Do artificial neural networks love sex? How the combination of artificial neural networks with evolutionary algorithms may help to identify gender influence in rheumatic diseases

E. Grossi

Villa Santa Maria Foundation, Tavernerio, Italy.

Please address correspondence to:
Enzo Grossi, MD,
Scientific Director,
Villa Santa Maria Foundation,
Via IV Novembre,
22038 Tavernerio (CO), Italy.
E-mail: enzo.grossi@bracco.com

Received on December 14, 2022; accepted in revised form on January 2, 2023.

Clin Exp Rheumatol 2023; 41: 1-5.

Key words: artificial neural networks, gender influence, evolutionary algorithms, rheumatic diseases, psoriatic arthritis

Competing interests: none declared.

ABSTRACT

Although medical research has been performed predominantly on men both in preclinical and clinical studies, continuous efforts have been made to overcome this gender bias.

Examining retrospectively 21 data sets containing sex as one of the descriptive variables, it was possible to verify how many times our AI protocol decided to keep gender information in the predictive model. The data sets pertained a vast array of diseases such as dyspeptic syndrome, atrophic gastritis, venous thrombosis, gastroesophageal reflux disease, irritable bowel syndrome, Alzheimer diseases and mild cognitive impairment, myocardial infarction, gastrointestinal bleeding, gastric cancer, hypercortisolism, AIDS, COVID diagnosis, extracorporeal membrane oxygenation in intensive therapy, among others. The sample size of these data sets ranged between 80 and 3147 (average 600). The number of variables ranged from 19 to 101 (average 41).

Gender resulted to be part of the heuristic predictive model 19 out of 21 times. This means that also for highly adaptive and potent tools like Artificial Neural Networks, information on sex carries a specific value.

In the field of rheumatology, there is a specific example in psoriatic arthritis that shows that the presence of gender information allows a significantly better accuracy of ANNs in predicting diagnosis from clinical data (from 87.7% to 94.47%).

The results of this study confirm the importance of gender information in building high performance predictive model in the field of Artificial Intelligence (AI). Therefore, also for AI, sex counts.

Background

Evidence of gender differences has been described in many chronic diseases such as diabetes, cardiovascular disorders, neurological diseases, mental disorders, cancer, and autoimmunity, as well as in complex physiological processes such as physical and cognitive ageing. In addition, gender differences in lifestyles, such as nutrition, physical activity, tobacco and alcohol use, are correlated with diseases epidemiological trends (1).

The increase in the use of artificial intelligence (AI) in healthcare has led to the increasing use of large clinical datasets for machine learning (ML). Despite significant scientific advances to date, most biomedical AI technologies currently in use take little account of the detection of sex-related bias (2). On the other hand, classification algorithms are increasingly used in healthcare to assist physicians in decisions regarding diagnosis, prognosis, and choice of the most appropriate therapy. If these algorithms are not carefully evaluated for the presence of bias, the expected improvement in care will be limited to a subset of patients and consequently inequalities in health care will increase (3). Attention to the selection of gender characteristics in medical datasets subjected to modelling with ML is important to mitigate biases related to the underrepresentation of women in medical science.

Artificial neural networks

Artificial Neural Networks (ANNs), the progenitors of all machine learning systems, are information processing paradigms inspired by the analytical processes of the human brain. Such systems can modify their internal op-

erating structure and, therefore, the resulting analysis in relation to a defined goal, question, or function objective. They learn to recognise the complex patterns existing between the input signals and the corresponding outputs. ANNs are particularly suited for solving problems of the non-linear type and to analyse complex datasets (4).

The mathematical structure of ANNs makes them able to simultaneously handle a very high number of variables collected in clinical field notwithstanding the underlying non-linearity. This represents a tremendous advantage in comparison with classical statistics models. They receive input data together with defined targets of interest (diagnosis, outcome, etc.) from an external "teacher", learning how to link the input pattern to the desired output. The quality and quantity of the examples is fundamental.

