frontiers in Psychiatry

OPINION
published: 24 September 2020
doi: 10.3389/fpsyt.2020.552262

At the Crossroads Between Psychiatry and Machine Learning: Insights Into Paradigms and Challenges for Clinical Applicability

Sarah Itani[1,2*] *and Mandy Rossignol*[3]

[1] Fund for Scientific Research (F.R.S.-FNRS), Brussels, Belgium, [2] Department of Mathematics and Operations Research, Faculty of Engineering, University of Mons, Mons, Belgium, [3] Department of Cognitive Psychology and Neuropsychology, Faculty of Psychology and Education, University of Mons, Mons, Belgium

Keywords: psychiatry, machine learning, diagnosis, open data, nosology and classification of mental disorders

OPEN ACCESS

Edited by:
Tetsuya Takahashi, University of Fukui, Japan

Reviewed by:
Cristina Scarpazza, University of Padua, Italy
Georgia Koppe, University of Heidelberg, Germany

****Correspondence:***
Sarah Itani
sarah.itani@umons.ac.be

Specialty section:
This article was submitted to Computational Psychiatry, a section of the journal Frontiers in Psychiatry

Received: *16 April 2020*
Accepted: *07 September 2020*
Published: *24 September 2020*

Citation:
Itani S and Rossignol M (2020) At the Crossroads Between Psychiatry and Machine Learning: Insights Into Paradigms and Challenges for Clinical Applicability. Front. Psychiatry 11:552262. doi: 10.3389/fpsyt.2020.552262

THE CURRENT PSYCHIATRIC DIAGNOSIS: A PRACTICE SUBJECT TO DEBATE

Good health and well-being feature among the development goals set by the United Nations members in their action plans to ensure peace and prosperity by 2030 (1). In this context, promoting mental health constitutes an important target. Additional efforts for an accurate, early and objective diagnosis of mental disorders can only contribute to move forward in this direction.

Classification systems—such as the Diagnostic and Statistical Manual of Mental Disorders (DSM), edited by the American Psychiatric Association (APA)—have been developed as a common language to conduct diagnosis in the most possible form of universality. However, these a-theoretical classifications lay down clinical descriptive criteria that can be open to subjective interpretation by clinicians. The APA tried to address this issue by fine-tuning the diagnostic criteria through the successive revisions of the DSM. However, each of these versions has always sparked lively debate in the community (2, 3). The explosion in diagnostic categories was notably heavily criticized. On the one hand, this multiplication was considered as a way of integrating the scientific progress of psychopathology research and offering more exhaustive descriptions (4). On the other hand, it has been argued that this explosion of diagnostic categories not only answers a commercial objective, but is also a way to satisfy the society's tendency to organize and annotate the mental phenomena (5).

A diagnosis based on neuromarkers may respond to the criticisms addressed to the psychiatric classification systems (3, 6). Here, the greatest challenge remains to identify the relevant and discriminative markers that would reach sufficient scientific consensus.

A SENSE OF PROGRESS THROUGH MACHINE LEARNING

Over the last decade, the application of Machine Learning (ML) to the psychiatric research has given the latter a new pulse (7). Indeed, in comparison to classical statistical approaches, ML provides outstanding capabilities for processing multivariate and multimodal data sets (8).

The ADHD-200 competition was probably the catalyst for inspiring such interdisciplinary research (9). This international contest was intended to accelerate the understanding of Attention Deficit Hyperactivity Disorder (ADHD), by inviting competitors to develop an imaging-based diagnostic classifier with the highest possible performance. The ADHD-200 collection is the first in a series of data sets released in the context of a large-scale project for open data sharing. This valuable culture was promoted by the 1000 Functional Connectomes Project (FCP), followed by the International Neuro-imaging Data sharing Initiative (INDI) (10). The will of opening research to the largest extent possible led to the sharing of software (11) and preprocessed data. The Autism Brain Imaging Data Exchange (ABIDE) is a notable example of the achievements of the INDI project. Related to Autism Spectrum Disorder (ASD), the ABIDE data set was released in two parts, including brain imaging data for over two thousand subjects aggregated across twenty-four worldwide imaging sites (12, 13). The analysis of such large data sets is expected to reduce inconsistency in research results, while the use of small sample sizes has demonstrated its limitations with variable levels of accuracy (14).

