

Pregnancy risk factors in autism: a pilot study with artificial neural networks

Enzo Grossi¹, Federica Veggo^{1,2}, Antonio Narzisi³, Angelo Compare⁴ and Filippo Muratori⁵

BACKGROUND: Autism is a multifactorial condition in which a single risk factor can unlikely provide comprehensive explanation for the disease origin. Moreover, due to the complexity of risk factors interplay, traditional statistics is often unable to explain the core of the problem due to the strong inherent nonlinearity of relationships. The aim of this study was to assess the frequency of 27 potential risk factors related to pregnancy and peri-postnatal period.

METHODS: The mothers of 45 autistic children and of 68 typical developing children completed a careful interview. Twenty-four siblings of 19 autistic children formed an internal control group.

RESULTS: A higher prevalence of potential risk factors was observed in 22 and 15 factors in external control and internal control groups, respectively. For six of them, the difference in prevalence was statistically significant. Specialized artificial neural networks (ANNs) discriminated between autism and control subjects with 80.19% global accuracy when the data-set was preprocessed with TWIST (training with input selection and testing) system selecting 16 out of 27 variables. Logistic regression applied to 27 variables gave unsatisfactory results with global accuracy of 46%.

CONCLUSION: Pregnancy factors play an important role in autism development. ANNs are able to build up a predictive model, which could represent the basis for a diagnostic screening tool.

Autism is a specific neurodevelopmental condition that typically displays qualitative socio-communicative impairment and restricted, stereotyped interests and activities (1). Although a large proportion of children with autism manifests abnormal development during the first year of life, 15–62% of them show a regression between 18 and 24 mo of age after a period of apparently typical development (2,3).

Approximately 70% of individuals with autism present a variable degree of intellectual disability (4) and expressive/receptive language can be absent or very insufficient (5). Other problems, not exclusive of autism, are attention-deficit and disturbed behaviors as hetero-autism. Thirty percent of

children manifest epileptic seizures by late childhood or adolescence and 10% of cases are associated with several genetic disorders such as tuberous sclerosis, Angelman syndrome, phenylketonuria, and fragile X syndrome (6).

The etiopathogenesis of autism is not yet understood; the prevalence is undoubtedly rising, and it is not clear if this increase is linked to the diagnostic improvement or to a greater susceptibility of the population to the disease. Many twins and family studies point out the importance of inherited predisposition to the disorder even though epidemiologic research suggest the strong contribution of prenatal and early postnatal environmental factors. In fact, genetic factors alone account for no more than 20–30% of all cases, whereas other 70–80% are the result of a complex interaction between environmental risk factors and inherited or *de novo* genetic susceptibility (7).

Many pregnancies and obstetric complications have been correlated in several studies to a greater risk of autism (8–11). Recent meta-analyses (12,13) underline the role of advanced maternal and paternal age, gestational infections, low birth weight, fetal hypoxia, preterm birth, and labor complications in the pathogenesis of autism. There is still a debate in international literature about some perinatal risk factors, for example, it is not clear if there is an association between autism and neonatal hyperbilirubinemia (13,14).

Many studies have investigated the role of prenatal maternal smoking (15), parental psychiatric conditions (16,17) including baby blues symptoms (8,15), and depressive/anxious traits of parents and siblings (18). Less prominence has been given to maternal stress and stressful life events during pregnancy or in first months after birth (15,19). Psychological stress during pregnancy increases the secretion of corticotropin-releasing hormone from the hypothalamus and regulates the hypothalamic–pituitary–adrenal axis. Elevated plasma levels of corticotropin-releasing hormone correlate to preterm labor and to anxiety perceived during pregnancy. Acute stress also elevates serum levels of interleukin-6 produced by mast cells, responsible for neuroinflammatory response and disruption of gut–blood barrier and blood–brain barrier permitting neurotoxic molecules to enter the brain, contributing to autism pathogenesis and symptoms (8). After delivery, maternal stress, in particular during the first 6 mo of life of

¹Department of Autism Research, Villa Santa Maria Institute, Tavernerio, Italy; ²Department of Child and Adolescent Psychiatry, University of Milano-Bicocca, Monza, Italy;

³Department of Developmental Neuroscience, IRCCS Stella Maris Foundation, Calambrone, Italy; ⁴Department of Psychology, University of Bergamo, Bergamo, Italy;

⁵Department of Developmental Neuroscience, University of Pisa, Calambrone, Italy. Correspondence: Enzo Grossi (enzo.grossi@bracco.com)

Received 29 January 2015; accepted 4 August 2015; advance online publication 23 December 2015. doi:10.1038/pr.2015.222

Table 1. Frequency distribution and odds ratio of factors under study: subjects affected by autism compared with typicals

