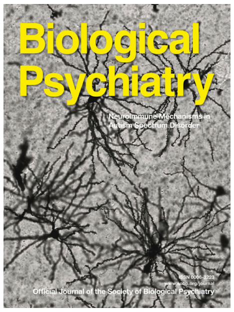


# Leveraging machine learning for gaining neurobiological and nosological insights in psychiatric research

Ji Chen, Kaustubh R. Patil, B.T. Thomas Yeo, Simon B. Eickhoff

PII: S0006-3223(22)01481-0

DOI: https://doi.org/10.1016/j.biopsych.2022.07.025

Reference: BPS 14947

To appear in: *Biological Psychiatry*

Received Date: 13 March 2022

Revised Date: 6 July 2022

Accepted Date: 28 July 2022

Please cite this article as: Chen J., Patil K.R., Yeo B.T.T. & Eickhoff S.B., Leveraging machine learning for gaining neurobiological and nosological insights in psychiatric research, *Biological Psychiatry* (2022), doi: https://doi.org/10.1016/j.biopsych.2022.07.025.

# Leveraging machine learning for gaining neurobiological and nosological insights in psychiatric research

Ji Chen[1,2,3,*], Kaustubh R. Patil[3,4], B.T. Thomas Yeo[5,6,7,8,9], Simon B. Eickhoff[3,4]

[1]Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, Zhejiang, China;
[2]Department of Psychiatry, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, Zhejiang, China;
[3]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany;
[4]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany;
[5]Centre for Sleep and Cognition& Centre for Translational MR Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore;
[6]Department of Electrical and Computer Engineering, National University of Singapore, Singapore;
[7]N.1 Institute for Health & Institute for Digital Medicine, National University of Singapore, Singapore;
[8]Integrative Sciences & Engineering Programme, National University of Singapore, Singapore;
[9]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA.

*Correspondence should be addressed to:

Dr. Ji Chen; Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 310028, China; E-mail: jichen.allen@hotmail.com.

**Short title:** Machine learning for insights in psychiatric research

**Keywords:** Machine learning, Psychiatric disorder, Biomarker, Nosology, Multiview integration, Heterogeneous dissection

## ABSTRACT

Much attention is currently devoted to developing diagnostic classifiers for mental disorders. Complementing these efforts, we highlight the potential of machine-learning to gain biological insights into the psychopathology and nosology of mental disorders. Studies to this end have mainly used brain imaging data, which can be obtained non-invasively from large cohorts and have repeatedly been argued to reveal potentially intermediate phenotypes. This may become particularly relevant in light of recent efforts to identify MRI derived biomarkers that yield insight into pathophysiological processes as well as to refine the taxonomy of mental illness. In particular, the accuracies of machine-learning models may be used as dependent variables to identify features relevant to pathophysiology. Moreover, such approaches may help to disentangle the dimensional (within diagnosis) and often overlapping (across diagnoses) symptomatology of psychiatric illness. We also point out a multi-view perspective that combines data from different sources, bridging molecular and system-level information. Finally, we summarize recent efforts toward a data-driven definition of subtypes or disease entities through unsupervised and semi-supervised approaches. The latter, blending unsupervised and supervised concepts, may represent a particularly promising avenue toward dissecting heterogeneous categories. Finally, we raise several technical and conceptual aspects related to the reviewed approaches. In particular, we discuss common pitfalls pertaining to flawed input data or analytic procedures that would likely lead to unreliable outputs.

Substantial efforts have been devoted to building machine learning models for aiding or automating clinical decisions related to diagnosis, treatment guidance and prognosis (1,2). Such models learn patterns relating the input features to targets from training data that are then expected to generalize to new subjects (3). This predictive ability combined with the current lack of objective means for diagnosing mental disorders created enthusiasm in the psychiatric neuroimaging community. Researchers eagerly, and rightly so, trained classification models *(4-7)* as potentially objective means to aid standard (interview-based) strategies. As an important benefit, explainable models may further provide biological markers that could add validity to the extant diagnostic system and treatment design.

These motivations combined with promising initial accuracies as well the increased data availability through large data-sharing initiatives, fueled a booming research area of psychiatric machine learning, mainly using brain MRI data. As a non-invasive, in vivo technique, MRI investigations have added valuable insights into the pathophysiology of psychiatric disorders by revealing brain structural, functional and metabolic abnormalities (8-11) as possible intermediate phenotypes (12).

While these topics have been covered multiple times, we here review several developing, complementary lines of research using machine learning to gain insights into the neurobiology underlying psychopathology and the nosological structures of mental illness. These studies have opened new windows into the neurobiological aspects underlying particular disorders, the relationships among diagnostic groups and potential heterogeneities within them.

In the following, we will cover three directions in the current literature that hold substantial promise for applying machine learning in psychiatric neuroimaging and beyond. I) Using prediction accuracy as a dependent variable to investigate the pathophysiological relevance of different feature sets in both single- and multi-view fashions. II) Gaining insights into nosology through differential accuracies between disorders. III) Data-driven consolidation to disentangle heterogeneous disorders and redefine clinical entities. We will concentrate on MRI to keep the scope manageable and focused. Notwithstanding, many of the proposed perspectives and considerations apply likewise to data from other imaging modalities and biological sources, which should complement the system-level surrogate markers of pathophysiology provided by MRI (13). In particular, molecular assays could potentially be closer to the actual pathological process, although they are likely less sensitive to systemic phenomena like dysregulated connectivity.

## 1.-Accuracies as the dependent variable to select informative brain features in single- and multi-view fashions

Supervised machine learning requires, for each subject, a set of input features (such as regional brain volume measurements or estimates of connectivity strengths) and a target label. In this context, we note that labels can be categorical for diagnostic classifiers or continuous, e.g., representing clinical or psychological phenotypes. The task of the algorithm is then to learn a relationship between the features and the target that optimally generalizes to new observations (3). This can be evaluated by providing the trained model with the features of new subjects and comparing the predicted labels against the true ones, which are known but not available to the model. When comparing various features in such an approach, it becomes evident that the observed accuracies directly provide information on the predictive power and hence relevance of a feature set (14). Conversely, features that yield the highest accuracies are likely to represent or relate to a core aspect of the target, i.e., pathophysiology. From a technical perspective, two approaches may be differentiated:

1) Comparing accuracies for different types of features. In this case, different *a priori feature*-sets, such as different types of MR images, connectivity values for different networks or parcellation schemes, may be presented. Classification accuracies can be compared across feature sets (in the same subjects, predicting the same target) to rank them according to their predictive capacity (8,15,16). Thus, we can infer which features provide the most generalizable and hence robust prediction. These are then interpreted as the most likely neurobiological

substrate of the investigated disorder among the investigated aspects of brain structure, function and connectivity.

2) Pruning a broad feature set. Here, all available features are provided to the algorithm, which is then tasked with extracting the most relevant features either by iteratively ranking features based on usefulness due to prediction power (wrapper approach)(17) or by estimating feature importance while training (embedded approach)(18). In essence, these approaches aim to find *a subset of the features that is not worse or even better – through a more favorable feature-sample ratio – than the full data.*

Both strategies have been employed to detect potential elements of pathophysiology for several psychiatric disorders (Figure 1A&1B). In schizophrenia, they repeatedly pointed toward abnormalities in the prefrontal cortex, the lateral temporal lobes, the striatum, and the thalamus as well as the posterior parietal cortex and the precuneus as discriminative aspects of neurobiology (16,19,20). Interestingly, in particular the latter two regions also seemed to be discriminative between schizophrenia patients with predominantly positive vs. negative psychopathology (8). Studies contrasting schizophrenia to other psychiatric illnesses revealed frontal and temporal cortices as well as cingulate, cerebellar and subcortical regions as the most salient features (21， 22). Informative MRI features to robustly distinguish between attention deficit hyperactivity disorder (ADHD) types (23) and predict the prognosis of late-life depression (24) have likewise been identified.

*Since mental illness likely represents multi-scale pathology spanning genetic, molecular, cellular and system levels* (27,28)*, data from other biological sources like genetic assays* *should contain complimentary information for the understanding of (individual presentations of) mental disorders (29,30). The aforementioned approaches could thus be applied to select respective features from each data type, and information fusion methods (31) like multi-kernel and stacked learning may then be used for a unified representation (Figure 1C; for extended reading please see https://www.sciencedirect.com/journal/information-fusion/special-issue/10MJ9Z5TNCP). Following this approach, the expression profile of genes related to autism spectrum disorder (ASD) was revealed to complement resting-state functional connectivity (rsFC) as robust features in classifying typically developing children versus ASD (31). Sub-space learning methods (32), such as canonical correlation analysis (CCA), can achieve similar goals which often assume that the input features from different views are related in a common subspace. The algorithms are then used to extract l*atent patterns of the input data while retaining correlations across views (33), as widely used in psychiatric studies (cf section 3 below).