There is no doubt that the availability of large, free and well-structured databases has been a positive incentive for their analysis through ML, for the purpose of both knowledge extraction and diagnosis prediction. These capabilities were described in terms of their application to psychiatry (15–19). We will here discuss the main technical challenges of this ML-guided research.

TOWARD A SUCCESSFUL CONJUNCTION OF PSYCHIATRY AND MACHINE LEARNING

Designing Explainable Solutions

Over the years, the ML algorithms have been improved to be more and more performant. In particular, deep learning methods advantageously capture complex patterns in data, therefore allowing to reach higher levels of accuracy (20). But concretely, clinicians expect more than just high accuracy from predictive systems, which opacity constitutes a constant criticism (16, 19). The emerging domain of *explainable Artificial Intelligence* (xAI) is of particular interest in this respect (21). Indeed, explainability allows clinicians to choose to trust, or not, the recommendations (22). Moreover, it is well established that ML models tend to reproduce the biases present in the training data sets, often caused by the unbalanced representation of the classification categories. Explainable decision chains thus allow to control that the outputs are conform to ethical standards, and notably unsupported by any form of discrimination (23, 24).

Most of the attempts in developing an xAI were focused on the design of *post-hoc systems*, i.e. black boxes completed by a component explaining predictions a posteriori (25). Post-hoc systems are thus thought as an interesting way of combining high accuracy and explainability. However, they remain questionable on (i) the veracity of their explanations, which are generated around the considered data point, and (ii) the consequent inability to give a comprehensive picture of the model behavior (21).

Concurrently, it has been shown that models, ranging from white to black boxes, all perform comparatively when trained on quality and meaningful data (21, 26). This observation suggests a double perspective on the development of explainable ML systems.

- *Data preprocessing* conditions the performance of any decision system. This initial phase in the ML process can consist of applying a transformation to the original training features, in order to make them more discriminative. The transformation is sometimes unavoidably achieved through the (complex) combination of the initial training features, which introduces some interaction effects. Such a combination should thus be understood for the interpretation (even simplified) of the resulting features (27, 28).
- There also remain *theoretical challenges* to the improvement of the current predictive ML algorithms. A modern research avenue involves the design of optimal logical models (21) such as decision trees, that may be algorithmically strengthened to perform similarly as black boxes.

Efforts should thus be made both on data preprocessing and model design, in order to better address the need for explainability and transparency required by medical applications.

Reconciling Theory and Data-Driven Approaches

Two main methodologies exist for scientific modeling (29).

- *Theory-based models*; that are grounded on known scientific laws based on some parameters, and a low amount of data is generally sufficient to fit these parameters.
- *Data-based models*; that require large data sets for an automatic training procedure which is expected to yield general models, able to describe the related phenomenon.

While theory-based methods are usually considered for the understanding of disorders, data-driven methods are rather considered for the design of clinical tools (16). A hybrid approach guided by data and theory would broaden the field of investigation, reconciling the existing scientific knowledge with elements extracted from data. The concept, which is not new, was highlighted in (16), and then properly formalized in (29) as the *Theory-Guided Data Science* (TGDS) paradigm. This principle should be encouraged in psychiatric research. Indeed, TGDS may be put in practice through the interaction with domain experts (i.e. psychiatrists, neuroscientists, neurologists) bringing their medical knowledge for feature selection (16), or more globally to refine ML models in the frame of an expert-in-the-loop mechanism (30). The aforementioned explainability naturally fosters the implementation of a TGDS.

Considering One-Class Classification

Though they are mostly considered in the development of decision aid systems, Multi-Class Classification (MCC) algorithms are criticized for several reasons. Indeed, MCC does not address comorbidity appropriately since it considers the different diagnostic categories as mutually exclusive (16). In addition, MCC becomes more challenging in presence of

unbalanced and noisy data sets (31). The domain of One-Class Classification (OCC) (32) covers a range of algorithms capable of describing a given class [e.g., a neuropathology (33, 34)], in such a way to reject cases that do not comply with this description. It is thus possible to use ensembles of one-class models in order to test a patient for several conditions simultaneously. One-class classification also gets rid of the need for a balanced data set including training instances from each class, as required by the MCC scheme. Finally, through its very nature, OCC can efficiently rule out noise, and specifically class noise which is usually located on the class boundaries (31, 35).