Factors	Autism group, <i>n</i> = 45	Typicals group, <i>n</i> = 68	Odds ratio	<i>P</i> value
	Frequency (number of cases)	Frequency (number of cases)		
Pregnancy order 1 (vs. pregnancy order 2–3 + pregnancy order >3)	46.70% (21)	60.30% (41)	0.8	0.514
Pregnancy order 2–3 (vs. pregnancy order 1 + pregnancy order >3)	48.90% (22)	39.70% (27)	1.4	0.3357
Pregnancy order >3 (vs. pregnancy order 1 + pregnancy order 2–3)	4.40% (2)	0.00% (0)	n.d.	n.d.
Father age >40 at conception	8.90% (4)	11.76% (8)	0.7	0.6282
Mother age > 40 at conception	4.40% (2)	1.47% (1)	3.1	0.3594
Smoking at conception	31.10% (14)	22.06% (15)	1.6	0.2827
Smoking during pregnancy	11.10% (5)	7.35% (5)	1.6	0.4939
Alcohol (pregnancy)	4.40% (2)	2.94% (2)	1.5	0.6742
Solvents/paints exposure (pregnancy)	24.40% (11)	4.41% (3)	7.0	0.0045
Polyvinyl chloride flooring	22.00% (10)	36.76% (25)	0.5	0.1048
Tap water (pregnancy)	31.10% (14)	27.94% (19)	1.2	0.7169
Average number of stressful events/pregnancy	0.48	0.14	-	0.0127 ^a
Death or severe disease of a relative	4.40% (2)	4.41% (3)	1.01	0.9934
Divorce, separation, or conjugal conflict	11.10% (5)	1.47% (1)	8.4	0.0563
Loss of house, eviction, or relocation	8.90% (4)	4.41% (3)	2.1	0.343
Abuse or violence	2.20% (1)	0.00% (0)	n.d.	n.d.
Difficulties with income or job	15.60% (7)	4.41% (3)	4.0	0.0544
Postpartum depression	22.00% (10)	11.8% (8)	2.1	0.1426
Fever (pregnancy)	15.60% (7)	14.7% (10)	1.0	0.9016
Use of potentially teratogenic drugs (pregnancy)	0.00% (0)	2.94% (2)	n.d.	n.d.
Pregnancy complications	51.10% (23)	26.5% (18)	2.9	0.0086
Dystocic (delivery)	6.70% (3)	4.41% (3)	1.5	0.6032
Cesarean section (delivery)	31.10% (14)	16.2% (11)	2.3	0.0649
Perinatal complications	37.70% (17)	20.6% (14)	2.3	0.0476
Low gestational age	20.00% (9)	13.2% (9)	1.6	0.339
No breastfeeding	33.30% (15)	16.2% (11)	2.6	0.0371
Early antibiotic therapy	26.70% (12)	11.8% (8)	2.7	0.0471

The 16 variables selected by TWIST system are given in bold.

n.d., not determined; TWIST, training with input selection and testing.

^at-test;

the child, can create important alterations in the early affective mother–child interaction conditioning emotional–behavioral responses of the offspring.

Recently, other studies have focused on the protective or toxic role of chemical substances on the brain of the fetus. Mothers of typically developed children have a greater intake of prenatal vitamins and, in particular, folic acid compared with mothers of children with autism (20). An early *in utero* exposure to traffic-related pollution and particulate matter (21) or other neurotoxic elements widespread in the environment as phthalates (endocrine-disrupting chemicals normally present in many plastic materials as polyvinyl chloride) is suspected to interfere with the neurodevelopment and favor the onset of developmental disorders such as autism (22).

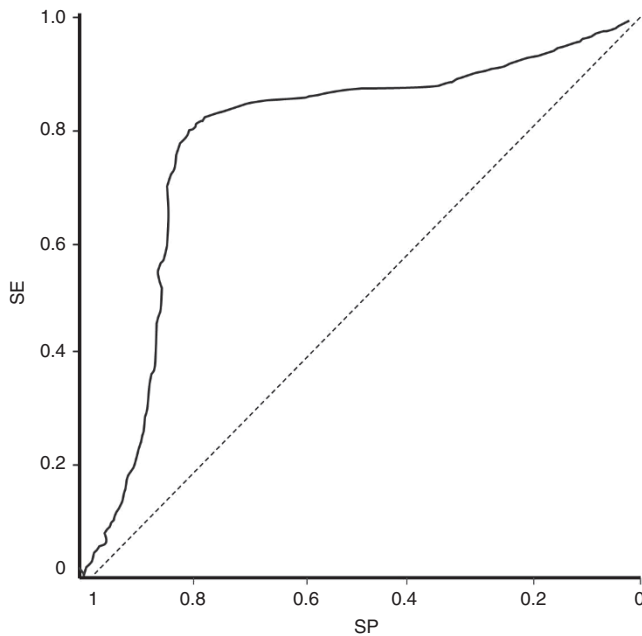
Autism is a multifactorial disease in which a single risk factor unlikely can provide comprehensive information. Moreover, due to the complexity of interplay of risk factors, traditional statistics is often unable to grasp the core of the problem due to the strong inherent nonlinearity of relationships.

Artificial neural networks (ANNs) are artificial adaptive systems, inspired by the functioning processes of the human brain. These computational systems are able to modify their internal structure in relation to a function objective. So, they are particularly suited for solving nonlinear problems, being able to reconstruct the fuzzy logic rules that govern the optimal solution for these problems. The ability to learn through an adaptive way (i.e., extracting from the available data the information needed to gather a specific task and to generalize

Table 2. Predictive results obtained in prediction with ANNs and ROC analysis

Neural network	Records in testing	Sensitivity (%)	Specificity (%)	Overall accuracy	ROC AUC
FF_Bp ab	53	85	75.64	81.82	0.795
FF_Bp ba	60	78.83	81.3	77.57	0.795
Mean		81.91	78.47	80.19	0.795

ab, training on subset (a) and testing on subset (b); ANN, artificial neural network; ba, training on subset b and testing on subset a; FF_BP, feed-forward back-propagation architecture; ROC, receiver-operating characteristic.

**Figure 1.** ROC AUC for artificial neural networks in identifying autism and typical. SE, sensitivity; SP, specificity.

the acquired knowledge) is a characteristic that makes the ANNs models a very powerful tool for data analysis even in presence of relatively small samples. The internal structure and functional organization of such systems can be updated and modified with respect to the environmental changes enabling the ANN to create its own representation of the information. Moreover, these models are interesting also for their noise tolerance that allows accurate performances in the presence of unreliability, wrong data, or measurement errors.