As data like multi-tracer PET maps and postmortem transcriptional profiles are not routinely attainable for individual patients, recent efforts integrated machine-learning-derived MRI features and information from other sources using spatial correlation in a post-hoc manner (Figure 1D). Employing this approach to schizophrenia, *Li et al.* (34) linked striatal activity abnormalities to dopamine D2/D3 receptor availability and dopamine synthesis capacity as well as genes encoding particularly dopamine receptor D2 and glutamate metabotropic receptor, while *Chen et al.* (8) bridged cognitive symptomatology, socio-affective brain network functionality, and the dopamine and serotonin systems.

However, caution and careful evaluations are needed to prevent overfitting in a finite amount of data (25) when the optimization objectives are related to the feature-selection criteria. This becomes particularly relevant as for different modalities, variable levels of overfitting could occur due to the size of the respective feature sets. Hence, evaluations using independent test samples are particularly relevant to avoid leakage, circularity and inflated performance (26). Under strict evaluation, a different number of features across modalities per se should not impair comparability. However, *complexities such as* interactions and nonlinearities among candidate features *continue to pose challenges to the identification of a small number of predictive features with explanatory value, especially in* multi-view scenarios. Consequently, different forms of regularization have been utilized for feature selection and fusion, although there is not yet a consensus on the best approach beyond recommendations of careful design and skeptical exploration (35-37).

Overall, leveraging feature selection methods in a stringent way and incorporating clean external validation may pave the way toward unraveling robust and generalizable neurobiological underpinnings of psychiatric illness (Figure 1E). Multi-view perspectives might be particularly necessary and promising to provide corroborative and complementary molecular information to MRI findings and may thus provide important leads for linking across multiple levels of description. Given the growing number of datasets that allow the type of investigations covered (such as the Brain Genomics Superstruct Project and the UK Biobank), they are now readily available, as well as approaches rapidly developed in the field that are widely implementable to fuse multiscale data. However, caution is warranted given the ensuing size and complexity of feature sets vis-à-vis the still finite amount of available cases.

## 2- Accuracies to inform nosological relationships between disorders

Current nosological systems such as the DSM-5 and the ICD-10 represent core pillars of psychiatric diagnosis and by extension research into the neurobiology of mental illness. However, it needs to be remembered that in addition to being based on patients' self-report and clinicians' assessment, inherently limiting objectivity and hence reliability, these systems fundamentally represent historically derived heuristics rather than well-defined biological classifications. This leads to two important challenges. First, diagnostic criteria for a disorder such as depression or schizophrenia comprise a range of symptoms out of which a certain number need to be present – often separated into core and accessory ones – leading to a large number of possible symptom-combinations resulting in the same diagnosis (38,39). This can be problematic as specific symptoms are likely causally linked and differ in the underlying neurobiology (40,41). Second, the current classification systems yield high rates of comorbidity between psychiatric disorders (42), potentially attributable to the misalignment between biological groupings and nosological entities. Corroborating this conjecture is a large body of literature indicating a substantial overlap across diagnostic groups in aspects ranging from genetics and molecular features to brain atrophy and network abnormalities (43-45). This poses a major challenge to differential diagnoses based on machine-learning classifiers as indicated by relatively poor performance even in cases where subjects were optimistically selected to be free of comorbidities. For example, a recent study applying state-of-the-art deep learning to differentiate patients with schizophrenia, schizoaffective and bipolar disorders (BDs) from each other and healthy controls yielded only a moderate overall accuracy of 46% (21). However, it also opens the door toward systematic investigations into relationships between current categories using classification accuracies as the dependent variable and potentially the redefinition of disease entities.

For example, some diagnostic categories are more easily “confused” on the neurobiological level than others, and this relationship does not necessarily follow clinical similarities. One particularly interesting case in this context is BD (Figure 2A). Patients with BD are almost indistinguishable from those with major depression (46) or schizophrenia (47) based on structural MRI metrics or connectivity patterns, as in both cases classification performance was below 60%. However, patients with major depression are much more robustly differentiated from those with schizophrenia based on structural MRI (76% accuracy)(22), and classification based on symptomatology significantly differentiates BD from unipolar depression despite a lack of differentiating MRI features in the same sample (48).

Reversing the conventional perspective on machine-learning, the inverse of these accuracy measures may hence be used as a proxy for the biological relationships between clinical labels. Indeed, there is converging evidence for genetic, molecular and physiological overlap of BD with both major depression and schizophrenia (28,49-51). Nevertheless, it bears mention that despite the shared biology, pharmacological therapy for BD usually requires a combination between antipsychotic medication akin to the therapy of schizophrenia and mood-stabilizers, i.e., drugs that are not commonly used for either schizophrenia or depression, while antidepressive (mono-) therapy is not recommended. Although this contrasts with the suggested intermediate position of BD, it resonates well with a recent out-of-category assessment (34). The idea behind this approach is to train a model to differentiate a group of patients (here, schizophrenia patients) from healthy controls and then apply that model to patients with other diagnoses under the idea that the distance of these patients from the binary decision boundary would reflect the degree to which these patients show an overlap with the "schizophrenia-signature" captured by the model. Importantly, among major psychiatric disorders, BD showed by far the highest similarity in (striatal) dysfunction with schizophrenia (34). Alternatively, we can assess the proportion of patients assigned to (different) categories used for classifier training. For example, individuals with generalized anxiety disorder (GAD, 69%) were much more likely than schizophrenia patients (<10%) to be labeled depressed by a classifier trained for depression diagnosis (52).

Despite the exciting potential of these approaches to unravel nosological relationships, we also need to present words of caution that extend beyond the indispensable requirement for unbiased methodology and careful out-of-sample validation. These relate to confounding effects that can easily be of the same magnitude as the differences under investigation. To illustrate this point, classification accuracies for male vs. female (53), young vs. older (54,55) or subjects from different scanners (56,57) easily exceed those for diagnostic classifiers. Unfortunately, these factors may co-vary with diagnoses due to aspects such as likely age of onset, sex distribution, and differential relationships with additional health issues, including drug abuse. In addition, sampling biases may be introduced by the recruitment of different populations in different hospitals and hence catchment areas. In contrast to univariate analyses where covariates of no interest can rather effectively be regressed out, treatment of confounding effects in multivariate settings is non-trivial (58,59,60).

Overall, we argue that the outlined perspective of mapping neurobiological similarity via classification performance together with the concept of identifying core features may represent a key avenue for redefining psychiatric nosology into neurobiologically grounded and hence actionable categories. Nonetheless, care must be taken to ensure that these indeed show added value over the likely crude but ultimately helpful clinical heuristics currently in use.

## 3-Data-driven consolidation to disentangle heterogeneous disorders and redefine clinical entities

Clinically or theoretically defined categories and their subtypes (61) have received substantial criticism for their poor diagnostic stability, validity, and utility (62,63,64). The need for defining subtypes inherently relates to the heterogeneity within a disorder, i.e., inter-individual variability in clinical phenotypes but likely also neurobiology (65,66). When the strategy outlined in the previous section was applied to clinically defined subtypes (23,67), it revealed accuracies similar to or exceeding those for cross-disorder classification (Figure 2B).

While supervised approaches may thus provide a more objective evaluation of the distinctiveness and relationship among clinically defined subtypes, an alternative avenue that has received much attention over the past years is the re-definition of disease subtypes in a data-driven manner using un-supervised approaches (68,69). The key idea behind these approaches is to algorithmically define subgroups within a large set of presumably heterogeneous patients (Figure 2C), aiming at finding latent or hidden structures based on individual-level features. In clustering approaches, this structure is binary-disjoint, i.e., subjects are each assigned into (only) one group, while groups are as homogeneous and distinct as possible. Alternatively, the structure could be continuous-overlapping, in which case the variance within the feature space is explained by a low-dimensional set of variables and each individual is then represented by how strongly their features load onto these dimensions (factorization). In any case, it needs to be remembered, although, that virtually all algorithms will find differences within the data at hand, even in the absence of natural subtypes or dimensions. This again highlights the need for assessing generalization in new data (70), as incidental patterns are unlikely to extend to new data whereas true subtypes should also remain discernable.