Being universal approximators, ANNs when applied to multiple variables can very easily approximate the global information available during the training phase. This is the strength, but unfortunately, also the weakness of ANNs. In fact, almost inevitably, several variables which do not contain specific information pertinent to the problem are processed in the training phase. These variables, inserted in the model, act as a sort of 'white noise' which can interfere with the network's generalisation ability in the testing phase reducing the external validity and, consequently, the overall accuracy.

The performance of a neural network can be improved by reducing the number of inputs, even sometimes at the cost of losing some input information.

The selection can be based on different criteria, but in nonlinear systems is difficult to establish a priori which of them are most useful. For example, a pair of variables can be of no value taken individually, but extremely useful in conjunction.

As per the Law of Parsimony of '*Occam's Razor'*, the best explanation to a problem is that which involves the fewest possible assumptions. Thus, feature selection becomes an indispensable part of building machine learning models. Feature selection methods are intended to reduce the number of input variables to those that are believed to be most useful to a model to predict the target variable with high accuracy.

Table I. The effect of variables reduction in improving the predictive performance of ANNs.

Papers	no. of variables	Overall accuracy (%)	no. of variables	Overall accuracy (%)
Andriulli	45	69.0	31	88.6
Pagano	24	74.4	9	89.4
Sato	199	75.0	60	83.5
Pace	101	77.0	44	100
Lanher	37	96.6	30	98.4

Table II. The twenty-one data sets containing gender analysed using the TWIST algorithm.

Disease	ref.	Study population	no. of variables	no. of variables selected	Gender selected	Gender not selected
Dyspeptic syndrome	8	860	45	20	X	
Atrophic gastritis	9	350	37	12	X	
Venous thrombosis genetics	10	450	64	14	X	
Gastroesophageal reflux disease	11	159	101	45		X
Irritable bowel syndrome	12	340	36	23	X	
Alzheimer	13	211	60	24	X	
Cardiovascular risk	14	949	31	9	X	
Gastrointestinal bleeding	15	807	50	22	X	
Atrophic gastritis and thyroid	16	253	29	14	X	
Gastric cancer	17	3147	19	11	X	
Mild cognitive impairment	18	80	22	14	X	
Hypercortisolism	19	518	20	12	X	
Energy expenditure children	20	561	24	9	X	
Determinants of well-being	21	1500	57	23	X	
COVID	22	199	74	34	X	
Anticoagulant prediction	23	377	48	22	X	
Obesity and smell impairment	*	59	22	10	X	
Autism genetics	24	58	32	14		X
HIV renal impairment	25	1212	50	22	X	
ECMO intensive therapy	*	244	32	13	X	
Psoriatic athritis	26	282	19	8	X	
Mean/Total		**600.76**	**41.52**	**18.35**	**19**	**2**

*data on file.

In the last twenty years, our group has applied artificial neural networks to address a variety of medical problems in different field of medicine. Since from the beginning in was clear that features selection before feeding ANNs with data for the training phase is a fundamental step to improve the predictive performance capability of these systems, we have applied systematically automatic procedures able to reduce redundant information from a data set by eliminating variables not providing an added value for the solution of the clinical problem under study. This task was accomplished using evolutionary algorithms in combination with ANN.

Evolutionary algorithms

At variance with neural networks which are adaptive systems able to discover the optimal hidden rules explaining a certain data set, evolutionary algorithms (EA) are adaptive systems able to find optimal data when fixed rules or constraints have to be respected. They are, in other words, optimisation tools which become fundamental when the space of possible states in a dynamic system tends toward infinitum. This is just the case of variables selection. Given a certain large number of dichotomous variables (for example 100) the problem to define the most appropriate subgroup of them to better solve the problem under examination, has a very large space of possible states and exactly: 2^{100-1}. The computational time required to sort all possible variables subsets in order to submit them to ANNs processing would be in the order of million years; a so-called NP (non-polynomial) hard mathematical problem.

The introduction of variable selection systems generally results in a dramatic improvement of ANNs performance.

Table III. Predictive performance of machine learning in psoriatic arthritis diagnosis with or without information regarding sex.