In recent years, ANNs have been used successfully in medicine. For example they have been used to investigate the predictive values of risk factors on the conversion of amnesic mild cognitive impairment to Alzheimer's disease (23), to identify placental determinants of fetal growth (24), to identify genetic variants essential to differentiate sporadic amyotrophic lateral sclerosis cases from controls, and to distinguish between Alzheimer's disease patients and controls, among others. Although ANNs offer promise for improving the predictive value of traditional statistical data analysis and have been successfully used in many areas of medicine, no reports have so far investigated the ability of ANNs in predicting autism risk.

The aims of this pilot study were to evaluate the differential frequency of several potential risk factors related to pregnancy, peri and early postnatal period investigated through a careful interview of mothers of autistic children in comparison with mothers of typicals, and to assess the predictive capacity of ANNs in distinguishing consistently these two different conditions employing rigorous validation protocols.

RESULTS

The mean Hollingshead index of parents of autistic children resulted significantly lower in comparison with that recorded in control group parents. *T*-test gave a *P* value < 0.0001.

Table 1 shows the comparison of frequencies of different risk factors in the autistic and external control groups. A statistically significant difference was found for the following six factors, which were all more frequent in autistic group: solvents/paints occupational exposure during pregnancy (*P* = 0.004), stressful events during pregnancy (*P* = 0.012), pregnancy complications (*P* = 0.008), perinatal complications (*P* = 0.047), deficient breastfeeding after delivery (*P* = 0.037), and early antibiotic therapy of the newborn (*P* = 0.047).

The results obtained with multivariable logistic regression were rather unsatisfactory, with very poor sensitivity rate (0.33) and specificity rate (0.59); overall accuracy = 0.46.

The application of TWIST algorithm to eliminate noisy variables before the main test of pattern recognition allowed the selection of 16 attributes (given in bold in **Table 1**). It is interesting to note that this evolutionary hybrid algorithm adaptively selected variables even with very low correlation index to build up the best predictive model to discriminate between cases and controls, confirming the well-known predisposition of ANNs to handle nonlinearity without major problems.

Table 2 shows the predictive results obtained with back-propagation ANN and **Figure 1** the relative receiver-operating characteristic (ROC) curve. ANNs reach an average global accuracy rate of 80.19%.

We also had the opportunity to recall the pregnancy history from 19 mothers with autistic children related to 25 siblings. In this way, a much matched case-control comparison was possible. **Table 3** shows the frequencies of different factors in these two groups. Beyond statistical significance, which was not reached in any of the factors due to the small sample size, it is interesting to note how most of the factors in absolute terms were more frequent in autistic group, in particular, solvents/paint exposure, stressful events, perinatal complications, low gestational age, deficient breastfeeding, and early antibiotic therapy, mirroring the situation of the comparison with typical group. Due to small sample size, ANNs application was not feasible; therefore, for this subgroup, nonmultivariate analysis has been performed.

DISCUSSION

This study represents a new model for epidemiological survey in the autism field and the first attempt to our knowledge to develop a predictive model for autism risk derived by factors collected during and immediately after the pregnancy

Table 3. Frequency distribution and odds ratio of factors under study: subjects affected by autism compared with siblings

Factors	Autism group, <i>n</i> = 19	Siblings, <i>n</i> = 24	Odds ratio	<i>P</i> value
	Frequency (number of cases)	Frequency (number of cases)		
Pregnancy order 1 (vs. pregnancy order 2–3 + pregnancy order >3)	47.4% (9)	42.1% (8)	1.8	0.3
Pregnancy order 2–3 (vs. pregnancy order 1 + pregnancy order >3)	52.6% (10)	62.5% (15)	0.7	0.5
Pregnancy order >3 (vs. pregnancy order 1 + pregnancy order 2–3)	0.0% (0)	4.2% (1)	n.d.	n.d.
Father age >40 at conception	0.0% (0)	8.3% (2)	n.d.	n.d.
Mother age >40 at conception	0.0% (0)	0.0% (0)	n.d.	n.d.
Smoking at conception	15.8% (3)	12.5% (3)	1.3	0.7
Smoking during pregnancy	0.0% (0)	0.0% (0)	n.d.	n.d.
Alcohol (pregnancy)	0.0% (0)	0.0% (0)	n.d.	n.d.
Solvents/paints exposure (pregnancy)	21.1% (4)	12.5% (3)	1.9	0.4
Polyvinyl chloride flooring	31.6% (6)	20.8% (5)	1.7	0.4
Tap water (pregnancy)	36.8% (7)	20.8% (5)	2.2	0.2
Average number of stressful events/pregnancy	0.47	0.29	—	0.3 ^a
Death or severe disease of a relative	0.0% (0)	8.3% (2)	n.d.	n.d.
Divorce, separation, or conjugal conflict	5.3% (1)	8.3% (2)	0.6	0.7
Loss of house, eviction, or relocation	10.5% (2)	4.2% (1)	2.7	0.4
Abuse or violence	0.0% (0)	0.0% (0)	n.d.	n.d.
Difficulties with income or job	21.1% (4)	8.3% (2)	2.9	0.2
Postpartum depression	15.8% (3)	16.7% (4)	0.9	0.9
Fever (pregnancy)	26.3% (5)	33.3% (8)	0.7	0.6
Use of potentially teratogenic drugs (pregnancy)	0.0% (0)	12.5% (3)	n.d.	n.d.
Pregnancy complications	68.4% (13)	45.8% (11)	2.6	0.1
Dystocic (delivery)	5.3% (1)	4.2% (1)	1.3	0.9
Cesarean section (delivery)	26.3% (5)	12.5% (3)	2.5	0.2
Perinatal complications	31.6% (6)	8.3% (2)	5.1	0.1
Low gestational age	21.1% (4)	0.0% (0)	n.d.	n.d.
No breastfeeding	31.6% (6)	29.2% (7)	1.1	0.9
Early antibiotic therapy	26.3% (5)	12.5% (3)	2.5	0.2

n.d., not determined.

^at-test.

with ANNs. In our opinion, some points of strength reserve consideration.