The probably best-studied application for these methods is schizophrenia, where both factorization and clustering methods have a long tradition (Figure 2c). Most of this literature converges on a distinction between positive-psychotic symptoms and negative-cognitive affections, a differentiation that may be found both in terms of continuous dimensions of psychopathology (71,72) as well as clustering (15,73-77). However, there are also reports that indicate alternative modes of variation, such as subtypes differing by structural covariance features of subcortical regions, posterior orbitofrontal, superior temporal and occipital gyri as well as anxious-depressed symptoms (78). This work resonates with factor models of schizophrenia psychopathology indicating more than two dimensions (15,79), raising the question of whether the positive vs negative distinction may be most salient but incomplete (80,81). Subtyping has likewise received substantial attention in major depression, particularly through the seminal study by Drysdale *et al.* (71), which proposed four “biotypes” based on rsFC patterns, symptom dimensions (Figure 2D) and differential responses to transcranial magnetic stimulation. A follow-up evaluation, however, casts doubts over this four-biotype differentiation (82), and other studies using similar albeit smaller samples proposed two (83-85) or three (86) subtypes.

More work is certainly needed to identify robust subtypes within diagnostic categories and characterize their clinical relevance such as differences in prognosis or therapy response. However, the probably even bigger challenge is to apply the approaches mentioned here to heterogeneous samples, i.e., across diagnoses. As outlined, traditional diagnostic boundaries are sometimes ambiguous. Overlap between diagnoses in terms of clinical and neurobiological features thus prompted transdiagnostic initiatives such as the National Institute of Mental Health’s “Research Domain Criteria” framework (87). In this context, data-driven approaches may offer a perspective for a re-definition of disease entities along biological categories. Some early work in this respect provides interesting leads. Clustering on a cohort of patients with major depression and BD based on cortical thickness measures revealed two new divisions that differ in the proportions of BD II and major depression diagnoses and positive family history of psychiatric disorders (48). When trying to separate a pool of patients diagnosed with schizophrenia, BD or major depression by patterns of functional imbalance between frontal and posterior brain regions, only two divisions emerged (76). Patients may also be represented by transdiagnostic dimensions of neurobiology in relation to low-rank psychopathological components through sub-space learning methods such as CCA. A recent study jointly analyzed healthy individuals and individuals with BD, ADHD, schizophrenia and schizoaffective disorder and identified three psychopathological dimensions related to dissociable functional connectivity signatures (88). Similarly, dissociable functional connectivity signatures for dimensions of psychopathology have been reported in an adolescent cohort (89), highlighting that connectivity patterns may relate to psychopathology in a specific yet diagnosis-overarching manner. An extensive overview of the application and caveats of categorical vs. dimensional approaches as well as single-view vs. multi-view approaches can be found in a recent review (90). Systematic overviews of subtyping ADHD and ASD, including a hybrid dimensional subtyping approach to allow each patient to associate with one or more categories to varying degrees (91) (Figure 2E), as well as details in the re-division of schizophrenia-spectrum, ASD and BD cohorts (92-95) may be found in a special issue of Biological Psychiatry on the topic of "Convergence and Heterogeneity in Psychopathology" (https://www.biologicalpsychiatryjournal.com/issue/S0006-3223(19)X0013-X).

Data-driven methods promise a high potential for a redefinition of psychiatric classification and ultimately care. However, they can only fulfill this of the ensuing patterns and are both robust and valid, which requires large samples and stringent validation in independent samples, as well as very careful handling of disease-immanent and practical confounders (96). Semi-supervised approaches such as HYDRA maximize the distance between subtypes and control subjects in a cross-validated procedure while trying to account for covariates (97). HYDRA may hence overcome the issue of subtyping by non-relevant demographic and clinical diversity (98,76). Likewise**,** hybrid approaches provided by normative modeling (99-101) and functional random forest (102) have provided interesting hypotheses on ASD and schizophrenia spectrum heterogeneity (65,103,104). Once established, however, the critical question then pertains to the clinical utility of the ensuing categories relative to the traditional and well-tested heuristics manifested in today’s diagnostic labels.

## Limitations and Concerns

Both the approaches outlined here and the more prevailing classification setup suffer from the “garbage-in, garbage-out” problem, a longstanding aphorism in computer science specifying unreliable outputs from flawed input data. These include but are likely not limited to the aspects summarized below:

**- Limits of phenotypical data**

Traditional rating forms may provide only coarse measures of behavior, encounter adherence issues and yield limited real-life validity. This has prompted research to improve individualized quantification of symptomatology and behaviors like the development of Extended Strengths and Weaknesses Assessment of Normal Behavior (E-SWAN). These may provide more reliable measures (105) to better characterize data-driven subtypes and help to maximize predictive accuracy (106,107). A more recent focus is digital phenotyping leveraging web, social media and mobile monitoring devices to obtain rich phenotypes continuously, objectively and passively through large-scale data collection (108,109). Digital phenotypes are likely to be clinically relevant and indicative of psychiatric syndromes (110-112).

**-Biased image quality**

Neuroimaging data may be degraded by factors such as distortions, poor signal-to-noise ratios and (particularly) excessive within-scanner head motion (113,114). These introduce artifacts to neuroimaging features that may then drive predictive algorithms (115-118). This is particularly critical if image quality is correlated with disease burden, diagnoses and other outcomes of interest (119,116). Unfortunately, this seems to be the case, as patients and among them in particular those with psychotic and manic symptoms systematically trend to differ from healthy controls in within-scanner movement, providing a hard-to-overcome confound.

**-Missing data and state assessment**

Missing data are pervasive in large-scale data but in particular in clinical settings where data are usually not missing at random (120,121). For example, more severely ill patients may not be able to compete parts of the assessment fail data-quality control (116) leading to missing data that are unevenly distributed among groups. While a majority of current psychiatric machine learning studies aim to identify stable markers capturing trait effects in cross-sectional data, we would argue that state-related effects, in particular of acute illness vs remission or chronic phases would be of greatest interest and promise for biomarker and/or therapeutic development (122,123). However, access to recovered patients living in their community setting as opposed to being inpatients in a university hospital, is often limited.

### -Overfitting and researcher solutions

Flawed analytic procedures can accidentally or artificially occur due to heedless or intended analysis optimization, leading to overfitting (7,26,125). While setting aside parts of the data from the whole process as a "lock-box" could effectively mitigate this (124), fresh holdout datasets are not always attainable, tempting researchers to access the "lock-box" multiple times, which can be problematic (though see (126)). Ultimately, though, pre-registration (127,128) following clinical trial standards where the trained model is deposited with a third party (e.g., a clinical trial center) before a single test-dataset is acquired would be needed before real-life use.

### -Sample size and effect size issues

Small samples often produce inflated effect sizes that are not replicable in independent datasets (129,130). Thus, it seems likely that thousands of subjects are needed to attain reliable associations (130) and sufficiently small cross-validation error bars (131,132). This will obviously pose challenges in clinical settings. Hence strategies such as meta-matching (133), which leverages very large cohort datasets to only require a more manageable number of clinical cases for fine-tuning, may represent a promising avenue.

### -Considerations for clinical translation

Which effect size or accuracy should be considered clinically useful remains to be discussed and is likely negotiated between the different stakeholders such as patients, practitioners and insurances. One critical aspect in this discussion and related legal and ethical considerations (134-136) is the lack of explainability in most machine-learning models (137-139). In this context, we note that this "black box" problem may moreover pose challenges to the fundamental idea of informed consent, if the treating physician cannot explain evidence for a clinical decision and the patient is hence forced into blind trust.

## SUMMARY AND CONCLUSIONS

Here, we outlined emerging trends to leverage machine-learning for neurobiological and nosological insights. We reviewed and summarized these new approaches using machine-learning accuracies as dependent variables to inform biological features related to psychopathology, diagnosis, and nosology. We also pointed to multi-view machine-learning to combine data from different sources for bridging molecular and system-level information. Finally, we highlighted unsupervised and semi-supervised approaches to disentangle psychopathological-neurobiological relationships within a diagnostic group or from a transdiagnostic perspective. While we expect these approaches to gather more attention in the future, we also highlighted the caveats and pitfalls that need to be overcome before machine-learning may contribute to biomarkers, pathophysiological understanding and ultimately precision psychiatry.

