Asian Journal of Psychiatry 44 (2019) A1–A4

Contents lists available at ScienceDirect

Asian Journal of Psychiatry

journal homepage: www.elsevier.com/locate/ajp

Machine learning in psychiatry- standards and guidelines

1. Introduction

Over the past decade, there has been a veritable explosion in the application of machine learning (ML) methods to a myriad of fields, including scientific research. Psychiatry has been somewhat of a late entrant, but the rapidly increasing number of ML-utilizing publications across Psychiatry scientific journals suggests that it is catching up (Bassett et al., 2018; Cornblath et al., 2019; Ferrante et al., 2019; Grabowski and Rappsilber, 2019; Huys et al., 2016; Kas et al., 2019; Tai et al., 2019; Tandon and Tandon, 2018; Zitnik et al., 2019). Exponential expansion of computing ability has fueled this dramatic increase in the use of ML techniques in research. In contrast to traditional statistical methods, ML techniques are able to process enormous volumes of "big data" and decipher the complexity of multidimensional information derived from different high-throughput technologies (connectomics, epigenetics, genomics, metabolomics, proteomics, transcriptomics, etc.). Since mental functions are generated by recursive, dynamic interactions between individuals' brains and their physical and social environment in the context of the individual's development and life history, elucidation of the nature of psychiatric disorders requires obtaining and making sense of vast amounts of interacting, multi-dimensional, multi-level information. ML appears well-suited to this task.

The proliferation of ML-using articles, however, runs the risk of generating of a lot of heat without much light (Tandon and Tandon, 2019), if article submissions are not subjected to proper scientific scrutiny by Journal editors/reviewers and the ability of readers to critically evaluate published articles is inadequate. Machine learning approaches are, however, poorly understood by psychiatry researchers, journal editors and reviewers, and the broader clinical and scientific community alike. Since all these groups are much more familiar with traditional statistical approaches, they tend to utilize the same approach to evaluating ML-using manuscripts that they have learned to effectively apply to scientific reports that utilize statistical methods. Some fundamental differences between ML and traditional statistical methodologies with regard to assumptions, procedures, outputs, and implications, however, necessitate a substantial modification in how the various stakeholders approach ML-using manuscripts. Furthermore, usual issues of rigorous study design (sampling frame, data quality, sample size, etc.) and threats to validity of inferences (Rankupalli and Tandon, 2010) also apply to studies employing ML methods; this need for equivalent methodological rigor is often overlooked. Finally, ML employs a different 'language' and sometimes uses familiar terms in deceptively different ways and this can complicate evaluation of ML-using articles.

With the objective of harnessing the incredible potential of ML in 'truly' advancing our field, we in the Asian Journal Psychiatry have adopted some policies, standards, and practices right from the outset. We describe our approach and then suggest some guidelines for researchers, readers, and gatekeepers of the burgeoning machine learning literature. We conclude with a new approach in our Journal to evaluating and presenting ML-using research articles to the field.

2. Machine learning and differences from traditional methods of statistical analysis

Deriving from the disciplines of artificial intelligence, computational statistics, and pattern recognition, ML is a family of computational techniques of data analytics based on learning from data for the purpose of data-driven prediction, description, or explanation (Dwyer et al., 2018; Rutledge et al., 2019). ML generates an algorithm or function that provides an optimized description of the patterns and relationships between the sets of observations and variables in the dataset. ML algorithms search through a large space of candidate programs, make initial informed guesses, and then adjust the model little by little to optimize performance. Model parameters are found through a series of back and forth steps, where model parameters are estimated, the model performance is evaluated, errors are identified and corrected, and then the process is repeated until model error is minimized (Dwyer et al., 2018; Tandon and Tandon, 2019).

A few critical differences from traditional statistical methods are worth noting:

i) The first step in traditional statistics is to establish a significant relationship between variables before generating an equation or function linking them. ML methods, on the other hand, presuppose a significant relationship between a set of independent variables and dependent variable and only seeks to find the path that most strongly links the two. Thus, one significant safeguard of first establishing a significant association (p-value) before then developing a function that describes the structure of the relationship is now absent. This results in an essential difference in study interpretation: ML will always delineate some simple pattern or data structure (because it presumes that there is such a relationship to be found); consequently, it is important not to over-read data and 'see' phenomena that are not really there.
ii) Traditional statistical approaches generate functions that are comprehensible and more easily interpretable, the algorithms generated by ML methods are more of "black boxes" with varying degrees of opacity. While this is not a problem by itself, it does make proper and independent validation of the generated model even more important. Furthermore, their black-box nature makes them much more difficult to debug;
iii) The term cross-validation in ML is not the same as independent

https://doi.org/10.1016/j.ajp.2019.09.009

1876-2018/ © 2019 Published by Elsevier B.V.