The first aspect is related to the selection of potential risk factors from an accurate review of published epidemiological case-control studies published in the literature. The selection has taken into account many factors beyond both their actual relevance in the published report from statistical point of view, considering their plausibility in relation to the most recent pathogenic theories. In fact, statistical relevance in epidemiology is often obscured by nonlinear functions, which hamper the execution of classical statistical tests. On the other hand, since we employed ANNs, we were aware of the possibility to build up efficient models also with variables having poor linear relation with the target variable, represented by diagnostic class.

ANNs in fact were able to build up a predictive model with high sensitivity and specificity with a nonparsimonious selection of variables despite the relatively small sample size.

ANNs are computerized systems that use nonlinear statistical analysis to reveal previously unrecognized and/or weak relationships between the input and output variables. These multilayered connectionist models are adaptive; that is, they dynamically find the connections between various datasets. ANNs also have the capability of generalization: when an ANN has been trained with proper data for determining the rules that better describe a certain phenomenon, it also can make some correct generalizations responding correctly to data it has not yet processed. In other words, after a learning period, the ANN can predict the output on inputs of further unknown cases.

Table 4. Variables on study

Demographics	Behavior/ environment	Stressful events	Health problems during pregnancy	Delivery problems	Postpartum
Pregnancy order 1	Smoking at conception	Death or severe disease of a relative	Fever	Dystocic delivery	Low gestational age
Pregnancy order 2–3	Smoking during pregnancy	Divorce, separation, or conjugal conflict	Use of potentially teratogenic medications	Cesarean section	Breastfeeding
Pregnancy order >3	Alcohol during pregnancy	Loss of house or evicted or relocation	Pregnancy complications	Perinatal complications	Early antibiotic therapy
Father age at conception	Occupational exposure to solvents/paints	Abuse or violences	—	—	Postpartum depression
Mother age at conception	Drinking tap water	Job strain	—	—	—
—	PVC flooring at home	Average number of stressful events	—	—	—

PVC, polyvinyl chloride.

ANNs have proved to be superior to classical logistic regression in cases in which excessive collinearity among independent variables or poor linear correlation indexes among independent and dependent variables impair the logistic modeling.

The second aspect is the use of a structured direct interview of the mothers performed by a medical doctor, always the same. Most studies published in this field have used data derived from computerized health records stored in large hospital databases. It is clear that these databases in case of retrospective studies cannot address efficiently delicate factors as life stressful events or behavioral lifestyles. The confidence posed by mothers in the professional skill of educators and therapists of the center and in particular to the medical doctor interviewing was so strong that none of the mothers refused the informed consent to cooperate in the one-to-one long interview. This direct interview with confident attitude can explain at last in part the unusually high frequency of some specific factors encountered in our study.

We observed a similarity in differential pattern of frequencies both in external case-control and in internal case-control comparison with a strong contribution of solvents/paint exposure, stressful events, perinatal complications, low gestational age, no breastfeeding, and early antibiotic therapy. This in our view is a remarkable sign pointing out the reliability of the findings, especially when the same mother is control of herself.

The parental occupational exposure has already been object of study (25) and a link between some chemicals and the presence of autism spectrum disorders in the offspring seems to be more and more realistic. In our interviews, we did not start from a specific list of chemical agents, but we first asked the mothers to describe what kind of job or task they had when they were pregnant and then we retraced the toxic agents involved. It is important to note that some women, who worked in factories in close contact to potentially toxic agents, left the place of work only after discovering the pregnancy, so the fetus during the first weeks of gestations was not protected.

The increased prevalence of breastfeeding deficiency after delivery, early antibiotic therapy of the newborn, and cesarean section in autistic group, both in external and internal control

comparison, deserves a special comment, since all these three factors are able to impair the gut microbiota of the newborn. There is a unanimous consensus about the fact that vaginal delivery and breastfeeding are essential in providing the appropriate starting bacterial substrate for the development of a physiological gut flora. Early antibiotic therapy obviously can influence the equilibrium of intestinal flora as well. The possibility that autism is the consequence of an imperfect development of gut flora is supported by a number of observations like: the frequent coexistence of gastrointestinal symptoms in autistic children, the appearance of the disease after an incidental antimicrobial therapy, and the increased levels of urinary biomarkers of specific pathogens of *Clostridium* spp. in the urine of autistic children (26).

In a recent study, the authors demonstrated that in a particular model of autistic mouse displaying behavioral symptoms relevant to autism spectrum disorders and other neurodevelopmental disorders, dysbiosis of the commensal microbiota, and alterations in serum metabolites are treated by the administration of a particular commensal (*Bacteroides fragilis*). These findings represent a major breakthrough in the microbiota hypothesis of autism spectrum disorder (27).

The role of stressful events in contributing to increased autism risk also deserves a special attention since very few studies have attempted to collect this kind of information. Adverse experiences during the prenatal period (a time of rapid growth and of heightened brain plasticity) have been demonstrated to induce significant effects on neurobiology, metabolism, and physiology that can persist across the life span (28). Generally, the more variable the stressor and the earlier the stressors occur in pregnancy, the more profound the effect on offspring development. Neurobiological studies on animals show, on male offspring, that stressful events in pregnancy affects prefrontal and temporal cortical regions. Studies in rats showed that the development of layer II/III pyramidal neurons in the dorsal anterior cingulate and orbitofrontal cortex is significantly affected in offspring of mothers exposed to stress during pregnancy. Moreover, in both cortical areas, a significant reduction of dendritic length was observed in males (29). Previous

Table 5. Mathematical and technical characteristics of back propagation ANN used to model the dataset composed of 16 variables selected by TWIST system

Main ANN technical features	Main ANN architecture	
Learning law: back propagation	Input: 16	
Topology: feed forward	Hidden: 1:12	D rate: 0.1000
Temperature: Constant		W range: 1.0000
Momentum: self-momentum		Function: 12 sigmoid
Rate rule: fixed rate	Output: 2	D rate: 0.1000
Output function: soft max		W range: 1.0000
Input output: indirect connection		Function: 2 sigmoid
Input scaling: input (0,1)	Training set file: set A training patterns: 55	
Target scaling: 1.0000	Testing set file: set B testing patterns: 43	
Number of weights: 2,016	Process strategy: free random	

ANN, artificial neural network; TWIST, training with input selection and testing.

