Psychiatria Danubina, 2019; Vol. 31, Suppl. 3, pp 261-264 *Conference paper*

IDENTIFICATION AND EVALUATION OF COGNITIVE DEFICITS IN SCHIZOPHRENIA USING "MACHINE LEARNING"

Antonella Vacca[1], Roberto Longo[1] & Corrado Mencar[2]

[1]'Città Solidale' Società Cooperativa Sociale, Latiano, Italy
[2]University of Bari 'Aldo Moro', Bari, Italy

SUMMARY

Background: *Schizophrenia can be interpreted as a pathology involving the neocortex whose cognitive dysfunctions represent a central and persistent characteristic of the disease, as well as one of the more important symptoms in relation to the impairment of psychosocial functioning and the resulting disabilities. Given the implication of cognitive functions in everyday life, they can better predict the degree of schizophrenia. The study proposes to use Machine Learning techniques to identify the specific cognitive deficits of schizophrenia that mostly characterize the disorder, as well as to develop a predictive system that can diagnose the presence of schizophrenia based on neurocognitive tests.*

Subjects and methods: *The study employs a dataset of neurocognitive assessments carried out on 201 people (86 schizophrenic patients and 115 healthy patients) recruited by the Neuroscience Group of the University of Bari "A. Moro". A data analysis process has been carried out, with the aim of selecting the most relevant features as well as to prepare data for training a number of "off-the-shelf" machine learning methods (Decision Tree, Random Forest, Logistic Regression, k-Nearest Neighbor, Neural Network, Support Vector Machine), which have been evaluated in terms of classification accuracy according to stratified 20-fold cross-validation.*

Results: *Among all variables, 14 were selected as the most influential for the classification problem. The variables with greater influence are related to working memory, executive functions, attention, verbal fluency, memory. The best algorithms turned out to be Support Vector Machine (SVM) and Neural Network, showing an accuracy of 87.8% and 84.8% on a test set.*

Conclusions: *Machine Learning provides "cheap" and non-invasive methods that potentially enable early intervention with specific rehabilitation interventions. The results suggest the need to integrate a thorough neuropsychological evaluation into the more general diagnostic evaluation of patients with schizophrenia disorder.*

Key words: *cognitive disorders - neuropsychological evaluation - schizophrenia - machine learning*

* * * * *

INTRODUCTION

Among psychiatric disorders, schizophrenia is one of the main causes of disability today, as delusions, hallucinations and negative symptoms lead to a cognitive impairment that represents the most significant deficit of this mental disorder (Marder & Fenton 2004). All these symptoms contribute to the deterioration of the personal and social functioning of the subject, causing the person to lose touch with reality.

Today the research interprets schizophrenia as a neocortex pathology whose main, purely cognitive, role is represented by information processing: cognitive dysfunctions therefore represent, in this perspective, a central and persistent characteristic of the disease, as well as one of the more important symptoms in relation to the impairment of psychosocial functioning and the resulting disabilities (Gold et al. 2002). These dysfunctions can be already detected in the premorbid period, are present at the onset of the disease and persist, even in the periods of clinical stability, after the resolution of the acute phases (Reichenberg & Harvey 2007). Due to these characteristics, cognitive disorders are currently among the most studied as possible endophenotypes in the psychiatric field (Bertolino et al. 2001, 2006, Weinberger 2001, Zhang et al. 2007). From published data it has been demonstrated that, on average, the cognitive performance is from 1 to 2 standard deviations below the healthy controls in several domains (Dickinson et al. 2007, Keefe et al. 2011).

Memory, learning, attention and processing speed, executive functions and abstraction ability have been highlighted as the most deficient areas (Green 2006, Heinrichs et al. 1998, Reichenberg et al. 2009). There is a wealth of evidence that shows how cognitive deficits are closely related to low levels of psychosocial functionality in the context of interpersonal relationships, in the field of work and in the dimension of the perception of Quality of Life (Marder & Fenton 2004, McGurk et al. 2004). Given the implication of cognitive functions in everyday life, they can better predict the degree of disability, especially in schizophrenia (Shamsi et al. 2011).

The aim of the study is to succeed in developing an accurate prediction model for the diagnosis of schizophrenia, through the identification of possible characteristic cognitive deficits already present in the initial phase or even before the onset of the disease.

SUBJECTS AND METHODS

Purpose of the study

Given the growing evidence in favor of neurocognitive impairment in schizophrenic disorders, the present study proposes to use machine learning techniques as a tool to:

- Identify the specific cognitive deficits of schizophrenia that characterize the disorder more markedly only through neurocognitive tests;
- Develop a predictive system that is based on the detection of the values of neurocognitive variables that can diagnose the presence of schizophrenia.

The study was conducted on two samples: 1) a sample of people diagnosed with schizophrenia who underwent a series of neuropsychological tests to measure cognitive functions; 2) a sample of healthy people evaluated with the same neuropsychological tests.

The hypothesis is that, through the comparison of the same variables in healthy subjects and those with schizophrenia, carried out through the application of machine learning techniques, it is possible to perform, on an empirical and predictive basis, a more accurate diagnosis of schizophrenia based on specific cognitive deficits.