**Acknowledgments:**

This study was supported by Grant No. 2022ZD0214000 [to JC], the National Key R&D Program of China (Grant No. 2021YFC2502200 [to JC]), Deutsche Forschungsgemeinschaft (Grant No. EI 816/4-1 [to SBE]), the National Institute of Mental Health (Grant No. R01-MH074457 [to SBE]), the Helmholtz Portfolio Theme "Supercomputing and Modeling for the Human Brain," and the European Union's Horizon 2020 Research and Innovation Programme (Grant Nos. 785907 [HBP SGA2] [to SBE] and 945539 [HBP SGA3] [to SBE]).

**Disclosures:**

The authors reported no biomedical financial interests or potential conflicts of interest.

## References:

1. Obermeyer Z, Emanuel EJ (2016): Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 375:1216-1219.
2. Lee CS, Lee AY (2020): Clinical applications of continual learning machine learning. Lancet Digit Health 2:e279-e281.
3. Jordan MI, Mitchell TM (2015): Machine learning: Trends, perspectives, and prospects. Science 349:255-260.
4. Pellegrini E, Ballerini L, Hernandez M, Chappell FM, Gonzalez-Castro V, Anblagan D, et al. (2018): Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. Alzheimers Dement (Amst) 10:519-535.
5. Myszczynska MA, Ojamies PN, Lacoste AMB, Neil D, Saffari A, Mead R, et al. (2020): Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat Rev Neurol 16:440-456.
6. Rashid B, Calhoun V (2020): Towards a brain-based predictome of mental illness. Hum Brain Mapp 41:3468-3535.
7. Arbabshirani MR, Plis S, Sui J, Calhoun VD (2017): Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage 145:137-165.
8. Chen J, Muller VI, Dukart J, Hoffstaedter F, Baker JT, Holmes AJ, et al. (2021): Intrinsic Connectivity Patterns of Task-Defined Brain Networks Allow Individual Prediction of Cognitive Symptom Dimension of Schizophrenia and Are Linked to Molecular Architecture. Biol Psychiatry 89:308-319.
9. Baker JT, Dillon DG, Patrick LM, Roffman JL, Brady RO, Jr., Pizzagalli DA, et al. (2019): Functional connectomics of affective and psychotic pathology. Proc Natl Acad Sci U S A 116:9050-9059.
10. Nazeri A, Schifani C, Anderson JAE, Ameis SH, Voineskos AN (2020): In Vivo Imaging of Gray Matter Microstructure in Major Psychiatric Disorders: Opportunities for Clinical Translation. Biol Psychiatry Cogn Neurosci Neuroimaging 5:855-864.
11. Schur RR, Draisma LW, Wijnen JP, Boks MP, Koevoets MG, Joels M, et al. (2016): Brain GABA levels across psychiatric disorders: A systematic literature review and

meta-analysis of (1) H-MRS studies. Hum Brain Mapp 37:3337-3352.

12. Gottesman II, Gould TD (2003): The endophenotype concept in psychiatry: Etymology and strategic intentions. Am J Psychiatry 160:636–645.
13. Zhang YD, Dong Z, Wang SH, Yu X, Yao X, Zhou Q, et al. (2020): Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. Inf Fusion 64:149-187.
14. Marx V (2019): Machine learning, practically speaking. Nat Methods 16:463-467.
15. Chen J, Patil KR, Weis S, Sim K, Nickl-Jockschat T, Zhou J, et al. (2020): Neurobiological Divergence of the Positive and Negative Schizophrenia Subtypes Identified on a New Factor Structure of Psychopathology Using Non-negative Factorization: An International Machine Learning Study. Biol Psychiatry 87:282-293
16. Yan WZ, Calhoun V, Song M, Cui Y, Yan H, Liu SF, et al. (2019): Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data. Ebiomedicine 47:543-552.
17. Kohavi R, John GH (1997): Wrappers for feature subset selection. Artif Intell 97:273-324.
18. Guyon I, Elisseeff A (2003): An introduction to variable and feature selection. J Mach Learn Res 3:1157-1182.
19. Moghimi P, Lim KO, Netoff TI (2018): Data Driven Classification Using fMRI Network Measures: Application to Schizophrenia. Front Neuroinform 12.
20. Wu Y, Ren P, Chen R, Xu H, Xu J, Zeng L, et al. (2022): Detection of functional and structural brain alterations in female schizophrenia using elastic net logistic regression. Brain Imaging Behav 16:281-290.
21. Yan W, Zhao M, Fu Z, Pearlson GD, Sui J, Calhoun VD (2021): Mapping relationships among schizophrenia, bipolar and schizoaffective disorders: A deep classification and clustering framework using fMRI time series. Schizophr Res 245:141-150.
22. Koutsouleris N, Meisenzahl EM, Borgwardt S, Riecher-Rossler A, Frodl T, Kambeitz J, et al. (2015): Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. Brain 138:2059-2073.

23. Qureshi MNI, Oh J, Min B, Jo HJ, Lee B (2017): Multi-modal, Multi-measure, and Multi-class Discrimination of ADHD with Hierarchical Feature Extraction and Extreme Learning Machine Using Structural and Functional Brain MRI. Front Hum Neurosci 11:157.

24. Lebedeva AK, Westman E, Borza T, Beyer MK, Engedal K, Aarsland D, et al. (2017): MRI-Based Classification Models in Prediction of Mild Cognitive Impairment and Dementia in Late-Life Depression. Front Aging Neurosci 9:13.

25. Loughrey J, Cunningham P (2004): Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets. International conference on innovative techniques and applications of artificial intelligence, pp 33-43.

26. Poldrack RA, Huckins G, Varoquaux G (2020): Establishment of Best Practices for Evidence for Prediction: A Review. JAMA Psychiatry 77:534-540.

27. Price JL, Drevets WC (2010): Neurocircuitry of mood disorders. Neuropsychopharmacology 35:192-216.

28. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. (2018): Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science 359:693-697.

29. Anderson KM, Collins MA, Chin R, Ge T, Rosenberg MD, Holmes AJ (2020): Transcriptional and imaging-genetic association of cortical interneurons, brain function, and schizophrenia risk. Nat Commun 11:2889.

30. Anderson KM, Collins MA, Kong R, Fang K, Li J, He T, et al. (2020): Convergent molecular, cellular, and cortical neuroimaging signatures of major depressive disorder. Proc Natl Acad Sci U S A 117:25138-25149.

31. Lu PX, Li X, Hu LT, Lu L (2021): Integrating genomic and resting State fMRI for efficient autism spectrum disorder classification. Multimed Tools Appl

32. De la Torre F, Black MJ (2003): A framework for robust subspace learning. Int J Comput Vis 54:117-142.

33. Wang HT, Smallwood J, Mourao-Miranda J, Xia CH, Satterthwaite TD, Bassett DS, et al.

(2020): Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. Neuroimage 216:116745.

34. Li A, Zalesky A, Yue W, Howes O, Yan H, Liu Y, et al. (2020): A neuroimaging biomarker for striatal dysfunction in schizophrenia. Nat Med 26:558-565.

35. Zhang YP, Wang SH, Xia KJ, Jiang YZ, Qian PJ, Initia ADN (2021): Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. Inf Fusion 66:170-183.

36. Peng J, An L, Zhu X, Jin Y, Shen D (2016): Structured Sparse Kernel Learning for Imaging Genetics Based Alzheimer's Disease Diagnosis. Med Image Comput Comput Assist Interv 9901:70-78.

37. Szafranski M, Grandvalet Y, Rakotomamonjy A (2010): Composite kernel learning. Mach Learn 79:73-103.

38 Olbert CM, Gala GJ, Tupler LA (2014): Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. J Abnorm Psychol 123:452-462.

39. Park SC, Kim JM, Jun TY, Lee MS, Kim JB, Yim HW, et al. (2017): How many different symptom combinations fulfil the diagnostic criteria for major depressive disorder? Results from the CRESCEND study. Nord J Psychiatry 71:217-222.

40. Fried EI, Nesse RM (2015): Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med 13:72.

41. Fried EI, van Borkulo CD, Cramer AOJ, Boschloo L, Schoevers RA, Borsboom D (2017): Mental disorders as networks of problems: a review of recent insights. Soc Psychiatry Psychiatr Epidemiol 52:1-10.

42. Jacobi F, Wittchen HU, Holting C, Hofler M, Pfister H, Muller N, et al. (2004): Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). Psychol Med 34:597-611.

43. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. (2013): Genetic

relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45:984.

44. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. (2015): Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry 72:305-315.
45. McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A (2017): Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. Am J Psychiatry 174:676-685.
46. Sacchet MD, Prasad G, Foland-Ross LC, Thompson PM, Gotlib IH (2015): Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. Front Psychiatry 6:21.
47. Rashid B, Arbabshirani MR, Damaraju E, Cetin MS, Miller R, Pearlson GD, et al. (2016): Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. Neuroimage 134:645-657.
48. Yang T, Frangou S, Lam RW, Huang J, Su Y, Zhao G, et al. (2021): Probing the clinical and brain structural boundaries of bipolar and major depressive disorder. Transl Psychiatry 11:48.
49. Green EK, Grozeva D, Jones I, Jones L, Kirov G, Caesar S, et al. (2010): The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. Mol Psychiatry 15:1016-1022.
50. Schulze TG, Akula N, Breuer R, Steele J, Nalls MA, Singleton AB, et al. (2014): Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. World J Biol Psychiatry 15:200-208.
51. McIntosh AM, Sullivan PF, Lewis CM (2019): Uncovering the Genetic Architecture of Major Depression. Neuron 102:91-103.
52. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. (2017): Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med 23:28-38.

53. Weis S, Patil KR, Hoffstaedter F, Nostro A, Yeo BTT, Eickhoff SB (2020): Sex Classification by Resting State Brain Connectivity. Cereb Cortex 30:824-835.

54. Cropley VL, Tian Y, Fernando K, Mansour LS, Pantelis C, Cocchi L, et al. (2021): Brain-Predicted Age Associates With Psychopathology Dimensions in Youths. Biol Psychiatry Cogn Neurosci Neuroimaging 6:410-419.

55. Plaschke RN, Patil KR, Cieslik EC, Nostro AD, Varikuti DP, Plachti A, et al. (2020): Age differences in predicting working memory performance from network-based functional connectivity. Cortex 132:441-459.

56. Chen AA, Beer JC, Tustison NJ, Cook PA, Shinohara RT, Shou H, et al. (2022): Mitigating site effects in covariance for machine learning in neuroimaging data. Hum Brain Mapp 43:1179-1195.

57. Pomponio R, Erus G, Habes M, Doshi J, Srinivasan D, Mamourian E, et al. (2020): Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage 208:116-450.

58. Snoek L, Miletic S, Scholte HS (2019): How to control for confounds in decoding analyses of neuroimaging data. Neuroimage 184:741-760.

59. More S, Eickhoff SB, Caspers J, Patil KR (2021): Confound Removal and Normalization in Practice: A Neuroimaging Based Sex Prediction Case Study. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp 3-18.

60. Dinga R, Schmaal L, Penninx BWJH, Veltman DJ, Marquand AF (2020): Controlling for effects of confounding variables on machine learning predictions. bioRxiv:2020.2008.2017.255034.

61. Fenton WS, McGlashan TH (1991): Natural history of schizophrenia subtypes. I. Longitudinal study of paranoid, hebephrenic, and undifferentiated schizophrenia. Arch Gen Psychiatry 48:969-977.

62. Braff DL, Ryan J, Rissling AJ, Carpenter WT (2013): Lack of use in the literature from the last 20 years supports dropping traditional schizophrenia subtypes from DSM-5 and ICD-11. Schizophr Bull 39:751-753.63.

63. Angst J, Sellaro R, Merikangas KR (2000): Depressive spectrum diagnoses. Compr Psychiatry 41:39-47.

64. Melartin T, Leskela U, Rytsala H, Sokero P, Lestela-Mielonen P, Isometsa E (2004): Co-morbidity and stability of melancholic features in DSM-IV major depressive disorder. Psychol Med 34:1443-1452.

65. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA (2019): The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. Trends Cogn Sci 23:584-601.

66. Kirkpatrick B, Buchanan RW, Ross DE, Carpenter WT (2001): A separate disease within the syndrome of schizophrenia. Arch Gen Psychiatry 58:165-171.

67. Nicholson AA, Densmore M, McKinnon MC, Neufeld RWJ, Frewen PA, Theberge J, et al. (2019): Machine learning multivariate pattern analysis predicts classification of posttraumatic stress disorder and its dissociative subtype: a multimodal neuroimaging approach. Psychol Med 49:2049-2059.

68. Chapelle O, Scholkopf B, Zien A (2009): Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Trans Neural Netw, 20:542.

69. Dike HU, Zhou YM, Deveerasetty KK, Wu QT (2018): Unsupervised Learning Based On Artificial Neural Network: A Review. 2018 Ieee International Conference on Cyborg and Bionic Systems (Cbs), pp 322-327.

70. Lombardo MV, Lai MC, Baron-Cohen S (2019): Big data approaches to decomposing heterogeneity across the autism spectrum. Mol Psychiatry 24:1435-1450.

71. Kirschner M, Shafiei G, Markello RD, Makowski C, Talpalaru A, Hodzic-Santor B, et al. (2020): Latent Clinical-Anatomical Dimensions of Schizophrenia. Schizophr Bull 46:1426-1438.

72. Levine SZ, Rabinowitz J (2007): Revisiting the 5 dimensions of the Positive and Negative Syndrome Scale. J Clin Psychopharmacol 27:431-436.

73. Dollfus S, Everitt B, Ribeyre JM, Assouly-Besse F, Sharp C, Petit M (1996): Identifying subtypes of schizophrenia by cluster analyses. Schizophr Bull 22:545-555.

74. Lykouras L, Oulis P, Daskalopoulou E, Psarros K, Christodoulou GN (2001): Clinical subtypes of schizophrenic disorders: a cluster analytic study. Psychopathology 34:23-28.

75. Rahaman MA, Turner JA, Gupta CN, Rachakonda S, Chen J, Liu J, et al. (2020): N-BiC: A Method for Multi-Component and Symptom Biclustering of Structural MRI Data: Application to Schizophrenia. IEEE Trans Biomed Eng 67:110-121.

76. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, et al. (2020): Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain 143:1027-1038.

77. Sun H, Lui S, Yao L, Deng W, Xiao, Y., Zhang, W., et al. (2015): Two patterns of white matter abnormalities in medication-naive patients with first-episode schizophrenia revealed by diffusion tensor imaging and cluster analysis. JAMA Psychiatry, 72: 678-686.

78. Liu ZW, Palaniyappan L, Wu XR, Zhang K, Du JN, Zhao Q, et al. (2021): Resolving heterogeneity in schizophrenia through a novel systems approach to brain structure: individualized structural covariance network analysis. Mol Psychiatry, 26:7719–7731.

79. Wallwork RS, Fortgang R, Hashimoto R, Weinberger DR, Dickinson D (2012): Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. Schizophr Res 137:246-250.

80. Xiao Y, Liao W, Long Z, Tao B, Zhao Q, Luo C, et al. (2022): Subtyping Schizophrenia Patients Based on Patterns of Structural Brain Alterations. Schizophr Bull 48:241-250.

81. Di Biase MA, Geaghan MP, Reay WR, Seidlitz J, Weickert CS, Pébay A, et al. (2022): Cell type-specific manifestations of cortical thickness heterogeneity in schizophrenia. Molecular Psychiatry, In Press: https://doi.org/10.1038/s41380-022-01460-7.

82. Dinga R, Schmaal L, Penninx B, van Tol MJ, Veltman DJ, van Velzen L, et al. (2019): Evaluating the evidence for biotypes of depression: Methodological replication and extension of. Neuroimage Clin 22:101796.

83. Price RB, Gates K, Kraynak TE, Thase ME, Siegle GJ (2017): Data-Driven Subgroups in Depression Derived from Directed Functional Connectivity Paths at Rest.

Neuropsychopharmacology 42:2623-2632.