Table 1
Check-list of Peer-Review Items for Evaluating Machine Learning Manuscripts in Psychiatry.

Component	Questions
1. Overall	a) What is the central question (what important gap in the existing literature does the paper seek to fill?)? b) Why choose ML approach to address this gap? c) Quality of the work. d) Ethical concerns, if any; e) Are the authors willing to send in their data and source code for true independent replication (either as supplementary material or for Journal repository)- if so, are they able to so in an auditable structure;
2. Title	(a) Does it explicitly identify: i nature (is it proof of concept of potential of ML to provide potentially meaningful information OR actual advance in knowledge/understanding); ii purpose (explanatory, predictive, descriptive) of model generated;
3. Abstract	(a) Does it appropriately summarize the manuscript? Background what is the specific area of psychiatry where a knowledge gap exists and where this study provides actual advance or promise of potential to provide future advance; Methods Data source or sources; ML type (supervised, unsupervised, etc.) Internal validation approach utilized (left-out test sample, or some form of cross-validation using entire sample) Results Performance metrics of generated model, if appropriate (e.g., of predictive models- accuracy, false-positives, etc.); Conclusion/Discussion Precisely what is the implication of the resultant model; Exactly, what are the next follow-up steps in the application or development of the generated model' (b) Are there discrepancies between the abstract and the remainder of the manuscript? (c) Can the abstract be understood without reading the remainder of the manuscript?
4. Introduction	a) Is the purpose of the study clearly laid out? It is NOT enough to merely state the purpose of the study as "it is novel that we applied ML to …..". Was the objective of the ML application a test of a hypothesized relationship, development of a "better" predictive model, or just a proof of concept- if so exactly what proof of concept and is that novel; b) Is the scope of the model being developed spelled out? c) Is a rationale for the study provided on the basis of a succinct review of the literature ("what gap in the existing literature does this study seeking to address")? d) What is the specific hypothesis being tested? e) Is the ML tool being developed aimed at improving mechanistic insight or optimizing predictive accuracy;
5. Materials and Methods	a) How were the utilized datasets selected; b) Are sources of data utilized in modelling fully described in terms of sampling frame/recruitment methods of data source/s and relevant inclusion/exclusion criteria; c) Are data elements spelled out and definitions/methods to quantify them spelled out? d) If there are multiple sources of data, are there differences in sampling frame, recruitment methods, definition and measurement of data elements, relevant methods of data collection; etc.? e) Are methods of data pre-processing (e.g., data transformation, feature selection, etc.) clearly outlined? f) How was the specific ML method or methods selected from those available; g) Is the extent of and constraints on parameter tuning clearly described; h) Are evaluation criteria of model being generated explicitly defined in terms of key quantitative performance metrics;
6. Results, Tables, and Figures	a) Are performance metrics of generated model provided ([predictive models- sensitivity and specificity, false positives and negatives, etc.]; performance metrics for descriptive and explanatory models also exist, although they can be more complex); b) Are results of sensitivity analyses provided; c) Are details of test- or cross- validation clearly outlined?
7. Discussion	a) Is the discussion concise and clear? b) Is there a clear statement about the principal study findings? c) Do the study conclusions clearly flow from the results and are NOT overstated or otherwise inappropriately stated? d) Is it clear what new knowledge the study has provided? e) Does the generated ML model really address the specific problem it was being developed to address; f) Exactly, how does the model generated by ML compare to or represent an improvement over existing models? g) How are discrepant findings explained? h) Are the strengths and weaknesses of the study noted? Specifically, does the discussion include potential pitfalls in model interpretation, potential biases of the data used in modelling, limitations in generalizability; i) Is there a clear and concise conclusion about the implications of the study and next steps, if appropriate?
8. References	a) Are important relevant references all included? Are there major omissions? b) Are salient points of cited articles accurately quoted?

validation. Much more is needed for true reproducibility of results (Lewis et al., 2016).

At the same time, the following important similarities of ML approaches to traditional statistical methods bear mention:

i) Resulting models encode correlation and association, not causation or ontological relationships;
ii) Data quality, sampling frame, sample size, and research designs of studies that are the source of the utilized data are just as consequential. Furthermore, data errors and sample biases often tend to be amplified in ML (Kolossa and Kopp, 2018).

3. Evaluating an ML study

What specific guidance can one provide to editors, reviewers, readers, and researchers about meticulous evaluation of any ML study in Psychiatry? Guidance for researchers has been provided in several publications (Cornblath et al., 2019; Grabowski et al., 2019; Luo et al., 2016; Nichols et al., 2017; Scheinost et al., 2019; Vu et al., 2018; Weber et al., 2019; Wilson et al., 2014, several others). It is, however, vital that an expert group of scientists and clinicians in psychiatry, neuroscience, machine learning, research methodology, neuroethics, and scientific publication convene to formulate a common set of guidelines for

scientific reporting of ML studies in psychiatry. Such guidelines have been developed for a whole range of other research methodologies and have significantly improved the quality and value of publications that require their utilization and application by authors/researchers (Equator Network, 2019; Martensson et al., 2016; McLeroy et al., 2016).

