$ is then rewritten taking into account the values of all the many-to-many relationships in a new fuzzy dataset $Z(n, p)$. The Z transformed dataset has been proved to be often more informative than the original.²⁶ $Z(n, p)$ is then split in training and testing set and used as input of a kNN classifiers. The complete methodology flow is shown in Figure 6.

It should be noted that K-CM is one application of the more general contractive map technique according which many other classification or regression methods can be applied on $Z(n, p)$, in the case of supervised task, and in the case of unsupervised dataset many exploratory data analysis methods can be used.

One of the further potentials of this method is the capability to draw a semantic graph of the relationships among the test samples. At first, the Euclidean distance between each test and training sample (record + target) is computed according to Equation (1).

$$D_{m,i} = \sqrt{\sum_{j=1}^p (z_{m,j} - z_{i,j})^2 + \sum_{g=1}^G (t_{m,g} - t_{i,g})^2} \quad (1)$$

where $m \in \{1, 2, \dots, n_{test}\}$, $i \in \{1, 2, \dots, n_{train}\}$ and $g \in \{1, 2, \dots, G\}$ denotes the possible targets. In this case, $t_{i,g}$ corresponds to the real class of the i th record of the training set while $t_{m,g}$ is the estimated class of the m th record in testing. Then, distance matrix $\mathbf{D} = \{\mathbf{D}_{m,i}\}_{m,i}$ is not a square matrix but it is a $n_{test} \times n_{train}$ rectangular matrix.

As a second step in the procedure to build the K-CM semantic graph, the distance among the \mathbf{D} distances have to be computed according to Equation (2).

$$M_{m,m'} = \sqrt{\sum_{j=1}^p (D_{m,i} - D_{m',i})^2} \quad (2)$$

where $m, m' \in \{1, 2, \dots, n_{test}\}$.

Since by means of equation (2) the distance matrix of distances is computed, M has been named *Meta Distance Matrix*. The Meta Distance Matrix is an $n_{test} \times n_{test}$ symmetric matrix.

Now, according to the graph theory, it is possible to use M as the weights matrix of a weighted graph $G(V, E, M)$ where the vertices V are the n_{test} records of the testing set, the edges

E are all the $\frac{n_{test} \cdot (n_{test} - 1)}{2}$ possible connections and M_{ij} corresponds to the weight of the (i, j) edge.

Many filters can be applied over the graph in order to visualize and highlight the most relevant connection, for example, the Minimum Spanning Tree (MST)³² or the Maximally Regular Graph (MRG).³³ The MST is a spanning tree, that is, a tree containing all the vertices, whose sum of edge weights is as small as possible. The MRG is a graph built starting from the MST by adding the skipped links until the maximum of the graph complexity is not reached. The complexity of the graph is computed by a specific function named H .³³ In this study, the MRG filter has been used. The resulting graph tends to put nearby elements whose distances from the training records are similar. Then, if K-CM performed a good classification it is reasonable to expect a clustering of the testing set according to the real belonging classes and a highlight of the misclassifications. Thus, the semantic graph serves to visualize if and how K-CM understood the problem.

Authors' Note

Final data sets are made available on request to corresponding author.

Acknowledgments

The author(s) thank Ms Judy Crawford for her invaluable help in English editing, and Prof Filippo Muratori and Dr. Federico Sicca from Stella Maris Institute of Tirrenia (Italy) for their kind provision of EEG data of the 10 cases belonging to external validation group.

Authors Contributions

EG: study plan and organization; manuscript preparation. MB: data management and processing; application of artificial adaptive systems. FDT: data management and processing; application of artificial adaptive systems; manuscript preparation. RJS: data collection; manuscript supervision.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Statement

The study has been carried out in agreement with Helsinki Declaration. The data collected were all part of the routine assessment of individuals suspected of having autism and parents signed an informed consent form to allow the scientific use of this information.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Enzo Grossi  <https://orcid.org/0000-0003-0346-2684>

