VIEWPOINT

Are Psychiatric Nosologies Limiting the Success of Clinical Prediction Models?

Isabelle Scott, PhD
Centre for Youth Mental Health, The University of Melbourne, Parkville, Melbourne, Victoria, Australia; and Orygen, Parkville, Melbourne, Victoria, Australia.

Barnaby Nelson, PhD
Centre for Youth Mental Health, The University of Melbourne, Parkville, Melbourne, Victoria, Australia; and Orygen, Parkville, Melbourne, Victoria, Australia.

In many instances across psychiatric research, binary or multiclass supervised machine learning (ML) or statistical classifiers are used to predict diagnostic outcomes. To train and test supervised classifiers, ground-truth labels identifying the true diagnostic outcomes for each individual are required. Consequently, the success of supervised prediction models is inherently linked to the degree to which diagnostic labels represent valid groupings that can be separated (linearly or nonlinearly) according to some set of biomarkers (**Figure**). These diagnostic labels are generally derived from current psychiatric nosologies. In the context of diagnostic marker discovery, labels may indicate the current presence or absence of a disorder, whereas, for early intervention, we are often interested in predicting future transition to a disease state in those at risk. Furthermore, prognostic and treatment-matching algorithms frequently seek to predict the future diagnostic status of individuals with a mental health condition, with diagnostic status defined by criteria attached to those used for initial diagnosis (eg, prediction of remission in *DSM-5*–defined schizophrenia[1]).

Critically, discrete diagnostic labels have been designed to guide clinical practice and decision-making, not to inform ML or because true discrete groups even necessarily exist. This is not just an issue of a single-dimensional continuous scale being discretized. Rather, the heterogeneity of symptoms, functioning, and duration that defines many diagnostic labels—extending from initial diagnosis to remission and recovery criteria—suggests that we are discretizing a highly multidimensional continuum. Thus, there is heterogeneity within and overlap across diagnostic groups.[2] For example, according to a recently developed instrument, an individual can transition to first-episode psychosis by exceeding a severity threshold for any 1 of 16 symptoms.[3] Consequently, 2 individuals exhibiting the same symptom but with a small difference in severity are allocated to separate diagnostic labels. Conversely, 2 individuals exhibiting distinct psychotic symptoms—such as perceptual abnormalities and disorganized speech—above or below a certain threshold are grouped together. For clinical decision-making, such severity-based groupings are likely necessary. However, whether these diagnostic groups exhibit distinct biomarkers and are fit for purpose for predictive modeling is another question. If the validity of the diagnostic classes is already questionable, ML models may overfit and achieve overly optimistic internal validation results because of the low signal to noise ratio and will subsequently fail external validation tests. A low signal to noise ratio is especially problematic in the setting of small datasets, which is often the case in psychiatry. Cross-validation can assist in identifying overfitting but can still achieve overoptimistic estimates.[1]

To address the lack of model generalizability in psychiatry, more representative samples and robust validation approaches have been suggested.[1] Although critical, these approaches do not mitigate the issue of poor ground-truth labels. Given the recurrent challenges to psychiatric nosologies in recent years, the effect of the poor validity of psychiatric nosologies on the success of clinical prediction models is a pertinent issue to consider. Many researchers are now in fact arguing for a transdiagnostic approach to mental health. However, although there are some promising possibilities on the horizon,[4,5] there does not yet appear to be sufficient agreement on a new nosology fit for purpose for implementation in clinical practice.[2] If the clinical and research community is in such strong disagreement regarding the categorization of mental disorders, is it really worthwhile to attempt to predict these categories or an individual's state within these categories (eg, initial diagnosis, transition, and remission)? If so, how do we go about doing this while incorporating the current limitations attached to diagnostic labels?

We suggest that to address the challenge that current nosologies pose to the success of prediction models, we should move beyond predicting discrete diagnostic classes. There are multiple options for achieving this. Multilabel classifiers and supervised regression models are readily available options that allow us to take into account comorbidities and symptomatology that overlaps between diagnostic classes and to remove the need to arbitrarily discretize continuous scales. A caveat in advocating for this approach is that more data are often required to implement these models, reinforcing the need for the research community to fully adopt open science. Unsupervised learning methods have also demonstrated utility in evaluating the relationship between clinical dimensions, in the form of symptoms that are present across multiple diagnostic categories, and continuous biomarkers. For example, canonical correlation analysis has been used to identify new brain-based dimensions of psychopathology.[6] Such approaches may provide the opportunity to identify new psychopathological dimensions that have the potential to guide biomarker-based treatment approaches.[7]