	Machine learning	Recs	Arthritis yes	Arthritis no	Sensitivity (%)	Specificity (%)	Overall accuracy (%)
Without sex	Back propagation ANN a-b sequence	144	29	115	89.66	89.57	89.61
	Back propagation ANN b-a sequence	138	26	112	80.77	91.07	85.92
	Mean/sum	282	55	227	**85.21**	**90.32**	**87.77**
With sex	Back propagation ANN a-b sequence	130	25	105	100	88.57	94.29
	Back propagation ANN b-a sequence	152	30	122	96.67	92.62	94.64
	Mean/sum	282	55	227	**98.33**	**90.6**	**94.47**

The input selection (IS) is an example of the adaptation of an evolutionary algorithm to this problem.

This is a selection mechanism of the variables of a fixed dataset, based on the evolutionary algorithm GenD. The IS system becomes operative on a population of ANNs, each of them with a capability of extract a different pool of independent variables. Through the GenD evolutionary algorithm, the different "hypotheses" of variable selection, generated by each ANNs, change over time, generation after generation. When the evolutionary algorithm no longer improves, the process stops, and the best hypothesis of the input variables is selected and employed on the testing subset. The goodness-of-fit rule of GenD promotes, at each generation, the best testing performance of the ANN model with the minimal number of inputs.

An input selection system (IS) based on the evolutionary algorithm GenD proved to be able to handle the relevance of the different variables of the dataset in an intelligent way (5) and therefore became a standard of our work.

In a previous paper we showed how the variables selection can improve the predictive performance of ANNs. Table I shows the results obtained (6).

After IS another sophisticated AI system called TWIST (Training With Input Selection and Testing) was developed by Semeion Research Institute in Rome to further improve this difficult task. TWIST finds dynamically the subset of variables which allows a neural network to build a mathematical model with the best predictive performance in terms of global accuracy allowing simultaneously the optimal repartition of records in training and testing subsets (7).

Many papers in the literature have been published using the TWIST algorithm in different fields of medicine (Table II).

Review of studies employing the TWIST algorithm

To answer the question "Is sex information useful for a neural network to build a predictive model?", we checked retrospectively how many times sex has been saved as part of informative variables subset in 21 data sets used from our group for scientific studies in which ANNs were employed to answer critical questions pertinent to a large array of diseases and conditions after preprocessing with TWIST algorithm.

Table II lists the twenty-one data sets containing sex which underwent predictive analyses using the same protocol with TWIST preprocessing.

The data sets pertained a vast array of diseases like dyspeptic syndrome, atrophic gastritis, venous thrombosis, gastroesophageal reflux disease (GERD), irritable bowel syndrome, Alzheimer diseases and mild cognitive impairment, myocardial infarction, gastrointestinal bleeding, gastric cancer, hypercortisolism, AIDS, Covid diagnosis, extracorporeal membrane oxygenation (ECMO) in intensive therapy among others. The sample size of these data set ranged between 80 and 3147 (average 600). The number of variables in the original data sets ranged from 19 to 101 (average 41). In average the reduction of variables after TWIST application was 55.8%.

Sex resulted to be part of the heuristic predictive model 19 out of 21 times.

The example of psoriatic arthritis

In the field of rheumatology, sex-specific differences have consistently been reported for various diseases, including rheumatoid arthritis (27, 28), axial spondyloarthritis (29), or ankylosing spondylitis (30). Despite this, bias based on sex persists in clinical rheumatology. Sex differences may have important implications for clinical research in Psoriatic arthritis and in terms of epidemiology (incidence, prevalence, lifetime risk, survival, and mortality), clinical, radiological, and laboratory features, and response to treatment (31-33).

To address this issue within the context of artificial intelligence we use a national data set made available for the analysis. Data for analysis were derived from a multicentre study was conducted in the dermatological outpatient clinics of Italian universities (26). The study was approved by the Ethical Committee of each participating centre.

The inclusion criteria were age ≥18 years, diagnosis of plaque psoriasis according to the Italian Guidelines, ability to read and understand the Italian language, signed written informed consent.