The tool chosen for the "data understanding" phase and the subsequent machine learning phase is Orange (https://orange.biolab.si/).

Subjects

The dataset consists of neurocognitive assessments carried out on people recruited from the research conducted, from 2006 to 2015, by the Neuroscience Group directed by Prof. Alessandro Bertolino of the University of Bari "A. Moro". A total of 201 subjects were randomly selected from all the participating subjects: n. 86 schizophrenic patients (SCZ) diagnosed with SCIID and 115 healthy controls (NC), 83 women and 118 men. The sample is aged between 18 and 63 and is homogeneous for the age and sex variables.

Methods

A battery of tests, taken from the Protocol created in the framework of research by the Neuroscience Group, was used for the neuropsychological evaluation, with the aim of detecting the socio-cultural level, the laterality and the different aspects related to attention, memory, praxic, visuospatial and executive functions. In Table 1 the domains and the relative variables are displayed.

In the data understanding phase, the domains associated with the 28 variables measured by neuropsychological tests were described. The data were prepared, taking decisions on the missing data, the feature selection was made, also taking into consideration the analysis of the correlation of the variables, and the features ranking was made with ReliefF algorithm (Kononenko 1994). The results are shown in Figure 1, where the rankings of 14 main variables are displayed.

The dataset was divided (holdout) into a training set (80%) and a test set (20%). Machine learning techniques were applied to the training set and various predictive models (Tree, Random Forest, Logistic Regression, k-Nearest Neighbor, Neural Network, Support Vector Machine) were analyzed (Berthold et al. 2010). Since all the records of the dataset already had a classification, this was supervised learning. The sample was validated with 20-fold Cross Validation and many tests were made, repeatedly changing the parameter settings of each prediction algorithm. Finally, based on the simulations, the best performing algorithm was chosen by evaluating the Classification Accuracy (CA). The best algorithms turned out to be Support Vector Machine (SVM) and Neural Network (Figure 2).

Table 1. Measured Domains and Related Tests

Measured domain	Test
Socio-economic condition	Hollingshead
Lateralization (right-left)	Edinburgh Handedness Inventory
Pre morbid intelligence	Brief Intelligence Test (TIB)
Executive functions	Wisconsin Card Sorting Test (WCST)
Working memory	N-Back 0, N-Back 1, N-Back 2
Verbal and non-verbal memory	Wechsler Memory Scale
Verbal fluency	Phonological Fluence Test Semantic Fluence Test
Executive functions	Trail Making Test A e B
Attention	Continuous Performance Test (CPT)

	#	ReliefF
N-Back_OneBackRT(msec)_Generic		0.072
WCST_Number of Perseverative Errors_Generic		0.071
N-Back_TwoBack accuracy%_Generic		0.063
Trail making Test A_Generic		0.049
CPT-FS_HIT_Generic		0.046
Semantic Fluency_Generic		0.044
WMS_Generic		0.042
CPT-DP_S_Generic		0.040
Trail making Test B_Generic		0.028
CPT-FS_tHIT_Generic		0.027
N-Back_ZeroBackRT(msec)_Generic		0.025
Phonolog Fluency_Generic		0.018
CPT-DPC_S_Generic		0.004
N-Back_ZeroBack accuracy%_Generic		0.003

Figure 2. Feature ranking

Evaluation Results

Model	AUC	CA	F1	Precision	Recall
SVM	0.933	0.910	0.860	0.949	0.787
Neural Network	0.950	0.910	0.860	0.949	0.787
Logistic Regression	0.940	0.895	0.844	0.884	0.809
Random Forest	0.947	0.880	0.818	0.878	0.766
Tree	0.861	0.850	0.778	0.814	0.745
kNN	0.894	0.812	0.658	0.923	0.511

Figure 3. Best performing training algorithms

RESULTS

For each of the two algorithms we set up the settings that maximized performance and we applied them to the test set. Comparing the results of the predictions made by the algorithms with the diagnosis made by the Neuroscience Group, it was found that:

- SVM misses the prediction of 4 records out of 33;
- Neural Network fails to predict 5 records out of 33.

Compared to the first objective of the study, the data set revealed that some cognitive variables, which can be evaluated rapidly through neurocognitive tests, can indicate the onset of schizophrenia early. As shown in Figure 1, the variables with greater weight identify the following cognitive domains: working memory, executive functions (flexibility of thought and visual-motor coordination), attention, verbal fluency, memory. The analysis through statistical methods and a machine learning tool (Orange) has shown that the distribution of the values of these variables separate the NC and SCZ classes in a defined manner.

With respect to the second objective of the study, two machine learning algorithms effective in predicting membership to one class or another have been identified, Support Vector Machine with an accuracy of 87.8% and Neural Network with an accuracy of 84.8%.

DISCUSSION

The machine learning analysis of cognitive variables can be a valid support for operators for the early diagnosis of schizophrenia. It is a "cheap" and non-invasive method that shortens the time because it allows early intervention with specific rehabilitation interventions, such as cognitive rehabilitation (Vita 2013).