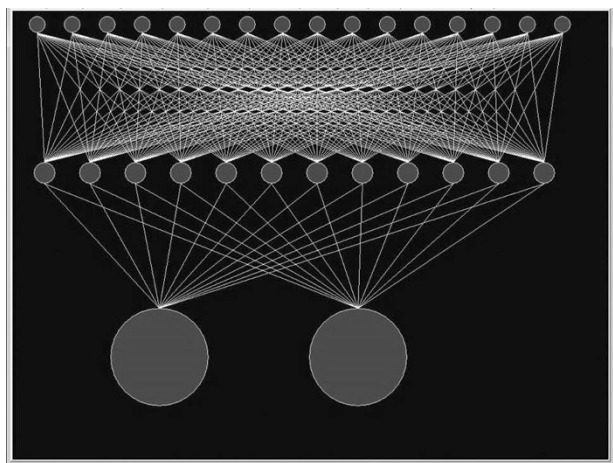


Figure 2. Architecture of back-propagation ANN used for prediction.

study (30) shows a significant effect of emotional stress experienced by rats before pregnancy on brain development of their offspring. In 21-d-old rats, layer V neuronal nuclei were enlarged, while the thickness of the parietal cortex tended to decrease. Thirty-day-old experimental rats demonstrated low exploratory activity in the plus-maze test. Moreover, offspring rats of mothers exposed to stress during pregnancy show that interleukin-1 β exposure had reduced estradiol content in hippocampus, medial prefrontal cortex, and diencephalon (31). The precise neurobiological mechanisms of action underlying these phenomena remain unknown.

Recently, in a small, explorative study, a focal disruption of cortical laminar architecture was found in prefrontal and temporal cortical tissue of a majority of young children with autism. These data support a probable dysregulation of layer

formation and layer-specific neuronal differentiation at prenatal developmental stages (32).

The third remarkable aspect is the use of ANNs to handle the data and to overcome the sample size limits. ANNs aim to understand natural processes and recreate those processes using automated models. These networks determine a noteworthy improvement, compared with traditional methods of analysis, and allow a method of forecasting with better and safer understanding of the relationship between variables, in particular, nonlinear relationships. ANNs function by initially learning a known set of data from a given problem with a known solution (training), and then the networks, inspired by the analytical processes of the human brain, are able to reconstruct the imprecise rules which may be underlying a complex set of data (testing). Using these tools has been possible to achieve a consistent predictive rate of pregnancy outcomes across different experiments with a ROC area under the curve (AUC) equal to 0.795. This value, following Sweet and Pickett criteria, can be considered as moderate diagnostic accuracy. The interpretation of the AUC values is traditionally as follows: an AUC < 0.7 suggests “low” diagnostic accuracy; an AUC from 0.7 to 0.9 suggests “moderate” diagnostic accuracy, and an AUC $\geq$ 0.9 suggests “high” diagnostic accuracy (33).

There are also some substantial weaknesses in the present study, key among them the understandably small sample size which makes difficult to be fully confident in the statistical estimates. In particular, the small sample size creates situations in which high odds ratio values do not convey necessarily significant *P* values (this occurs for example when the odds ratio of postpartum depression was calculated in [Table 1](#) or pregnancy complications in [Table 3](#)). Another limitation is the imperfect matching of cases and external controls in terms of geographical distribution. Nevertheless, mothers could and did complete the survey, even if the recall sometimes was hard, and this is meaningful for them. This implies that the protocol of this pilot study is feasible and can easily be replicated in larger sample. An additional limitation could be the retrospective self-report of occupational exposure to toxic agents, which may result in bias and misclassification. Anyway this potential memory-based bias is equal in the three groups (clinical sample, internal control, and external control).

Conclusions

The authors feel confident to draw the following conclusions:

1. the frequency of some risk factors like exposure to solvents or paints, stressful events, pregnancy complications, perinatal complications, among others, is significantly higher in autistic group compared with external and internal control children;
2. ANNs are able to build up a predictive model with high sensitivity and specificity with a nonparsimonious selection of variables despite the relatively small sample size;
3. the use of ANNs may be an important advance research on complex disorder as autism.

METHODS

Participants

Samples. The clinical sample included a cases group of 45 autistic children and adolescents (4 females and 41 males; mean age: 12.88, SD 3.49) and two control groups: (i) an external control group formed by 68 typical developing children (29 females and 39 males; mean age: 9.13, SD 4.20) and (ii) an internal control group formed by 24 siblings (11 females and 13 males; mean age 13.3 y) of 19 autistic children (19 males; mean age 12.08 y).