The dermatologists collected socio-demographic and clinical information, scored the clinical severity at baseline according to the 7-point Physician Global Assessment (PGA), ranging from 0 (no skin involvement) to 6 (very severe involvement), and calculated the Psoriasis Area and Severity Index (PASI). They also recorded whether a diagnosis of psoriatic arthritis had already been made by a rheumatologist.

There were 17 input variables available for machine learning: sex (male/female); age; weight; BMI; psoriasis familiarity, familiarity for psoriasis arthritis; nail involvement; osteoarthritis; use of NSAIDs; PGA, PASI (six variables). There were 282 patients' records;

55 of them were affected by psoriatic arthritis and 227 were not affected.

Machine learning systems were trained to predict diagnosis from input variables.

Two experiments have been carried out: one with sex information available, after automatic selection of variables with TWIST system; one without sex information available. The same validation protocol was adopted with training-testing procedure with cross-over. The experiments with sex and without sex were conducted in a blind and independent manner in two directions: training with sub-sample A and blind testing with sub-sample B *versus* training with sub-sample B and blind testing with sub-sample A. The best results obtained using classical Back Propagation artificial neural network as classifier are reported in Table III.

Information on gender allowed the machine learning system to reach an overall accuracy of 94.47% while the absence of this information was associated with a lower level of overall accuracy (87.77%). The improvement in overall accuracy resulted to be statistically significant ($p<0.05$).

Comment

Neural networks can input multiple factor values simultaneously, combining and recombining them in different ways according to specific equations which are generally non-linear. In comparison with classical statistics, neural networks allow for the building up of a high number of independent models which, have different predictive capacity in classifying patients according to certain targets, due to slight differences in their architecture, topology and learning laws. Overall, neural networks belonging to specific settings do not provide a unique solution, because their performance is determined by several factors, such as the initial randomised incidence of interconnections between nodes, the order of presentation of cases during the training cycle and the number of training cycles. Other variables pertaining to the mathematical attributes of a specific neural network will also affect the final state of a trained neural network, allowing for a very high number of different possible combinations. Evolutionary algorithms by the way have been proposed to find the most suitable design of neural networks, to allow a better prediction, given the high number of possible combinations of parameters. In this paper we have shown that combining ANNs with EA the variables models developed for a vast range of chronic diseases contain sex almost always.

The historical absence of women from the health professions and clinical research has led to medical knowledge that focuses on the male body and neglects female physiological differences. To ensure that gender-based inequalities do not manifest themselves in AI applied to medicine, great care is needed to incorporate gender information into predictive models in order to avoid disparities in predictive performance in the two genders. Is therefore crucial that this information is available in all data sets undergoing predictive modelling to better assist doctors and health professionals with decision-making support.

In summary the results obtained mean that also for highly adaptive and potent tools like ANNs, the information of gender carries a specific value, being almost always selected, and plays an important role in the predictive model. The example of psoriatic arthritis seems emblematic in this context. In conclusion we can say that also for AI sex counts.

Acknowledgements

The author thanks Dr Damiano Abeni for providing the data set used for the analysis with machine learning systems