A recent article (Tandon and Tandon, 2019) outlined a strategy for readers of the growing ML literature in Psychiatry that included asking the following questions of any such report:

(i) What is the precise question that the study addresses or exactly what is the problem that the model or algorithm seeks to solve?
(ii) How well have the results been validated?
(iii) Exactly what data went into the "black box" in which the model or algorithm was developed?
(iv) What is the exact purpose of the model that was generated? Is it to describe, explain, or predict?
(v) Are the results reproducible?
(vi) Are the results transferable?
(vii) Are they results actionable? If not, what are the next steps towards making this happen?
(viii) Is it clear as to what problem or knowledge gap the ML-applying study was designed to address and how does the solution provided compare to that of other approaches?

There is no corresponding guidance for Editors of scientific journals and reviewers of ML articles. These quality gatekeepers of scientific publications play an important role in the promotion and dissemination of biomedical research. With reference to ML submissions, they are expected to evaluate the replicability, validity, robustness, relevance, utility, and applicability of the algorithm or model generated by the study. We utilize a set of standards in our Journal, which are summarized in Table 1.

3.1. Role of the editor and reviewers

Machine learning is an essential and incredibly powerful tool in our efforts to better understand the nature of psychiatric disorders and develop more effective treatments. ML also has significant constraints that warrant its thoughtful utilization and careful application. All segments of the scientific community have an important role in ensuring appropriate and efficient utilization of ML to allow it to live up to its promise of helping advance the practice of Psychiatry. Funding and regulatory entities must ensure rigorous data management and sharing, transparency, and "good science" (Moher et al., 2016). Researchers in the field need to be more self-critical in their construction and presentation of their models and make available their data and source code to enable independent replication. Readers need to improve their understanding of computational psychiatry and ML (Goldman and Fee, 2017). Journal editors and reviewers have a difficult, but crucial responsibility in the critical evaluation of ML articles and ensuring the dissemination of meaningful and comprehensible scientific information to their readership.

There is a crisis of reproducibility in biomedical research (Bzdok and Ioannidis, 2019) and the fields of psychiatry and neuroscience share this significant problem (Dacrema et al., 2019). As summarized by Lewis et al. (2016), reproducibility encompasses (i) replicability ("other people get the same result when doing exactly the same thing"), (ii) true reproducibility ("something similar will happen in other researchers' hands when they study the same phenomenon", albeit with different methods, and competing explanations of observations are excluded), and (iii) extensibility (application of research findings to the broader real-world). Machine learning methods present unique challenges with regard to reproducibility in addition to those posed by traditional statistical approaches (Lewis et al., 2016). Our field is also being flooded by isolated "findings" that are never reproduced (but not discarded) that are overwhelming our ability to synthesize or make sense of (Nasrallah et al., 2011; Tandon, 1999). We are certain to be inundated by increasing reports of ML findings in Psychiatry, a trend exacerbated by the explosion of predatory publishing in our field (Gogtay and Bavdekar, 2019; Nuland and Rogers, 2017).

4. Machine learning articles in the Asian journal of psychiatry

With the objective of disseminating research that is credible (coherent, consistent, rigorous, and transparent), contributory (original, relevant, and generalizable), communicable (understandable, accessible, and searchable), and conforming (regulatory aligned, ethical, sustainable) (Martensson et al., 2016), we in the Asian Journal of Psychiatry will continue to utilize our designated ML section editor and our guidelines for review (Table 1) that have been disseminated to our reviewers and readership. Additionally, we will begin adding a brief expert analysis to all future ML articles in our Journal with a view to better fulfilling the Journal's responsibilities to its readership and our field at large.