Case features. Forty-five children and adolescents received independent diagnoses of autism according to DSM-5 criteria, then confirmed by a qualified child and adolescent psychiatrist at Villa Santa Maria, Tavernerio (Italy) where the patients reside. The 45 cases selected are those children whose mothers gave their informed consent. Among the pregnancies of the patients selected for the study, 37 occurred in Lombardy (a northern region of Italy), 5 in central and southern Italian regions, and 3 in European countries (e.g., Romania, Albania, Spain). All the participants showed a comorbid intellectual disability. All subjects were assessed with the Childhood Autism Rating Scale, and the scores had a range from a minimum of 30 to a maximum of 60 points; in particular, 8 subjects were assessed as affected by mild-to-moderate autism and 37 by severe autism. No autistic child was affected by genetic conditions, cerebral malformations documented by neuroimaging, and epilepsy. The socioeconomic status was calculated with the Hollingshead index (AA Hollingshead, personal communication), a measure of social status of an individual based on educational attainment and occupational prestige. The child participant's parent's education code is rated on a 7-point scale that lists highest grade completed, in which 7 = graduate/professional training, 6 = standard college or university graduation, 5 = partial college, at least 1 y of specialized training, 4 = high school graduate, 3 = partial high school, 10th or 11th grade, 2 = junior high school, including 9th grade, 1 = less than 7th grade, 0 = not applicable or unknown. The child participant's parent's occupational code is rated on a 9-point scale, for which the Hollingshead manuscript provides a more detailed list: 9 = higher executive, proprietor of large businesses, major professional, 8 = administrators, lesser professionals, proprietor of medium-sized business, 7 = smaller business owners, farm owners, managers, minor professionals, 6 = technicians, semi-professionals, small business owners (business valued at \$50,000–70,000), 5 = clerical and sales workers, small farm and business owners (business valued at \$25,000–50,000), 4 = smaller business owners (<\$25,000), skilled manual laborers, craftsmen, tenant farmers, 3 = machine operators and semi-skilled workers, 2 = unskilled workers, 1 = farm laborers, menial service workers, students, housewives, (dependent on welfare, no regular occupation), 0 = not applicable or unknown. socioeconomic status of the clinical sample was equal to 2.55 (SD 1.07).

The recruitment of controls started with a public announcement spread in the geographical area of Tavernerio, where Villa Santa Maria is set. Sixty-eight mothers of typical developing children answered and decided to participate (the participation rate was 100%). In particular, 31 subjects were children of the staff members of Villa Santa Maria and 37 were children of women living in the area. The control subjects recruited were 29 females and 39 males, with a mean age of 9.13 y (SD 4.20). All the pregnancies occurred in Lombardy region. None of these subjects claimed symptoms related to autism spectrum disorders or learning disabilities. The mean socioeconomic status of the control group was equal to 3.51 (SD 1.01).

Parental Interview

Mothers of autistic children who met the inclusion criteria for the research were contacted by Villa Santa Maria staff. If the mother agreed to participate, she was invited to an individual structured interview about early risk factors summarized in Table 4 after an explanation of the scope of the study and after having signed an informed consent. Mothers with nonautistic siblings were interviewed in separate sessions (two or three) each dedicated to a specific pregnancy. The first interview was always dedicated to the autistic kid. The same procedure (contact, meeting, explanation, agreement, and interview) was performed for the mothers of typical developing children. The interview was always performed by the same person, a female medical

doctor specialized in child neuropsychiatry (F.V.). The interview used as check list a questionnaire with 27 items with possible answers as yes/no or as single choice from a closed list. The answers were then collected in a electronic dataset where every column represents a single risk factor (e.g., exposure to solvents; smoking at conception) and then coded as absent (0) or present (1). In particular, we asked the mothers to describe their jobs and tasks during the pregnancies, and “exposure to toxics” was coded when a continuous and relevant professional exposure to particular substances was present. In particular, in this study, we focused on paints, lacquers, paint thinner solvents, detergents, disinfectants, toluene, benzene, phenol, and metal (in particular chromium) powders that could represent a toxic insult for the fetus. “Polyvinyl chloride flooring” means the presence of this kind of pavement in the work place or the house. We chose polyvinyl chloride flooring as one of the most representative and widespread fabric containing phthalates because polyvinyl chloride furniture or window treatments are less common in the considered geographic area.

Conditions as placental diseases, preeclampsia, infections, oligodramnios, etc. were included in the category “pregnancy complications”; fetal respiratory distress, cyanosis, neonatal hyperbilirubinemia, etc. were included in “perinatal complications”; low gestational age is considered a birth occurred before 37th wk of gestation; every breastfeeding suspended before the first month of life of the baby was considered “no breast feeding”; every antibiotic therapy administered from birth to the 3 mo of life was considered “early antibiotic therapy.”

Statistical Analyses

Data were presented as number with percentages or as means with SDs for nominal and continuous variables, respectively. We performed frequencies comparison among risk factors under study with odds ratio, confidence intervals, and chi-square test. Multivariable modeling of data was carried out with logistic regression and supervised ANNs using data from autistic and external controls interviews. Sibling's data were only used for frequency analysis.

Logistic regression. Multivariable logistic regression has been applied by means of WEKA data mining software (34) using the same training and testing protocol.

Supervised ANNs. To evaluate the possibility to predict outcome (autistic vs. typical external controls) starting from risk factors on study, we used supervised ANNs developed in Semeion Research Centre using the training and testing validation protocol.

Training and testing validation protocol implies the following steps:

1. Subdivision of the dataset into two subsamples, subset A and subset B. In the first run, subset A is used as the training set and the subset B as the testing set.
2. Application of ANN on the training set. In this phase, the ANN learns to associate the input variables with those indicated as targets.
3. At the end of the training phase, the weights matrix produced by the algorithm is saved and frozen together with all of the other parameters used for the training.
4. The testing set is then shown to a virgin twin (same architecture and base parameters) ANN with the same weights matrix of the trained ANN, acting as final classifier so that for each case of the testing set, new for the ANN, the classifier expresses a classification in one of the two targets based on the learning previously carried out during training. This operation takes place for all records in testing each and result (right or wrong classification) is not communicated to the classifier. This allows to assess the generalization ability of trained ANN.
5. In a second run, another virgin ANN is applied to subset B which is used as training subset and then to subset A which is used as a testing subset.
6. Therefore, the results are relevant to two sequences of training testing protocol: A-B and B-A. Results are expressed in terms of sensitivity (correct classification of autism), specificity (correct classification of typical), and global accuracy (arithmetic mean between sensitivity and specificity). Overall results are expressed as average of the two experiments.