Several OCC tools are already available for clinical use, despite being mainly targeted towards neurological disorders (36). Hence, OCC deserves greater attention to be further developed in psychiatric research. Additional efforts for algorithmic improvements would be particularly worth considering in the context of explainable AI.

Addressing the Question of Heterogeneity in Data

Though outstanding, the efforts for large-scale data gathering across several sites yield disparities in terms of demographics and experimental protocols (22). The homogenization of experimental protocols, and the design of appropriate validation procedures are respectively thought as ways for achieving and assessing generalizability (18, 19). The question of the extent to which this generalizability needs to be achieved deserves to be discussed, and probably requires a scientific consensus to provide a clear research direction. Indeed, the available financial and technical means for medical assessment and data may differ from a region to another.

Furthermore, psychiatric conditions can be characterized by clinical and/or neurobiological heterogeneity, the latter being also established in healthy controls (37). In this case, a thorough analysis may help to consider the best modeling strategy. For example, a ML framework can be implemented to perform diagnosis prediction in different levels, i.e., to detect a disorder first, and the disorder subtype then (38, 39). In (33), the authors focused on the description of ASD through OCC, since controls showed high neurobiological disparity. Moreover, ahead of the ML process, the experimental protocol should not necessarily be aligned with the DSM diagnostic categories. Indeed, these diagnostic labels are heterogeneous and derived from traditional assessments conducted by clinicians (40, 41). The Research Domain Criteria (RDoC) framework was introduced by the National Institute of Mental Health to alleviate this issue (42, 43). The RDoC orients the study of mental illnesses towards *domains of human functioning* described at different levels, rather than towards *symptoms*. The lowest level relates to *units of analysis*, suggesting relevant biological, genetic and physiological investigation markers (43).

Encouraging Scientific Reproducibility

The psychiatric domain has witnessed a significant increase in ML-based studies, along with a diversification of the modalities considered for data processing and modeling (44). It is therefore imperative to apply guidelines that ensure reproducibility; recommendations are provided in (44). The appropriate choice of procedures for model training and evaluation, as well as the availability of source code/data should notably be encouraged. Yet, a recent study highlighted that these aspects are often lacking: it appeared that 50% of studies do not share software, while 36% do not give access to data (45).

More specifically in the context of open data sharing, a standard segmentation of the data collections into training and test sets would reinforce reproducibility (40). On the occasion of the ADHD-200 competition, training and test sets were kept separately in order to allow respectively the development and the assessment of the predictive models developed by the competing teams. Since then, these data subsets have mostly been used in their initial form, which makes it easy to report the evolution of the progress achieved on the prediction of ADHD. The same cannot be said for other INDI data sets such as the ABIDE collection, where the segmentation of data is a choice made for each research study. This great disparity in the definition of the data subsets therefore makes it difficult to track the research progress on a given mental disorder.

CONCLUSION

Through the present perspective, we wished to draw attention on key principles for the design of Machine Learning (ML) solutions able to help clinicians to diagnose mental disorders. Our consideration addressed some main criticisms found in the literature about ML-based systems for clinical applications. It appears that a form of explainable and knowledge-guided data science will certainly help in the design of transparent mechanisms making sense to clinicians. The use of one-class classification algorithms allows to describe each neuropathological condition separately, and may better take into consideration comorbidity aspects. These practices are worth being encouraged, even though they are currently timidly implemented. Amid these capabilities, research will undoubtedly accelerate in addressing the question of heterogeneity in data and in encouraging scientific reproducibility. All these endeavors are definitely promising for the future of psychiatric research.

AUTHOR CONTRIBUTIONS

Both authors contributed to the article and approved the submitted version.

FUNDING

This work is funded by the Fund for Scientific Research (Fonds pour la Recherche Scientifique, F.R.S.-FNRS), Brussels (Belgium), through a research fellowship granted to SI.

ACKNOWLEDGMENTS

The authors would like to thank Kendra Kandana Arachchige (University of Mons, Belgium) for proofreading.