References

1. MAUVAIS-JARVIS F, BAIREY MERZ N, BARNES PJ *et al*.: Sex and gender: modifiers of health, disease, and medicine. *Lancet* 2020; 396(10250): 565-82. https://doi.org/10.1016/s0140-6736(20)31561-0
2. CIRILLO D, CATUARA-SOLARZ S, MOREY C *et al*.: Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digit Med* 2020; 3: 81. https://doi.org/10.1038/s41746-020-0288-5
3. KRIEGER N, FEE E: Man-made medicine and women's health: the biopolitics of sex/gender and race/ethnicity. *Int J Health Serv* 1994; 24(2): 265-83. https://doi.org/10.2190/lwlh-nmcj-uacl-u80y
4. RUMELHART DE, HINTON GE, WILLIAMS RJ: Learning internal representations by error propagation. *In*: RUMELHART DE, MCCLELLAND JL (Eds.): Parallel distributed processing, Foundations, Explorations in the Microstructure of Cognition. Cambridge: The MIT Press, 1986, pp 318-362.
5. BUSCEMA M: Genetic doping algorithm (GenD): theory and applications. *Expert Systems* 2004; 21(2): 63-79. https://doi.org/10.1111/j.1468-0394.2004.00264.x
6. GROSSI E, MANCINI A, BUSCEMA M: International experience on the use of artificial neural networks in gastroenterology. *Dig Liver Dis* 2007; 39(3): 278-85. https://doi.org/10.1016/j.dld.2006.10.003
7. 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 Learn Syst Appl* 2013; 5: 29-38.
8. ANDRIULLI A, GROSSI E, BUSCEMA M, PILOTTO A, FESTA V, PERRI F: Artificial neural networks can classify uninvestigated patients with dyspepsia. *Eur J Gastroenterol Hepatol* 2007; 19(12): 1055-8. https://doi.org/10.1097/meg.0b013e3282f198b2
9. LAHNER E, INTRALIGI M, BUSCEMA M *et al*.: Artificial neural networks in the recognition of the presence of thyroid disease in patients with atrophic body gastritis. *World J Gastroenterol* 2008; 14(4): 563-8. https://doi.org/10.3748/wjg.14.563
10. PENCO S, GROSSI E, CHENG S *et al*.: Assessment of the role of genetic polymorphism in venous thrombosis through artificial neural networks. *Ann Hum Genet* 2005; 69(Pt 6): 693-706. https://doi.org/10.1111/j.1529-8817.2005.00206.x
11. PACE F, BUSCEMA M, DOMINICI P *et al*.: Artificial neural networks are able to recognize gastro-oesophageal reflux disease patients solely on the basis of clinical data. *Eur J Gastroenterol Hepatol* 2005; 17(6): 605-10. https://doi.org/10.1097/00042737-200506000-00003
12. GROSSI E, ASTEGIANO M, DEMARCHI B *et al*.: Dolore addominale e alterazione dell'alvo: è possibile semplificare l'iter diagnostico attraverso l'uso di Sistemi di Intelligenza Artificiale? *Sistemi Artificiali Adattivi in Biomedicina* 2005; 1: 66-102.
13. COPPEDÈ F, GROSSI E, BUSCEMA M, MIGLIORE L: Application of artificial neural networks to investigate one-carbon metabolism in Alzheimer's disease and healthy matched individuals. *PLoS One* 2013; 8(8): e74012. https://doi.org/10.1371/journal.pone.0074012
14. BALDASSARRE D, GROSSI E, BUSCEMA M *et al*.: Recognition of patients with cardiovascular disease by artificial neural networks. *Ann Med* 2004; 36(8): 630-40. https://doi.org/10.1080/07853890410018880
15. 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(2): 218-26, 226.e1-2. https://doi.org/10.1016/j.gie.2010.10.006