References

Bassett, D.S., Zurn, P., Gold, J.I., 2018. On the nature and use of models in network neuroscience. Nat. Rev. Neurosci. 19, 566–578.
Bzdok, D., Ioannidis, J.P.A., 2019. Exploration, inference, and prediction in neuroscience and biomedicine. Trends Neurosci. Educ. 42, 251–262.
Cornblath, E.J., Lydon-Staley, D.M., Bassette, D.S., 2019. Harnessing networks and machine learning in neuropsychiatric care. Curr. Opin. Neurobiol. 55, 32–39.
Dacrema, M.F., Cremonisi, P., Jannach, D., 2019. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. arXiv 1907, 06902.
Dwyer, D.B., Falkai, P., Koutsouleris, N., 2018. Machine learning approaches for clinical psychology and psychiatry. Annu. Rev. Clin. Psychol. 14, 1–28.
Equator Network, 2019. Reporting Guidelines. . http://www.equator-network.org/library-guidelines/.
Ferrante, M., Redish, A.D., Oquendo, M.A., Averbeck, B.B., Kinnane, M.E., Gordon, J.A., 2019. Computational psychiatry: a report from the 2017 NIMH workshop on opportunities and challenges. Mol. Psychiatry 24, 479–483.
Gogtay, N.J., Bavdekar, S.B., 2019. Predatory journals- can we stem the rot? J. Postgrad. Med. 65, 129–131.
Goldman, M.S., Fee, M.S., 2017. Computational training for the next generation of neuroscientists. Curr. Opin. Neurobiol. 46, 25–30.
Grabowski, P., Rappsilber, 2019. A primer on data analytics in functional genomics: how to move from data to insight. Trends Biochem. Sci. 44, 21–32.
Huys, Q.J.M., Maia, T.V., Frank, M.J., 2016. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat. Neurosci. 19, 404–413.
Kas, M.J., Penninx, B., Sommer, B., et al., 2019. A quantitative approach to neuropsychiatry: the why and how. Neurosci. Behav. Rev. 97, 3–9.
Kolossa, A., Kopp, B., 2018. Data quality over data quantity in computational cognitive neuroscience. Neuroimage 172, 775–785.
Lewis, J., Breeze, C., Charlesworth, J., et al., 2016. Where next for the reproducibility agenda in computational biology? BMC Syst. Biol. 10, 52.
Luo, W., Phung, D., Tran, T., et al., 2016. Guidelines for developing and reporting machine learning predictive models in biomedical research. J. Med. Internet Res. 18, e323.
Martensson, P., Fors, U., Wallin, S.-B., Zander, U., Nilsson, G.H., 2016. Evaluating research: a multi-disciplinary approach to assessing research practice and quality. Res. Policy 45, 593–603.
McLeroy, K.R., Garney, W., Mayo-Wilson, E., Grant, S., 2016. Scientific reporting: raising the standards. Health Educ. Behav. 43, 501–508.
Moher, D., Glasziou, P., Chalmers, I., et al., 2016. Increasing value and reducing waste in biomedical research: who's listening? Lancet 387, 1573–1586.
Nasrallah, H.A., Tandon, R., Keshavan, M.S., 2011. Beyond the facts in schizophrenia. Epidemiol. Psychiatric Serv. 20, 317–327.
Nichols, T.E., Das, S., Eickhoff, S.B., et al., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci. 20, 299–303.
Nuland, S.E.V., Rogers, K.A., 2017. Academic nightmares: predatory publishing. Anat. Sci. Educ. 10, 392–394.
Rankupalli, B., Tandon, R., 2010. Practicing evidence-based psychiatry- threats to validity approach. Asian J. Psychiatry 3, 35–40.
Rutledge, R.B., Chekroud, A.M., Huys, Q.J.M., 2019. Machine learning and big data in psychiatry: toward clinical applications. Curr. Opin. Neurobiol. 55, 152–159.
Scheinost, D., Noble, S., Horien, C., et al., 2019. Ten simple rules for predictive modeling of individual differences in neuroimaging. NeuroImage 193, 35–45.
Tai, A.M.Y., Albuquerque, A., Carmona, N.E., et al., 2019. Machine learning and big data: implications for disease modeling and therapeutic discovery in psychiatry. Artif. Intell. Med. 99, 101704.
Tandon, N., Tandon, R., 2018. Will machine learning enable us to finally cut the Gordian knot of schizophrenia? Schizophr. Bull. 44, 939–941.
Tandon, N., Tandon, R., 2019. Using machine learning to explain the heterogeneity of

schizophrenia. Realizing the promise and avoiding the hype. Schizophr. Res In Press.
Tandon, R., 1999. Moving beyond findings: concepts and model building in schizophrenia. J. Psychiatr. Res. 33, 467–471.
Vu, M.-A.T., Adali, T., Ba, D., et al., 2018. A shared vision for machine learning in neuroscience. J. Neurosci. 38, 1601–1607.
Weber, L.M., Saelens, W., Cannoodt, R., et al., 2019. Essential guidelines for computational method benchmarking. Genome Biol. 20, 125.
Wilson, G., Aruliah, D.A., Brown, C.T., et al., 2014. Best practices for scientific computing. PLoS Biol. 12, e1001745.
Zitnik, M., Nguyen, F., Wang, B., et al., 2019. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. Inf. Fusion 50, 71–91.

Neeraj Tandon, Rajiv Tandon*
Department of Psychiatry, WMU Homer Stryker School of Medicine, Kalamazoo, MI, United States
E-mail address: rajiv.tandon@med.wmich.edu (R. Tandon).

* Corresponding author at: Department of Psychiatry, WMU Homer Stryker School of Medicine, 1000 Oakland Drive, Kalamazoo, MI, 49008, United States.