REFERENCES

1. United Nations Department of Economic and Social Affairs. *Sustainable development goals: sustainable development knowledge platform* (2015). Available at: https://bit.ly/3bCIAma (Accessed March 30, 2020).
2. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: Approaches to identify psychiatric subtypes. *Trends Cogn Sci* (2019) 23(7):584–601. doi: 10.1016/j.tics.2019.03.009
3. Kapadia M, Desai M, Parikh R. Fractures in the framework: limitations of classification systems in psychiatry. *Clin Neurosci* (2020) 22(1):17. doi: 10.31887/DCNS.2020.22.1/rparikh
4. Kawa S, Giordano J. A brief historicity of the diagnostic and statistical manual of mental disorders: issues and implications for the future of psychiatric canon and practice. *Philos Ethics Humanit Med* (2012) 7:2.
5. Adam C. Jalons pour une théorie critique du Manuel diagnostique et statistique des troubles mentaux (DSM). *Déviance société* (2012) 36(2):137–69. doi: 10.3917/ds.362.0137
6. Wiecki TV, Poland J, Frank MJ. Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. *Clin Psychol Sci* (2015) 3(3):378–99. doi: 10.1177/2167702614565359
7. Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci* (2017) 20(3):365. doi: 10.1038/nn.4478
8. Walter M, Alizadeh S, Jamalabadi H, Lueken U, Dannlowski U, Walter H, et al. Translational machine learning for psychiatric neuroimaging. *Prog Neuropsychopharmacol Biol Psychiatry* (2019) 91:113–21. doi: 10.1016/j.pnpbp.2018.09.014
9. Milham MP, Fair D, Mennes M, Mostofsky SH. The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Front Syst Neurosci* (2012) 6:62. doi: 10.3389/fnsys.2012.00062
10. Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: the FCP/INDI experience. *Neuroimage* (2013) 82:683–91. doi: 10.1016/j.neuroimage.2012.10.064
11. Milham MP. Open neuroscience solutions for the connectome-wide association era. *Neuron* (2012) 73(2):214–8. doi: 10.1016/j.neuron.2011.11.004
12. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry* (2014) 19(6):659. doi: 10.1038/mp.2013.78
13. Di Martino A, O'connor D, Chen B, Alaerts K, Anderson JS, Assaf M, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Sci Data* (2017) 4:170010. doi: 10.1038/sdata.2017.10
14. Vieira S, Gong QY, Pinaya WH, Scarpazza C, Tognin S, Crespo-Facorro B, et al. Using machine learning and structural neuroimaging to detect first episode psychosis: reconsidering the evidence. *Schizophr Bull* (2020) 46(1):17–26. doi: 10.1093/schbul/sby189
15. Oquendo MA, Baca-Garcia E, Artes-Rodriguez A, Perez-Cruz F, Galfalvy HC, Blasco-Fontecilla H, et al. Machine learning and data mining: strategies for hypothesis generation. *Mol Psychiatry* (2012) 17(10):956–9. doi: 10.1038/mp.2011.173
16. Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* (2016) 19(3):404. doi: 10.1038/nn.4238
17. Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med* (2016) 46(12):2455–65. doi: 10.1017/S0033291716001367
18. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry: Cogn Neurosci Neuroimaging* (2018) 3(3):223–30. doi: 10.1016/j.bpsc.2017.11.007
19. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol* (2018) 14:91–118. doi: 10.1146/annurev-clinpsy-032816-045037
20. Moustafa AA, Diallo TM, Amoroso N, Zaki N, Hassan M, Alashwal H. Applying big data methods to understanding human behavior and health. *Front Comput Neurosci* (2018) 12:84. doi: 10.3389/fncom.2018.00084
21. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* (2019) 1(5):206–15. doi: 10.1038/s42256-019-0048-x
22. Itani S, Lecron F, Fortemps P. Specifics of medical data mining for diagnosis aid: a survey. *Expert Syst Appl* (2019) 118:300–14. doi: 10.1016/j.eswa.2018.09.056
23. Yuste R, Goering S, Bi G, Carmena JM, Carter A, Fins JJ, et al. Four ethical priorities for neurotechnologies and AI. *Nat News* (2017) 551(7679):159. doi: 10.1038/551159a
24. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *New Engl J Med* (2018) 378(11):981. doi: 10.1056/NEJMp1714229
25. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?: Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. New-York, United States: Association for Computing Machinery (2016). p. 1135–44.
26. Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. *Big Data* (2015) 3(4):277–87. doi: 10.1089/big.2015.0020
27. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* (2014) 87:96–110. doi: 10.1016/j.neuroimage.2013.10.067
28. Itani S, Thanou D. Combining anatomical and functional networks for neuropathology identification: a case study on autism spectrum disorder. *arXiv* (2019).
29. Karpatne A, Atluri G, Faghmous JH, Steinbach M, Banerjee A, Ganguly A, et al. Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Trans Knowledge Data Eng* (2017) 29(10):2318–31. doi: 10.1109/TKDE.2017.2720168
30. Itani S, Rossignol M, Lecron F, Fortemps P. Towards interpretable machine learning models for diagnosis aid: a case study on attention deficit/hyperactivity disorder. *PloS One* (2019) 14(4):e0215720. doi: 10.1371/journal.pone.0215720
31. Krawczyk B, Galar M, Woźniak M, Bustince H, Herrera F. Dynamic ensemble selection for multi-class classification with one-class classifiers. *Pattern Recog* (2018) 83:34–51. doi: 10.1016/j.patcog.2018.05.015
32. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. *Knowledge Eng Rev* (2014) 29(3):345–74. doi: 10.1017/S026988891300043X
33. Retico A, Gori I, Giuliano A, Muratori F, Calderoni S. One-class support vector machines identify the language and default mode regions as common patterns of structural alterations in young children with autism spectrum disorders. *Front Neurosci* (2016) 10:306. doi: 10.3389/fnins.2016.00306
34. Itani S, Lecron F, Fortemps P. A one-class classification decision tree based on kernel density estimation. *Appl Soft Comput* (2020) 91:106250. doi: 10.1016/j.asoc.2020.106250
35. Sáez JA, Galar M, Luengo J, Herrera F. Analyzing the presence of noise in multi-class problems: alleviating its influence with the one-vs-one decomposition. *Knowledge Inf Syst* (2014) 38(1):179–206. doi: 10.1007/s10115-012-0570-1
36. Scarpazza C, Ha M, Baecker L, Garcia-Dias R, Pinaya WHL, Vieira S, et al. Translating research findings into clinical practice: a systematic and critical review of neuroimaging-based clinical tools for brain disorders. *Trans Psychiatry* (2020) 10(1):1–16. doi: 10.1038/s41398-020-0798-6
37. Scarpazza C, Baecker L, Vieira S, Mechelli A. Applications of machine learning to brain disorders. In: *Machine Learning*. Academic Press (2020). p. 45–65.
38. Colby JB, Rudie JD, Brown JA, Douglas PK, Cohen MS, Shehzad Z. Insights into multimodal imaging classification of ADHD. *Front Syst Neurosci* (2012) 6:59. doi: 10.3389/fnsys.2012.00059
39. Itani S, Lecron F, Fortemps P. A multi-level classification framework for multi-site medical data: application to the ADHD-200 collection. *Expert Syst Appl* (2018) 91:36–45. doi: 10.1016/j.eswa.2017.08.044
40. Itani S, Lecron F, Fortemps P. Data mining for ADHD & ASD prediction based on resting-state fMRI signals: a literature review. *CEUR Workshop Proc* (2019) 2491.
41. Rashid B, Calhoun V. Towards a brain-based predictome of mental illness. *Hum Brain Mapp* (2020) 41(12):3468–535. doi: 10.1002/hbm.25013
42. Carcone D, Ruocco AC. Six years of research on the national institute of mental health's research domain criteria (RDoC) initiative: a systematic

review. *Front Cell Neurosci* (2017) 11:46. doi: 10.3389/fncel.2017.00046
43. National Institute of Mental Health. *Research Domain Criteria* (2020). Available at: https://bit.ly/2Yd859r (Accessed August 18, 2020).
44. Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. *Trans Psychiatry* (2019) 9(1):1–12. doi: 10.1038/s41398-019-0607-2
45. Littmann M, Selig K, Cohen-Lavi L, Frank Y, Hönigschmid P, Kataka E, et al. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nat Mach Intell* (2020) 2:18–24. doi: 10.1038/s42256-019-0139-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.