84. Price RB, Lane S, Gates K, Kraynak TE, Horner MS, Thase ME, et al. (2017): Parsing Heterogeneity in the Brain Connectivity of Depressed and Healthy Adults During Positive Mood. Biol Psychiatry 81:347-357.
85. Feder S, Sundermann B, Wersching H, Teuber A, Kugel H, Teismann H, et al. (2017): Sample heterogeneity in unipolar depression as assessed by functional connectivity analyses is dominated by general disease effects. J Affect Disord 222:79-87.
86. Tokuda T, Yoshimoto J, Shimizu Y, Okada G, Takamura M, Okamoto Y, et al. (2018): Identification of depression subtypes and relevant brain regions using a data-driven approach. Sci Rep 8:14082.
87. Cuthbert BN (2014): The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 13:28-35.
88. Kebets V, Holmes AJ, Orban C, Tang S, Li J, Sun N, et al. (2019): Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. Biol Psychiatry 86:779-791.
89. Xia CH, Ma Z, Ciric R, Gu S, Betzel RF, Kaczkurkin AN, et al. (2018): Linked dimensions of psychopathology and connectivity in functional brain networks. Nat Commun 9:3003.
90. Kaczkurkin AN, Moore TM, Sotiras A, Xia CH, Shinohara RT, Satterthwaite TD (2020): Approaches to Defining Common and Dissociable Neurobiological Deficits Associated With Psychopathology in Youth. Biol Psychiatry 88:51-62.
91. Tang S, Sun N, Floris DL, Zhang X, Di Martino A, Yeo BTT (2020): Reconciling Dimensional and Categorical Models of Autism Heterogeneity: A Brain Connectomics and Behavioral Study. Biol Psychiatry 87:1071-1082.
92. Dias TGC, Iyer SP, Carpenter SD, Cary RP, Wilson VB, Mitchell SH, et al. (2015): Characterizing heterogeneity in children with and without ADHD based on reward system connectivity. Dev Cogn Neurosci 11:155-174.

93. Karalunas SL, Nigg JT (2020): Heterogeneity and Subtyping in Attention-Deficit/Hyperactivity Disorder-Considerations for Emerging Research Using Person-Centered Computational Approaches. Biol Psychiatry 88:103-110.

94. Clementz BA, Sweeney JA, Hamm JP, Ivleva EI, Ethridge LE, Pearlson GD, et al. (2016): Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. Am J Psychiatry 173:373-384.

95. Stefanik L, Erdman L, Ameis SH, Foussias G, Mulsant BH, Behdinan T, et al. (2018): Brain-Behavior Participant Similarity Networks among Youth and Emerging Adults with Schizophrenia Spectrum, Autism Spectrum, or Bipolar Disorder and Matched Controls. Neuropsychopharmacology. 43: 1180–1188

96. Snoek L, Miletić S, Scholte HS. (2019): How to control for confounds in decoding analyses of neuroimaging data. Neuroimage 184:741-760.

97. Varol E, Sotiras A, Davatzikos C, Neuroimaging AD (2017): HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage 145:346-364.

98. Baller EB, Kaczkurkin AN, Sotiras A, Adebimpe A, Bassett DS, Calkins ME, et al. (2021): Neurocognitive and functional heterogeneity in depressed youth. Neuropsychopharmacology 46:783-790.

99. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF (2019): Conceptualizing mental disorders as deviations from normative functioning (vol 24, pg 1415, 2019). Mol Psychiatry 24:1565-1565.

100. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. (2019): Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. Biol Psychiatry Cogn Neurosci Neuroimaging 4:567-578.

101. Wolfers T, Doan NT, Kaufmann T, Alnaes D, Moberget T, Agartz I, et al. (2018): Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. JAMA Psychiatry 75:1146-1155.

102. Voineskos AN, Jacobs GR, Ameis SH (2020): Neuroimaging Heterogeneity in Psychosis:

Neurobiological Underpinnings and Opportunities for Prognostic and Therapeutic Innovation. Biol Psychiatry 88:95-102.

103. Feczko E, Balba NM, Miranda-Dominguez O, Cordova M, Karalunas SL, Irwin L, et al. (2018): Subtyping cognitive profiles in Autism Spectrum Disorder using a Functional Random Forest algorithm. Neuroimage 172:674-688.

104. Feczko E, Fair DA (2020): Methods and Challenges for Assessing Heterogeneity. Biol Psychiatry 88:9-17.

105. Alexander LM, Salum GA, Swanson JM, Milham MP (2020): Measuring strengths and weaknesses in dimensional psychiatry. J Child Psychol Psychiatry 61:40-50.

106. Genon S, Bernhardt BC, La Joie R, Amunts K, Eickhoff SB (2021): The many dimensions of human hippocampal organization and (dys)function. Trends Neurosci 44:977-989.

107. Genon S, Reid A, Langner R, Amunts K, Eickhoff SB (2018): How to Characterize the Function of a Brain Region. Trends Cogn Sci 22:350-364.

108. Ressler KJ, Williams LM (2021): Big data in psychiatry: multiomics, neuroimaging, computational modeling, and digital phenotyping. Neuropsychopharmacology 46:1-2.

109. Dunster GP, Swendsen J, Merikangas KR (2021): Real-time mobile monitoring of bipolar disorder: a review of evidence and future directions. Neuropsychopharmacology 46:197-208.

110. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC (2016): The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 4.

111. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. (2015): Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 1:15030.

112. Insel TR (2017): Digital Phenotyping: Technology for a New Science of Behavior. JAMA 318:1215-1216.

113. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012): Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142-2154.

114. Alexander-Bloch A, Clasen L, Stockman M, Ronan L, Lalonde F, Giedd J, et al. (2016): Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. Hum Brain Mapp 37:2385-2397.

115. Makowski C, Lepage M, Evans AC (2019): Head motion: the dirty little secret of neuroimaging in psychiatry. J Psychiatry Neurosci 44:62-68.

116. Power JD, Schlaggar BL, Petersen SE (2015): Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105:536-551.

117. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014): Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84:320-341.

118. Krause F, Benjannins C, Eck J, Luhrs M, van Hoof R, Goebel R (2019): Active head motion reduction in magnetic resonance imaging using tactile feedback. Hum Brain Mapp 40:4026-4037.

119. Yao N, Winkler AM, Barrett J, Book GA, Beetham T, Horseman R, et al. (2017): Inferring pathobiology from structural MRI in schizophrenia and bipolar disorder: Modeling head motion and neuroanatomical specificity. Hum Brain Mapp 38:3757-3770.

120. Thomas RM, Bruin W, Zhutovsky P, van Wingen G (2020): Chapter 14 - Dealing with missing data, small sample sizes, and heterogeneity in machine learning studies of brain disorders. In: Mechelli A, Vieira S, editors. Machine Learning. Academic Press, pp 249-266.

121. Le Morvan M, Josse J, Scornet E, Varoquaux G (2021): What'sa good imputation to predict with missing values? Adv Neural Inf Process Syst 34:11530-11540.

122. Wang Y, Gao Y, Tang S, Lu L, Zhang L, Bu X, et al. (2020): Large-scale network dysfunction in the acute state compared to the remitted state of bipolar disorder: A meta-analysis of resting-state functional connectivity. EBioMedicine 54:102742.

123. da Silva RA, Tancini MB, Lage R, Nascimento RL, Santana CMT, Landeira-Fernandez J, et al. (2021): Autobiographical Memory and Episodic Specificity Across Different Affective States in Bipolar Disorder. Front Psychiatry 12:641221.

124. Hosseini M, Powell M, Collins J, Callahan-Flintoft C, Jones W, Bowman H, et al. (2020): I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data. Neurosci Biobehav Rev 119:456-467.
125. Varoquaux G, Cheplygina V (2022): Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digit Med 5:48.
126. Dwork C, Feldman V, Hardt M, Pitassi T, Reingold O, Roth A (2015): STATISTICS. The reusable holdout: Preserving validity in adaptive data analysis. Science 349:636-638.
127. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018): The preregistration revolution. Proc Natl Acad Sci U S A 115:2600-2606.
128. Chambers CD, Forstmann B, Pruszynski JA (2017): Registered reports at the European Journal of Neuroscience: consolidating and extending peer-reviewed study pre-registration. Eur J Neurosci 45:627-628.
129. Abraham A, Milham MP, Di Martino A, Craddock RC, Samaras D, Thirion B, et al. (2017): Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. Neuroimage 147:736-745.
130. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. (2022): Reproducible brain-wide association studies require thousands of individuals. Nature 603:654-660.
131. Flint C, Cearns M, Opel N, Redlich R, Mehler DMA, Emden D, et al. (2021): Systematic misestimation of machine learning performance in neuroimaging studies of depression. Neuropsychopharmacology 46:1510-1517.
132. Varoquaux G (2018): Cross-validation failure: Small sample sizes lead to large error bars. Neuroimage 180:68-77.
133. He T, An LJ, Chen PS, Chen JZ, Feng JS, Bzdok D, et al. (2022): Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. Nat Neurosci 25:1-10.
134. Vayena E, Blasimme A, Cohen IG (2018): Machine learning in medicine: Addressing ethical challenges. PLoS Med 15:e1002689.