This crossover procedure allows to classify blindly all the records with the trained algorithm ensuring the generalization capability of the model on records never seen before.

The split of original data set in two subsets has not been obtained with a simple random split but with an evolutionary algorithms called TWIST, recently described (35) and developed in a special noncommercial research software at the Semeion Research Center in Rome, Italy.

TWIST algorithm. TWIST algorithm is a complex evolutionary algorithm able to look for the best distribution of the global dataset divided in two optimally balanced subsets containing a minimum number of input features useful for optimal pattern recognition. TWIST is an evolutionary algorithm based on a seminal paper about genetic doping systems, already applied to medical data with very promising results (36,37). Usually, TWIST evolutionary system is constituted by a population of multilayer perceptrons. Each ANN has to learn a subset of the global dataset and has to be tested in a blind way with another subset. TWIST selected some of the original attributes and generated two optimal subsets for training and testing with a reduced number of variables. We then applied the training testing protocol to the reduced dataset to verify the predictive capacity of supervised back-propagation neural network on the attributes selected by TWIST.

Predictive results and ROC curve. Predictive results have been expressed with sensitivity, specificity, and overall accuracy rates. ROC AUC has also been calculated. The ROC curve was first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battlefields and was soon introduced to psychology to account for perceptual detection of stimuli. ROC analysis since then has been used in medicine, radiology, biometrics, and other areas for many decades and is increasingly used in machine learning and data mining research.

In statistics, a ROC, or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true-positive rate against the false-positive rate at various threshold settings. The true-positive rate is also known as sensitivity in biomedical informatics, or recall in machine learning. The false-positive rate is also known as the fall-out and can be calculated as 1-specificity. The ROC plot shows sensitivity (true-positive fraction) on the horizontal axis against 1-specificity (false-positive fraction) on the vertical axis over all possible decision thresholds. A diagnostic test able to perfectly identify subjects with and without the condition produces a curve that passes through the upper left corner (0, 1) of the plot. A diagnostic test with no ability to discriminate better than chance produces a diagonal line from the origin (0, 0) to the top right corner (1) of the plot. Most tests lie somewhere in between these extremes. If a curve lies below the diagonal line (0, 0 to 1, 1), it can be inverted to produce a curve above the line, by swapping the decision criteria. ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making.

Description of back propagation neural network. This type of ANN belongs to a very large family of ANNs that normally uses a specific kind of law of learning named feed forward. In the feed forward ANNs, the signal proceeds from the input to the output of the ANN, crossing all of the nodes only once. The architecture of these networks is characterized by different layers of interconnected nodes (input, hidden, and output nodes), which processes the input signal according to a nonlinear function (generally, of sigmoid type). The fundamental equation that characterizes the activation of a single node and, therefore, the signal transfer from one layer to another is:

$$x_j^{[s]} = f\left(\sum_{i=0}^n w_{ji}^{[s]} \cdot x_i^{[s-1]}\right)$$

Learning, *i.e.*, the modeling of the input–output relation represented by data, occurs through minimization of the error in output and

retropropagation of this to the internal nodes, one hidden units, using the algorithm of the descending gradient in the majority of cases. In particular, each weight is corrected by the formula:

$$\Delta w_{ji}^{[s]} = -lcoef \cdot \frac{\partial E}{\partial w_{ji}^{[s]}} = lcoef \cdot e_j^{[s]} \cdot x_i^{[s-1]}$$

where for the retropropagated error $e_j^{[s]}$, we have:

$$e_j^{[out]} = f'(I_j^{[out]}) \cdot (t_j - x_j^{[out]})$$

for the last layer and:

$$e_j^{[s]} = f'(I_j^{[s]}) \cdot \sum_k (e_k^{[s+1]} \cdot w_{kj}^{[s+1]})$$

for all the other layers.

In theory, a Back Propagation having a sufficient number of hidden units is able to reconstruct any $y = f(x)$ function.

The architecture of ANN back-propagation feed forward employed consisted of 16 nodes in an input layer according to the number of selected variables (see Table 1 and the Results section), one hidden layer with 12 nodes, and one output layer with 2 nodes corresponding to the targets (autism; control). The architecture of this neural network is shown in Figure 2. Table 5 describes its mathematical and technical characteristics.

Ethics

Every mother was informed about the aim and methods of the study and gave a written consent to data treatment. Ethics Committee of Como Provincial Health authority approved the study. All investigations have been conducted according to the principles expressed in the Declaration of Helsinki and its later revisions.

STATEMENT OF FINANCIAL SUPPORT

No financial assistance was received to support this study.

Disclosure: The authors declare that they have no financial ties to any products discussed in this review and have no conflicts of interest.

REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th edn. Washington, DC: American Psychiatric Publications, 2013.
2. Stefanatos GA. Regression in autistic spectrum disorders. *Neuropsychol Rev* 2008;18:305–19.
3. Ji L, Chauhan V, Flory MJ, Chauhan A. Brain region-specific decrease in the activity and expression of protein kinase A in the frontal cortex of regressive autism. *PLoS One* 2011;6:e23751.
4. Pagnamenta AT, Khan H, Walker S, et al. Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. *J Med Genet* 2011;48:48–54.
5. Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M; American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry* 2014;53:237–57.
6. Hertz-Picciotto I, Croen LA, Hansen R, Jones CR, van de Water J, Pesah IN. The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environ Health Perspect* 2006;114:1119–25.
7. Lai MC, Lombardo MV, Baron-Cohen S. Autism. *Lancet* 2014;383:896–910.
8. Angelidou A, Asadi S, Alysandratos KD, Karagkouni A, Kourembanas S, Theoharides TC. Perinatal stress, brain inflammation and risk of autism-review and proposal. *BMC Pediatr* 2012;12:89.
9. Langridge AT, Glasson EJ, Nassar N, et al. Maternal conditions and perinatal characteristics associated with autism spectrum disorder and intellectual disability. *PLoS One* 2013;8:e50963.
10. Leavey A, Zwaigenbaum L, Heavner K, Burstyn I. Gestational age at birth and risk of autism spectrum disorders in Alberta, Canada. *J Pediatr* 2013;162:361–8.