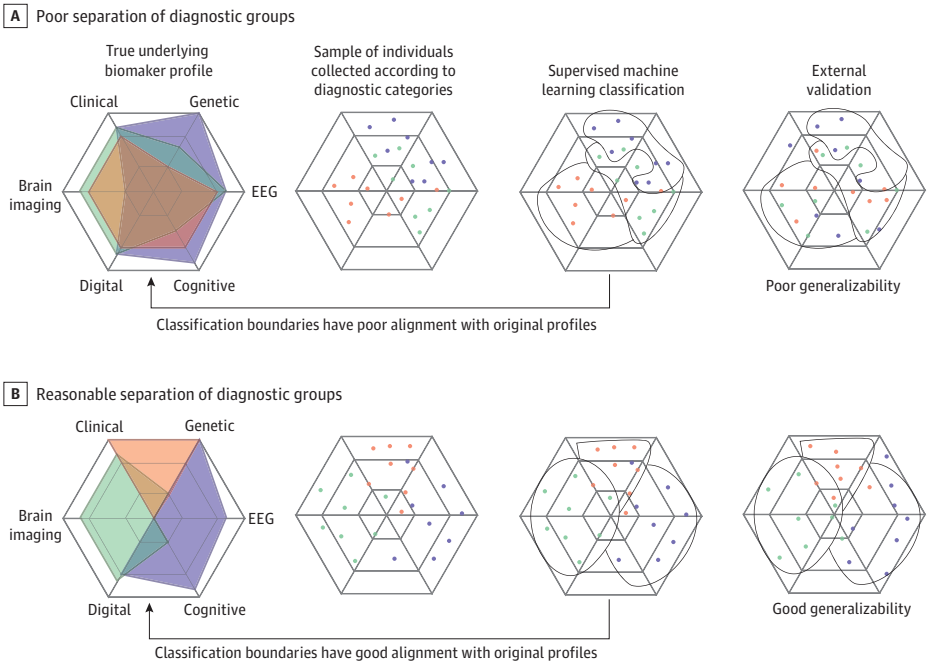
Recognizing the clinical utility of discrete diagnostic classes, ML also offers the potential for data-driven refinement of diagnostic boundaries. Discretizing continuous psychopathological processes may be beneficial if we can identify diagnostic classes with sufficiently distinct biological signatures to guide treatment.[8] To leverage the insight psychiatric nosologies provide, without treating them as the criterion standard, weakly supervised learning presents a potential yet relatively unexplored field. One form, termed *inaccurate*

Corresponding Author: Isabelle Scott, PhD, Orygen, 35 Poplar Rd, Parkville, Melbourne, VIC 3052, Australia (isabelle.scott@orygen.org.au).

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Helsinki Univeristy Library user on 10/03/2025

Figure. Supervised Machine Learning (ML) Used to Classify Individuals Into Discrete Diagnostic Classes



Poor separability of the biomarkers between groups results in a low signal to noise ratio, leading to overfitting of the training data by the ML algorithm.

supervision, performs prediction and label refinement, treating the original labels as partially inaccurate. For example, joint optimization frameworks find the optimal model parameters and training labels for the input data (biomarkers in the current setting) by iterating between parameterizing the model and relabeling the data.[9] Thus, weakly supervised learning provides a potential opportunity to refine diagnostic classes rather than a complete reconceptualization. Recently, researchers have started to consider the inverse relationship of ML with psychiatric nosologies, suggesting that the success of prediction models could inform the degree to which diagnostic classes exhibit unique biological signatures.[10]

ML and the conceptualization of mental disorders are 2 evolving fields, and the interplay between the 2 needs to be considered. The generalizability of supervised prediction models is inherently related to the validity of psychiatric nosologies. Before asking the question "How can we predict this outcome?" researchers might want to reflect on the questions "What do I want to predict? Is this outcome valid?" Less restrictive prediction models that move beyond traditional supervised prediction of diagnostic classes or status provide a promising future direction to address the limitations of our current conceptualization of mental disorders.

ARTICLE INFORMATION

Published Online: September 18, 2024. doi:10.1001/jamapsychiatry.2024.2662

Conflict of Interest Disclosures: Dr Nelson reported receiving grants from the National Institute of Mental Health, from the Wellcome Trust, and from the National Health and Medical Research Council outside the submitted work. No other disclosures were reported.

Additional Contributions: We would like to express our gratitude to the reviewer whose comprehensive and thoughtful feedback improved the quality of the article.

REFERENCES

1. Chekroud AM, Hawrilenko M, Loho H, et al. Illusory generalizability of clinical prediction models. *Science*. 2024;383(6679):164-167. doi:10.1126/science.adg8538

2. Dalgleish T, Black M, Johnston D, Bevan A. Transdiagnostic approaches to mental health problems: current status and future directions. *J Consult Clin Psychol*. 2020;88(3):179-195. doi:10.1037/ccp0000482

3. Woods SW, Parker S, Kerr MJ, et al; Accelerating Medicines Partnership Schizophrenia. Development of the PSYCHS. *Early Interv Psychiatry*. 2024;18(4):255-272. doi:10.1111/eip.13457

4. Jonas KG, Cannon TD, Docherty AR, et al. Psychosis superspectrum, I. *Mol Psychiatry*. 2024; 29(4):1005-1019. doi:10.1038/s41380-023-02388-2

5. Scott J, Iorfino F, Capon W, et al. Staging 2.0: refining transdiagnostic clinical staging frameworks to enhance reliability and utility for youth mental health. *Lancet Psychiatry*. 2024;11(6):461-471. doi:10.1016/S2215-0366(24)00060-9

6. Xia CH, Ma Z, Ciric R, et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun*. 2018;9(1):3003. doi:10.1038/s41467-018-05317-y

7. Kotov R, Krueger RF, Watson D, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP). *J Abnorm Psychol*. 2017;126(4):454-477. doi:10.1037/abn0000258

8. Insel TR, Cuthbert BN. Medicine: brain disorders? precisely. *Science*. 2015;348(6234):499-500. doi:10.1126/science.aab2358

9. Tanaka D, Ikami D, Yamasaki T, Aizawa K. Joint optimization framework for learning with noisy labels. Paper presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, Utah.

10. Chen J, Patil KR, Yeo BTT, Eickhoff SB. Leveraging machine learning for gaining neurobiological and nosological insights in psychiatric research. *Biol Psychiatry*. 2023;93(1):18-28. doi:10.1016/j.biopsych.2022.07.025

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Helsinki Univeristy Library user on 10/03/2025