References

- Hutsler JJ, Casanova MF. Review: cortical construction in autism spectrum disorder: columns, connectivity and the subplate. *Neuropathol Appl Neurobiol.* 2016;42:115-134. doi:10.1111/nan.12227
- Wang J, Barstein J, Ethridge LE, Mosconi MW, Takarae Y, Sweeney JA. Resting state EEG abnormalities in autism spectrum disorders. *J Neurodev Disord.* 2013;5:24. doi:10.1186/1866-1955-5-24
- Belmonte MK, Allen G, Beckel-Mitchener A, Boulanger LM, Carper RA, Webb SJ. Autism and abnormal development of brain connectivity. *J Neurosci.* 2004;24:9228-9231. doi:10.1523/JNEUROSCI.3340-04.2004
- Assaf M, Jagannathan K, Calhoun VD, et al. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage.* 2010;53:247-256. doi:10.1016/j.neuroimage.2010.05.06
- Minshew NJ, Keller TA. The nature of brain dysfunction in autism: functional brain imaging studies. *Curr Opin Neurol.* 2010;23:124-130. doi:10.1097/WCO.0b013e32833782d4
- Sato JR, Vidal MC, de Siqueira Santos S, Massier KB, Fujita A. Complex network measures in autism spectrum disorders. *IEEE/ACM Trans Comput Biol Bioinform.* 2018;15:581-587. doi:10.1109/TCBB.2015.2476787
- Nelson CA 3rd. Introduction to special issue on the role of connectivity in developmental disorders: genetic and neural network approaches. *Dev Sci.* 2016;19:523. doi:10.1111/desc.12477
- Bosl W, Tierney A, Tager-Flusberg H, Nelson C. EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Med.* 2011;9:18.
- Gurau O, Bosl WJ, Newton CR. How useful is electroencephalography in the diagnosis of autism spectrum disorders and the delineation of subtypes: a systematic review. *Front Psychiatry.* 2017;8:121.
- Grossi E, Olivieri C, Buscema M. Diagnosis of autism through EEG processed by advanced computational algorithms: a pilot study. *Comput Methods Programs Biomed.* 2017;142:73-79. doi:10.1016/j.cmpb.2017.02.002
- Buscema M, Vernieri F, Massini G, et al. An improved I-FAST system for the diagnosis of Alzheimer's disease from unprocessed electroencephalograms by using robust invariant features. *Artif Intell Med.* 2015;64:59-74. doi:10.1016/j.artmed.2015.03.003
- Buscema M, Rossini P, Babiloni C, Grossi E. The IFAST model, a novel parallel nonlinear EEG analysis technique, distinguishes mild cognitive impairment and Alzheimer's disease patients with high degree of accuracy. *Artif Intell Med.* 2007;40:127-141.
- Rossini PM, Buscema M, Capriotti M, et al. Is it possible to automatically distinguish resting EEG data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy? *Clin Neurophysiol.* 2008;119:1534-1545. doi:10.1016/j.clinph.2008.03.026
- Buscema M, Grossi E, Capriotti M, Babiloni C, Rossini P. The I.F.A.S.T. model allows the prediction of conversion to Alzheimer disease in patients with mild cognitive impairment with high degree of accuracy. *Curr Alzheimer Res.* 2010;7:173-187.
- Kohonen T. *Self-Organizing Maps*. Berlin, Germany: Springer-Verlag; 1995.
- Thuraisingham RA, Gottwald GA. On multiscale entropy analysis for physiological data. *Physica A.* 2006;366:323-332.
- Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2005;71(2 pt 1):021906.
- Buscema M. Supervised ANNs and Organism [computer software]. Version 29.1. Rome, Italy: Semeion Software 12; 1999-2018.
- Buscema M. I.F.A.S.T.: AutoAssociative Metatasks [computer software] Version 9.1. Rome, Italy: Semeion Software 32; 2005-2013.
- Buscema M. *T&T: A New Pre-Processing Tool for Non-Linear Data Set (s)*. Rome, Italy. Semeion Technical Paper; 2001; 25.
- Buscema M, Breda M, Lodwick W. Training With Input Selection and Testing (TWIST) algorithm: a significant advance in pattern recognition performance of machine learning. *J Intell Learning Syst Appl.* 2013;5:29-38.
- Vomweg TW, Buscema M, Kauczor HU, et al. Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Med Phys.* 2003;30:2350-2359.
- Buscema M, Terzi S, Breda M. Using sinusoidal modulated weights improve feed-forward neural network performances in classification and functional approximation problems. *WSEAS Trans Inform Sci Appl.* 2006;3:885-893.
- Buscema PM, inventor; Applicant Semeion Research Centre, assignee. Sine Net: an artificial neural network. European Patent (Application n. 03425582.8 deposited 09-09-2003). USA Patent No US 7,788,196 B2. International Patent: Application PCT/EP2004/05189 deposited 08-28-2004. August 31, 2010.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986;323:533-536.

26. Buscema M, Consonni V, Ballabio D, Mauri A, Massini G, Breda M, Todeschini R. K-CM: a new artificial neural network. Application to supervised pattern recognition. *Chemometrics Intell Lab Syst.* 2014;138:110-119.
27. Kowalski BR, Bender CF, The K-Nearest Neighbor classification rule (Pattern Recognition) applied to nuclear magnetic resonance spectral interpretation. *Anal Chem.* 1972;44:1405-1411.
28. Landa R. Early communication development and intervention for children with autism. *Ment Retard Dev Disabil Res Rev.* 2007;13:16-25.
29. Reichow B. Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *J Autism Dev Disord.* 2012;42:512-520.
30. Rogers SJ. Brief report: early intervention in autism. *J Autism Dev Disord.* 1996;26:243-246.
31. Buscema PM, Massini G, Breda M, Lodwick WA, Newman F, Asadi-Zeydabadi M. *Artificial Adaptive Systems Using Auto Contractive Maps: Theory, Applications and Extensions.* New York, NY: Springer; 2018.
32. Cormen TH, Leiserson CE, Rivest RL, Stein C. *Introduction to Algorithms.* Cambridge, MA: MIT Press; 2009.
33. Buscema M, Sacco PL. Auto-contractive maps, the H function, and the maximally regular graph (MRG): a new methodology for data mining. In: Capecchi V, Buscema M, Contucci P, D'Amore B, eds. *Applications of Mathematics in Models, Artificial Neural Networks and Arts.* Dordrecht, Netherlands: Springer; 2010:227-275.