11. Zerbo O, Iosif AM, Walker C, Ozonoff S, Hansen RL, Hertz-Picciotto I. Is maternal influenza or fever during pregnancy associated with autism or developmental delays? Results from the CHARGE (CHildhood Autism Risks from Genetics and Environment) study. *J Autism Dev Disord* 2013;43:25–33.
12. Sandin S, Hultman CM, Klevzon A, Gross R, MacCabe JH, Reichenberg A. Advancing maternal age is associated with increasing risk for autism: a review and meta-analysis. *J Am Acad Child Adolesc Psychiatry* 2012;51:477–486.e1.
13. Mamidala MP, Polinedi A, P T V PK, et al. Prenatal, perinatal and neonatal risk factors of autism spectrum disorder: a comprehensive epidemiological assessment from India. *Res Dev Disabil* 2013;34:3004–13.
14. Gardener H, Spiegelman D, Buka SL. Perinatal and neonatal risk factors for autism: a comprehensive meta-analysis. *Pediatrics* 2011;128:344–55.
15. Ronald A, Pennell CE, Whitehouse AJ. Prenatal maternal stress associated with ADHD and autistic traits in early childhood. *Front Psychol* 2010;1:223.
16. Daniels JL, Forssen U, Hultman CM, et al. Parental psychiatric disorders associated with autism spectrum disorders in the offspring. *Pediatrics* 2008;121:e1357–62.
17. Morgan VA, Croft ML, Valuri GM, et al. Intellectual disability and other neuropsychiatric outcomes in high-risk children of mothers with schizophrenia, bipolar disorder and unipolar major depression. *Br J Psychiatry* 2012;200:282–9.
18. Eriksson MA, Westerlund J, Anderlid BM, Gillberg C, Fernell E. First-degree relatives of young children with autism spectrum disorders: some gender aspects. *Res Dev Disabil* 2012;33:1642–8.
19. Kinney DK, Miller AM, Crowley DJ, Huang E, Gerber E. Autism prevalence following prenatal exposure to hurricanes and tropical storms in Louisiana. *J Autism Dev Disord* 2008;38:481–8.
20. Schmidt RJ, Tancredi DJ, Ozonoff S, et al. Maternal periconceptional folic acid intake and risk of autism spectrum disorders and developmental delay in the CHARGE (CHildhood Autism Risks from Genetics and Environment) case-control study. *Am J Clin Nutr* 2012;96:80–9.
21. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-related air pollution, particulate matter, and autism. *JAMA Psychiatry* 2013;70:71–7.
22. Testa C, Nuti F, Hayek J, et al. Di-(2-ethylhexyl) phthalate and autism spectrum disorders. *ASN Neuro* 2012;4:223–9.
23. Tabaton M, Odetti P, Cammarata S, et al. Artificial neural networks identify the predictive values of risk factors on the conversion of amnesic mild cognitive impairment. *J Alzheimers Dis* 2010;19:1035–40.
24. Street ME, Grossi E, Volta C, Faleschini E, Bernasconi S. Placental determinants of fetal growth: identification of key factors in the insulin-like growth factor and cytokine systems using artificial neural networks. *BMC Pediatr* 2008;8:24.
25. McCanlies EC, Fekedulegn D, Mnatsakanova A, et al. Parental occupational exposures and autism spectrum disorder. *J Autism Dev Disord* 2012;42:2323–34.
26. Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. *Physiol Rev* 2010;90:859–904.
27. Hsiao EY, McBride SW, Hsien S, et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* 2013;155:1451–63.
28. Weinstock M. The long-term behavioural consequences of prenatal stress. *Neurosci Biobehav Rev* 2008;32:1073–86.
29. Murmu MS, Salomon S, Biala Y, Weinstock M, Braun K, Bock J. Changes of spine density and dendritic complexity in the prefrontal cortex in offspring of mothers exposed to stress during pregnancy. *Eur J Neurosci* 2006;24:1477–87.
30. Ryzhavskaia BY, Sokolova TV, Uchakina RV, et al. Effect of emotional stress experienced by female rats before pregnancy on brain development in their offspring. *Bull Exp Biol Med* 2002;134:126–9.
31. Paris JJ, Brunton PJ, Russell JA, Frye CA. Immune stress in late pregnant rats decreases length of gestation and fecundity, and alters later cognitive and affective behaviour of surviving pre-adolescent offspring. *Stress* 2011;14:652–64.
32. Stoner R, Chow ML, Boyle MP, et al. Patches of disorganization in the neocortex of children with autism. *N Engl J Med* 2014;370:1209–19.
33. Sweet JA, Pickett RM. Evaluation of Diagnostic Systems: Methods From Signal Detection Theory. New York: Academic Press New York, 1982.
34. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. *Bioinformatics* 2004;20:2479–81.
35. 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 Intelligent Learning Syst Appl* 2013; 5: 29–38.
36. 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;7:173–87.
37. Rotondano G, Cipolletta L, Grossi E, et al.; Italian Registry on Upper Gastrointestinal Bleeding (Progetto Nazionale Emorragie Digestive). Artificial neural networks accurately predict mortality in patients with nonvariceal upper GI bleeding. *Gastrointest Endosc* 2011;73:218–26, 226.e1–2.