16. LAHNER E, GROSSI E, INTRALIGI M *et al*.: Possible contribution of artificial neural networks and linear discriminant analysis in recognition of patients with suspected atrophic body gastritis. *World J Gastroenterol* 2005; 11(37): 5867-73. https://doi.org/10.3748/wjg.v11.i37.5867
17. BURI L, ZULLO A, HASSAN C *et al*.; SIED APPROPRIATENESS WORKING GROUP: Upper GI endoscopy in elderly patients: predictive factors of relevant endoscopic findings. *Intern Emerg Med* 2013; 8(2): 141-6. ttps://doi.org/10.1007/s11739-011-0598-3
18. TABATON M, ODETTI P, CAMMARATA S *et al*.: Artificial neural networks identify the predictive values of risk factors on the conversion of amnestic mild cognitive impairment. *J Alzheimers Dis* 2010; 19(3): 1035-40. https://doi.org/10.3233/jad-2010-1300
19. MORELLI V, PALMIERI S, LANIA A *et al*.: Cardiovascular events in patients with mild autonomous cortisol secretion: analysis with artificial neural networks. *Eur J Endocrinol* 2017; 177(1): 73-83. https://doi.org/10.1530/eje-17-0047
20. SPOLIDORO GCI, D'ORIA V, DE COSMI V *et al*.: Artificial neural network algorithms to predict resting energy expenditure in critically ill children. *Nutrients* 2021; 13(11): 3797. https://doi.org/10.3390/nu13113797
21. GROSSI E, TAVANO BLESSI G, SACCO PL, BUSCEMA M: The interaction between culture, health and psychological well-being: data mining from the Italian Culture and Well-Being Project. *J Happiness Stud* 2011; 13: 129-48. https://doi.org/10.1007/s10902-011-9254-x
22. LANGER T, FAVARATO M, GIUDICI R *et al*.: Development of machine learning models to predict RT-PCR results for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in patients with influenza-like symptoms using only basic clinical data. *Scand J Trauma Resusc Emerg Med* 2020; 28(1): 113. https://doi.org/10.1186/s13049-020-00808-8
23. GROSSI E, PODDA GM, PUGLIANO M *et al*.: Prediction of optimal warfarin maintenance dose using advanced artificial neural networks. *Pharmacogenomics* 2014; 15(1): 29-37. https://doi.org/10.2217/pgs.13.212
24. STOCCORO A, GALLO R, CALDERONI S *et al*.: Artificial neural networks reveal sex differences in gene methylation, and connections between maternal risk factors and symptom severity in autism spectrum disorder. *Epigenomics* 2022; 14(19): 1181-95. https://doi.org/10.2217/epi-2022-0179
25. GROSSI E, GALLI L, POLI A *et al*.: Advanced machine learning systems in HIV-1 infected patients: a promising tool to predict renal impairment 11° ICAR Congress; 5-7 June 2019
26. PIASERICO S, GISONDI P, AMERIO P *et al*.: Validation and field performance of the Italian version of the Psoriatic Arthritis Screening and Evaluation (PASE) Questionnaire. *Acta Derm Venereol* 2016; 96(217): 96-101. https://doi.org/10.2340/00015555-2429
27. VAN VOLLENHOVEN RF: Sex differences in rheumatoid arthritis: more than meets the eye. *BMC Med* 2009; 7: 12. https://doi.org/10.1186/1741-7015-7-12
28. MAYNARD C, MIKULS TR, CANNON GW *et al*.: Sex differences in the achievement of remission and low disease activity in rheumatoid arthritis. *Arthritis Care Res* (Hoboken) 2020; 72(3): 326-33. https://doi.org/10.1002/acr.23873
29. RUSMAN T, VAN BENTUM RE, VAN DER HORST-BRUINSMA IE: Sex and gender differences in axial spondyloarthritis: myths and truths. *Rheumatology* (Oxford) 2020; 59(Suppl. 4): iv38-46. https://doi.org/10.1093/rheumatology/keaa543
30. CHEN HH, CHEN YM, LAI KL *et al*.: Gender difference in ASAS HI among patients with ankylosing spondylitis. *PloS One* 2020; 15(7): e0235678. https://doi.org/10.1371/journal.pone.0235678
31. GLADMAN DD, BRUBACHER B, BUSKILA D, LANGEVITZ P, FAREWELL VT: Psoriatic spondyloarthropathy in men and women: a clinical, radiographic, and HLA study. *Clin Invest Med* 1992; 15(4): 371-5.
32. EDER L, THAVANESWARAN A, CHANDRAN V, GLADMAN DD: Gender difference in disease expression, radiographic damage and disability among patients with psoriatic arthritis. *Ann Rheum Dis* 2013; 72(4): 578-82. https://doi.org/10.1136/annrheumdis-2012-201357
33. QUEIRO R, SARASQUETA C, TORRE JC, TINTURÉ T, LÓPEZ-LAGUNAS I: Comparative analysis of psoriatic spondyloarthropathy between men and women. *Rheumatol Int* 2001; 21(2): 66-8. https://doi.org/10.1007/s002960100135