135. Starke G, De Clercq E, Borgwardt S, Elger BS (2021): Computing schizophrenia: ethical challenges for machine learning in psychiatry. Psychol Med 51:2515-2521.

136. de Miguel I, Sanz B, Lazcoz G (2020): Machine learning in the EU health care context: exploring the ethical, legal and social issues. Inform Commun Soc 23:1139-1153.

137. Eickhoff SB, Heinrichs B (2021): The predictable human : Possibilities and risks of AI-based prediction of cognitive abilities, personality traits and mental illnesses. Nervenarzt 92:1140-1148.

138. Bzdok D, Ioannidis JPA (2019): Exploration, Inference, and Prediction in Neuroscience and Biomedicine. Trends Neurosci 42:251-262.

139. Rudin C (2019): Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206-215.

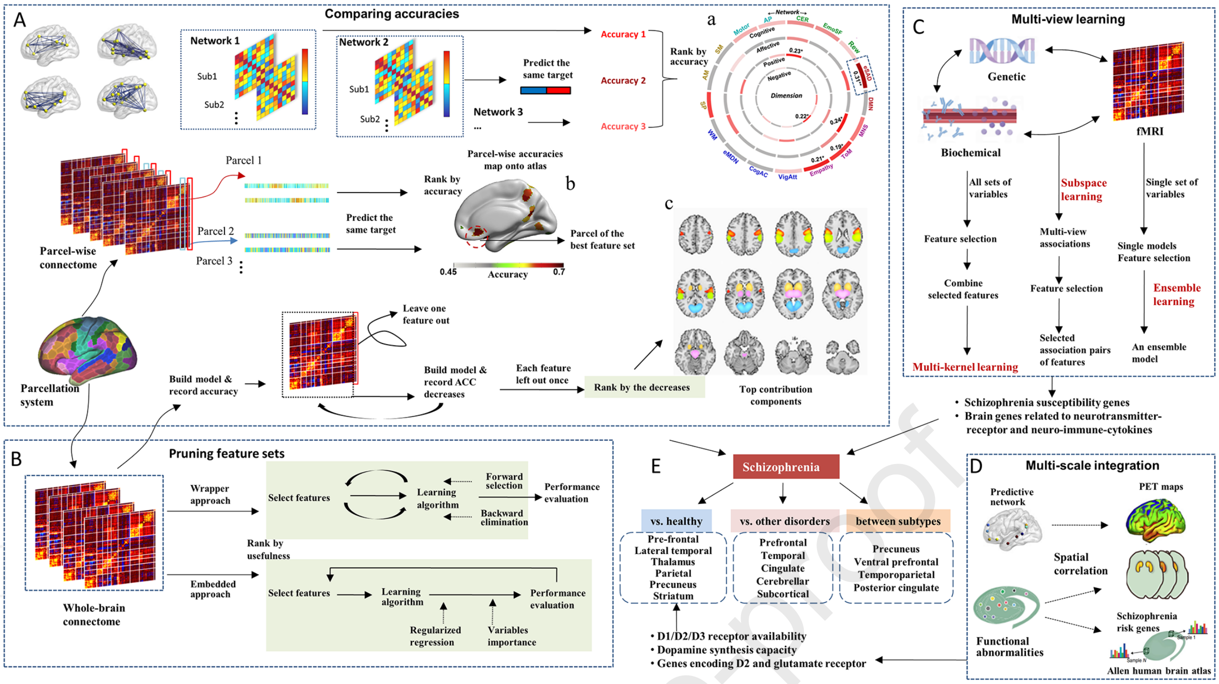
## Figure Legends:

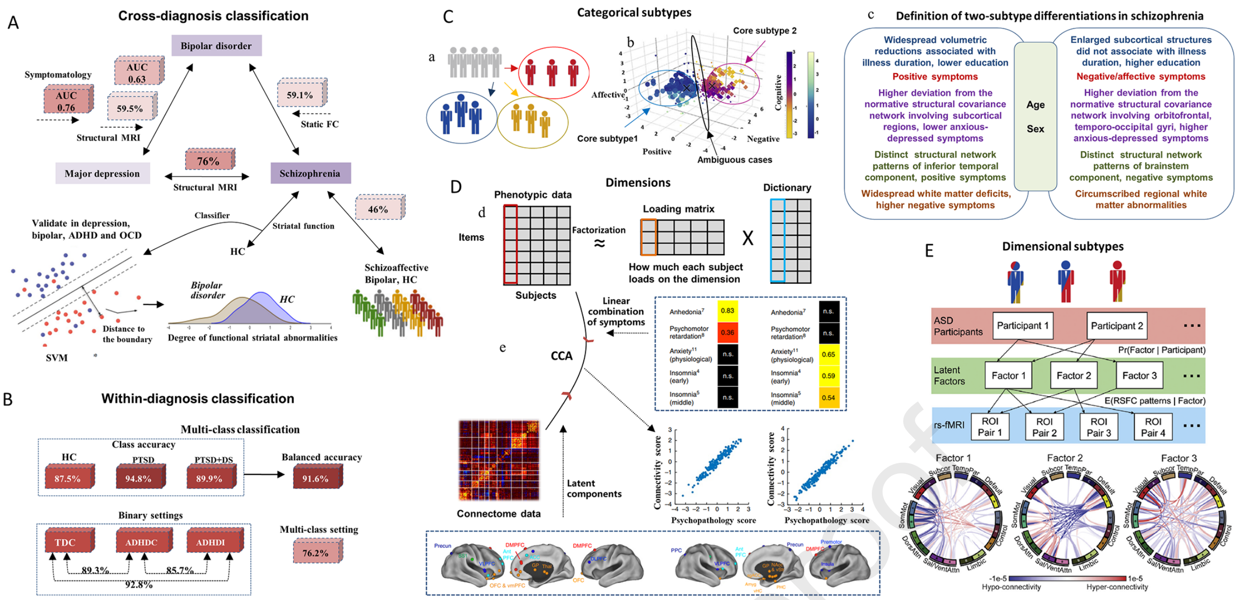
**Figure 1** Visual summary of the "perspective in leveraging accuracies to inform the pathophysiological relevance of feature-sets in single- and multi-view fashions". A) Selecting informative features via comparing their achievable accuracies (8,15) or based on the accuracy drops due to the exclusion of each feature from the model (16,21). Here the *a priori* sets of features are the connectome of a single network or a single parcel (i.e., region). Then a machine learning model can be built to assess the performance given each feature set individually. After each feature set is tested, sets of features with top performance can be selected; In *Chen et al.,* (8), connectivity within socio-affective network was found to be the best predictor of the cognitive symptomatology in schizophrenia patients (panel a); In *Chen et al.,* (15), the connectivity profiles of parcels seated in ventromedial frontal area were identified to best represent the neurobiological distinctions between the patients with schizophrenia dominated by negative or positive symptomatology (panel b). In *Yan et al.,* (21), network components involving hippocampus, supplementary motor area, paracentral lobule, precentral, and insular regions were found to be the top discriminative features across major psychiatric disorders (panel c); B) Informative features obtained after pruning a broad feature set based on predictive power with all available features input to the algorithm. Here we differentiate two approaches: the wrapper and the embedded settings; C) Multi-view machine learning to select informative features from different views of data like genetic and biochemical assays, complementing the biological insights provided by *in vivo* neuroimaging; D) Post-hoc integration of machine learning derived neuroimaging features with data from other biological sources through spatial correlation analysis. Here we show prior studies to integrate MRI derived features (network importance [8] and straital functional connectome [34]) with the distribution pattern of multiple neurotransmitter-receptors from multi-tracer PET maps as well as the gene expression patterns from the Allen human brain atlas (34) to reveal the associated molecular substrates of schizophrenia; E) Overview of the pathophysiological insights according to the new perspectives as informed by previous MRI-based machine learning studies in schizophrenia patients and complemented by those molecular and genetic features selected from multi-view classification experiments and multi-scale integration analytics. Integrant figures partly taken or adapted from Figure 4 in (15) with permission under CC-BY-NC-ND, from Figures 1,2,4 in (8) under CC-BY license, from Figure 3 in (21) under the Copyright Clearance Center's *RightsLink* (sso.copyright.com) license (*NO. 5340820701746*), from Figure 1 in (52) under *RightsLink* license (*NO. 5340821139538*), from Figure 1 in (34) under *RightsLink* license (*NO. 5340830193233*).