As future developments, it would be useful to validate the experiment with a larger dataset. It would also be useful to study the impact on the outcome of the study of the other variables that were excluded from the experiment, namely Age, Gender, Hollingshead, TIB QI and Edinburgh Scale. Future studies may investigate the existence of relationships between schizophrenia and other factors such as socio-economic status, level of education and the familiarity of the disorder. Having available neuroimaging data, it will be possible to relate cognitive aspects and biological and functional aspects.

CONCLUSION

This study demonstrates how Data Analysis and Machine Learning can be useful tools that have great potential in the study and rehabilitative treatment of mental disorders. The reproducibility and strength of these results suggest the need to integrate a thorough neuropsychological evaluation into the more general diagnostic evaluation of patients with this disorder.

Acknowledgements:

Data were granted by Prof. Alessandro Bertolino who heads the Neuroscience Group at the "A. Moro" University of Bari.

Conflict of interest: *None to declare.*

Contribution of individual authors:

All three authors conceived and designed the study.

Antonella Vacca made the clinical and psychological writing.

Roberto Longo made data analysis and applied machine learning algorithms.

Corrado Mencar supervised the final draft.

References

1. *Berthold MR, Borgelt C, Höppner F, Klawonn F: Guide to Intelligent Data Analysis, Springer, 2010*
2. *Bertolino A, Brudaglio F, Nardini M: Pathophysiology of schizophrenia and the neurodevelopmental hypothesis: evidence from an animal model. Psychiatric Networks 2001; 3:10-21*
3. *Bertolino A, Caforio G, Petruzzella V, Latorre V, Rubino V, Dimalta S, Torraco A, Blasi G, Quartesan R, Mattay VS, Callicott JH, Weinberger DR, Scarabino T: Prefrontal dysfunction in schizophrenia controlling for COMT val158met genotype and working memory performance. Psychiatry Research: Neuroimaging 2006; 147:221-6*
4. *Dickinson D, Ramsey MB & Gold JM: Overlooking the obvious: A meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. Archives of General Psychiatry 2007; 64:532–542*
5. *Gold J, De Girolamo G, Brambilla L, Cappa S, Mazzi F, O'Donnell K, Scala V, Pioli R: Schizofrenia e funzioni cognitive: una revision critica della letteratura, Rivista di psichiatria 2002;37*
6. *Green MF: Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. J Clin Psychiatry 2006; 67:e12*
7. *Heinrichs RW & Zakzanis KK: Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. Neuropsychology 1998; 12:426-45*
8. *Keefe RS, Fox KH, Harvey PD et al.: Characteristics of the MATRICS consensus cognitive battery in a 29 site antipsychotic schizophrenia clinical trial. Schizophr Res 2011; 125:161-168*
9. *Kelleher A, Kelleher A: Machine Learning in Production, Addison Wesley, 2019*
10. *Kononenko I: Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano F., De Raedt L., editors. Machine Learning: ECML-94. Springer; Berlin/Heidelberg, Germany, 1994; pp. 171–182*
11. *Marder SR, Fenton W: Measurement and treatment research to improve cognition in schizophrenia: NIMH MATRICS initiative to support the development of agents for improving cognition in schizophrenia. Schizophrenia research 2004; 72:5-9*
12. *McGurk SR, Twamley EW, Sitzer DJ, McHugo KT, Mueser: A meta-analysis of cognitive remediation in*

schizophrenia. American Journal of Psychiatry 2007; 164:1791-1802

13. *Reichenberg A, Harvey PD, Bowie CR, Mojtabai R, Rabinowitz J, Heaton RK, Bromet E: Neuropsychological function and disfunction in schizophrenia and psychotic affective disorders, Schizophrenia bulletin 2009; 35:1022-29*
14. *Reichenberg A, Harvey PD: Neuropsychological impairments in schizophrenia: Integration of performance-based and brain imaging findings. Psychol bull 2007; 133:833-58*
15. *Shamsi S, Lau A, Lencz T et al.: Cognitive and symptomatic predictors of functional disability in schizophrenia. Schizophr Res 2011; 126:257-264*
16. *Vita A: La riabilitazione cognitiva della schizofrenia, Springer-Verlag Italia, 2013*
17. *Weinberger DR, Egan MF, Bertolino A, Callicott JH, Mattay VS, Lipska BK, KF Berman KF, Goldberg TE: Prefrontal neurons and the genetics of schizophrenia. Biol Psychiatry 2001; 50:825-44*
18. *Zhang Y, Bertolino A, Fazio L, Blasi G, Rampino A, Romano R, Lee ML, Xiao T, Papp A, Wang D, Sadee W: Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. Proc Natl Acad Sci USA 2007; 1 04:20552-7*

Correspondence:

Antonella Vacca, PhD
Psychologist, Psychiatric Rehabilitation Community Consultant
Città Solidale Società Cooperativa Sociale
Via Anania Lamarina, 75, 72022 Latiano (BR), Italy
E-mail: antonellavacca@yahoo.com