Abbreviations: ACC, accuracy; Sub, subject; PET, positron emission tomography.

**Figure 2** Summary of the perspectives in utilizing accuracies from differential classification to inform nosological relationship and to disentangle heterogeneous disorders through semi-supervised and unsupervised approaches. A) Overview of prior machine learning studies for differential diagnosis; Patients with major depression and schizophrenia diagnoses would likely be more differentiable than major depression versus bipolar disorder and schizophrenia versus bipolar disorder diagnoses at the neurobiological level (22,46-48). Comparing to other major psychiatric disorders, patients with bipolar disorder present more closely related the striatal dysfunction patterns with schizophrenia patients as indicated by testing the binary support vector machine classifier in independent cohorts (34); B) Comparatively, within diagnosis classification of clinical subtypes demonstrated much higher accuracies as indicated by two studies focused on PTSD (67) and ADHD (23); C) Most of previous subtyping work employed various hard-clustering methods which resulted in each patient to be part of one group as illustrated in panel (a); Panel (b) in turn applied fuzz-clustering algorithms for a soft separation of schizophrenia patients based on their expressive patterns along four data-driven psychopathological dimensions. Here each patient allows for shared cluster-memberships, and thus patients with ambiguous cluster-assignments could be identified and removed to form more compact, and hence core subgroups of a patient cohort (15); In panel (c), we summarized the definitions of schizophrenia two-subtype differentiations in recent studies as depicted in different colors (dark blue: *Chand et al.,* [76]; red: *Chen et al.,* [15]; purple: *Liu et al.,* [78]; green: *Rahaman et al.,* [75]; orange: *Sun et al.,* [77] age and gender are comparable between two-subgroup differentiations); D) Describing the structure of disorders via continuous dimensions: in panel (d), we show a conceptual schematic for phenotypic data factorization to derive psychopathological dimensions; in panel (e), we show an example from *Drysdale et al.,* (52) to depict the bidirectional compression approach of CCA which searches for a linear combination of symptoms to maximize its correlation with neural components; E) The use of Bayesian models for deriving dimensional subtypes of ASD participants. In particular, the model allows for each subject to express one or more patterns (i.e., factors), and each factor is formed by distinct functional connectivity patterns (91). Integrant figures partly taken or adapted from Figure 3 in (15) and Figures 1&2 in (91) with permission under CC-BY-NC-ND, from Figure 1 in (21) under the Copyright Clearance Center's *RightsLink* (sso.copyright.com) license (*NO. 5340820701746*), from Figure 1 in (52) under *RightsLink* license (*NO. 5340821139538*), from Figure 3 and Extended Data Figure 6 in (34) under *RightsLink* license (*NO. 5340830193233*).

Abbreviations: ASD, autism spectrum disorder; PTSD, post-traumatic stress disorder; TDC, typical developing control; HC, healthy control; ADHD, attention deficit hyperactivity disorder; CCA: canonical correlation analysis. Pr(Factor l Participant), the probability of a participate associates with each of the factors; E(RSFC patterns l Factor), the expectation of resting-state functional connectivity associates specifically with a factor.


A
Comparing accuracies
Network 1
Sub1
Sub2
Network 2
Sub1
Sub2
Predict the same target
Network 3
...
Accuracy 1
Accuracy 2
Accuracy 3
Rank by accuracy
a
Network
Motor
AP
CER
EmoSF
Rew
eSAD
DMN
MNS
ToM
Empathy
VigAtt
CogAC
eMDN
WM
SP
AM
SM
Cognitive
Affective
Positive
Negative
Dimension
0.23*
0.31**
0.22*
0.24*
0.19*
0.21*
Parcel 1
Parcel 2
Parcel 3
Parcel-wise connectome
Predict the same target
Rank by accuracy
Parcel-wise accuracies map onto atlas
b
Parcel of the best feature set
0.45
Accuracy
0.7
c
Parcellation system
Build model & record accuracy
Leave one feature out
Build model & record ACC decreases
Each feature left out once
Rank by the decreases
Top contribution components
B
Pruning feature sets
Whole-brain connectome
Wrapper approach
Select features
Learning algorithm
Forward selection
Backward elimination
Performance evaluation
Rank by usefulness
Embedded approach
Select features
Learning algorithm
Performance evaluation
Regularized regression
Variables importance
C
Multi-view learning
Genetic
fMRI
Biochemical
All sets of variables
Feature selection
Combine selected features
Multi-kernel learning
Subspace learning
Multi-view associations
Feature selection
Selected association pairs of features
Single set of variables
Single models Feature selection
Ensemble learning
An ensemble model
• Schizophrenia susceptibility genes
• Brain genes related to neurotransmitter-receptor and neuro-immune-cytokines
E
Schizophrenia
vs. healthy
Pre-frontal
Lateral temporal
Thalamus
Parietal
Precuneus
Striatum
vs. other disorders
Prefrontal
Temporal
Cingulate
Cerebrellar
Subcortical
between subtypes
Precuneus
Ventral prefrontal
Temporoparietal
Posterior cingulate
• D1/D2/D3 receptor availability
• Dopamine synthesis capacity
• Genes encoding D2 and glutamate receptor
D
Multi-scale integration
Predictive network
PET maps
Spatial correlation
Functional abnormalities
Schizophrenia risk genes
Sample 1
Sample N
Allen human brain atlas


A
Cross-diagnosis classification
Bipolar disorder
Symptomatology
AUC 0.76
AUC 0.63
59.5%
Structural MRI
59.1%
Static FC
Major depression
76%
Structural MRI
Schizophrenia
46%
Validate in depression, bipolar, ADHD and OCD
Classifier
Striatal function
HC
Schizoaffective Bipolar, HC
Distance to the boundary
SVM
Bipolar disorder
HC
Degree of functional striatal abnormalities
B
Within-diagnosis classification
Multi-class classification
Class accuracy
HC
87.5%
PTSD
94.8%
PTSD+DS
89.9%
Balanced accuracy
91.6%
Binary settings
TDC
ADHDC
ADHDI
89.3%
85.7%
92.8%
Multi-class setting
76.2%
C
Categorical subtypes
a
b
Core subtype 2
Affective
Cognitive
Core subtype1
Positive
Negative
Ambiguous cases
c
Definition of two-subtype differentiations in schizophrenia
Widespread volumetric reductions associated with illness duration, lower education
Positive symptoms
Higher deviation from the normative structural covariance network involving subcortical regions, lower anxious-depressed symptoms
Distinct structural network patterns of inferior temporal component, positive symptoms
Widespread white matter deficits, higher negative symptoms
Age
Sex
Enlarged subcortical structures did not associate with illness duration, higher education
Negative/affective symptoms
Higher deviation from the normative structural covariance network involving orbitofrontal, temporo-occipital gyri, higher anxious-depressed symptoms
Distinct structural network patterns of brainstem component, negative symptoms
Circumscribed regional white matter abnormalities
D
Dimensions
d
Phenotypic data
Items
Subjects
Factorization
≈
Loading matrix
How much each subject loads on the dimension
X
Dictionary
e
Linear combination of symptoms
CCA
Latent components
Connectome data
Anhedonia7 0.83
Psychomotor retardation8 0.36
Anxiety11 (physiological) n.s.
Insomnia4 (early) n.s.
Insomnia5 (middle) n.s.
Anhedonia7 n.s.
Psychomotor retardation8 n.s.
Anxiety11 (physiological) 0.65
Insomnia4 (early) 0.59
Insomnia5 (middle) 0.54
Connectivity score
Psychopathology score
Precun
SS1
VLPFC
DMPFC
Ant PFC
ACC
GP
Thal
OFC & vmPFC
DLPFC
OFC
PPC
Precun
GP NAcc & vSt
sgACC
Amyg
vHC
PHC
Premotor
Insula
E
Dimensional subtypes
ASD Participants
Participant 1
Participant 2
Pr(Factor | Participant)
Latent Factors
Factor 1
Factor 2
Factor 3
E(RSFC patterns | Factor)
rs-fMRI
ROI Pair 1
ROI Pair 2
ROI Pair 3
ROI Pair 4
Factor 1
Factor 2
Factor 3
Subcor
TempPar
Default
Visual
SomMot
DorsAttn
Sal/VentAttn
Limbic
Control
-1e-5
1e-5
Hypo-connectivity
Hyper-connectivity